UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

École doctorale IAEM Lorraine

Analyse prédictive des données d'apprentissage, en situation d'enseignement à distance

THÈSE

présentée et soutenue publiquement le 5 Décembre 2022

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention Informatique)

par

Amal BEN SOUSSIA

Composition du jury

Président :	Laurent VIGNERON	Professeur, Université de Lorraine
Rapporteurs :	Sébastien IKSAL	Professeur, Université du Mans
	Marie-Hélène ABEL	Professeure, Université Technologique de Compiègne
Examinateur :	Leandro WIVES	Professeur, Université Fédérale du Rio Grande do Sul
Directrice de thèse :	Anne BOYER	Professeure, Université de Lorraine
Co-Directeur de thèse :	Azim ROUSSANALY	Maître de Conférences, Université de Lorraine
Invitée :	Rozenn JARNOUEN	Directrice des formations, CNED

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503
LORIA

Mis en page avec la classe thesul.

Résumé

Pendant les dernières décennies, l'adoption de l'apprentissage en ligne a rapidement évolué et son utilisation a été encore plus poussée avec la pandémie de la COVID-19. L'objectif de ce mode d'enseignement est de garantir la continuité du processus d'apprentissage. Cependant, ce mode d'apprentissage connaît plusieurs défis, dont le plus répandu est les taux élevés d'échec. Ce problème est dû à de nombreuses raisons comme l'hétérogénéité des apprenants et la diversité de leurs comportements d'apprentissage, leur totale autonomie, le manque et/ou l'inefficacité du suivi pédagogique fourni.... Par conséquent, les enseignants ont besoin d'un système basé sur des méthodes analytiques leur permettant une prédiction précise et au plus tôt des apprenants à risque d'échec. Si cette solution est communément adaptée dans l'état de l'art, les travaux réalisés ne répondent pas à certaines particularités de l'apprentissage (la continuité et l'évolution de l'apprentissage, la diversité des apprenants et leur totale autonomie) et attentes des enseignants comme la génération d'alerte.

Cette thèse s'inscrit dans le domaine de l'analytique de l'apprentissage (*Learning Analytics*) et exploite les traces numériques des apprenants en ligne pour concevoir un système prédictif (*Early Warning Systems* (EWS)) dédié aux enseignants des établissements en ligne. L'objectif de cet EWS est d'identifier au plus tôt les apprenants à risque pour alerter les enseignants. Afin d'atteindre cet objectif, nous avons traité plusieurs sous-problématiques qui ont permis l'élaboration de quatre contributions scientifiques.

Nous commençons par proposer une méthodologie en profondeur qui repose sur les étapes de l'apprentissage automatique (ML) et qui permet l'identification de quatre indicateurs d'apprentissage : la performance, l'engagement, la réactivité et la régularité. Cette méthodologie met aussi en valeur l'importance des données temporelles pour l'amélioration des performances de prédiction. De plus, cette méthodologie a permis de définir le modèle avec la meilleure capacité à identifier les apprenants à risque.

Afin de répondre aux limites de l'état de l'art concernant l'importance de la dimension temporelle dans l'évaluation des systèmes prédictifs éducatifs, nous proposons des métriques temporelles qui mesurent la précocité des prédictions et la stabilité des systèmes. À partir de ces deux métriques, nous étudions les compromis qui existent entre les métriques de précision de ML et les métriques temporelles.

Les apprenants en ligne se caractérisent par la diversité de comportements d'apprentissage. Ainsi, un EWS doit répondre à cette diversité en assurant un fonctionnement équitable entre les différents profils types d'apprenants. Nous proposons une méthodologie d'évaluation qui se base sur l'identification des profils types d'apprenants et utilise des métriques temporelles et de précision.

En utilisant un EWS, les enseignants s'attendent à une génération d'alerte. Pour cette raison, nous concevons un algorithme qui s'appuie sur les résultats de prédiction, les métriques temporelles et la notion des règles d'alerte pour proposer une méthode automatique de génération d'alerte.

Le contexte applicatif de cette thèse est le Centre National d'Enseignement à Distance (CNED). Nous exploitons les traces numériques d'une population de collégiens inscrits en classe 3ème pendant les années scolaires 2017-2018 et 2018-2019.

Mots clés : analytique de l'apprentissage, apprentissage automatique, EWS, prédiction précoce,traces numériques, indicateurs d'apprentissage, évaluation, métriques temporelles, profils d'apprenants, génération d'alerte

Abstract

Over the past few decades, the adoption of e-learning has evolved rapidly and its use has been pushed even further with the COVID-19 pandemic. The objective of this learning mode is to guarantee the continuity of the learning process. However, the online learning is facing several challenges, and the most widespread is the high failure rates among learners. This issue is due to many reasons such as the heterogeneity of the learners and the diversity of their learning behaviors, their total autonomy, the lack and/or the inefficiency of the pedagogical provided follow-up. . .. Therefore, teachers need a system based on analytical and intelligent methods allowing them an accurate and early prediction of at-risk of failure learners. This solution is commonly adopted in the state of the art. However, the work carried out does not respond to some particularities of the learning process (the continuity and evolution of learning, the diversity of learners and their total autonomy) and to some teachers expectations such as the alert generation.

This thesis belongs to the field of learning analytics and uses the numeric traces of online learners to design a predictive system (*Early Warning Systems* (EWS)) dedicated to teachers in online establishments. The objective of this EWS is to identify learners at risk as soon as possible in order to alert teachers about them. In order to achieve this objective, we have dealt with several sub-problems which have allowed us to elaborate four scientific contributions.

We start by proposing an in-depth methodology based on the Machine Learning (ML) steps and that allows the identification of four learning indicators among : performance, engagement, reactivity and regularity. This methodology also highlights the importance of temporal data for improving prediction performance. In addition, this methodology allowed to define the model with the best ability to identify at-risk learners.

The 2nd contribution consists in proposing a temporal evaluation of the EWS using temporal metrics which measure the precocity of the predictions and the stability of the system. From these two metrics, we study the trade-offs that exist between ML precision metrics and temporal metrics.

Online learners are characterized by the diversity of their learning behaviors. Thus, an EWS should respond to this diversity by ensuring an equitable functioning with the different learners profiles. We propose an evaluation methodology based on the identification of learner profiles and that uses a wide spectrum of temporal and precision metrics.

By using an EWS, teachers expect an alert generation. For this reason, we design an algorithm which, based on the results of the prediction, the temporal metrics and the notion of alert rules, proposes an automatic method for alert generation. This algorithm targets mainly at-risk learners.

The context of this thesis is the French National Center for Distance Education (CNED). In particular, we use the numeric traces of k-12 learners enrolled during the 2017-2018 and 2018-2019 school years.

Keywords : learning analytics, machine learning, EWS, early prediction, learning traces, learning indicators, evaluation, temporal metrics, learning profils, alert generation

Table des matières

Résumé **i**

Abstract **iii**

Introduction

1 Contexte général 1
2 Problématique 4
2.1 Question de recherche 1 : comment représenter au mieux l'activité d'un apprenant inscrit dans un établissement d'apprentissage en ligne ? 4
2.2 Question de recherche 2 : comment évaluer l'évolution temporelle des systèmes de prédiction ? 5
2.3 Question de recherche 3 : les systèmes de prédiction sont-ils aussi performants sur l'ensemble de la population ? 6
2.4 Question de recherche 4 : comment alerter au plus tôt les enseignants des apprenants à risque ? 6
3 Contributions 6
3.1 Contribution 1 : Méthodologie en profondeur 7
3.2 Contribution 2 : Évaluation temporelle des systèmes prédictifs 7
3.3 Contribution 3 : Identification des profils types d'apprenants pour l'évaluation des systèmes prédictifs 8
3.4 Contribution 4 : Génération précoce d'alerte 8
4 Organisation de la thèse 8

Chapitre 1
État de l'art

1.1 Introduction 11
1.2 Notions issues des sciences de l'éducation 12
1.2.1 Définition de la réussite 12
1.2.2 Notion d'apprenant à risque 13

1.2.3 Apprenant en ligne à risque . . . 13
1.2.4 Discussions . . . 14
1.3 Analytique de l'apprentissage . . . 15
1.4 La prédiction précoce et précise du risque d'échec . . . 17
1.4.1 Définition . . . 17
1.4.2 La précocité dans l'état de l'art . . . 18
1.4.3 Systèmes d'alerte précoce . . . 18
1.4.3.1 Définition et objectifs . . . 19
1.4.3.2 Exemples . . . 19
1.4.4 Discussions et Limites . . . 20
1.5 Les données utilisées dans les projets des LA . . . 20
1.5.1 Les données et les traces numériques utilisées dans la littérature . . . 20
1.5.2 Les données temporelles . . . 22
1.5.3 Indicateurs d'apprentissage . . . 23
1.5.3.1 L'indicateur de la performance . . . 23
1.5.3.2 L'indicateur de l'engagement . . . 23
1.5.3.3 L'indicateur de la régularité . . . 24
1.5.4 Discussions . . . 25
1.6 La modélisation de la problématique du risque dans l'état de l'art . . . 25
1.6.1 Apprentissage automatique . . . 26
1.6.1.1 Définition . . . 26
1.6.1.2 Apprentissage supervisé . . . 26
1.6.1.3 Apprentissage non supervisé . . . 26
1.6.2 L'apprentissage automatique en éducation . . . 27
1.6.3 Modèles de ML utilisés dans l'état de l'art . . . 27
1.6.4 Comment traiter la problématique de la prédiction du risque ? . . . 29
1.6.5 Aperçu sur les méthodologies de l'état de l'art . . . 30
1.6.6 Synthèse . . . 31
1.7 Évaluation des modèles et systèmes de l'état de l'art . . . 31
1.7.1 Métriques de performance proposées par le ML . . . 32
1.7.1.1 Métriques de régression . . . 32
1.7.1.2 Métriques de classification . . . 32
1.7.2 Les métriques de performance algorithmique dans l'état de l'art . . . 33
1.7.2.1 Les métriques de précision dans les LA . . . 34
1.7.2.2 Les métriques temporelles . . . 34
1.7.3 Discussions et Limites . . . 35
1.8 Tableau récapitulatif des travaux de l'état de l'art . . . 36

1.9 Conclusion 40

Chapitre 2
Contexte applicatif

2.1 Introduction 43
2.2 Le contexte du CNED 44
2.2.1 Historique du CNED 44
2.2.2 Le CNED... un centre d'enseignement assez particulier 45
2.3 Analyse descriptive des données 46
2.3.1 Description des sources de données 46
2.3.2 Description de jeu de données 47
2.3.3 Description de la population 48
2.3.4 Description des matières 49
2.3.5 Description des devoirs 52
2.3.6 Description des traces numériques 52
2.4 Analyse diagnostic : étude de corrélation 53
2.4.1 Définitions de base 53
2.4.1.1 Matrice de corrélation 53
2.4.1.2 Test de significativité (p-values) 54
2.4.2 Définitions des groupes d'élèves 54
2.4.2.1 Groupes d'élèves selon le seuil 12 54
2.4.2.2 Groupes d'élèves selon le seuil 10 55
2.5 Construction de la base de données des expérimentations 57
2.6 Conclusion 58

Chapitre 3
Une méthodologie en profondeur

3.1 Introduction 59
3.2 Description de la méthodologie proposée 60
3.2.1 Extraction des caractéristiques 62
3.2.2 Sélection des caractéristiques 63
3.2.3 Approche de modélisation 64
3.2.4 Intégration d'une caractéristique temporelle dans le vecteur de l'apprenant . . . 64
3.3 Application de la méthodologie sur le contexte du CNED 65
3.3.1 Extraction des caractéristiques de la base de données CNED 65
3.3.2 Application de l'approche de modélisation au contexte du CNED 66
3.4 Implantation de la méthodologie : expérimentations et résultats 67

3.4.1 Protocole expérimental 67
3.4.2 Modélisation et évaluation de l'exactitude 68
3.4.2.1 1ère étape du protocole expérimental 68
3.4.2.2 2ème étape du protocole expérimental 69
3.4.3 Processus de la sélection des caractéristiques 70
3.4.3.1 Application des méthodes filtrantes 70
3.4.3.2 Application des méthodes enveloppantes 72
3.4.3.3 Application des méthodes intégrées 73
3.4.3.4 Analyse des résultats de la première phase d'évaluation de la méthodologie 73
3.4.4 Ré-évaluation : résultats de la mesure FPR 74
3.4.4.1 Suivi des résultats de la mesure FPR 75
3.4.4.2 Analyse des résultats de la 2ème phase d'évaluation de la méthodologie 76
3.4.5 Résultats de l'intégration de la caractéristique temporelle dans le vecteur d'activité de l'apprenant 77
3.5 Validation : Application des résultats de la méthodologie sur d'autres matières 78
3.5.1 1ère matière : les mathématiques 79
3.5.2 2ème matière : Histoire-géographie 82
3.6 Conclusion 86

Chapitre 4
Évaluation temporelle des systèmes prédictifs

4.1 Introduction 87
4.2 Description du problème 88
4.2.1 Formalisation du problème 89
4.2.2 Suivi régulier des prédictions 89
4.3 1ère métrique temporelle : La précocité 89
4.3.1 La précocité : première date d'une bonne prédiction 90
4.3.1.1 Description de l'algorithme 90
4.3.1.2 Exemple de calcul de la précocité 91
4.3.2 La précocité : un ensemble de dates de bonnes prédictions 91
4.3.2.1 Description du 1er sous-algorithme 92
4.3.2.2 Description du 2ème sous-algorithme 92
4.3.2.3 Exemple de calcul de la précocité 93
4.4 2ème métrique temporelle : la stabilité 94
4.5 Étude des compromis 94
4.5.1 Le compromis Précocité-Exactitude : EAS 95

4.5.2 Le compromis Précocité-Stabilité : ESS . . . 95
4.6 Description des systèmes . . . 96
4.7 1ère partie expérimentale . . . 98
4.7.1 Résultats des métriques temporelles sur les 12 premières semaines . . . 98
4.7.1.1 Résultats de précocité, d'exactitude et de EAS sur 12 semaines . . . 98
4.7.1.2 Résultats de la stabilité sur 12 semaines . . . 99
4.7.2 Résultats des métriques temporelles sur toute l'année . . . 100
4.7.2.1 Résultats de précocité, d'exactitude et de EAS sur toute l'année . . . 100
4.7.2.2 Résultats de la stabilité sur toute l'année . . . 102
4.8 2ème partie expérimentale . . . 103
4.8.1 Analyse des mesures de compromis . . . 103
4.8.1.1 Analyse des résultats de la mesure EAS . . . 104
4.8.1.2 Analyse des résultats de la mesure ESS . . . 105
4.8.1.3 Comparaison entre EAS et ESS : . . . 106
4.8.2 Discussion . . . 106
4.9 Conclusion . . . 107

Chapitre 5
Identification des profils types d'apprentissage pour évaluer les systèmes de prédiction

5.1 Introduction . . . 109
5.2 La méthodologie d'évaluation . . . 110
5.2.1 Description de la méthodologie . . . 110
5.2.2 Description des données . . . 111
5.3 1ère étape de la méthodologie : identification des profils types d'apprenants . . . 112
5.3.1 État de l'art . . . 112
5.3.2 Identification des indicateurs d'apprentissage . . . 113
5.3.3 Identification des profils types d'apprenants . . . 113
5.3.4 Personas : des exemples . . . 114
5.4 2ème étape de la méthodologie : Métriques d'évaluation . . . 115
5.5 3ème étape de la méthodologie : une évaluation en profondeur d'un système prédictif . . . 115
5.5.1 Une brève description du modèle de prédiction . . . 115
5.5.2 Résultats . . . 116
5.5.2.1 Résultats de la mesure d'exactitude . . . 116
5.5.2.2 Résultats de la précocité et la stabilité temporelle . . . 118
5.5.3 Discussions . . . 120
5.6 Conclusion . . . 120

Chapitre 6
Génération précoce d'alerte

6.1 Introduction . . . 123
6.2 Formalisation du problème d'alerte . . . 124
6.3 Règle d'alerte . . . 124
6.4 Algorithme d'alerte . . . 126
6.4.1 Les pré-requis de l'alerte : la précocité de prédiction et la stabilité . . . 126
6.4.2 Description de l'algorithme d'alerte . . . 127
6.5 Description des EWS . . . 129
6.6 Expérimentations et résultats . . . 129
6.6.1 Les résultats de ESS des quatre systèmes . . . 130
6.6.2 Application des règles d'alerte et résultats . . . 133
6.7 Conclusion . . . 136

Conclusion générale

1 Rappel du cadre et des objectifs de la thèse . . . 137
2 Principales contributions . . . 138
2.1 Une méthodologie en profondeur . . . 138
2.2 Évaluation temporelle des systèmes prédictifs . . . 138
2.3 Identification des profils d'apprenants pour l'évaluation des systèmes prédictifs . 138
2.4 Génération précoce d'alerte . . . 138
3 Limites des travaux effectués . . . 139
4 Perspectives . . . 139

Bibliographie **141**

Bibliographie **141**

Annexe A Les caractéristiques extraites de la base de données CNED **151**
A.1 Les caractéristiques démographiques . . . 151
A.2 Les caractéristiques de l'indicateur de performance . . . 151
A.3 Les caractéristiques de l'indicateur d'engagement . . . 152
A.4 Les caractéristiques de l'indicateur de régularité . . . 156
A.5 Les caractéristiques de l'indicateur de réactivité . . . 156

Annexe B Tableau de bord avec les indicateurs d'apprentissage identifiés **157**
B.1 Vue de toute la classe . . . 157
B.2 Vue d'un élève . . . 160

Table des figures

1 Données multi-sources, hétérogènes et de différents types 2

1.1 Schéma représentatif des parties de l'état de l'art 12
1.2 Cycle de l'analytique de l'apprentissage - Figure reproduite de [1] 16
1.3 Fréquence d'utilisation des techniques de ML dans la littérature - Figure reproduite de [2] 28
1.4 Le processus standard des prédiction des résultats -Figure reproduite de [3]- 30
1.5 Matrice de confusion 33

2.1 Phases du modèle de processus CRISP-DM pour l'exploration de données - Figure reproduite de [4] - 43
2.2 Les sources de données du CNED 46
2.3 Les sources de données du CNED 53

3.1 Les étapes de la méthodologie proposée 62
3.2 Première étape de l'évaluation de l'exactitude 68
3.3 Deuxième étape de l'évaluation de l'exactitude 69
3.4 L'évolution des valeurs d'*exactitude* suite à la sélection par le test khi-deux 71
3.5 L'évolution des valeurs d'*exactitude* suite à la sélection par le test ANOVA 71
3.6 L'évolution des valeurs d'exactitude suite à la sélection par la technique de RFECV . . . 73
3.7 L'évolution des valeurs d'exactitude suite à la sélection par la technique intégrée 74
3.8 L'évolution des valeurs de FPR suite à la sélection par la technique de khi-deux 75
3.9 L'évolution des valeurs de FPR suite à la sélection par la technique de ANOVA 76
3.10 L'évolution des valeurs de *FPR* suite à la sélection par la technique de RFECV 76
3.11 L'évolution des valeurs de *FPR* suite à la sélection par la technique intégrée 77
3.12 L'évolution des valeurs d'exactitude et de FPR avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant 78
3.13 L'évolution de l'exactitude en utilisant les traces numériques des mathématiques - 1ère étape du protocole expérimental 80
3.14 L'évolution de l'exactitude en utilisant les traces numériques des mathématiques - 2ème étape du protocole expérimental 80
3.15 L'évolution de FPR en utilisant les traces numériques des mathématiques 81
3.16 L'évolution de l'exactitude et FPR avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant- Les mathématiques- 82
3.17 L'évolution de l'*exactitude* en utilisant les traces numériques effectuées sur la matière histoire-géographie - 1ère étape du protocole expérimental 83
3.18 L'évolution de l'*exactitude* en utilisant les traces numériques effectuées sur la matière histoire-géographie - 2ème étape du protocole expérimental 84

3.19 L'évolution de *FPR* en utilisant les traces numériques effectuées sur la matière histoire-géographie - . . . 85
3.20 L'évolution de l'*exactitude* et *FPR* avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant-Histoire géographie- . . . 85

4.1 Un exemple de suivi régulier des prédictions pendant un intervalle de temps T . . . 90
4.2 EWS_1 VS EWS_2 EWS_3 EWS_4 en terme d'exactitude . . . 97
4.3 Comparer EWS_1, EWS_2, EWS_3 and EWS_4 en terme de stabilité temporelle . . . 103
4.4 Taux de la mesure de EAS par étiquette de classe -tous les systèmes- . . . 104
4.5 Taux de la mesure de ESS par chaque étiquette de classe -tous les systèmes- . . . 105

5.1 Les étapes de la méthodologie d'évaluation . . . 111
5.2 Exemple de deux personas [5] . . . 115
5.3 Évolution de la mesure d'exactitude avec toute la population et les deux sous-ensembles de données . . . 117
5.4 Évolution de la mesure d'exactitude des profils types d'apprenants des sous-ensembles de données *traces (A)* et *pas_traces (B)* . . . 117

6.1 Tableau de génération d'alerte . . . 125
6.2 Matrice de la confusion adaptée à l'évaluation des règles d'alerte . . . 126
6.3 Comparaison des EWS_1, EWS_2, EWS_3 et EWS_4 en terme de stabilité temporelle . . . 130
6.4 Les points de précocité et leurs ESS correspondants - EWS_1 . . . 131
6.5 Les points de précocité et leurs ESS correspondants - EWS_2 . . . 132
6.6 Les points de précocité et leurs ESS correspondants - EWS_3 . . . 132
6.7 Les points de précocité et leurs ESS correspondants - EWS_4 . . . 133

B.1 Le tableau de bord vue de classe . . . 158
B.2 Évolution de trois indicateurs d'apprentissage pour toute la classe . . . 158
B.3 Pouls d'activité des élèves pendant une période choisie . . . 159
B.4 Proximité éducative des élèves- Indicateur d'engagement- . . . 159
B.5 Proximité éducative des élèves -Indicateur de réactivité- . . . 160
B.6 Le tableau de bord vue de classe . . . 160
B.7 Évolution de trois indicateurs d'apprentissage pour toute la classe . . . 161
B.8 Évolution de trois indicateurs d'apprentissage pour toute la classe . . . 161
B.9 Évolution de trois indicateurs d'apprentissage pour toute la classe . . . 162

Liste des tableaux

1.1 Groupes de données 21
1.2 Les métriques de performance utilisées pour la classification et la régression 33
1.3 Tableau récapitulatif des travaux de l'état de l'art 39

2.1 Tableau récapitulatif des types de données fournies par chaque source du CNED 47
2.2 Dates de scolarité de deux trains scolaires 48
2.3 Localisation des inscrits 48
2.4 Statistiques de la réussite scolaire pour chaque année scolaire 49
2.5 Description générale des matières de la classe 4-131 50
2.6 Caractéristiques des matières du train 7 51
2.7 Caractéristiques des matières du train 8 52
2.8 Nombre d'inscrits de SP31 appartenant à chaque catégorie de la première définition d'un groupe d'élèves 54
2.9 Résultats des corrélations entre une moyenne supérieure strictement à 12 et toutes les variables 55
2.10 Résultats des corrélations entre une moyenne comprise entre 8 et 12 et toutes les variables 55
2.11 Résultats des corrélations entre une moyenne inférieure strictement à 8 et toutes les variables 56
2.12 Nombre d'inscrits de SP31 appartenant à chaque catégorie de la deuxième définition des groupes d'élève 56
2.13 Résultats des corrélations entre une moyenne inférieure ou égale à 10 et toutes les variables 56
2.14 Résultats des corrélations entre une moyenne inférieure à 10 et toutes les variables . . . 57

3.1 Nombre d'apprenants par classe 67
3.2 Exactitude des modèles en utilisant uniquement les caractéristiques démographiques . . 69
3.3 Nombre d'apprenants total et par étiquette de classe inscrits dans la matière des Mathématiques 79
3.4 Nombre d'apprenants total et par étiquette de classe inscrits dans la matière Histoire-Géographie- 79

4.1 Mesures de précocité, exactitude et EAS-EWS_1 VS EWS_2 99
4.2 Mesures de précocité, exactitude et EASA-EWS_3 VS EWS_4 99
4.3 Mesure de stabilité 100
4.4 Mesures de précocité et EAS -EWS_1 VS EWS_2 101
4.5 Mesures de précocité et EAS -EWS_3 VS EWS_4 101

5.1 Nombre d'apprenants dans chaque classe de chaque sous-ensemble de données 112
5.2 Résultats du *clustering* par sous-ensembles de données et par classes 114

5.3 Mesure de la précocité et de la stabilité par chaque classe d'un groupe de profils 119
5.4 Les mesures de la précocité et stabilité pour chaque persona 119

6.1 ESS des quatre EWS et leurs précocités correspondantes 133
6.2 Application des cinq règles d'alerte sur les quatre systèmes 134

A.1 Tableau des caractéristiques démographiques 151
A.2 Les caractéristiques extraites de la base du CNED correspondantes à l'indicateur d'engagement 155

Introduction

1 Contexte général

Les premières traces d'enseignement ont été trouvées dans l'Égypte des pharaons et en Inde [6]. Les méthodes et moyens d'enseignement ont connu des évolutions à travers les époques jusqu'à l'école d'aujourd'hui qui se transforme par le numérique.
Étant donné l'importance de l'éducation pour le développement des peuples et dans l'objectif d'avoir une éducation pour tous, l'enseignement à distance est apparu. Par enseignement à distance, nous entendons tout processus qui se déroule entre un formateur et un apprenant éloignés géographiquement. En suivant les époques et le développement de nouveaux supports médiatiques, l'enseignement à distance a pris plusieurs formes et est passé par l'utilisation du support imprimé, l'acheminement des supports par la poste, la diffusion par la radio ensuite la télévision, les cassettes vidéos jusqu'à l'enseignement assisté par l'ordinateur et, plus récemment, avec l'apparition de l'Internet, l'enseignement en ligne (ou *e-learning*) sous la forme, entre autres, des cours en ligne ouverts et massifs (Massive Open Online Courses MOOCs) et la création des universités et écoles en ligne [7]. À titre d'exemple, la pandémie de la COVID-19 a montré l'importance et l'explosion de l'adoption de l'enseignement en ligne par les différentes institutions à travers le monde. C'est un moyen efficace de maintenir la continuité du processus d'apprentissage en permettant l'accès à l'école de n'importe où et à tout moment, à condition de disposer d'une infrastructure technologique. **Cette thèse se situe dans le contexte de l'enseignement à distance, plus particulièrement l'enseignement en ligne.**

L'enseignement en ligne demande une infrastructure numérique solide se composant essentiellement des plateformes éducatives et administratives dont le premier objectif est d'assurer la bonne diffusion du contenu pédagogique. En plus de leurs applications internes, les établissements scolaires et universitaires en ligne optent pour *les systèmes de gestion d'apprentissage* ou en anglais *Learning Management Systems (LMS)* en les adaptant à leurs besoins et objectifs d'enseignement. Tout LMS contient un espace numérique interactif qui fournit des contenus et des activités pédagogiques ainsi que des outils de communication : cours, vidéos, tests de positionnement, zones de soumission de devoirs, quizs, forums, espaces de discussion ….

En utilisant une plateforme éducative, l'apprenant laisse des *traces numériques*. Dans ce contexte, une trace numérique est toute activité ou consultation de ressource effectuée par l'apprenant, c'est-à-dire ses interactions avec le système d'apprentissage. [8] définit la trace numérique comme une suite d'actions et de réactions entre un humain et un ordinateur. Les LMS et les applications éducatives sauvegardent ces traces au cours du temps.
Dans le cadre d'un établissement d'enseignement en ligne, les données sont généralement multi-sources. En effet, les institutions s'appuient en général sur plus d'une application pour l'apprentissage et/ou la gestion administrative des ressources et des apprenants. Ces applications génèrent des données hété-

rogènes et de différents types. En effet, l'hétérogénéité est expliquée par la disponibilité des données administratives décrivant la démographie des profils des apprenants, des traces numériques ainsi que des données sur les performances académiques, les notes et les évaluations.
De plus, ces applications fournissent des données de types différents : stables et constantes dans le temps ou dépendantes du temps. En effet, elles peuvent être, par exemple, des données démographiques (date de naissance, sexe, lieu de résidence, bourse…) qui sont invariantes dans le temps. Par ailleurs, d'autres données générées par ces applications sont étroitement dépendantes du temps ; citons par exemple, les notes et les traces numériques [9]. La Figure 1 illustre la multitude de sources de données dont peut disposer un établissement d'enseignement en ligne. Chacune de ces sources correspond à un service de cet établissement.

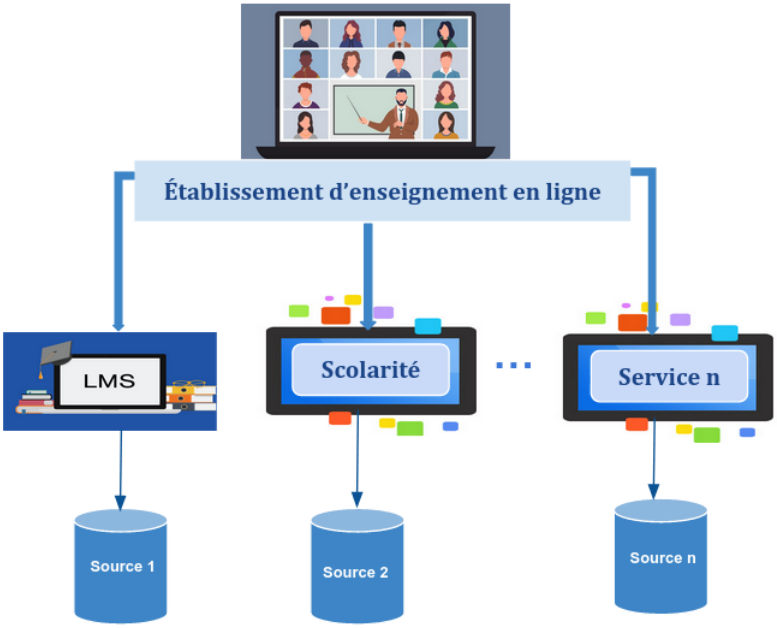


FIGURE 1 – Données multi-sources, hétérogènes et de différents types

L'évolution des technologies et outils d'apprentissage et leur généralisation est à l'origine de l'explosion de la quantité de données et de leur croissante accessibilité [10]. Dans le contexte d'enseignement en ligne, en l'absence de contact direct entre l'apprenant et le formateur, la compréhension de ces données et leur analyse sont des éléments clef pour avoir un aperçu sur le comportement des apprenants et anticiper les éventuelles difficultés d'apprentissage. **Cette thèse s'intéresse au traitement des données multi-sources, hétérogènes et diverses pour répondre à ces défis de l'enseignement en ligne.**

Les principaux problèmes détectés dans les environnements d'apprentissage en ligne sont les taux élevés d'absentéisme, les abandons précoces et les faibles taux d'achèvement des cours [11] qui conduisent à un échec total de l'apprenant. Certes, l'échec dans les études est un phénomène qui a toujours existé et auquel l'enseignement traditionnel a toujours cherché à trouver des solutions. Mais, il est encore plus accentué et répandu parmi les apprenants inscrits dans un établissement scolaire ou universitaire en ligne. Les statistiques annuelles de plusieurs institutions en ligne à travers le monde, notamment l'université ouverte au Royaume Uni (OU) [12], montrent l'importance de ce phénomène.

C'est dans ce cadre que le domaine de ***l'analytique de l'apprentissage*** ou en anglais *Learning Ana-*

lytics (LA) a vu le jour. Selon la 1ère conférence de « Learning Analytics and Knowledge », les LA sont la discipline consacrée à la mesure, la collecte, l'analyse et la présentation de rapports basés sur des données des apprenants en contexte d'apprentissage dans le but de comprendre et d'optimiser l'apprentissage et l'environnement dans lequel il s'opère [13]. Les LA se présentent comme une branche du domaine de l'apprentissage assisté par la technologie (*Technologie-Enhanced Learning* ou TEL) qui a émergé au cours de la dernière décennie. C'est une approche résolument multidisciplinaire basée sur le traitement de données, la fouille de données, l'intelligence artificielle, l'exploration des données éducatives (*Educational Data mining*) et la visualisation. [14] passe en revue les applications potentielles du LA et résume leurs objectifs de la manière suivante :

1. détection des élèves "*à risque*" afin de proposer des interventions positives conçues pour améliorer leur parcours
2. recommandation aux apprenants de support de cours et d'activités d'apprentissage
3. identification des besoins d'amélioration pédagogique et mesure des résultats obtenus
4. parcours d'apprentissage personnalisés
5. assistance aux enseignants en leur proposant des outils et méthodes d'enseignement
6. assistance au processus de recrutement des étudiants.

En bref, l'objectif des LA est d'adapter les opportunités éducatives aux besoins et aux capacités de chaque apprenant, en particulier d'intervenir auprès des élèves à risque.
Par apprenant à risque, nous entendons tout apprenant qui montre des difficultés d'apprentissage et/ou des comportements défaillants susceptibles de l'empêcher d'atteindre les objectifs pédagogiques prédéfinis par son institution.

Les différents travaux de LA sont essentiellement dédiés aux deux principaux acteurs de l'apprentissage : les apprenants et les enseignants.
Les travaux pour les apprenants présentent des solutions pour que ces derniers mettent en oeuvre des stratégies qui leur permettent de compenser leur retard, leurs erreurs et leurs différentes lacunes d'apprentissage [15].
L'intention des travaux pour les enseignants est généralement d'informer ces derniers sur la situation académique de chacun de leurs apprenants en temps réel et de manière évolutive. Ces travaux visent à suivre les progrès d'apprentissage de plusieurs élèves et à aider les enseignants à jouer efficacement leur rôle sur plusieurs axes de l'enseignement notamment la gestion de classe, la facilitation de l'apprentissage, le tutorat, le conseil, l'évaluation et la notation . . .[15].

Les travaux dans cette thèse sont dédiés aux enseignants des établissements d'apprentissage en ligne. En effet, elle s'inscrit dans le domaine des LA et traite le problème des taux élevés d'échec dans l'enseignement en ligne en suivant la 1ère application des LA qui est la détection précise et au plus tôt des apprenants défaillants et à risque.
Par détection **précise**, nous entendons un processus de prédiction qui identifie de manière fiable un nombre maximal d'apprenants à risque. Une détection **au plus tôt** permet une intervention précoce de l'enseignant.

Dans la prochaine section, nous présentons la problématique ainsi que ses enjeux et définissons plusieurs sous-problématiques permettant d'y apporter des réponses.

2 Comment prédire au plus tôt et avec précision les apprenants à risque dans un contexte d'enseignement en ligne ?

Nous cherchons à proposer un **système** aux enseignants d'un établissement d'apprentissage en ligne permettant d'être alertés au plus tôt et de manière précise des apprenants à risque d'échec, pour pouvoir intervenir au bon moment auprès de ces derniers. Dans cet objectif, nous avons choisi de nous appuyer sur les techniques d'apprentissage automatique (*Machine Learning (ML)*).

Dans la littérature, ces systèmes sont souvent appelés ***les systèmes d'alerte précoce*** ou en anglais *Early Warning Systems (EWS).* [1] Les EWS s'appuient à la fois sur l'ensemble de données d'apprentissage et les modèles prédictifs de ML pour prédire au plus tôt les apprenants à risque [16].

Lors de la conception d'un tel système, nous nous sommes trouvés face à plusieurs défis, notamment :
- la complexité de la compréhension et de traitement des traces numériques générées par les plateformes éducatives.
- la nature évolutive et temporelle du processus d'apprentissage.
- la diversité des profils d'apprenants exprimant chacun des besoins et attentes différents et nécessitant un accompagnement personnalisé.
- l'importance d'une intervention efficace et au bon moment qui cible au maximum les apprenants à risque.

Afin de trouver des solutions à chacun de ces défis, nous posons les questions de recherche suivantes que nous détaillons davantage dans les sections à venir :

1. **Comment représenter au mieux l'activité d'un apprenant inscrit dans un établissement d'apprentissage en ligne ?**
2. **Comment évaluer l'évolution temporelle des systèmes de prédiction ?**
3. **Les systèmes de prédiction sont-ils aussi performants sur l'ensemble de la population ?**
4. **Comment alerter au plus tôt les enseignants des apprenants à risque ?**

2.1 Question de recherche 1 : comment représenter au mieux l'activité d'un apprenant inscrit dans un établissement d'apprentissage en ligne ?

Dans le cadre d'un apprentissage traditionnel, les experts pédagogiques se sont intéressés à l'élaboration des critères académiques pour évaluer les apprenants qui peuvent être différents selon le contexte et le pays. Cependant, ces critères ne restent pas forcément valables pour l'enseignement en ligne. Ce mode d'apprentissage est particulier et se caractérise par des dimensions différentes, notamment la disparition du contact direct, la totale autonomie (pour certains cas), la flexibilité du rythme d'apprentissage,. . .qui rendent la représentation des apprenants plus complexe.
Les établissements d'enseignement en ligne n'imposent généralement pas un rythme bien déterminé à leurs apprenants. Au moment du suivi, les enseignants se trouvent avec une variété de profils d'apprenants tellement différents et hétérogènes qui ne facilitent pas leur accompagnement.
Les traces numériques d'apprentissage restent le seul moyen pour avoir un aperçu sur l'activité des apprenants. Mais, l'utilisation fréquente de ces plateformes et espaces numériques de travail par un nombre important d'apprenants au quotidien génèrent des données massives.

1. Dans la suite de ce manuscrit, nous utilisons EWS pour désigner les sytèmes d'alerte précoce.

Bien que chaque individu soit spécifique, les apprenants en ligne peuvent partager certains comportements d'apprentissage. L'analyse des traces numériques devient alors une clef d'observation de ces comportements. Nous exploitons ces données pour définir des critères d'apprentissage (présence, implication, performance académique, assiduité ...). Chacun de ces critères constitue un indicateur pédagogique. Nous nous appuyons sur l'ensemble de ces indicateurs pour définir le comportement pédagogique d'un apprenant.
Nous cherchons, à travers l'analyse des données historiques à définir automatiquement les indicateurs d'apprentissage qui représentent le mieux l'activité d'un apprenant inscrit en ligne.

En plus des indicateurs d'apprentissage, le système doit reposer sur tout un processus qui garantit une prédiction précise des apprenants à risque pour permettre une intervention pédagogique. C'est le résultat de toute une méthodologie qui, en se basant sur les traces numériques, les indicateurs pédagogiques et les modèles de ML, permet un suivi régulier de l'apprentissage et des apprenants. Dans ce contexte, nous cherchons à établir l'approche de travail basée sur les modèles de ML la mieux adaptée à nos objectifs.

Afin de proposer une solution de représentation précoce et précise du risque d'échec des apprenants en ligne, et étant conscients de la particularité du contexte relativement à l'apprentissage traditionnel, les toutes premières questions que nous nous posons sont les suivantes :

1. Quelles traces numériques faudrait-il collecter pour avoir un aperçu sur l'activité d'un apprenant en ligne ?
2. Quels sont les indicateurs d'apprentissage les plus pertinents permettant au mieux la représentation de l'activité d'un apprenant inscrit en ligne ?
3. Quelle approche de représentation serait la mieux adaptée à notre problématique et permettant un suivi régulier de l'activité des apprenants ?

2.2 Question de recherche 2 : comment évaluer l'évolution temporelle des systèmes de prédiction ?

L'objectif des EWS dédiés aux enseignants est de prédire précisément et au plus tôt les apprenants à risque. En adoptant ces systèmes, l'enseignant se réfère à leurs résultats de prédiction dans son suivi académique. De ce fait, les algorithmes exploités par les EWS doivent satisfaire les conditions de la précocité et la précision des résultats fournis ; d'où, l'importance de l'étape de l'évaluation des résultats de ces systèmes.
Les données générées par les plateformes d'éducation en ligne sont disponibles au fur et à mesure des activités des apprenants, car elles dépendent fortement du moment auquel les apprenants interagissent avec le contenu éducatif ; d'où l'aspect évolutif et dynamique des EWS. Cet aspect dynamique des EWS peut aussi impacter la stabilité de leurs prédictions au cours du temps.

Afin d'évaluer l'efficacité d'un EWS dans la réalisation de ses objectifs, des métriques communes sont utilisées. D'une part, des mesures de performance sont utilisées pour évaluer la complexité temporelle et spatiale des systèmes par rapport aux données utilisées. D'autre part, les métriques de précision de ML sont utilisées pour qualifier la capacité du système à prédire correctement. Malgré la diversité des métriques d'évaluation existantes, aucune d'entre elles n'est dédiée à mesurer la précocité, la continuité et l'évolution des prédictions dans le temps. Cependant, le temps est une dimension importante qui doit être prise en compte lors de l'évaluation d'un EWS à mesure que l'apprentissage et la prédiction évoluent dans le temps.

Afin de concevoir un EWS qui répond aux attentes de l'enseignant, nous nous posons les questions suivantes :

1. Quelles métriques permettent une évaluation approfondie des performances évolutives des EWS ?
2. Comment mesurer la précocité des EWS ?
3. Est-ce que les EWS sont stables dans le temps ?
4. Y aura-t-il un compromis à établir entre la précocité et la stabilité des résultats des EWS et leur précision ?

2.3 Question de recherche 3 : les systèmes de prédiction sont-ils aussi performants sur l'ensemble de la population ?

Le constat des taux élevés d'échec dans l'apprentissage en ligne est dû à de nombreuses raisons notamment la grande hétérogénéité et diversité de profils d'apprenants qui existent dans une même cohorte. En effet, même si les apprenants inscrits dans un programme en ligne partagent le même parcours et le même objectif final, chacun d'entre eux peut avoir des comportements et habitudes d'apprentissage différents. De plus, chaque apprenant en ligne peut exprimer des besoins et attentes différents du reste de la cohorte.
De ce fait, les solutions et notamment les EWS conçus pour la prédiction précoce du risque doivent tenir compte de la diversité des comportements et des profils d'apprentissage existants. Cependant, l'évaluation de ces systèmes n'est pas suffisante puisqu'elle ne tient pas compte des différents comportements et profils d'apprentissage.
Une évaluation du système prédictif par profil d'apprenants est pertinente pour un conseil et un un retour d'informations adapté à la situation académique de chacun des apprenants à risque.

Pour répondre à cet objectif, nous nous posons les questions suivantes :

1. Comment identifier les différents profils d'apprenants qui existent dans une même cohorte ?
2. Comment ces profils sont-ils représentés ?
3. Comment évaluer les performances d'un système prédictif selon les profils d'apprenants identifiés ?

2.4 Question de recherche 4 : comment alerter au plus tôt les enseignants des apprenants à risque ?

Comme son nom l'indique, en utilisant un EWS, l'enseignant s'attend à être alerté automatiquement des apprenants à risque. La génération précoce d'alerte pour les enseignants est le résultat final d'un EWS. Malgré l'importance de l'alerte dans les LA, la littérature ne va pas au-delà de l'identification précoce du risque d'échec et ne propose pas de méthodes particulières pour la génération d'alertes. De ce fait, il serait d'un grand intérêt de proposer un système aux enseignants dont la finalité est l'alerte précoce et précise des apprenants à risque.

Les questions qui se posent à cet égard sont les suivantes :

1. Quand l'alerte peut-elle être générée ?
2. Comment l'alerte est-elle générée ?
3. Comment l'alerte est-elle évaluée et sur la base de quels critères ?

3 Contributions

Les contributions de cette thèse ont fait l'objet des publications suivantes :

— Amal BEN SOUSSIA, Azim ROUSSANALY, Anne BOYER. An in-depth methodology to predict at-risk learners. *European Conference on Technology Enhanced Learning (EC-TEL)*, pages 193-206. Springer 2021.
— Amal BEN SOUSSIA, Chahrazed LABBA, Azim ROUSSANALY, Anne BOYER. Assess Performance Prediction Systems : Beyond Precision Indicators. *International Conference on Computer Supported Education (CSEDU)*, pages 489-496. 2022.
— Amal BEN SOUSSIA, Azim ROUSSANALY, Anne BOYER. Toward An Early Risk Alert In A Distance Learning Context. *IEEE International Conference on Advanced Learning Technologies (ICALT)*, pages 206-208. IEEE 2022.
— Amal BEN SOUSSIA, Célina TREUILLIER, Azim ROUSSANALY, Anne BOYER. Learning Profiles to Assess Educational Prediction Systems. *International Conference on Artificial Intelligence in Education (AIED)*, pages 41-52, Springer 2022.
— Amal BEN SOUSSIA, Chahrazed LABBA, Azim ROUSSANALY, Anne BOYER. Time-Dependent Metrics to Assess Performance Prediction Systems. *International Journal of Information and Learning Technology (IJILT)*. 2022.

Dans la suite de cette section, nous donnons un aperçu plus détaillé de chaque contribution.

3.1 Contribution 1 : Méthodologie en profondeur

Afin de répondre à la première question de recherche (Section 0.2.1) liée à la représentation de l'activité des apprenants inscrits en ligne, nous définissons une méthodologie approfondie basée sur les traces numériques et les techniques de ML. L'objectif de cette méthodologie est à la fois la détermination de la (des) catégorie(e) de données les plus pertinentes corrélées à l'activité d'un apprenant et la conception des meilleurs modèles assurant la prédiction précoce et précise des apprenants à risque.

La particularité de cette méthodologie est expliquée par le processus d'étapes qui la composent depuis le traitement des données brutes jusqu'à la phase de la modélisation et l'évaluation des résultats. En croisant les résultats trouvés dans l'état de l'art avec ceux de l'analyse des données réelles disponibles et les notions des sciences de l'éducation, la méthodologie commence par l'identification des indicateurs d'apprentissage les plus pertinents parmi **la performance, l'engagement, la régularité et la réactivité**. Ensuite, nous commençons à définir les axes de notre approche prédictive. D'une part, nous définissons le type de l'apprentissage automatique (supervisé ou non supervisé) à suivre et qui répond le plus à notre besoin. D'autre part, nous fixons la granularité temporelle pour l'application de notre approche. Les dernières étapes de cette méthodologie sont pour la modélisation et l'évaluation qui permettent d'identifier le modèle de ML le plus adéquat à nos objectifs d'identification des apprenants à risque d'échec.

3.2 Contribution 2 : Évaluation temporelle des systèmes prédictifs

L'évaluation est une étape primordiale pour évaluer un EWS et comparer ses résultats. Dans ce contexte, la 2ème contribution répond à la question de recherche posée dans la section 0.2.2.

Pour répondre aux limites des travaux de la littérature et des mesures de performance existantes, nous proposons des nouvelles métriques dépendantes du temps, dites métriques temporelles, pour évaluer les systèmes prédictifs dans le temps. Ces métriques temporelles sont **la précocité** et **la stabilité** et elles montrent l'importance de la dimension temporelle quand il s'agit d'un système des LA. Pour la précocité et la stabilité, nous définissons respectivement un algorithme et une formule de calcul.

Un EWS est dit efficace lorsqu'il fournit des prédictions précises et précoces. C'est pour cette raison que nous établissons une mesure qui permet l'étude de compromis entre l'exactitude des prédictions et

leur précocité. De plus, nous définissons une autre mesure qui permet d'étudier un 2ème compromis entre la précocité des prédictions d'un système et sa stabilité.

3.3 Contribution 3 : Identification des profils types d'apprenants pour l'évaluation des systèmes prédictifs

Avant de délivrer leurs résultats aux enseignants, les EWS doivent s'assurer d'avoir tenu compte de la diversité et multitude de profils d'apprenants existants dans une même cohorte en ligne. Cette 3ème contribution répond à la question de recherche posée dans la section 0.2.3, en proposant une évaluation plus fine et précise des performances du système prédictif. Elle prend en compte les différents profils d'apprentissage qui existent et se base sur plusieurs métriques temporelles et de précision.

Premièrement, cette méthodologie identifie les différents profils d'apprenants qui existent dans une cohorte. Ces profils sont identifiés grâce à la définition d'un ensemble d'indicateurs d'apprentissage caractérisant les comportements des apprenants. Ensuite, l'exactitude prédictive du système ainsi que sa précocité et stabilité sont évaluées pour chaque profil type d'apprenants.

3.4 Contribution 4 : Génération précoce d'alerte

La génération d'alerte est un axe de recherche très peu développé dans la littérature des LA et rares sont les travaux qui proposent des méthodes particulières d'alerte. Cette contribution répond à la question de recherche 4 posée dans la section 0.2.4 et propose un algorithme permettant la génération automatique des alertes précoces et précises pour les enseignants concernant les apprenants à risque. Par conséquent, l'objectif de ces alertes est de cibler au mieux les apprenants avec des difficultés d'apprentissage.
Cet algorithme d'alerte utilise les notions des métriques temporelles afin de fixer le temps auquel une génération d'alerte est pertinente. Ensuite, il a recours à une notion originale de règle d'alerte pour définir la méthode à suivre pour générer l'alerte en question.
Les alertes générées par cet algorithme sont généralement de deux différents types : bonne alerte et fausse alerte. Afin d'évaluer la pertinence de toute alerte générée, nous adaptons la matrice de confusion habituelle de ML à la problématique d'alerte.

4 Organisation de la thèse

Ce manuscrit de thèse est organisé autour de six chapitres.

Le premier chapitre *"État de l'art"* recouvre les différents travaux et approches de la littérature qui s'apparentent à notre problématique. Il passe en revue aussi toutes les notions et les sous-problématiques traitées dans cette thèse.

Le deuxième chapitre *"Contexte applicatif"* présente le contexte du travail et sa spécificité, le jeu de données exploité et l'analyse de ces données.

Le troisième chapitre *"Méthodologie en profondeur"* présente la première contribution scientifique qui est la méthodologie adoptée et suivie permettant d'identifier les indicateurs d'apprentissage représentant un apprenant en ligne. La partie expérimentale de ce chapitre conclut sur le modèle ML qui prédit au mieux les apprenants à risque.

Le quatrième chapitre *"Évaluation temporelle des systèmes prédictifs"* présente le travail sur la deuxième contribution de cette thèse qui concerne l'évaluation temporelle des performances des modèles ML et étudie les différents compromis entre les métriques temporelles proposées et les métriques de précision.

Le cinquième chapitre *"Identification des profils d'apprentissage pour évaluer les systèmes de prédiction"* traite la diversité des profils d'apprenants existants dans une cohorte et l'importance de leur identification pour une meilleure évaluation des systèmes prédictifs.

Le sixième chapitre *"Génération précoce d'alerte"* s'intéresse à l'importance de la génération d'alerte pour les enseignants.

Nous clôturons ce manuscrit avec une *"Conclusion"* qui fait le bilan sur le travail effectué et présente un ensemble de perspectives liées notamment à sa poursuite.

Chapitre 1

État de l'art

1.1 Introduction

L'apprentissage en ligne est un mode d'enseignement actuellement en pleine croissance. Les plateformes d'apprentissage en ligne telles que les systèmes de tutorat intelligents (*Intelligent Tutoring Systems ITS*), les systèmes de gestion de l'apprentissage (*Learning Management Systems LMS*) et les cours en ligne ouverts et massifs (*Massive Open Online Courses MOOC*) bénéficient de plus en plus des techniques de l'apprentissage automatique et de la fouille de donnés pour développer et créer des systèmes de recommandation, des systèmes de classement automatique, des systèmes adaptatifs, ainsi que des systèmes d'alertes.... Ces plateformes utilisent des outils intelligents qui collectent des informations utilisateur telles que la fréquence d'accès d'un apprenant à l'environnement d'apprentissage, la précision des réponses de l'apprenant aux questions et le nombre d'heures passées à lire des textes et à regarder des tutoriels vidéos [17].

Bien que l'apprentissage en ligne soit largement considéré comme une forme d'enseignement présentant plusieurs avantages (par exemple, moins coûteux en temps et plus flexible que le mode traditionnel [18]), il est toujours considéré comme un environnement d'apprentissage difficile vu qu'il n'y a pas toujours d'interaction directe entre les étudiants et les instructeurs [19]. À ce sujet, un des trois principaux défis associés à l'apprentissage en ligne est les taux d'abandon et d'échec plus élevés que ceux enregistrés par l'apprentissage en présentiel [17].
Afin de faire face à ce problème inhérent à l'éducation en ligne, les travaux de la littérature proposent de multiples méthodologies et solutions différentes dont l'objectif est d'améliorer l'expérience de l'apprentissage des apprenants inscrits en ligne. Ainsi, l'***identification des apprenants à risque*** est l'approche la plus adoptée dans l'état de l'art après celle de la prédiction de performance [17].

L'objectif des chercheurs intéressés par la problématique de l'identification du risque s'oriente vers l'exploitation des traces numériques d'apprentissage générées par les plateformes en ligne pour implanter des méthodes analytiques. Afin de minimiser ces taux d'échec élevés, les parties prenantes du domaine d'éducation et les établissements en ligne coopèrent pour proposer des solutions intelligentes permettant la prédiction précise et précoce des apprenants à risque pour pouvoir intervenir auprès d'eux.

Dans cette perspective, les techniques de l'analytique de l'apprentissage (*Learning Analytics (LA)*) prennent de l'ampleur dans la littérature, en particulier pour la prédiction du risque pour les apprenants inscrits en ligne.

Les projets de l'état de l'art des LA, s'intéressant à la prédiction précoce du risque, s'appuient sur la puissance de la donnée et des modèles de ML pour proposer les *systèmes d'alerte précoce (EWS)* les plus performants. Selon les données disponibles et la nature des résultats attendus, chaque projet suit une stratégie et une (ou des) approche(s) de traitement de la problématique.

Certes chaque projet de LA se distingue par son objectif final et son contexte d'application, mais, il partage avec les autres des étapes clefs, notamment la collecte des données, la modélisation et l'évaluation des résultats de tout le système prédictif à proposer. Ce sont les étapes majeures dans tout projet de LA se basant sur le ML.

Dans ce chapitre, nous passons en revue toutes les définitions liées à la problématique de **la prédiction précoce et précise du risque chez les apprenants en ligne**. Ensuite, nous présentons les solutions de l'état de l'art pour répondre aux différentes questions de recherche posées dans l'introduction de ce manuscrit. Ce travail nous permettra de discuter sur l'existant.

La Figure 1.1 présente les éléments de l'état de l'art sur lesquelles nous allons mettre l'accent ainsi que l'arborescence que nous suivons pour les exposer. À travers cette Figure, nous suivons un schéma composé de deux parties majeures : théorique et automatique. La 1ère partie est dédiée aux notions issues des sciences de l'éducation et liées à notre problématique. La 2ème partie expose les travaux des LA de l'état de l'art qui s'approchent de notre problématique. Chacune de ces parties sera accompagnée des discussions sur les limites de l'existant et les axes de recherche peu développés, à notre connaissance, et dont l'exploration pourrait permettre une amélioration des résultats obtenus.

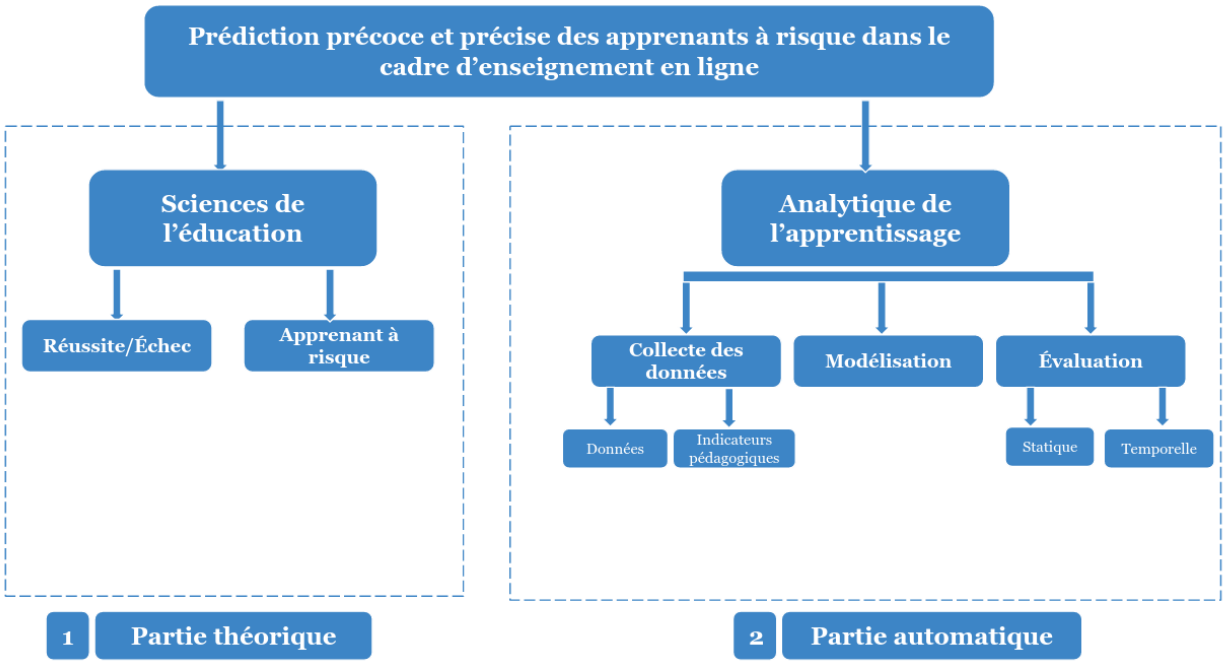


FIGURE 1.1 – Schéma représentatif des parties de l'état de l'art

1.2 Notions issues des sciences de l'éducation

Dans cette section, nous présentons les définitions liées à la réussite ainsi qu'à l'apprenant à risque selon les experts et les revues des sciences de l'éducation.

1.2.1 Définition de la réussite

Selon le Conseil Régional de Prévention de l'Abandon Scolaire du Saguenay–Lac-Saint-Jean au Québec (CRÉPAS) [2], le Centre de Transfert pour la Réussite Éducative du Québec (CTREQ)[20] ainsi que

2. https ://crepas.qc.ca/

l'Institut Français de l'Éducation (IFÉ)[21], nous distinguons deux types de réussite :

- **Réussite scolaire :** c'est l'achèvement d'un parcours scolaire avec succès et l'atteinte des objectifs d'apprentissage liés à la maîtrise des savoirs. Selon le CRÉPAS, ce terme est porteur d'une idée de rendement et de performance.
- **Réussite éducative :** c'est un concept multidimensionnel et plus vaste que la réussite scolaire. La réussite éducative concerne l'instruction, la socialisation et la qualification. Selon [22], « comme processus, la réussite éducative est l'ensemble des initiatives prises et des actions mises en œuvre par ses parents, son entourage ou des professionnels pour permettre à l'enfant ou à l'adolescent de se rapprocher et d'atteindre la réussite éducative, et la progressive appropriation par l'intéressé de ce qui lui est fourni ».

Étant donné l'objectif de notre travail et les définitions ci-dessus, nous pouvons constater que nous nous intéressons plutôt à la réussite scolaire dans cette thèse (l'opposé étant l'échec scolaire). Par conséquent, un apprenant qui n'a pas atteint les objectifs pédagogiques fixés au préalable est considéré en échec scolaire. Ce dernier se caractérise par un ensemble de traits de comportements d'apprentissage pendant toute l'année scolaire / universitaire. Cela nous amène à nous interroger sur : *ce qu'est un apprenant à risque d'échec ?* et *quelles sont ses caractéristiques ?*

1.2.2 Notion d'apprenant à risque

Selon [23], « un élève à risque est un jeune qui démontre certaines difficultés d'apprentissage ou manifeste des comportements qui sont susceptibles de l'empêcher d'atteindre les objectifs d'apprentissage et de socialisation poursuivis par l'école ».

La Fédération des syndicats de l'enseignement définit les élèves à risque dans [24] comme suit : ce sont « des élèves du préscolaire, du primaire et du secondaire qui présentent des facteurs de vulnérabilité susceptibles d'influer sur leur apprentissage ou leur comportement peuvent ainsi être à risque, notamment au regard de l'échec scolaire ou de leur socialisation, si une intervention rapide n'est pas effectuée ». De plus, selon cette même fédération, les élèves à risque ont les caractéristiques suivantes :

- sur le plan d'apprentissage ([25], [26], [27], [23]) : manque d'autonomie, des capacités intellectuelles normales, faible performance scolaire, manque d'engagement et de persévérance dans les activités…
- sur le plan comportemental ([26], [28], [23]) : avoir des difficultés liées à une cause ponctuelle (événement personnel, problèmes familiaux, toxicomanie, etc), comportements perturbateurs ou non conformes aux règles établies.

La littérature mentionnée ci-dessus est le résultat des études menées par des experts, centres et associations qui s'intéressent aux problématiques des élèves à risque et du décrochage scolaire. Certes ce sont les résultats des études menées dans le cas des établissements scolaires traditionnels, mais ce sont aussi une base pour adapter les indicateurs des élèves à risque et les profils dominants dans un contexte d'*e-learning*.

Nous retenons que ces études convergent sur certains indicateurs du risque, comme l'engagement, la (sous-)performance, les (bons/mauvais) résultats, la persévérance, la participation, la présence, le comportement, l'autonomie.

1.2.3 Apprenant en ligne à risque

Un apprenant à risque est décrit comme un apprenant avec une forte probabilité d'abandonner le cours en raison d'un échec scolaire, de décrochage scolaire ou d'une expulsion pour des raisons comportementales [29].

Selon [30], un apprenant en ligne à risque ne s'attend pas à réussir ; il abandonne tôt dans l'année ou il finit par échouer dans le cours. En effet, un apprenant en ligne à risque reçoit généralement un renforcement intermittent et incohérent pour ses accomplissements personnels et a tendance à montrer des degrés de persévérance inférieurs à ceux de ses pairs [31]. De plus, la procrastination est un autre problème majeur de l'apprentissage en ligne. Ainsi, les étudiants potentiellement à risque dans l'enseignement en ligne cumulent les devoirs jusqu'à la dernière minute. De ce fait, il n'est pas surprenant que ces apprenants soient à risque et aient des faibles résultats [32] [33].

D'autres facteurs sont aussi liés aux apprenants en ligne à risque et ils comprennent la dyslexie, la faible estime de soi, les faibles compétences en technologies de l'information et de la communication (TIC), le manque d'orientation et de soutien du tuteur, la rareté de la connexion et de la communication, la répétition d'un module après un échec et l'absence d'une évaluation complète et adaptée [31]. De plus, [34] a constaté que les étudiants à risque changent fréquemment d'environnement d'étude, tout en démontrant une motivation plus faible, moins de confiance en l'ordinateur, moins d'encouragement à suivre le cours et des responsabilités inattendues. Selon [34], ces facteurs sont la principale raison pour laquelle les apprenants en ligne à risque ont un taux d'échec et d'abandon nettement élevé.

Pour réussir dans les cours en ligne, les apprenants doivent démontrer certaines caractéristiques, notamment la gestion du temps, la motivation, la participation active, l'apprentissage indépendant, l'efficacité technologique, la communication et l'intégrité [35]. Étant donné le nombre croissant de cours dispensés en ligne, les établissements d'enseignement supérieur devraient aider les étudiants à se préparer à ce nouveau format d'enseignement. À cette fin, l'identification des étudiants qui pourraient ne pas être préparés à l'apprentissage en ligne est une première étape nécessaire [31].

1.2.4 Discussions

Les notions et définitions des sciences de l'éducation sont plus disponibles lorsqu'il s'agit d'un enseignement dans un établissement scolaire/universitaire traditionnel. Ce mode d'apprentissage se caractérise de celui en ligne par la présence physique et la possibilité du contact direct entre l'enseignant et l'apprenant ; ce qui permet une meilleure perception des comportements des apprenants par les enseignants.
Selon les définitions données dans la Section 1.2.3, ces notions restent, en grande majorité, valables pour le cadre d'un enseignement en ligne avec des ajustements pour s'adapter à ce mode d'apprentissage. Nous remarquons que les définitions données par les Sections 1.2.2 et 1.2.3 se croisent au niveau de certaines caractéristiques de l'apprenant à risque telles que : les faibles performances académiques, le manque de motivation, l'absentéisme, les faibles taux de participation aux activités du cours. . ..

Lorsqu'il s'agit d'un apprentissage en ligne, et compte-tenu de l'absence de la communication directe entre les apprenants et les enseignants et dans un contexte d'échange virtuel via une interface numérique, les notions de réussite/échec scolaire, profil d'apprenants à risque . . .prennent d'autres dimensions.
Contrairement à l'apprentissage sur les bancs de l'école, les institutions en ligne et entièrement à distance ou les MOOC n'imposent pas à leurs apprenants un emploi du temps bien précis à respecter. Chacun de ces derniers suit son propre rythme pour achever son programme pédagogique. C'est pour cette raison que la Section 1.2.3 met l'accent sur l'importance de la gestion du temps pour que les apprenants en ligne réussissent. De plus, dans le contexte d'enseignement en ligne, le changement de comportement n'est pas facilement détecté vu l'absence de la présence physique. En effet, les seules informations disponibles sur les apprenants inscrits dans ce mode d'apprentissage sont leurs traces numériques sur les différentes plateformes éducatives. C'est l'exploitation de ces traces, en se basant sur les nouvelles technologies et les notions d'analyse des données, qui participera à la résolution des problèmes éducatifs notamment celui des taux élevés d'échec.

1.3 Analytique de l'apprentissage

Dans cette partie, nous présentons en détails les notions sur lesquelles s'appuie l'analytique de l'apprentissage ainsi que ses outils et ses différentes utilisations dans le domaine de l'éducation.
L'analytique de l'apprentissage est un ensemble d'étapes permettant de comprendre et d'améliorer le processus d'apprentissage, ainsi que l'environnement dans lequel il se déroule en utilisant des instruments et des outils d'analyse [36]. C'est un domaine de recherche émergent et qui couvre plusieurs disciplines, bien établies. L'analyse de l'apprentissage s'appuie, en effet, sur les théories et les méthodes issues de l'apprentissage automatique et des sciences des données, des sciences de l'éducation, de la psychologie cognitive, des statistiques, de l'informatique, des neurosciences et des sciences sociales et de l'apprentissage et bien d'autres [37] [36].

L'analytique de l'apprentissage diffère des analyses de l'éducation plus traditionnelles à plusieurs égards. Premièrement, en raison de sa forte concentration quantitative, la taille des ensembles de données exploitables a tendance à être grande, ce qui permet un plus grand niveau de confiance dans la généralisabilité des résultats [38]. Deuxièmement, comme les données sont principalement collectées à partir des systèmes techniques, il existe un niveau de granularité très fin des variables disponibles qui ne peut pas être capturé par des études d'observation, des entretiens ou des auto-rapports [38]. Enfin, les données tendent à être longitudinales. Autrement dit, la manière dont les données sont collectées et les processus utilisés pour la collecte permettent d'inclure une forte dimensionnalité temporelle dans les études de recherche [38].
De ces faits, le LA est un domaine particulier qui se caractérise par plusieurs axes dont les plus importants sont les données à collecter, le problème à résoudre, les acteurs concernés ….. En réponse à ceci, "Chatti" dans [39] propose un modèle de référence pour les LA reposant sur quatre dimensions :

- ***quoi ? :*** quel type de données le système collecte-t-il ? gère-t-il ? et utilise-t-il pour l'analyse ? cette première dimension considère l'approche basée sur les données à la fois dans les systèmes centralisés et distribués.
- ***qui ? :*** qui est la cible de l'analyse ? cette dimension prend en compte tous les acteurs de l'environnement d'apprentissage : étudiants, enseignants, tuteurs et établissements d'enseignement. Chacun de ces derniers est engagé dans son rôle pour améliorer le processus d'apprentissage et minimiser les apprenants à risque.
- ***pourquoi ? :*** pourquoi le système analyse-t-il les données collectées ? c'est la dimension qui aborde les objectifs du processus des LA : suivi et analyse, prédiction et intervention, tutorat et mentorat, évaluation et rétroaction, recommandation ….
- ***comment ? :*** comment le système effectue-t-il l'analyse des données collectées ? cette dernière dimension prend en compte les techniques pour détecter les modèles intéressants existants dans les données éducatives telles que les statistiques, la visualisation de l'information, l'exploration de données, la classification, le clustering, les règles d'association ….

Comme montré par la Figure 1.2, [1] propose le cycle de vie des LA composé essentiellement de quatre grandes parties : *l'environnement d'apprentissage* où les acteurs produisent les données, *les grands volumes de données* qui composent des jeux de données massives, les différentes *techniques d'analyse* et *d'action* où les objectifs sont atteints pour optimiser l'environnement d'apprentissage. Nous présentons chaque entité du cycle :

- acteurs : il existe différents groupes qui sont engagés dans les LA. Chaque groupe peut obtenir des avantages en fonction de ses visions et missions [39]. **Dans cette thèse, nous nous intéressons aux LA pour les enseignants. Ces derniers ont recours aux outils des LA pour améliorer leurs méthodes d'enseignement et pour avoir un retour en temps réel sur les activités de leurs apprenants** [1].

— bases de données : les apprenants laissent beaucoup de données derrière eux quand ils utilisent n'importe quel environnement d'apprentissage. Ces données peuvent être de plusieurs types : interactions avec l'environnement d'apprentissage en ligne (espaces de discussions, téléchargement de ressources d'apprentissage...), notes ....
— techniques : il existe différentes méthodes pour analyser les données dans l'éducation. Ces méthodes d'analyse cherchent à découvrir des modèles intéressants cachés dans les ensembles de données éducatives. Les techniques d'analyse de l'apprentissage utilisent divers types de méthodologies analytiques. Ces méthodologies se divisent en deux catégories principales : Analyse quantitative (analyse statistique, visualisation, analyse quantitative des réseaux sociaux) et qualitative (intelligence émotionnelle, analyse qualitative des réseaux de neurones). Dans cette thèse, nous appliquons des méthodes d'analyse statistique pour la découverte du contexte applicatif (Voir chapitre 2).
— objectifs : le résultat de l'analyse est interprété pour atteindre les objectifs des LA. Parmi ces objectifs, nous citons la prédiction, l'intervention, la recommandation, la personnalisation, le Benchmarking .... Dans cette thèse, nous nous intéressons à **la prédiction** qui est l'exploration d'une valeur numérique/continue inconnue telle que la performance, les connaissances, le score ou la note [40]. Grâce à la prédiction, les activités de l'apprenant et les performances futures peuvent être révélées.

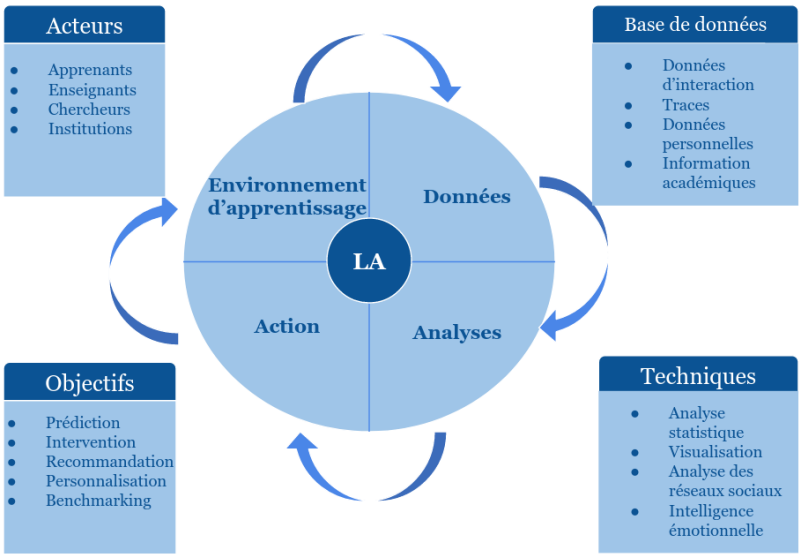


FIGURE 1.2 – Cycle de l'analytique de l'apprentissage - Figure reproduite de [1]

Les intentions de recherche nécessitent des combinaisons interdisciplinaires liées à la fois à la compréhension et à l'optimisation du processus d'apprentissage [37]. Dans ce contexte, [36] définit les cinq étapes suivantes pour un projet de LA dont l'objectif est la prédiction :

— **collecte :** c'est l'étape de la collecte de données et la réponse à quelques questions concernant ces dernières comme : quelle est la fréquence de la collecte ? où les données vont-elles être stockées ? sous quel format vont-elle être stockées et puis utilisées ? ...
— **reporting :** cette étape aborde le flux d'informations collectées lors de la première étape dans le but de produire des rapports sur les données. Les questions auxquelles l'étape de reporting devrait répondre sont les suivantes : qui recevra les rapports et à quelle fréquence ? quel type d'information doit être signalé et comment les rapports seront accessibles ?

— **prédiction :** c'est l'étape qui concerne les algorithmes de prédiction : leur précision, leur prédictions particulières, leurs données d'entrée et finalement quelles prédictions seront rapportées aux parties prenantes.
— **action :** c'est l'étape qui concerne les actions à déployer dans l'environnement d'apprentissage pour améliorer son processus global.
— **raffinement :** c'est l'étape de l'analyse des résultats obtenus des étapes précédentes, y compris leur utilité.

Les enseignants des institutions en ligne sont les acteurs concernés par les travaux de cette thèse. Chacune des sections de la suite de ce chapitre apporte une / des réponses(s) aux questions restantes du modèle de Chatti [39]comme suit :
— Pourquoi ? : la prédiction des apprenants à risque d'échec à travers les EWS.
— Quoi ? : les données utilisées dans les projets des LA de la prédiction du risque.
— Comment ? : les modèles et les algorithmes des projets des LA pour modéliser le risque d'échec.

Nous rajoutons au cycle de la Figure 1.2 un point supplémentaire qui expose les méthodes et métriques d'évaluation des performances des LA.

1.4 La prédiction précoce et précise du risque d'échec

Face aux taux élevés d'échec et d'abandon enregistrés annuellement, en particulier dans le contexte des enseignements en ligne, la prédiction précoce du risque d'échec et/ou d'abandon est un sujet d'actualité des LA.
Les chercheurs scientifiques de ce domaine s'intéressent, de leur côté, à proposer des moyens et solutions pour que cette prédiction s'effectue au plus tôt dans l'intervalle du temps d'apprentissage et avec une précision maximale.
Étant donné la variété de facteurs pouvant influencer le statut final d'un apprenant, la prédiction précoce se présente comme une tâche difficile dans le domaine des LA [2]. C'est une question cruciale en éducation car elle concerne de nombreux élèves de tous les niveaux (primaire, secondaire, supérieur) et dans les écoles et les universités du monde entier [2].
Dans cette section, nous répondons à la question du modèle de Chatti qui concerne les objectifs de l'utilisation des LA. En effet, nous présentons les définitions de l'état de l'art pour la prédiction précoce et exposons les résultats de la littérature sur la précocité de prédiction du risque. Ensuite, nous présentons les systèmes d'alerte précoce comme outil des LA. Enfin, nous discutons sur l'existant.

1.4.1 Définition

Selon [41], la prédiction précoce est définie comme l'application des modèles prédictifs utilisant des variables clés pour prédire avec précision l'échec ou le décrochage des élèves le plus tôt possible.

La détection au plus tôt des apprenants à risque est un moyen pour permettre aux enseignants une intervention et/ou une prise en charge précoce et efficace des élèves à risque et les aider à réussir leurs cours [2]. Selon [42], la prédiction est une tâche essentielle pour identifier les apprenants à risque, et donc, mettre en place des stratégies appropriées et efficaces de conseil et de recommandation des actions correctives.

1.4.2 La précocité dans l'état de l'art

Dans l'état de l'art, plusieurs travaux cherchent à trouver une mesure assurant à la fois une prédiction précise et une intervention efficace.

En se basant sur l'importance de la première évaluation, cette étude [43] introduit une nouvelle approche dont l'objectif est l'identification précoce des apprenants à risque. Cette approche s'appuie sur l'extraction des comportements des apprenants ayant déjà soumis leur première évaluation plus tôt que leurs pairs.

Pour identifier au plus tôt les étudiants à risque d'abandon inscrits dans un cours en ligne, [44] divise la période d'apprentissage en six périodes : début du cours, 20%, 40%, 60%, 80% et 100% de la période totale du cours. Pour comparer la performance des algorithmes de ML, ces derniers ont été entraînés et testés à différents pourcentages de la durée du cours. Les résultats montrent qu'à 20% de la longueur du parcours, le modèle des forêts aléatoires donne des résultats prometteurs avec un pourcentage d'*exactitude* (accuracy) [3] égal à 79%. À 60% de la longueur du parcours, les performances de ce même modèle sont significativement meilleures. Les résultats du modèle prédictif ont démontré une efficacité dans la prédiction la plus précoce possible de la performance des élèves à risque. De telles études peuvent aider les administrateurs et les instructeurs dans la formulation du cadre d'apprentissage en ligne, donc contribuer au processus de prise de décision [44].

Cette étude [45] présente une approche simple pour identifier les apprenants à risque d'abandon ou d'échec avant la moitié du semestre. Les résultats des expérimentations réalisées montrent que le principal facteur pour réussir l'examen final est l'apprentissage continu acquis pendant au moins la première moitié du cours.

Les travaux menés à l'université ouverte du Royaume-Uni prouvent que la première évaluation est un bon prédicteur du résultat final d'un apprenant [12].

1.4.3 Systèmes d'alerte précoce

Les systèmes d'alerte précoce sont surtout connus pour leur utilisation pour générer des alertes sur les catastrophes naturelles [46] et donc réduire leur impact. Lors de la détection des signes indiquant qu'une catastrophe pourrait se produire, les membres de la population potentiellement touchée sont alertés et reçoivent des instructions pour prévenir ou minimiser les dommages [47]. Cependant, ces systèmes ont été exploités dans une variété d'autres contextes. Par exemple, ils sont utilisés dans les environnements financiers pour prévoir les ralentissements économiques à un stade précoce et offrir de meilleures opportunités pour atténuer leurs effets négatifs [48]. Dans le domaine de la santé, les EWS sont utilisés par les équipes de soins hospitaliers pour reconnaître les premiers signes de détérioration clinique, permettant d'initier une intervention et une prise en charge précoces [49]. Le domaine de l'éducation a aussi pris sa partie de ces systèmes et les a exploité pour proposer des solutions à ses problèmes les plus inhérents, notamment la prédiction du risque d'échec. À travers ces systèmes, les chercheurs en éducation modélisent cette problématique avec l'objectif de générer des alertes au plus tôt concernant les apprenants à risque pour assurer une intervention efficace.

Les projets de LA proposent plusieurs types d'outils en faveur des différents acteurs de l'enseignement : universités, enseignants, apprenants, administrations. Ces outils sont employés dans plusieurs contextes et ont différents objectifs éducatifs et pédagogiques :

— alerter les apprenants et les enseignants des éventuels problèmes de performance et d'apprentissage.
— identifier les apprenants en besoin du soutien pédagogique et suivi académique.

3. C'est une mesure qui évalue l'exactitude des prédictions.

— prédire les comportements futurs des apprenants.
— recommander des ressources ou des activités appropriées pour les apprenants à risque.

Au cours du traitement du problème des taux élevés d'échec et après avoir prédit les apprenants à risque d'échec, les enseignants ont besoin d'un outil technique qui les alerte de l'existence de telles situations académiques défaillantes pour pouvoir intervenir efficacement.
Dans cette perspective, les systèmes d'alerte précoce sont communément utilisés dans la littérature pour répondre à cette demande.

1.4.3.1 Définition et objectifs

Un système d'alerte précoce est un outil utilisé par les éducateurs pour identifier les apprenants à risque d'échec dans un cours ou abandonner un programme [50]. Les EWS sont utilisés pour mieux accompagner les élèves en fonction de leur comportement et de leur performance [16]. En effet, les EWS sont capables d'anticiper les risques potentiels en s'appuyant sur les informations présentes et d'envoyer ainsi des alertes à la personne ou au groupe de personnes pouvant être concerné(e) par ces risques et/ou capable de les prévenir [46].
Les EWS se servent de plusieurs sources de données et des informations historiques et en temps réel pour analyser le profil académique des apprenants et détecter ceux qui sont défaillants au bon moment [16].

1.4.3.2 Exemples

La littérature est riche d'exemples d'EWS ayant pour objectif la prédiction précoce des apprenants à risque.

Un des exemples piliers dans le domaine des LA est celui de l'Université de Purdue connu sous le nom de « Course Signals (CS) ». Cet EWS a introduit un système de feux de circulation permettant aux étudiants de voir leur niveau de réussite/échec selon un code de couleur ; par exemple, la couleur verte indique une forte probabilité de réussite [51].

Un autre EWS bien connu dans l'état de l'art est le « Student Explorer» [52]. Cet EWS explore les données d'effort et de performance extraites du LMS afin d'évaluer la probabilité d'échec scolaire des apprenants. Ces derniers sont classés en trois classes par ordre croissant du risque. Ainsi, les conseillers pédagogiques utilisent ces informations pour prendre des mesures correctives.

Dans le but de réduire le décrochage scolaire et minimiser les retards d'obtention du diplôme, [53] propose la conception d'un EWS appliqué dans le contexte d'une université en Chine. Cet EWS classe les étudiants en fonction du type du risque auquel ils sont exposés comme : le risque d'obtenir de mauvaises notes, un retard d'obtention du diplôme, le risque de décrochage.

En se basant sur des données de lecture et dans le but d'identifier les étudiants à risque d'échec scolaire, [54] propose un EWS destiné à l'exploitation dans le cadre de l'utilisation des livres électroniques (e-book) comme matériel d'apprentissage. Chaque semaine, une analyse est effectuée pour étiqueter les élèves comme étant peu performants ou très performants, en essayant plusieurs algorithmes de prédiction.

[55] propose un EWS utilisant les traces d'activité des étudiants sur le LMS. Ce travail suit une approche mensuelle pour détecter des baisses d'activité significatives chez un élève, qui sera par la suite signalé comme à risque.

En utilisant les techniques d'exploration des données (ou en anglais *data mining*), [56] analyse les données d'interaction dans un cours en ligne pour développer un système d'alerte précoce pour prédire les performances d'apprentissage des apprenants dans ce même cours. Les auteurs de ce travail se sont

appuyés sur divers modèles de classification pour identifier le statut de réussite/échec de ces étudiants en utilisant les données des semaines 4, 8 et 13 d'apprentissage.

1.4.4 Discussions et Limites

La prédiction précoce et précise des apprenants à risque d'échec est une problématique importante dans l'état de l'art des LA. C'est l'une des approches les plus adoptées pour trouver des solutions au problème des taux d'échec élevés enregistrés dans l'enseignement en ligne. En suivant cette approche, les travaux de l'état de l'art confirment que la prédiction précoce permet une intervention constructive en laissant le temps à l'enseignant pour intervenir et à l'apprenant pour travailler sur ses défaillances. Plus explicitement, la précocité est une mesure du temps qui assure à la fois une prédiction précise des résultats des apprenants et une intervention efficace auprès de ceux qui sont à risque. Dans la majorité des résultats de l'état de l'art, le premier devoir (sa soumission ou pas, la note obtenue ...) s'avère décisif pour les résultats finaux. De plus, la première partie de l'année (trimestre par exemple) est aussi une période suffisante pour la prédiction précoce et précise.
Les systèmes d'alerte précoce sont bien exploités dans l'état de l'art et ils montrent des résultats prometteurs lorsqu'ils sont utilisés pour la détection précoce des apprenants à risque et des comportements défaillants d'apprentissage.

Cependant, en parlant de la précocité, les travaux de l'état de l'art ne proposent pas de mesures qui qualifient et / ou quantifient cette notion. Ils se restreignent à donner juste des indications temporelles et donc ne mesurent pas exactement la précocité des bonnes prédictions d'un système. Certes, ces indications sont importantes dans le processus d'apprentissage, mais sont-elles suffisantes pour qualifier la précocité ? De plus, ces indications temporelles données comme des résultats sur la précocité ne représentent pas forcément le temps d'apprentissage approprié pour une prédiction précise et donc une intervention efficace.
En bref, bien que la prédiction précoce du risque prenne de l'ampleur dans l'état de l'art, les études sur la notion de la précocité est un axe de recherche très peu développé.

Alerter l'enseignant des cas des apprenants défaillants pour qu'il intervienne auprès d'eux au bon moment est la définition même d'un EWS. Contrairement à cette définition et à ce que leur nom indique, les EWS de la littérature ne proposent pas de méthodes particulières pour la génération d'alertes. Ils s'arrêtent, en effet, à la phase de la modélisation prédictive et ne vont pas au-delà de l'identification des apprenants à risque. En utilisant les EWS éducatifs, les enseignants s'attendent à des générations d'alertes leur permettant de se diriger directement vers les apprenants défaillants.

1.5 Les données utilisées dans les projets des LA

Afin d'atteindre les objectifs de tout projet LA, notamment celui de la prédiction du risque, la 1ère étape est la collecte des traces numériques des apprenants ainsi que toutes leurs informations. Dans cette partie, nous présentons les données utilisées dans l'état de l'art pour représenter les comportements des apprenants en ligne. Nous passons en revue aussi les indicateurs d'apprentissage les plus abordés et identifiés dans les projets des LA qui travaillent sur la définition de l'activité de l'apprenant.
À travers cette partie, nous répondons à la question du modèle de Chatti qui concerne les données à utiliser dans les LA.

1.5.1 Les données et les traces numériques utilisées dans la littérature

Les LA promettent la compréhension et l'optimisation de l'apprentissage et son environnement. Pour cette raison, les solutions des LA se basent sur des données extraites de plusieurs sources d'apprentissage.

[57] rapporte les différents types de sources de données exploitées dans les travaux des LA. Les systèmes de gestion d'apprentissage ou les *LMS* arrivent en tête de la liste de ces sources de données. De plus, nous trouvons les systèmes d'informations sur les étudiants, les MOOC, les listes de présence, les résultats d'examens, les systèmes d'alerte, les outils de *eye tracker*, les tests de performance, les plateformes adaptatives …. Pour des campus encore plus avancés, une étude plus récente [58] montre l'importance des sources de données basées sur l'Internet des objets ou en anglais *Internet of Things (IoT)* (les capteurs par exemple) et des réseaux sociaux dans la génération des données sur les sentiments des apprenants.

Toutes ces sources fournissent des données très variées qui pourraient être exploitées de différentes manières dans les travaux des LA. Selon l'angle d'observation, ces données peuvent être catégorisées selon plusieurs critères.
[59] présente une étude d'un nombre assez important d'articles scientifiques qui a abouti à relever 200 données importantes pour les LA. Ces données ont été classées en six différents groupes comme le montre le tableau 1.1 :

Nom du groupe	Question à laquelle la donnée répond	Exemples
Lié à l'apprenant	Qui sont les apprenants ?	Compétences, âge, cours déjà suivis, note d'entrée à l'université
Lié à l'action	Que font-ils en apprenant ?	Nombre de visites d'une page, nombre de téléchargements, temps passé sur les tâches, temps de connexion, Agenda
Lié au contenu	Quel est le contenu impliqué dans leur apprentissage ?	Analyse des sentiments des messages, sujets abordés et omis dans le rapport
Lié au résultat	Quel est le résultat de leur apprentissage ?	Note moyenne, distribution des notes dans un groupe
Lié au contexte	Dans quel contexte l'apprentissage a-t-il lieu ?	Localisation géographique
Donnée sociale	Comment interagissent-ils avec les autres pendant l'apprentissage ?	Réseau montrant la communication avec les autres dans un forum de groupe

TABLE 1.1 – Groupes de données

[29] étudie plusieurs travaux de LA et classe les données exploitées pour la prédiction des apprenants à risque en six groupes comme suit :

— données sur le comportement d'apprentissage : nombre d'actions faites sur le LMS, la fréquence de visualisation des activités, informations sur la présence dans le LMS, nombre de clics et de téléchargements et nombre de jours pendant lesquels l'apprenant a été actif.
— données des réseaux d'apprentissage : nombre de posts, de réponses et de commentaires dans les forums de discussion…
— données du niveau d'apprentissage : notes des examens, résultats finaux, résultats des quizs et tests, résultats de mi-parcours…
— données d'apprentissage émotionnel : évaluation non-cognitive, auto-efficacité, niveau d'auto-évaluation…
— autres : données démographiques, caractéristiques des apprenants….

Afin de prédire les apprenants à risque, [17] divise les données utilisées dans l'état de l'art en deux types comme suit :

— données dynamiques : ce sont des données recueillies au fur et à mesure que les apprenants

interagissent avec les environnements éducatifs [17]. Les actions d'interaction des apprenants avec les plateformes d'apprentissage en ligne sont un exemple de données dynamiques car les caractéristiques de l'ensemble de données changent avec le temps.

— données statiques : ce sont des données acquises une seule fois et qui ne changent pas de valeurs dans le temps. Les informations d'inscription des apprenants et leurs données démographiques sont des exemples de données statiques [17].

À partir de ces différentes classifications, nous concluons que les traces numériques des apprenants en ligne peuvent être de plusieurs types dont chacun représente une dimension de l'apprentissage en ligne. De ce fait, l'ensemble de données qui représentent l'activité des apprenants en ligne est hétérogène.

1.5.2 Les données temporelles

L'apprentissage est un processus qui se déroule au fil du temps : l'apprenant développe des compétences et une compréhension de son environnement et change de perspectives au cours du temps. Ainsi, l'étude de la nature temporelle de l'apprentissage et des traces numériques d'apprentissage est au cœur de la recherche et des applications de l'analytique de l'apprentissage [60].

[61] classifie les caractéristiques temporelles, souvent considérées ensemble, en deux groupes :

— la 1ère catégorie concerne les caractéristiques du passage du temps, en considérant des questions telles que à quelle fréquence ou pendant combien du temps une action se produit, ou à quelle vitesse cette dernière change au fil du temps. Par exemple, à un niveau de base, nous pouvons examiner le temps qu'un apprenant consacre à des tâches d'apprentissage et quand un comportement d'apprentissage se produit.

— la 2ème catégorie des caractéristiques temporelles concerne l'ordre temporel, en considérant les questions sur la façon dont les événements ou les états sont organisés séquentiellement. Il est important de noter que les analyses axées sur l'ordre temporel ignorent généralement les différences de durée entre les événements, et peuvent donc présenter une image très différente de l'activité d'apprentissage.

[56] propose un système d'alerte précoce pour prédire les performances des élèves à risque dans un environnement d'apprentissage en ligne. L'un des objectifs de ce travail est la détermination des impacts des données dépendantes du temps sur les performances des algorithmes. Les résultats obtenus indiquent la nécessité de ce type de données comme facteur essentiel pour déterminer la performance des apprenants dans les LMS.

[62] présente un modèle basé sur la classification pour prédire la performance des apprenants en ligne et particulièrement de ceux qui sont à risque. Pour ce faire, deux types de caractéristiques ont été collectées : comportementales (dynamiques dans le temps) et d'apprentissage (statiques). En sélectionnant un type de caractéristiques ou combinant les deux types, plusieurs expérimentations ont été menées pour tester la performance de prédiction du classifieur. Les résultats montrent que les caractéristiques comportementales (dynamiques) et d'apprentissage (statiques) combinées permettent d'atteindre de meilleurs résultats qu'en utilisant chaque type de caractéristiques seul.

Afin de prédire au plus tôt les apprenants à risque, [63] montre l'importance des caractéristiques temporelles pour améliorer les performances des modèles testés et surtout avoir des prédictions correctes plus tôt.

Ainsi, nous concluons sur l'importance des caractéristiques temporelles pour représenter l'activité des apprenants en ligne. Ce type de caractéristiques impacte positivement les prédictions des modèles.

1.5.3 Indicateurs d'apprentissage

Sébastien Iksal [64] définit un indicateur comme suit : « Un indicateur est un observable signifiant sur le plan pédagogique, calculé ou établi à l'aide d'observés, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un Environnement Informatique pour l'Apprentissage Humain (EIAH). Il est défini en fonction d'un objectif d'observation et motivé par un objectif pédagogique. »
Dans le domaine de l'éducation, les indicateurs d'apprentissage ont pour objectif de définir les activités des apprenants. Chaque indicateur est un ensemble de données définissant un axe de l'apprentissage. [65] passe en revue plusieurs indicateurs d'apprentissage communs entre les travaux des LA traitant des problématiques d'apprentissage traditionnel et en ligne. Dans la liste citée, nous trouvons l'indicateur sur le comportement, la cognition, la méta-cognition, les émotions, l'interaction, la performance, l'engagement ….
Afin de proposer un modèle de prédiction de décrochage, [66] propose de suivre l'activité de leurs apprenants suivant, en plus des indicateurs démographiques, trois grandes classes d'indicateurs pédagogiques : comportement, motivation et performance. [66] divise les indicateurs de comportement en trois sous-groupes : autonomie, persévérance et engagement.
Dans les sous-parties qui vont suivre, nous introduisons des indicateurs communs dans l'état de l'art [65] [67] qui sont : la performance, l'engagement et la régularité.

1.5.3.1 L'indicateur de la performance

Dans l'apprentissage traditionnel comme dans celui en ligne, la *performance* académique est traduite par les notes dans les examens, les moyennes obtenues, les résultats dans les quizzs et les tests d'évaluation continue. Cet indicateur d'apprentissage est communément utilisé par les enseignants pour positionner leurs apprenants et évaluer leurs résultats partiels et/ou finaux. Ce critère d'évaluation est le plus répandu et quasiment le plus utilisé par les enseignants.
Dans l'état de l'art des travaux des LA cherchant à prédire les apprenants à risque, les données qui concernent les notes, les examens, les tests sont les premières à être sélectionnées pour présenter l'apprentissage de l'apprenant et surtout prédire son résultat final [12] [68].

Certes, l'apprentissage se caractérise par la disponibilité d'autres types de données qui pourraient avoir plus ou moins la même importance que celles des notes mais la performance reste un indicateur d'apprentissage très pertinent pour évaluer les performances des apprenants et surtout détecter ceux qui sont les plus défaillants. C'est un indicateur qui a une étroite relation avec la notion de la réussite scolaire (introduite dans la Section 1.2.1)

1.5.3.2 L'indicateur de l'engagement

L'engagement est un autre indicateur d'apprentissage dont l'utilisation est de plus en plus répandue dans les travaux des LA [69] [70] [71] [12]. L'engagement ou l'implication dans l'apprentissage a été initialement défini comme la quantité d'énergie physique et psychologique que l'étudiant consacre à l'expérience académique [72]. Dans [71], les auteurs définissent l'engagement comme la qualité des efforts que les élèves eux-mêmes consacrent à des activités à vocation éducative qui contribuent directement aux résultats souhaités. Au fil des années, cette définition a évolué et l'engagement des apprenants a pris plusieurs dimensions : comportementale, émotionnelle et cognitive [69].
Dans le contexte de l'apprentissage en ligne, l'engagement n'est plus mesuré par la présence dans la classe, la participation et l'interaction directe. En effet, l'engagement des apprenants en ligne est qualifié ainsi que quantifié par l'utilisation des plateformes d'apprentissage en ligne [69]. En d'autres termes, les plateformes d'apprentissage en ligne se présentent comme des sources principales pour la génération des données permettant de suivre l'engagement des apprenants. C'est l'activité en ligne sur le LMS qui

définit le taux d'engagement d'un apprenant. La majorité des systèmes répondant aux problématiques de l'enseignement en ligne utilisent les données de l'interaction de l'apprenant avec le LMS [2]. [70] définit l'engagement d'un apprenant en ligne comme l'effort que ce dernier consacre pour les processus d'apprentissage du contenu d'un cours spécifique. [70] montre que l'accès aux ressources et activités des environnements d'apprentissage virtuels traduit l'engagement des apprenants envers leurs études.
[73] analyse les profils d'engagement des apprenants inscrits dans des MOOC et en distingue trois formes :

- engagement actif : représente les apprenants activement engagés dans le cours en participant aux forums de discussion. Les données liées à ce type d'engagement sont : la participation aux forums de discussionet la réalisation des quizzs.
- engagement passif : représente les apprenants qui suivent les cours mais qui sont absents sur les forums de discussion. Les données liées à ce type d'engagement sont : soumission des cours, consultation / vote / souscription à des publications des forums de discussion
- désengagement : signifie que les étudiants cessent de suivre le cours de manière active et passive. Les données temporelles comme la dernière connexion sont des signes de désengagement.

Compte-tenu des définitions de l'engagement dans l'apprentissage en ligne ainsi que ses différentes formes, un apprenant qui est en risque d'échec est un apprenant peu engagé dans ses études. L'engagement est l'un des indicateurs d'apprentissage les plus employés dans les projets de LA pour la détection des apprenants défaillants. À titre d'exemple, ce projet [70] réalisé avec les données d'un cours de l'université ouverte (en anglais Open University OU) au Royaume-Uni et dont l'objectif est l'identification des étudiants désengagés en estimant que ceci est un indicateur pour la prédiction du risque. Le flux du nombre de clics sur le VLE est l'une des données les plus utilisées dans la littérature pour qualifier l'engagement d'un apprenant en ligne [12] [74] [75] [76].

1.5.3.3 L'indicateur de la régularité

Dans l'état de l'art, *la régularité* est, de plus en plus, considérée comme un indicateur assez pertinent et puissant pour l'explication de l'effort constant que l'apprenant fournit tout au long de son apprentissage.
En effet, les premiers psychologues scolaires ont émis l'hypothèse que l'auto-régularité est un facteur clé de la réussite scolaire des élèves [77]. De plus, plusieurs chercheurs du domaine éducatif ont montré que la régularité de l'apprentissage participe à la prédiction des résultats finaux d'un apprenant [78].
[79] définit la régularité de l'apprentissage comme la mesure dans laquelle les apprenants s'engagent régulièrement dans l'apprentissage, et a été reconnue comme l'une des stratégies de gestion du temps.
[67] distingue deux types de régularité de l'apprentissage :

- la régularité temporelle : des patterns qui se répètent dans le calendrier des sessions d'étude d'un apprenant ; par exemple, un apprenant qui se connecte et étudie des MOOC à des jours ou à des heures particuliers.
- la régularité de l'action : des patterns qui se répètent dans la séquence d'actions d'un apprenant ; par exemple, lire le cours → faire un test → relire le cours → refaire le test → passer au cours suivant.

Dans le cadre d'un MOOC, [67] s'intéresse à l'étude de la relation prédictive entre la régularité temporelle de l'apprenant et ses résultats finaux. Les auteurs de cet article proposent des mesures quantitatives de cet indicateur et en étudient trois dimensions clefs :

- intra-cours : c'est la participation répétitive aux cours et la réponse aux événements qui lui sont liés.
- intra-semaine : c'est la participation aux cours pendant les mêmes jours de la semaine.

— intra-journalière : correspond à un profil de comportement quotidien.

Dans le cadre de ce même travail, les auteurs ont fixé six motifs de régularité dans le temps :
— étudier à certaines heures de la journée.
— étudier en certains jours de la semaine.
— étudier pendant les mêmes jours de la semaine sur des semaines de cours.
— même répartition du temps d'étude entre les jours de la semaine sur des semaines du cours.
— quantité particulière du temps d'étude chaque jour de la semaine sur plusieurs semaines du cours.
— suivre le plan du cours

[80] travaille à identifier les indicateurs du LMS significatifs à l'apprentissage en ligne. Les résultats de cette étude montrent que la régularité est l'indicateur le plus puissant pour la prédiction des résultats finaux.
[81] montre que la participation quotidienne et régulière au forum de discussion et l'accès hebdomadaire régulier au matériel du cours sont les prédicteurs les plus significatifs des résultats d'apprentissage des apprenants en ligne. Ce même travail affirme que les données sur la régularité de l'apprentissage sont peu exploitées dans la littérature.

1.5.4 Discussions

L'identification des données et des traces numériques les plus représentatives de l'activité de l'apprenant et les plus corrélées à son résultat final est une étape cruciale dans le traitement de tout projet de LA, et particulièrement celui de l'identification des apprenants à risque. De plus, nous constatons que le recours à la définition des indicateurs d'apprentissage pour représenter l'activité d'un apprenant en ligne est de plus en plus répandu dans l'état de l'art. En effet, ces indicateurs servent au diagnostic des comportements d'apprentissage. Ayant des objectifs d'intervention auprès des apprenants à risque en leurs proposant des actions correctives, ces indicateurs pourraient identifier les points faibles chez ces apprenants.

Cependant, peu nombreux sont les travaux de l'état de l'art qui proposent un indicateur qui met l'accent sur l'importance de la gestion du temps chez l'apprenant en ligne, pourtant c'est un facteur important pour réussir en ligne (voir la Section 1.2.3). Ajoutons à cela, dans le contexte d'un apprentissage en totale autonomie, l'activité de l'apprenant devrait être, d'une manière ou d'une autre, comparée à celle de ses pairs ; ceci aide à qualifier le rythme d'apprentissage de l'apprenant. En plus des motifs des actions et des calendriers d'activités, la régularité de l'apprentissage pourrait avoir d'autres définitions qui mettent en valeur le progrès d'un apprenant ; un tel indicateur permet de valoriser l'activité de l'apprenant.
En plus des indicateurs d'apprentissage, il est d'un grand intérêt d'intégrer les données temporelles dans les vecteurs d'activité des apprenants. En effet, lorsqu'il s'agit d'une prédiction qui se produit régulièrement dans le temps, les changements de prédiction d'un temps d'observation à un autre est une donnée temporelle qui peut influencer sur les résultats futurs de prédiction.

1.6 La modélisation de la problématique du risque dans l'état de l'art

Dans les deux sections 1.4 1.5, nous avons répondu au deux premières questions du modèle de Chatti : pourquoi (à travers l'exposition de la problématique de la prédiction précoce du risque et les systèmes d'alerte précoce) et quoi (à travers la collecte des données, leur traitement et l'identification des indicateurs d'apprentissage). Dans cette section, nous répondons à la dernière question qui est la suivante : **comment modéliser la problématique de la prédiction du risque en utilisant l'ensemble**

de données collectées ?. Nous présentons dans cette partie les modèles utilisés, les approches adoptées ainsi que les méthodologies suivies par les travaux de l'état de l'art pour résoudre cette problématique.

1.6.1 Apprentissage automatique

Les projets des LA, surtout ceux dont l'objectif est l'identification des apprenants à risque, reposent en grande partie sur les notions et les modèles proposés par l'apprentissage automatique (*Machine Learning (ML) en anglais*)

1.6.1.1 Définition

D'après [82], l'apprentissage automatique s'intéresse à la question de savoir comment construire des programmes informatiques qui s'améliorent automatiquement avec l'expérience.
L'apprentissage automatique s'appuie sur plusieurs algorithmes pour résoudre les problèmes de données. Tous ces algorithmes partagent une même caractéristique qui est l'apprentissage à partir des données historiques pour trouver des modèles, des patterns cachés et prédire les comportements futurs [36].
Les algorithmes de ML, bien qu'ils partagent la même définition et le principe d'utilisation, sont répartis en deux grandes familles : l'apprentissage supervisé et l'apprentissage non supervisé.

1.6.1.2 Apprentissage supervisé

L'apprentissage supervisé est la tâche de ML consistant à apprendre une fonction qui attribue une sortie à une entrée en fonction d'exemples de paires entrée-sortie existantes. Cette fonction est déduite à partir des données d'apprentissage *étiquetées* formant l'ensemble d'apprentissage. L'ensemble de données d'entrée est divisé en deux sous-ensembles : un 1er sous-ensemble qui correspond aux données d'apprentissage ou d'entraînement et de validation et un 2ème sous-ensemble de test. L'ensemble de données d'entraînement a une variable de sortie qui doit être *prédite* ou *classée (étiquetée)*. Tous les algorithmes apprennent un modèle à partir de l'ensemble de données d'apprentissage et l'appliquent à l'ensemble de données de test pour la prédiction ou la classification [83].
Selon la valeur de la sortie, il existe deux types d'algorithmes d'apprentissage supervisé :

- régression : lorsque la sortie y à prédire est une valeur numérique dans un ensemble continu de réels $y \in \mathbb{R}$.
- classification : lorsque la sortie y à prédire appartient à un ensemble I fini de n valeurs $I = \{y_1, y_2, .., y_n\}$. Chaque entrée possède une étiquette dans I.

1.6.1.3 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, en suivant une approche d'apprentissage non supervisé, il n'y a pas de bonnes réponses. Les algorithmes d'apprentissage non supervisé se chargent d'eux-mêmes de la découverte et la présentation de la structure intéressante des données. L'apprentissage non supervisé est souvent utilisé dans les problèmes de regroupement (*clustering*) ou la réduction des caractéristiques [83].

Les algorithmes de l'apprentissage automatique (supervisé ou non supervisé) sont largement employés dans le LA. Leur utilisation est centrale dans le cycle de LA et dans l'achèvement de ses objectifs 1.2.

1.6.2 L'apprentissage automatique en éducation

L'étude menée par le Center for Digital Technology and Management (2015), montre l'utilisation croissante de l'apprentissage automatique dans l'éducation. Cette tendance est expliquée par l'explosion de la disponibilité des données éducatives grâce à la digitalisation de l'éducation [84]. En effet, diverses écoles utilisent des moyens technologiques avancés dans les salles de classe pour créer des expériences d'apprentissage personnalisées. Encore plus, les écoles et les universités convergent vers l'adaptation complète ou partielle du numérique et des plateformes digitales pour distribuer les contenus des programmes pédagogiques. De plus, les MOOC attirent des milliers voire des millions d'apprenants. Cette orientation vers le numérique permet d'appliquer et développer les méthodes d'apprentissage automatique pour analyser les données collectées et améliorer par la suite les résultats d'apprentissage académique des apprenants [85]. Grâce à l'augmentation de la quantité, qualité et disponibilité des données d'éducation, les techniques d'apprentissage automatique ont été appliquées pour atteindre plusieurs objectifs pédagogiques et académiques : le traçage des connaissances, l'amélioration du matériel d'apprentissage et les systèmes d'alerte précoce. [86]. L'une des premières applications de l'apprentissage automatique dans l'éducation était de passer des réponses parmi les choix multiples aux cases vides des quizzs et tests [86]. L'évaluation de ce processus par rapport aux réponses des étudiants était basée sur le traitement automatique du langage naturel (TAL) et l'apprentissage automatique [86].
Certes, l'utilisation de l'apprentissage automatique dans le domaine éducatif est en croissance continue. Mais, elle connaît aussi plusieurs défis et rencontre de multiple obstacles notamment l'accès libre aux bases de données [86].

1.6.3 Modèles de ML utilisés dans l'état de l'art

Pour modéliser la problématique de la prédiction précoce des résultats finaux d'un apprenant en ligne, une récente revue [29] montre que les deux types d'apprentissage automatique (supervisé et non supervisé) sont souvent employés dans la littérature. Comme le montre la figure 1.3, l'apprentissage supervisé est le plus adopté dans la littérature. Les techniques de la classification sont les plus courantes suivies par celles de la régression. La plupart des travaux qui s'intéressent à cette problématique de prédiction du risque ont abordé le problème comme une tâche de classification. Alors que, très peu d'études se sont basées sur des algorithmes de régression pour prédire des valeurs numériques spécifiques de performance. Encore moins de travaux ont traité la problématique comme un problème de *clustering* pour détecter les classes d'apprenants dans l'ensemble de données [17].
De plus, la problématique est majoritairement traitée comme une classification binaire [17]. Les étiquettes de classe les plus identifiées dans la littérature sont celles de : Réussite/Échec, poursuivre le programme/abandonner . . .[2].

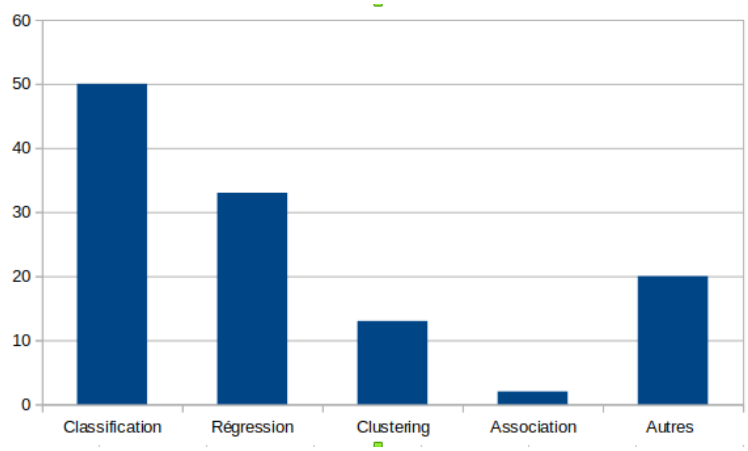


FIGURE 1.3 – Fréquence d'utilisation des techniques de ML dans la littérature - Figure reproduite de [2]

Pour chaque technique d'apprentissage automatique, [2] liste les modèles les plus utilisés dans la littérature pour traiter la prédiction précoce des résultats finaux d'un apprenant :

- Classification : arbre de décision (Decision Tree (DT)), forêts aléatoires (Random Forest (RF)), machine à vecteurs de support (Support Vector Machine (SVM)), classification naïve bayésienne (NB), K-plus proches voisins (K-Nearest Neighbour (K-nn)), ...
- Régression : régression logistique (Logistic Regression (LR)), régression linéaire, ...
- Clustering : le partitionnement k-moyennes, ...
- Association : règle d'association de classe ...

[29] passe en revue les modèles de ML les plus utilisés dans la littérature pour identifier les apprenants à risque. Les modèles de la régression logistique, des arbres de décision et des arbres de régression et classification arrivent en tête de la liste.

Pour des problèmes de classification de décrochage scolaire, plusieurs travaux existants se sont concentrés sur les algorithmes d'apprentissage supervisés tels que Naive Bayes, les règles d'association, les réseaux de neurones artificiels, la régression logistique, CART, C4.5, J48, (BayesNet), SimpleLogistic, JRip, les forêts aléatoires, Logistic regression analysis, ICRM2 [86]. Cependant, pour la prédiction des performances des apprenants, le réseau de neurones (*Artificial Neural Network (ANN)*) et l'arbre de décision sont les deux méthodes les plus largement utilisées dans l'état de l'art [86].

L'avantage des réseaux de neurones est la capacité de détecter toutes les interactions possibles entre les variables prédictives. Ils sont également capables d'effectuer une détection complète même dans une relation non linéaire complexe entre variables dépendantes et indépendantes [87].
Les arbres de décision sont utilisées en raison de leur simplicité et compréhensibilité pour analyser toute structure de données (simple ou complexe) et prédire la valeur de la variable dépendante [88].

[89] examine des modèles ML pour prédire les difficultés rencontrées par les apprenants en ligne. Dans ce travail, plusieurs caractéristiques d'interaction avec la plateforme en ligne et de performance académique sont sélectionnées. Ensuite, en utilisant ces données, cinq algorithmes ont été modélisés : ANN, LR, SVM, Naive Bayes (NB) et DT. En se basant sur les résultats de la phase évaluation, les auteurs de ce travail ont fait valoir l'importance des algorithmes de SVM et ANN.

Un autre projet très bien connu dans la littérature est celui de l'université ouverte (Open University (OU)) au Royaume-Uni. L'objectif de ce projet est de prédire au plus tôt possible les étudiants à risque inscrits dans un cours en ligne afin que les interventions des instructeurs soient efficaces [12]. Pour la réalisation de ce projet, des données démographiques et d'engagement ont été collectées afin de modéli-

ser quatre différents modèles de ML : K-nn, arbre de classification et de régression et NB. Les résultats de prédiction de ces algorithmes sont prometteurs [12] [90].

Afin de prédire les performances des apprenants en ligne à risque [56] collecte les données issues de l'interaction des apprenants avec le LMS pour modéliser trois modèles de classification qui sont : C4.5 Classification and Regression Tree (CART), Logistic Regression (LGR) et Adaptive Boosting (AdaBoost). Les résultats de l'évaluation de ces modèles montrent que CART est le plus performant.

En bref, la classification des apprenants en catégorie selon des étiquettes correspondantes chacune à l'intervalle dans lequel se trouve le résultat final obtenu a montré son efficacité dans la littérature. De plus, les modèles d'arbre de décision, SVM, k-nn, malgré la simplicité de leurs architectures comparant aux autres algorithmes, montrent aussi des bonnes performances prédictives. Pour toutes ces raisons, ce sont les modèles des parties expérimentales de cette thèse.

1.6.4 Comment traiter la problématique de la prédiction du risque ?

Selon le contexte et le résultat attendu, chacun des travaux de l'état de l'art adopte une approche pour traiter cette problématique de prédiction précoce du risque. Par approche, nous voulons dire la granularité temporelle à laquelle les données sont collectées et les modèles de ML sont appliquées.

Le suivi d'une approche de prédiction hebdomadaire des apprenants à risque, inscrits dans des établissements scolaires ou universitaires en ligne ou dans des MOOC, est très répandu dans l'état de l'art. À titre d'exemple, afin de prédire le risque d'abandon au plus tôt, l'OU a suivi une approche de prédiction hebdomadaire du risque [12]. [75] suit aussi cette même approche hebdomadaire pour prédire les apprenants inscrits dans un établissement en ligne. Les résultats de cette étude sont prometteurs.

[91] et [92] sont deux travaux réalisés dans le cadre des MOOC et qui ont adopté l'approche de la prédiction hebdomadaire pour pouvoir identifier au plus tôt les apprenants à risque d'abandonner. Cette approche de prédiction correspond au système d'apprentissage des MOOC où les cours sont disponibles progressivement : semaine après semaine.
Pour prédire au plus tôt les étudiants inscrits dans un établissement universitaire en ligne et à risque de décrochage, [44] divise le cours en six périodes comme suit : le début de cours, 20%, 40%, 60%, 80% et 100% du cours étudié. Par conséquent, les modèles de ML appliqués ont été entraînés et testés à différents pourcentages de la durée du cours.

Pour prédire au plus tôt la performance des apprenants dans un LMS, [76] utilise un modèle de séries temporelles, l'entraîne et le teste à quatre intervalles de temps différents : après 28 jours, 42 jours, 56 jours et 70 jours de l'année scolaire.

Afin d'identifier les apprenants à risque, [93] construit un modèle de prédiction en cinq étapes en incorporant le flux de données formatives et sommatives en cours : 1ère semaine, 3ème semaine, 6ème semaine, 9ème semaine, 12ème semaine.

[55] utilise les traces numériques pour suivre une approche mensuelle permettant la détection des baisses d'activité chez un élève, qui sera signalé par la suite comme à risque. Les expérimentations conduites et les résultats obtenus ont montré que les deux tiers des étudiants signalés finiraient en effet par abandonner leurs cours ou leurs diplômes.

En utilisant les données d'interaction avec l'environnement en ligne et en comparant les résultats de plusieurs modèles de classification, l'objectif de cette étude [94] est de développer un modèle permettant la prédiction au plus tôt des performances des étudiants à la fin du trimestre. Pour ce faire, les données ont été collectées et les modèles ont été entraînés toutes les trois semaines : la semaine 3, 6, 9, 12 et 14.

Cette section présente les différentes approches adoptées par les travaux de l'état de l'art pour prédire les performances des apprenants et prédire ceux qui sont à risque. Selon le contexte et le besoin, ces approches sont différentes d'un projet à l'autre. Afin d'identifier avec précocité les apprenants, une des approches les plus courantes dans la littérature des LA est la prédiction précoce du risque.

1.6.5 Aperçu sur les méthodologies de l'état de l'art

Les travaux qui s'intéressent à la résolution de la problématique de la prédiction précoce et précise des apprenants à risque commencent par définir leurs méthodologies et stratégies du travail.

Afin de traiter cette problématique, les méthodologies proposées et décrites par l'état de l'art se croisent au niveau de quelques phases en mettant l'accent sur l'importance du travail effectué pour la collecte des données et leur nettoyage, l'extraction et la sélection des caractéristiques ainsi que la modélisation des algorithmes de ML. Par la Figure 1.4, [3] résume le déroulement de ce processus se composant de 5 étapes par lesquelles les projets de l'état de l'art passent. Pour s'adapter au contexte et répondre aux besoins attendus, chacun de ces projets peut avoir quelques modifications sur ce processus. Afin d'avoir une idée plus vaste sur les stratégies du travail suivies de l'état de l'art, nous passons en revue, dans la suite de cette partie, quelques unes permettant la prédiction du risque d'échec/décrochage dans le cas d'un apprentissage en ligne (établissement en ligne / MOOC).



FIGURE 1.4 – Le processus standard des prédiction des résultats -Figure reproduite de [3]-

[95] propose une méthodologie et fixe la structure de travail pour prédire le décrochage dans le cas des MOOC. Cette structure comprend essentiellement trois étapes : la génération des caractéristiques, la sélection des caractéristiques et la prédiction d'abandon. Au cours de la 2ème étape de cette structure, les auteurs de ce travail adoptent une méthode d'ensemble pour sélectionner les caractéristiques les plus importantes. Ces méthodes sont connues par leur indépendance des algorithmes d'apprentissage pour classer les caractéristiques selon leur ordre d'importance.

Ce travail [96] propose une stratégie d'analyse pour le LMS *MOODLE* mettant en oeuvre les éléments les plus pertinents des modèles de prédiction. Cette stratégie passe par les différentes étapes de l'analyse des données brutes en fixant les caractéristiques indépendantes et la variable cible jusqu'à la modélisation et l'analyse des résultats de prédiction des difficultés potentielles des apprenants.

Afin de prédire la performance des apprenants inscrits dans un établissement en ligne et particulièrement ceux qui ont des difficultés d'apprentissage, [97] propose une méthodologie qui combine la puissance du modèle CRISP-DM [4] dans son processus d'exploration des données tout en introduisant une approche hybride d'utilisation d'algorithmes de classification et de modèles d'ensemble. La stratégie proposée commence par l'étape de la collecte des données (qui, pour ce travail, regroupe l'analyse des données et l'extraction des caractéristiques). À ce niveau, cette méthodologie divise le jeu de données en un jeu d'entraînement et un jeu de test et passe ensuite à l'étape du nettoyage des données du bruit et des valeurs manquantes. L'étape suivante est appelée extraction des meilleures caractéristiques, qui est plus connue par le processus de la sélection des caractéristiques. La dernière étape est celle de la modélisation et validation des modèles de classification et d'ensemble.

Afin de prédire les apprenants en difficulté et ayant besoin d'un soutien pédagogique, [98] suit une méthodologie de classification très standard et basique. Ce travail commence par choisir le jeu de données d'apprenants, puis passe par le traitement de ces données pour pouvoir sélectionner les caractéristiques les plus importantes à l'entraînement des modèles de classification. La dernière étape de ce travail est la prédiction.

4. Cross Industry Standard Process for Data Mining

Ces travaux de l'état de l'art proposent des stratégies pour la prédiction des apprenants à risque inscrits dans des institutions en ligne ou des MOOC. Toutes les stratégies se croisent au niveau de l'importance de la collecte des données et leur nettoyage pour avoir les meilleurs résultats. De plus, toutes les méthodologies évoquent la pertinence de passer par le processus de la sélection des caractéristiques pour garantir des meilleurs résultats. En effet, [97] montre que l'élimination du bruit, valeurs manquantes et variables peu pertinentes améliorent remarquablement les prédictions.
Cependant, moins d'importance est accordée aux tâches de l'ingénierie des caractéristiques (*feature engineering* en anglais), où les types des caractéristiques peuvent influencer les performances des classificateurs [17]. Ajoutons à cela que les méthodologies proposées n'illustrent pas le cas le plus général, où le jeu de données d'expérimentation est à l'origine multi-source et le travail qui doit être réalisé au niveau de la fusion des différentes tables de données. De plus, toutes les stratégies proposées sont très générales et les types des résultats retournés ne mettent pas en avant l'objectif primaire qui est celui de la prédiction des apprenants à risque. En d'autres termes, la phase de l'évaluation des prédictions (qui sont souvent des classifications) est très générale et s'intéressent majoritairement à la performance globale du classificateur ; les classes d'apprenants à risque sont peu ciblées. En effet, les méthodologies adoptées par l'état de l'art ne mettent pas assez en valeur l'importance de valider les capacités des modèles à détecter exactement les apprenants à risque.

1.6.6 Synthèse

L'analytique de l'apprentissage propose plusieurs outils pour répondre aux problèmes et défis de l'enseignement en ligne. Ces outils permettent la découverte des patterns d'apprenants, le regroupement des apprenants selon leur profils/comportements/résultats, la classification en classes d'apprenants et la prédiction des moyennes finales. Pour ce faire, ces outils des LA s'appuient en grande majorité sur les techniques de l'apprentissage automatique, la fouille des données et l'analyse des corrélations pour répondre aux attentes des utilisateurs surtout celles qui concernent la minimisation des taux d'échec des apprenants en ligne.
Cette section a répondu à la question suivante : comment les outils et techniques de LA ont contribué à la résolution de la problématique de l'identification précise et précoce de apprenant à risque ?

La modélisation prédictive est très importante pour avoir des résultats appropriés au contexte et aux besoins. C'est pour cette raison que le choix du (ou des) modèle(s) de ML ainsi que la granularité temporelle suivant laquelle ce dernier est appliquée jouent un rôle très important dans la qualité des résultats obtenus.
D'après les Sections 1.6.3 et 1.6.4, la majorité des travaux de l'état de l'art formalise la problématique de la prédiction du risque en un problème de classification binaire. Cependant, plusieurs autres classes pourraient être introduites et qui peuvent aider les tuteurs à fournir des interventions plus efficaces et surtout plus ciblées. La classification binaire est tellement catégorique qu'elle ne tient pas compte des différences de comportements qui pourraient exister au sein d'une même classe ; particulièrement au niveau des frontières qui définissent les intervalles des classes.

1.7 Évaluation des modèles et systèmes de l'état de l'art

Après avoir répondu aux questions posées dans le modèle de Chatti, nous discutons, dans cette section, les méthodes et mesures d'évaluation utilisées dans l'état de l'art pour évaluer les performances des modèles et approches adaptées pour la prédiction précoce et précise des apprenants à risque.
Dans un premier temps, nous commençons par introduire les mesures de performance proposées par le ML. Ensuite, nous passons en revue les métriques les plus utilisées dans la littérature des LA. Enfin,

nous nous référons aux travaux d'autres domaines de recherche scientifique pour introduire la notion des métriques d'évaluation temporelle.

1.7.1 Métriques de performance proposées par le ML

Dans cette partie, nous présentons les métriques de performance de ML les plus utilisées pour les différentes tâches de prédiction (régression et classification).

Lors de l'évaluation des performances d'un modèle de ML, deux concepts sont essentiels : le résultat réel généralement appelé «y» et le résultat prédit (généralement appelé « $\hat{y}$ »).
Selon la nature du problème (régression ou classification), les métriques d'évaluation de performance changent. Ainsi pour chacun de ces deux problèmes, nous trouvons plusieurs métriques avec différents objectifs.

1.7.1.1 Métriques de régression

Les métriques de performance utilisées dans les évaluations des performances des algorithmes de régression permettent l'évaluation du taux d'erreur dans les résultats fournis. Ci-dessous, les trois métriques les plus utilisées dans les problèmes de régression :

- $\mathbf{R}^2$: montre à quelle point le modèle correspond aux résultas réels. Cette métrique compare le modèle actuel avec une valeur de référence constante (c'est-à-dire la moyenne) et indique si le modèle est meilleur.
- **Erreur moyennne quadratique (*Mean Square Error (MSE)*)** : mesure la moyenne des erreurs au carré. Cette métrique calcule essentiellement la différence entre la valeur estimée et la valeur réelle, élève ces résultats au carré, puis calcule leur moyenne.
- **Erreur moyenne absolue (*Mean Absolute Error (MAE)*)** : cette mesure prend la valeur absolue vu qu'elle ne s'intéresse pas au sens dans lequel les valeurs cibles estimées et réelles diffèrent (estimées > réelles ou inversement) mais à la distance absolue.

1.7.1.2 Métriques de classification

Les métriques utilisées pour l'évaluation des performances des algorithmes de classification comparent généralement si la classe prédite $\hat{y}$ est égale à la classe réelle y. La majorité des métriques de classification sont déduites de la matrice de confusion (ou *Confusion Matrix (CM)* en anglais) qui contient toutes les informations de la prédiction. La matrice de confusion est un tableau qui enregistre le nombre d'occurrences entre deux évaluateurs, la classification vraie/réelle et la classification prédite [99]. Une illustration d'une CM est représentée par la Figure 1.5. Les cases de CM sont expliquées ci-dessous :

- VP : la donnée, réellement en classe " Positif", est correctement prédite dans la classe " Positif ".
- FN : la donnée, réellement en classe " Positif", est prédite par erreur dans la classe " Négatif ".
- FP : la donnée, réellement en classe "Négatif", est prédite par erreur dans la classe "Positif".
- VN : la donnée, réellement en classe "Négatif", est correctement prédite en classe "Négatif".

		Prédite	
	Classes	Positif	Négatif
Réelle	Positif	Vrai Positif (VP)	Faux Négatif (FN)
	Négatif	Faux Positif (FP)	Vrai Négatif (VN)

FIGURE 1.5 – Matrice de confusion

Nous listons, ci-dessous, les métriques de performance de classification les plus populaires :
— **exactitude (*accuracy*)** : mesure la probabilité que la prédiction du modèle soit correcte.
— **précision (*precision*)** : mesure la proportion de Vrais Positifs par rapport à tous les individus de la classe "Positif".
— **rappel (*Recall ou Sensitivity*)** : mesure la précision prédictive du modèle pour la classe positive. Intuitivement, cette métrique définit la capacité du modèle à trouver toutes les unités positives dans l'ensemble de données.
— **F-mesure (*F-score*)** : cette mesure calcule le compromis entre le rappel et la précision à travers leur moyenne harmonique.
— **spécificité (*specificity*)** : cette mesure cherche à trouver le taux d'individus négatifs correctement prédits par le modèle.
— **fonction d'efficacité du récepteur (*Receiver Operating Characteristic (ROC)*)** : cette fonction permet de tracer l'ensemble de valeurs du couple (spécificité, rappel) selon le seuil de classification.

Le Tableau 1.2 résume les métriques de la classification et la régression et donne à chacune sa formule mathématique. Pour les formules des mesures de régression, y_i correspond à la valeur réelle et $\hat{y}_i$ correspond à la valeur prédite.

Tâche de prédiction	Métrique	Formule
Classification	Précision (P)	$VP/(VP+FP)$
	Rappel (R)	$VP/(VP+FN)$
	Exactitude	$(VP+VN)/(VP+VN+FN+FP)$
	F-mesure	$2*R*P/(R+P)$
	Spécificité	$VN/(FP+VN)$
Régression	MSE	$\frac{1}{n}\times\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$
	MAE	$\frac{\sum_{i=1}^{n}\lvert y_i-x_i\rvert}{n}$
	R^2	$1-\frac{\sum_i(y_i-\hat{y}_i)^2}{\sum_i(y_i-\bar{y})^2}$

TABLE 1.2 – Les métriques de performance utilisées pour la classification et la régression

1.7.2 Les métriques de performance algorithmique dans l'état de l'art

Pour l'identification des apprenants à risque et le suivi des parcours académiques, les enseignants se basent sur les résultats fournis par les systèmes d'alerte précoce et les modèles prédictifs. C'est pour cette raison que l'évaluation des performances algorithmiques des modèles et algorithmes de prédiction

est une étape primordiale dans tout processus d'analyse d'apprentissage.
L'objectif de la recherche étant la prédiction précise et précoce de risque, l'évaluation des résultats des algorithmes comporte au moins deux volets, en plus de la complexité spatiale et d'exécution, et doit permettre de :

— minimiser l'erreur et augmenter l'exactitude des prédictions pour que l'alerte et donc l'intervention soient plus ciblées.
— suivre temporellement les résultats du système pour pouvoir mesurer la précocité ainsi que la stabilité des prédictions.

Dans cette section, nous passons en revue les outils et métriques utilisés dans l'état de l'art pour évaluer les systèmes d'alerte précoce appliqués dans l'éducation ainsi que dans d'autres domaines de la recherche scientifique.
Pour ce faire, nous nous intéressons aux métriques de précision et temporelles. Le premier type se réfère aux différentes mesures de performance des algorithmes de l'apprentissage automatique. Les métriques temporelles sont déduites à partir de suivi continu des prédictions au cours du temps.

1.7.2.1 Les métriques de précision dans les LA

Les mesures statiques de l'apprentissage automatique sont communément utilisées dans l'état de l'art pour évaluer les performances prédictives des systèmes éducatifs à bien identifier les apprenants à risque.

[44] compare les performances de différents modèles de ML dans l'analyse des problèmes rencontrés par les apprenants à risque inscrits dans des universités en ligne. Cette comparaison ainsi que la conclusion sur le modèle le plus performant se basent sur les résultats des mesures d'exactitude, de précision, de rappel, de support et de f-score.

[100] propose un modèle basé sur les notes des étudiants pour prédire ceux qui sont susceptibles d'échouer dans un cours. Pour évaluer les performances de ce modèle, ce travail se sert de la métrique d'exactitude.

Les travaux de cet article [101] visent à améliorer les performances d'un système d'alerte précoce d'abandon. Pour atteindre cet objectif et après avoir été entraîné, les modèles de classification ont été évalués en utilisant les métriques de précision-rappel (PR) et des courbes ROC.

En se basant sur un modèle d'ensemble utilisant une combinaison d'algorithmes de ML, l'objectif de [102] est d'identifier les étudiants à risque d'échec scolaire pendant la pandémie. Afin d'établir une classification dans laquelle les élèves présentant des risques scolaires peuvent être prédits avec plus de précision, les performances du modèle adopté ont été évaluées avec la mesure de spécificité.

L'objectif de ce travail [94] est de comparer plusieurs modèles de classification à bien prédire les performances des apprenants inscrits dans un cours en ligne à la fin du trimestre. Cette comparaison des résultats de prédiction a été établie en s'appuyant sur les métriques : exactitude, rappel, spécificité, F-mesure.

Cette étude [103] propose un modèle de classification pour prédire la performance scolaire des étudiants à la fin du semestre. La performance du modèle a été évaluée en suivant les métriques de précision de ML suivantes : exactitude, rappel, spécificité, F-mesure.

1.7.2.2 Les métriques temporelles

Les métriques temporelles permettent de qualifier et quantifier la performance des algorithmes de ML quand il s'agit d'un processus de prédiction évolutif dans le temps.
Bien que la prédiction précoce du risque d'échec soit un sujet très populaire dans la littérature, les métriques de performance utilisées dans l'état de l'art de domaine de l'éducation ne tiennent pas en considé-

ration l'évolution temporelle de l'apprentissage et ainsi donc de la prédiction. De plus, tous ces travaux ne proposent pas un moyen pour mesurer la précocité de la prédiction ou évaluer la stabilité des prédictions au cours du temps.
Cependant, les chercheurs dans d'autres domaines, notamment en *process mining* et le traitement du signal, sont plus conscients de l'importance de la dimension temporelle pour la valorisation de leurs travaux et proposent des définitions et mesures pour calculer la précocité.

Mesures de la précocité dans d'autres domaines
[104] donne un aperçu sur les approches de classification des séries chronologiques, en particulier celles utilisées pour des prédictions précoces. Les auteurs de cet article définissent la classification précoce comme le problème consistant à prendre des décisions de classification avec le moins d'observations possible d'une série chronologique tout en veillant à ne pas trop perdre en précision.

Afin d'évaluer les performances de prédiction d'un processus de monitoring axé sur les résultats, [105] utilise majoritairement deux métriques : l'exactitude et la mesure de précocité. Dans ce travail, la précocité est définie comme la plus petite longueur d'un cas d'étude en cours où un classificateur atteint un seuil de précision prédéfini.

[106] présente un cadre de diagnostic des défauts structurels du rotor avec une stratégie de classification précoce. Comme la politique de décision est cruciale pour le paradigme proposé, la précocité est mesurée. La précocité des tests fait référence au nombre de points de données utilisés pour la prédiction de classe. Par conséquent, la précocité d'un classificateur, dans ce contexte, est le pourcentage moyen de la longueur prévue par rapport à la durée totale du test.

Les auteurs de [107] proposent un classificateur de séries chronologiques pour la classification précoce des signaux électroniques. Pour évaluer les performances de ce modèle, l'importance de la précision des prédictions ainsi que leur précocité sont prises en compte. La précocité, dans ce travail, est définie comme le nombre moyen de points de données après lequel une étiquette est attribuée.

Mesure de la stabilité dans l'état de l'art
Étant donné les changements des comportements des apprenants au long de la période d'apprentissage, le modèle pourrait avoir, à chaque temps de prédiction, une performance différente ; ce qui rend tout le système prédictif instable.
La stabilité est un autre facteur important à considérer surtout lorsqu'il s'agit d'un processus des prédictions évoluant dans le temps. Dans l'état de l'art, la stabilité des prédictions a été abordée dans plusieurs contextes.

Une définition commune du terme de la stabilité est toujours lié aux petits changements dans les résultats du modèle en changeant la base de données de l'entraînement [108].

En se basant sur plusieurs mesures de performance notamment la stabilité, les auteurs de [109] proposent un processus pour évaluer l'interprétabilité des modèles de ML. Dans ce contexte, la stabilité est définie comme l'écart de la distribution des mesures sur les modèles appris à partir de différents échantillons de la population.

Cependant, pour l'évaluation des performances d'un modèle de prédiction pour un système éducatif, nous nous intéressons à la stabilité temporelle définie comme la capacité du modèle à retourner un bon résultat au cours du temps en entraînant la même base de données [110]. Cette mesure caractérise la capacité du système de maintenir la même performance à travers les temps de prédiction.

1.7.3 Discussions et Limites

Les travaux de l'état de l'art du domaine éducatif s'intéressent plutôt à l'évaluation statique des performances des modèles employés pour la prédiction précoce et précise des apprenants à risque. Cette

évaluation se base essentiellement sur les mesures statiques et standards proposées par l'apprentissage automatique.
Certes, ces métriques sont importantes pour une évaluation générale de la performance algorithmique à un moment donnée de l'apprentissage. Mais, leur utilisation seule est insuffisante et ne reflète pas la continuité de la prédiction au cours du temps.

En effet, l'apprentissage ainsi que la prédiction sont deux processus évolutifs dans le temps dont les résultats dépendent étroitement des changements et fluctuations dans les comportements des apprenants pendant toute la période de l'apprentissage. L'utilisation des métriques temporelles et l'étude des compromis entre elles gagnent l'intérêt des chercheurs de plusieurs domaines dont les principaux processus sont évolutifs dans le temps.
L'investigation de la dimension temporelle a aussi la même importance pour une évaluation plus profonde et complète des modèles et systèmes éducatifs visant à identifier au plus tôt les apprenants à risque pour une intervention pédagogique et académique efficace.

1.8 Tableau récapitulatif des travaux de l'état de l'art

Le Tableau 1.3 résume les projets de l'état de l'art qui s'intéressent à la problématique de la prédiction du risque dans un contexte d'apprentissage / enseignement en ligne. Pour chacun de ces projets, nous présentons le contexte pour lequel il a été réalisé, les données utilisées, les modèles adoptés ainsi que les métriques d'évaluation des performances des prédictions.

D'après ce tableau, la plupart des études réalisées appartient au contexte des universités en ligne et MOOC. De plus, cette revue [111] confirme que les travaux de l'état de l'art traitant de la problématique de la prédiction du risque sont majoritairement dédiés à l'enseignement supérieur et aux niveaux d'études avancés. Cependant, peu de travaux s'intéressent aux cas des écoliers, collégiens et lycées ($\approx 1\%$) [111]. De plus, nous constatons que la littérature des LA utilise largement les données issues de la base de données fournie par l'université ouverte du Royaume-Uni (connue par son nom *OULAD*). Cette base de données se caractérise par son accessibilité, sa documentation claire et sa facilité de compréhension.

Les travaux de l'état de l'art pour la prédiction de risque dans le contexte de l'apprentissage en ligne s'appuient sur plusieurs types de données et traces numériques. D'après le Tableau 1.3, les données démographiques restent toujours exploitées dans le contexte de l'éducation en ligne. Ensuite, les données sur les devoirs, les notes, les évaluations continues, les tests de positionnement. . .gardent leur importance et impact pour identifier les apprenants en ligne à risque. Tout travail réalisé dans un contexte d'apprentissage en ligne (MOOC ou établissement en ligne) met l'accent sur la pertinence des données des activités effectuées sur la plateforme en ligne. Il s'agit des traces d'interaction entre l'apprenant et son environnement numérique d'apprentissage. Les données de temps de connexion sur la plateforme, les traces de flux de clics, les traces par type d'action effectuée (télécharger, soumettre, lire, voir, consulter . . .), les traces des publications dans les forums de discussion, les traces des quizzs effectuées . . .sont parmi les données les plus utilisées dans la littérature. Ces données montrent la participation, l'engagement et l'implication de l'apprenant dans son contenu de formation. Cependant, l'état de l'art ne montre pas assez d'intérêt aux données qui représentent la gestion du temps de la part de l'apprenant en ligne et aux données qui illustrent le progrès continu qu'un apprenant pourrait avoir.

Afin de modéliser la problématique de la prédiction du risque, le Tableau 1.3 confirme que l'apprentissage automatique et les modèles de classification de ML dominent l'état de l'art. Cette revue [111] confirme que RF et DT sont parmi les modèles les plus performants. Étant donné leurs architectures complexes et difficilement adaptables aux différents contextes et structures de données, les modèles de réseau de neurones sont très peu exploitables dans la littérature.

La partie évaluation des performances algorithmiques est primordiale pour mesurer la pertinence

des données, indicateurs d'apprentissage et modèles sélectionnés. Le Tableau 1.3 montre que, pour cette phase, il y a une domination complète de l'utilisation des métriques de précision de ML. Ces mesures s'intéressent à la performance générale du modèle sans pour autant se focaliser sur les résultats d'une classe bien définie d'apprenants. Ces métriques sont statiques et n'illustrent pas la continuité et l'évolutivité de deux processus d'apprentissage et de prédiction.

Titre	Contexte	Données	Modèle(s)	Métrique(s) d'évaluation
[112]	Programme de maîtrise en ligne en administration publique	Données démographiques et d'inscription, données d'interaction sur la plate-forme en ligne (publications dans les forums, accès au cours en ligne, nombre de lecture)	Régression logistique	Rappel, taux de faux positive
[12]	Université ouverte du Royaume-Uni	Données démographiques, données de flux de clics	K-NN, NB, CART	Précision, rappel, F-score
[53], [43], [44], [66], [75], [113]	Cours en ligne (Base de données OULAD)	Données démographiques, données d'activités sur la plate-forme en ligne, données des évaluations, données des cours	ANN, RF, SVM, NB, LR, XGBoost, LSTM	accuracy, rappel, précision, F-score, support
[89]	Environnement virtuel d'apprentissage	données démographiques, données d'interaction avec la plate-forme en ligne,	ANN, SVM, LR, DT, NB	RMSE, ROC, accuracy, precision, recall, F-score
[96], [114]	MOOC	Données de connexion, données des quiz et forums, nombre de vues sur les vidéos, données de flux de données (stop, pause, retourner, avancer)	ANN, RNN, GRU	Accuracy, ROC, F-score, MAE, RMSE
[54], [63]	Enseignement et apprentissage basés sur les livres en ligne	Données de flux de clics et d'interaction avec le contenu du cours, Temps moyen passé sur la plate-forme, données des activités à la maison	RF, K-NN, NB, CART, SVM , XGBoost, AdaBoost, LSTM	Accuracy, Kappa, précision, rappel, F-score
[76]	Apprentissage via un LMS	Nombre de clics, temps de connexion, interactions avec les composantes du LMS	LSTM, ANN, LR, SVM, DT, K-NN, RF, NB	AUC

[94], [115]	Cours d'informatique hybride (Majoritairemet en ligne)	connexion aux sessions, participations au forum de discussion, performance à la fin du trimestre, Données de log	K-NN, DT, NB, RF, SVM, ANN, CN2rules, LR	Accuracy, rappel, F-score, specificité, Rsquare, MSE, MAE
[100], [116]	Enseignement via un LMS	Données historiques, données des sessions actuelles	NB	TNR, TPR, acuuracy, F-score
[117], [118]	MOOC	Données d'intercation avec l'environnement en ligne, résulltat final, Certifié ou non, flux de données de clics	Règles d'association, ANN, SVM, K-NN	Support, confiance, lift, accuracy
[119]	Programme d'études supérieures en ligne	Fréquence d'accès au programme en ligne, données d'interaction sur les forums, des publications des discussions, données démographiques	k-means (clustering), DT, LR, Boosting, induction de règles	Accuracy
[120]	Apprentissage via un LMS	Données sur le temps passé sur le LMS, données d'intercation avec le LMS (nombre de téléchargements, de posts dans le forum), nombre d'exercices effectués	LR	R square
[121]	- Cours en ligne de l'Université de Harvard et du Massachusetts - Cours en ligne de l'université ouverte	données démographiques, type et nombre d'activités journalières, données d'interaction avec la plateforme en ligne, dates de scolarité (inscription et désinscription)	RF, ANN, modèle linéaire généralisé, Gradient boosting machine	Accuracy, F-score, sensibilité, F-score, AUC

TABLE 1.3: Tableau récapitulatif des travaux de l'état de l'art

1.9 Conclusion

L'explosion de l'utilisation de l'Internet et son accessibilité ainsi que la disponibilité des plateformes numériques dédiées à l'apprentissage et à la gestion administrative ont favorisé l'adoption de l'éducation en ligne par plusieurs institutions et organismes éducatifs à travers le monde.
Certes, l'apprentissage en ligne a amélioré plusieurs aspects de l'enseignement comme : la disponibilité des ressources éducatives de n'importe où et à n'importe quel moment, la disponibilité des espaces dédiés pour poser des questions et lancer des discussions, la facilité du contact tuteur-apprenant. Mais, ce mode d'apprentissage se trouve face à plusieurs défis et problèmes, notamment les taux élevés d'échec et d'abandon, desquels les acteurs du domaine de l'éducation se soucient et cherchent à trouver des solutions.
En effet, la littérature du domaine d'éducation utilise les techniques et outils de l'analyse de l'apprentissage et suit l'approche de la prédiction précoce des apprenants en ligne à risque pour faire face à ce problème.
En commençant tout d'abord par établir les notions des sciences de l'éducation liées à la réussite et l'apprenant à risque, ce chapitre a passé en revue toutes les notions liées au traitement de cette problématique.

Dans un premier temps, nous avons présenté le domaine des LA et avons introduit le modèle de référence pour les LA proposé par Chatti l'un des pionniers de ce domaine. Nous avons ensuite organisé ce chapitre de l'état de l'art selon ce modèle.

Ayant besoin d'une solution qui prédit les apprenants à risque au plus tôt, les travaux de cette thèse sont dédiés aux enseignants des institutions en ligne qui souhaitent être alertés des situations académiques défaillantes.

En plus de leurs informations démographiques et académiques, en se connectant à une plateforme d'apprentissage en ligne, les apprenants laissent des traces numériques : ce sont les données à collecter et à comprendre. Cette étape est importante pour la compréhension du contexte et des comportements des apprenants. De plus, la littérature propose d'identifier des indicateurs d'apprentissage pour définir le comportement d'apprenant. Parmi ces indicateurs, nous trouvons la performance, l'engagement et l'implication de l'apprenant vers la plateforme en ligne, la régularité, l'effort.... La définition de ces indicateurs est très utile pour le diagnostic des comportements et des résultats finaux. Cependant, à notre connaissance, nous ne trouvons pas dans la littérature un indicateur qui sert à comparer l'activité de l'apprenant et son intérêt aux cours par rapport à ses pairs. De plus, la littérature ne propose pas un indicateur qui compare régulièrement l'avancement fait par un apprenant avec celui de ses pairs.

La prédiction précoce et précise du risque est l'approche la plus adaptée dans la littérature des LA pour faire face au problème des taux élevés d'échec et d'abandon. Pour ce faire, les techniques de l'apprentissage automatique sont très employées, notamment les arbres de décision, les forêts aléatoires, les réseaux de neurones, les machines à vecteurs de support, les k-plus proches voisins. Nous remarquons que les modèles de classification sont les plus utilisés et que la formalisation de la problématique en un modèle de classification binaire est la plus répandue dans la littérature. Or, la classification binaire est tellement générale que nous pouvons ignorer des sous-classes dans une mêmes classe.

D'un autre côté, en reposant sur les modèles de ML, les EWS sont très utilisés dans la littérature et leur objectif est la prédiction précoce du risque afin d'alerter les enseignants des éventuels cas d'échec. Par conséquent, en utilisant ces outils, les enseignants s'attendent à ce qu'ils soient alertés automatiquement des apprenants à risque. Or, cela n'est pas la finalité des travaux sur les systèmes d'alerte. Autrement dit, la littérature sur les systèmes d'alerte s'arrête au niveau de la prédiction du risque et ne propose pas de méthodes particulières pour la génération d'alerte.

L'évaluation des performances des systèmes et modèles proposés pour la prédiction précoce et précise du risque est une étape très importante. En effet, tous les travaux de l'état de l'art du domaine de l'éducation utilisent les métriques de précision proposées par le ML pour qualifier leur modèles. Or,

l'apprentissage ainsi que la prédiction sont deux processus évolutifs et continus dans le temps. De plus, la notion de la précocité est très utilisée, sans pour autant qu'il y ait des travaux qui la mesure, qualifie ou quantifie. Malgré le fait que la dimension temporelle soit très importante dans les processus d'apprentissage et d'enseignement, les travaux de l'état de l'art des LA ne s'intéressent pas à l'évaluation temporelle des performances des modèles prédictifs.

Dans les chapitres suivants, nous répondons aux limites de l'état de l'art et y proposons des solutions.

Chapitre 2

Contexte applicatif

2.1 Introduction

Un projet de LA repose essentiellement sur les données ; c'est la dimension qui définit le mieux le contexte du travail. La compréhension de jeu de données, comme le montre le processus CRISP-DM [5] représenté par la Figure 2.1, est une étape importante dans un projet de LA basé sur l'exploration de données. Cette phase commence par une collecte de données et se poursuit par des activités afin de se familiariser avec les données, d'identifier les problèmes de qualité des données, de découvrir les premiers aperçus des données ou de détecter des sous-ensembles intéressants pour former des hypothèses. La compréhension des données et leur analyse est très importante pour la compréhension du contexte applicatif [123] [4].

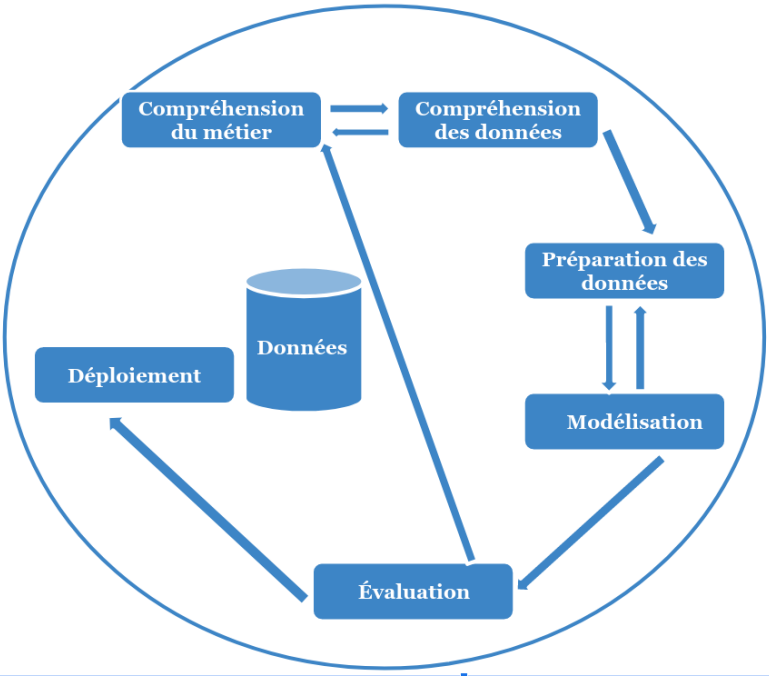


FIGURE 2.1 – Phases du modèle de processus CRISP-DM pour l'exploration de données - Figure reproduite de [4] -

5. Cross Industry Standard Process for Data Mining (CRISP-DM) est un modèle de processus d'exploration de données qui décrit une approche utilisée pour résoudre les problèmes de l'analyse, de l'extraction et des sciences des données [122].

Les instituts d'enseignement en ligne possèdent généralement plus d'une plateforme éducative et administrative pour fournir un contenu pédagogique et académique à tous leurs apprenants. Les activités de ces derniers au sein de ces plateformes sont généralement sauvegardées dans des grands entrepôts de données. Ces traces numériques sont les données à collecter, analyser et exploiter pour proposer des solutions aux problèmes inhérents à l'enseignement en ligne, notamment les taux élevés d'échec scolaire.

Le Centre National d'Enseignement à Distance (CNED) en France ne fait pas l'exception et enregistre annuellement des taux élevés d'échec parmi ses apprenants. En effet, les enseignants du CNED ont, aussi, besoin des outils technologiques leur permettant une identification au préalable des apprenants dont les comportements sont défaillants ou à risque de devenir défaillants.

Cette thèse est un projet de collaboration entre le CNED et le LORIA. Le CNED est le cadre applicatif de tous les travaux effectués.
Afin de répondre à la problématique de la prédiction précoce et précise du risque d'échec, le CNED a fourni, tout au long de cette thèse, les données d'apprentissage et administratives de ses apprenants.

Dans ce présent chapitre, nous décrivons le cadre du CNED et nous faisons la description des données reçues ainsi que leur diagnostic à travers une étude de corrélation.

2.2 Le contexte du CNED

Actuellement, le CNED est le plus grand centre d'enseignement public à distance en France qui se veut complémentaire à l'enseignement traditionnel.
Dans cette section, nous présentons l'historique du CNED et exposons ses nombreuses caractéristiques qui font de lui un contexte très particulier.

2.2.1 Historique du CNED

Le CNED est un organisme public de formation à distance créé en 1939, date de début de la guerre, avec initialement seulement 900 inscriptions. Étant une « création temporaire », ce service a été crée pour faire face aux interruptions et restrictions apportées du fait de la guerre aux études des élèves. Déclaré en 1944 comme un Centre National d'Enseignement par Correspondance, le CNEPC a fait appel à des professeurs contraints d'abandonner leur enseignement pour des raisons de santé ou de famille. Les années 50 sont la période d'or de la radio et avec la disponibilité de la télévision chez la majorité des français, le CNEPC continue à faire évoluer ses services et se transforme en 1953 en un Centre National d'Enseignement par Correspondance, Radio et Télévision (CNEPCRT). Le CNEPCRT met en chantier, dès l'année académique 1952-1953, deux émissions hebdomadaires destinées à soutenir les enseignements scientifiques notamment : télévision scolaire et télévision éducative. Et le CNEPCRT continue à enregistrer une forte croissance d'inscriptions pour devenir le plus grand établissement scolaire de France. Selon le décret de 4 mai 1959, le centre s'appelle désormais « Centre National de Télé-Enseignement » (CNTE). En 1962, l'enseignement via les disques a vu le jour au sein du CNTE et la première série de ceux-ci a été réalisée pour l'apprentissage de l'anglais en première. En 1967, les machines à calculer apparaissent en France et en 1970 enfin, la réflexion sur l'« apport de l'informatique et de l'ordinateur » conduit les « pédagogues du CNTE » à prôner la « conjonction des moyens d'enseignement à distance » et à « envisager le schéma pédagogique multimédia ». Le 31 décembre 1979, le CNTE devient le Centre National d'Enseignement par Correspondance (CNEC), « établissement public national à caractère administratif, doté de la personnalité civile et de l'autonomie financière » qui devient à son tour en février 1986 le Centre National d'Enseignement à Distance (CNED). Le développement de l'Internet conduit le CNED à se préparer à l'ère d'e-learning. Et depuis 2005, le CNED ne se contente pas des simples envois des devoirs et des copies et continue à suivre les évolutions technologiques en allant plus loin avec l'ex-

tension de ses services et la création des forums et des espaces d'échange et de tutorat [124]. Le CNED offre aujourd'hui pas moins de 200 formations regroupant un très large nombre d'inscrits qui vont de la maternelle au supérieur [124]. À titre d'exemple, les lycéens au sein du CNED sont très nombreux leur suivi est assuré par seulement une douzaine de conseillers pédagogiques.

2.2.2 Le CNED... un centre d'enseignement assez particulier

De nos jours, le CNED reçoit des milliers d'inscriptions annuellement ; le CNED compte actuellement près de 170 000 inscrits [6]. De plus, le CNED propose 200 formations[6] et divers programmes éducatifs dédiés à tous les niveaux d'apprentissage.

Les apprenants du CNED sont hétérogènes et physiquement dispersés. De plus, chaque apprenant du CNED est unique et est inscrit au CNED pour une raison particulière. En général, les apprenants s'inscrivent au CNED parce qu'ils ne peuvent pas accéder à l'école / université traditionnelle pour des raisons familiales (des familles qui se déplacent tout le temps à cause des professions des parents par exemple, ou à l'étranger loin d'une école française), de santé (incapacité de se déplacer au quotidien sur les bancs de l'école), de disponibilité (étant des athlètes professionnels ou ayant d'autres occupations en parallèle avec les études..) . . .. Par conséquent, chacun de ces apprenants a son propre rythme pour suivre le contenu de son programme.

Outre la diversité et l'hétérogénéité des apprenants du CNED, l'apprentissage est aussi assez particulier. D'une part, l'inscription au CNED demeure possible pendant toute l'année. De ce fait, la date de début d'activité n'est pas commune entre les apprenants et diffère de l'un à l'autre. D'autre part, l'apprentissage au CNED est multi-modal. Le CNED est en phase d'évolution de ses services et veut s'adapter aux nouvelles tendances technologiques du marché éducatif en se transformant en un organisme d'enseignement en ligne. Cependant, afin d'assurer l'accès quotidien à l'information à tous leurs apprenants, qu'ils soient habitants des zones rurales ou urbaines, le CNED garde toujours l'utilisation du support papier en parallèle avec l'environnement d'apprentissage digital qu'il propose. Plus explicitement, en s'inscrivant au CNED, l'apprenant reçoit à la fois ses coordonnées pour accéder à son programme en ligne en se connectant au LMS du CNED ainsi qu'une version papier de ses cours et exercices à son domicile. De ce fait, le LMS n'est pas la seule source d'information pour les apprenants inscrits au CNED qui peuvent être à jour dans leurs programmes sans pour autant être très actifs sur la plateforme.

Parallèlement, le CNED dispose d'une multitude d'applications et de plateformes dédiées aux enseignants et aux apprenants dont les usages permettent de capturer un volume important de données qui pourraient servir à concevoir des outils de suivi de l'avancement du cursus académique des inscrits par les conseillers concernés. La nécessité de tels outils est d'autant plus importante que, d'une part, le CNED prend en charge un nombre très élevé d'apprenants et, d'autre part, il a abondamment recours à l'enseignement en ligne.

Face au nombre important d'élèves à gérer au quotidien, à la particularité du contexte académique ainsi qu'à la disponibilité de plusieurs plateformes d'enseignements et administration, il devient de plus en plus difficile pour les conseillers pédagogiques de suivre de manière fiable les performances respectives des apprenants et de leur fournir un accompagnement personnalisé adéquat. Les chargés de conseil et tutorat ne sont plus en mesure d'identifier les apprenants avec des comportements d'apprentissage défaillants et en besoin d'aide et de soutien pédagogique. L'objectif est, par conséquent, d'améliorer le suivi des élèves, à commencer par l'identification au plus tôt de ceux qui sont susceptibles de rencontrer des difficultés qui les conduiraient à un échec ou un abandon de leurs cursus.

Dans les parties qui vont suivre, nous procédons à une étude descriptive et diagnostic du cadre du CNED en entier.

6. https ://www.cned.fr/decouvrir-le-cned/le-cned-en-bref

2.3 Analyse descriptive des données

Dans cette section, nous procédons à l'analyse descriptive du contexte. Parmi toutes les plateformes que le CNED possède, nous décrivons les deux sources de données qui sont les plus intéressantes et informatives pour notre problématique. Ensuite, nous passons à la description de toutes les dimensions qui construisent le contexte d'apprentissage du CNED : caractéristiques de scolarité, population, matières, devoirs, actions sur la plateforme en ligne.

2.3.1 Description des sources de données

Comme pour la majorité des institutions académiques, les données du CNED sont multi-sources.
Le CNED dispose de plusieurs plateformes d'enseignement, d'apprentissage et d'administration. Certaines de ces applications contiennent des informations et données plus intéressantes que d'autres pour notre problématique.
Dans cette thèse, nous avons collecté les traces numériques et les données provenant essentiellement de deux sources de données comme le montre la Figure 2.2.



FIGURE 2.2 – Les sources de données du CNED

Afin de construire notre jeu de données, nous avons fusionné les données de deux applications du CNED suivantes :

— ***LMS (MOODLE) :*** c'est l'application qui se charge de la proposition et l'affichage du contenu pédagogique. En se connectant à cette plateforme, l'apprenant accède à son programme de formation, peut effectuer des tests et quizzs, discute dans les forums avec ses pairs, pose des questions à ses enseignants . . .. Dans la base de données de ce LMS, nous trouvons toutes les données de tous les composants de l'apprentissage au CNED de toutes les années scolaires (ou les trains, dans le vocabulaire CNED) : les actions datées et typées, les cours et leurs catégories, les données des forums, les tables des utilisateurs, ...

— ***GAEL :*** c'est l'application pour la gestion administrative des élèves. Cette application est accessible seulement par les responsables administratifs et les enseignants. C'est dans cette application que les enseignants et les correcteurs des examens soumettent les données des évaluations, notamment les notes. De plus, cette plateforme stocke les données démographiques des apprenants :

âge, genre, lieu de résidence, bourse, redoublant ou pas ….

Le Tableau 2.1 récapitule les types d'informations fournies par chacune de deux sources de données du CNED.

Application / Type de données	LMS	GAEL
Traces d'apprentissage	✓	
Nomenclatures des modules		✓
Identifiants des modules	✓	
Notes et examens		✓
Données démographiques	✓	✓
Identifiants des utilisateurs	✓	✓

TABLE 2.1 – Tableau récapitulatif des types de données fournies par chaque source du CNED

2.3.2 Description de jeu de données

Le CNED propose des formations et programmes pédagogiques à plusieurs niveaux d'apprentissage allant de la maternelle jusqu'à l'enseignement supérieur.
Dans cette thèse en concertation avec le CNED, nous avons spécifié le niveau d'apprentissage sur lequel nous travaillons et investiguons les caractéristiques ; il s'agit de la classe troisième du niveau collège. Le choix sur ce niveau a été établi en nous basant sur maintes raisons :

— cette population est rarement explorée dans l'état de l'art. La majorité des travaux de l'état de l'art est dédiée aux étudiants de l'enseignement supérieur ou aux inscrits dans les MOOC.
— les traces numériques de cette population sont les plus disponibles.
— le mode d'apprentissage adopté par la grande majorité de cette population illustre la particularité de l'apprentissage au CNED, notamment la multi-modalité de l'apprentissage et la totale autonomie des apprenants.

Le CNED propose à ses apprenants différents modes, modalités et types de classes. En effet, une cohorte d'apprenants peut choisir une formation dont le programme est composé seulement d'une matière ou une formation pédagogique complète. Dans ce dernier cas, nous parlons d'une classe complète dans le vocabulaire du CNED.

Afin d'avoir plus de données, nous travaillons, dans cette thèse, avec les traces numériques d'une population inscrite dans une classe complète. Il s'agit de la classe de 3ème du site de Rouen pendant les deux années scolaires 2017-2018 et 2018-2019 appelées dans le vocabulaire CNED trains scolaires (train 7 et train 8 respectivement). Les nomenclatures complètes de cette classe correspondantes à chacune de deux années scolaires respectivement sont : 4-131-7 et 4-131-8. La 1ère partie de cette nomenclature correspond au numéro du site de Rouen (4). Le libellé 131 est le niveau d'apprentissage de la classe. Et la dernière partie (7 ou 8) correspond au numéro du train.
Le tableau 2.2 déclare les dates de début et fin d'inscription pour les deux trains scolaires 7 et 8 respectivement ainsi que les dates de début et fin de scolarité. Nous constatons que l'inscription, pour les deux années scolaires, commence deux mois avant le début de la scolarité et dure jusqu'à la fin du mois de mai. La période de scolarité au CNED est prolongée de dix jours.

Avant tout nettoyage des données et extraction de la population la plus intéressante et active, les inscrits pendant les trains 7 et 8 sont respectivement au nombre de 1586 et 1523.

	Début d'inscription	**Fin d'inscription**	**Début de scolarité**	**Fin de scolarité**	**Prolongé à**
Train 7	01-07-2017	30-05-2018	01-09-2017	07-07-2018	18-07-2018
Train 8	01-07-2018	30-05-2019	01-09-2018	07-07-2019	18-07-2019

TABLE 2.2 – Dates de scolarité de deux trains scolaires

2.3.3 Description de la population

La classe 4-131 est une classe de niveau collège et se caractérise, comme toutes les classes au CNED, par la diversité de ses inscrits. À travers le tableau 2.3, nous décrivons le nombre d'inscrits dans chaque train, leur pays de connexion et les fuseaux horaires auxquels ils appartiennent.

Train	**Nombre d'inscrits**	**Pays**	**Fuseau horaire**
Deux trains	3109	AU (1), CA (1), CH (2), CO(1), DE(2), FR(2009), IQ(1), MA(1),MQ(1), RE(4), US(1), YT(1), NULL(1084)	99(2015), Africa/Casablanca(1), America/Denver(1), America/Martinique(1), America/New_York(1), Asia/Baghdad(1), Australia/Sydney(1), Europe/Paris(1), Europe/Sofia(1), Indian/Reunion(2), NULL (1084)
Train 7	1586	AU(1), CA(1), CH(2), CO(1), DE(1), FR(1542), IQ(1), MA(1), MQ(1), RE(1), US(1), YT(1),NULL(32)	99(1545), Africa/Casablanca(1), America/Denver(1), America/Martinique(1), America/New_York(1), Asia/Baghdad(1), Australia/Sydney(1), Europe/Paris(1), Europe/Sofia(1), Indian/Reunion(1), NULL(32)
Train 8	1523	DE(1), FR(467), RE(3), NULL(1052)	99(470), Indian/Reunion(1)

TABLE 2.3 – Localisation des inscrits

Le Tableau 2.3 fournit en outre les informations suivantes :

— la colonne *Pays* ne correspond pas à la nationalité des apprenants mais plutôt à leurs pays de résidence. Plus explicitement, c'est l'ensemble de pays depuis lesquels la population étudiée se connecte.
— le nombre entre parenthèse, qui se trouve après chaque code du pays, représente le nombre d'inscrits qui se connectent depuis ce pays.
— une localisation non définie dans la base de données fournie est représentée par NULL dans la colonne *Pays*
— dans la dernière colonne du tableau, le nombre entre parenthèse et qui se trouve après chaque fuseau représentes le nombre d'inscrits correspondant.
— 99 représente le fuseau horaire par défaut celui de la France
— un fuseau horaire non défini est représenté par NULL dans la dernière colonne

Par ailleurs, une intersection entre les deux trains 7 et 8 montre que nous avons 72 élèves en commun entre les deux années scolaires. Ces inscrits se trouvent en France métropolitaine, en Allemagne et à la Réunion.

Le Tableau 2.3 met l'accent sur la dispersion physique des apprenants du CNED et leur localisation un peu partout dans le monde. Le décalage horaire entre les différentes localisations favorise encore plus l'autonomie de l'apprentissage au sein du CNED, et plus particulièrement au sein de cette population.

Le tableau 2.4 présente les résultats obtenus par les élèves. Ce tableau présente respectivement la moyenne des notes finales de la classe, l'écart-type de ces notes, le nombre d'élèves qui ont réussi leur année et le nombre de ceux qui ont échoué.
À ce stade, par réussite, nous entendons l'obtention d'une note finale supérieure à 10. Le cas contraire est comptabilisé comme un échec. La note finale est calculée à partir des moyennes des matières. En effet, chaque élève passe un nombre d'examens et de devoirs dans chaque matière. Nous calculons la moyenne des notes obtenues dans tous les devoirs passés dans chaque matière. Et la moyenne finale se calcule à partir de la somme des moyennes de toutes les matières divisée par le nombre de matières suivies par l'élève.

Train	**Moyenne**	**Écart-type**	**Réussite**	**Échec**
Train 7	14.00	4.17	1245	190
Train 8	13.88	4.34	1143	199

TABLE 2.4 – Statistiques de la réussite scolaire pour chaque année scolaire

2.3.4 Description des matières

La classe complète 4-131 propose à ses inscrits une formation se composant en total de 26 matières. À savoir, les inscrits de cette classe ne suivent pas tous forcément le même nombre de matières.
La description des matières est donnée par le Tableau 2.5 :

Code de la matière	**Libellé**	**Type**	**Discipline**
FR31	FRANÇAIS	obligatoire	littéraire
MA31	MATHÉMATIQUES	obligatoire	scientifique
AN31	ANGLAIS	optionnelle	littéraire
AP31	ARTS PLASTIQUES	obligatoire	artistique
EPI3	EPI	obligatoire	interdisciplinaires
ES31	ESPAGNOL	optionnelle	littéraire
GH31	HIST-GEO	obligatoire	histoire-géographie
MU31	EDUCAT-MUSICALE	obligatoire	éducation musicale
SP31	PHYSIQUE CHIMIE	obligatoire	scientifique
SV31	SC-VIE-ET-TERRE	obligatoire	scientifique
TE31	TECHNOLOGIE	obligatoire	technique
AL32	ALLEMAND	optionnelle	littéraire
IT32	ITALIEN	optionnelle	littéraire
LA32	LATIN	optionnelle	littéraire
AR32	ARABE	optionnelle	littéraire
PO32	PORTUGAIS	optionnelle	littéraire
RU32	RUSSE	optionnelle	littéraire
ABSR	SÉCURITÉ ROUTIÈRE	obligatoire	sécurité routière
DEM3	POUR BIEN DÉMARRER 3E	obligatoire	initiation
AR31	ARABE	optionnelle	littéraire
GR32	GREC	optionnelle	littéraire
CH31	CHINOIS 3	obligatoire	littéraire
HA31	HISTOIRE DES ARTS	obligatoire	artistique
HE32	HÉBREU	optionnelle	littéraire

TABLE 2.5 – Description générale des matières de la classe 4-131

D'après ce tableau, chaque matière possède un code. Certaines matières sont optionnelles et d'autres, obligatoires. Elles sont aussi classées en discipline : littéraires, scientifiques, histoire-géographie, art plastique, éducation musicale, initiation, sécurité routière et histoire de l'art. Nous constatons que les matières optionnelles sont toutes littéraires et que les matières scientifiques sont forcément obligatoires.

Chaque matière possède ses propres caractéristiques. Le nombre d'évaluations diffère notamment d'une matière à l'autre. Les deux tableaux 2.6 et 2.7 décrivent les matières suivies par les apprenants du train 7 et 8 respectivement. Le nombre d'inscrits correspond au nombre d'élèves ayant déposé au moins un devoir à une matière.

Les colonnes de ces deux tableaux correspondent respectivement à :

— la matière
— le nombre d'inscrits,
— le nombre maximal de devoirs effectués par un élève,
— le nombre minimal de devoirs effectués par un élève,
— le nombre moyen de devoirs effectués par élève,
— l'écart-type,
— le nombre d'élèves qui ont réalisé un nombre de devoirs supérieur ou égal à la moyenne,
— le nombre d'élèves qui ont réalisé un nombre de devoirs inférieur ou égal à la moyenne.

Matière	Inscrits	Max	Min	Moyenne	Écart-type	>moyenne	<=moyenne
FR31	1373	152	1	6.9534	4.86	843	529
MA31	1377	145	1	7.14	4.74	744	632
AN31	1281	124	1	7.07	4.71	585	695
AP31	1113	61	1	3.41	2.43	506	606
EPI3	939	13	1	1.53	0.623	493	445
ES32	3	9	1	4	3.55	2	2
GH31	1389	180	1	7.19	5.41	751	637
MU31	1053	49	1	3.98	2.29	619	433
SP31	1312	82	1	4.16	2.70	610	701
SV31	1318	76	1	4.19	2.58	621	696
TE31	1206	44	1	3.09	1.57	553	652
AL32	122	19	1	7.39	3.99	62	59
IT32	104	15	1	7.5	3.55	52	51
LA32	34	8	1	4.79	2.86	18	15
AL31	31	11	1	5.87	3.49	16	15
AN32	29	12	1	6.51	4.19	15	14
AR32	45	12	1	9.2	3.05	27	17
GR32	16	8	1	4.62	2.64	9	7
ES31	18	11	1	6.61	4	11	6
ES51	955	61	1	6.94	3.85	527	427
PO32	10	12	1	9	3.37	5	5
RU32	10	12	3	8.2	3.18	5	5
MA2R	1	1	1	1	0	1	1
CH32	11	12	5	9.45	2.27	5	5
HE32	4	7	3	4.5	1.65	2	2

TABLE 2.6 – Caractéristiques des matières du train 7

Matière	Inscrits	Max	Min	Moyenne	Écart-type	>moyenne	<=moyenne
FR31	1277	201	1	6.39	6.24	674	602
MA31	1265	256	1	6.57	7.65	704	560
AN31	1137	225	1	6.34	7.31	489	647
AP31	920	123	1	3.16	4.28	958	306
EPI3	767	23	1	1.48	0.92	351	415
ES32	824	112	1	4.78	4.40	448	375
GH31	1285	252	1	6.56	7.48	694	590
MU31	907	61	1	3.54	2.66	428	478
SP31	1180	171	1	3.92	5.15	676	503
SV31	1177	104	1	3.89	3.37	690	486
TE31	1167	177	1	3.42	5.29	570	596
AL32	136	9	1	3.98	2	72	63
IT32	96	24	1	5.56	3.18	52	43
LA32	26	8	1	3.9615	2.71	13	12
AL31	26	11	1	5.65	3.36	15	10
AN32	27	8	1	4.37	2.43	13	13
AR32	14	9	4	6.85	1.64	8	5
GR32	10	8	1	5.30	2.28	5	5
ES31	17	11	1	5.82	3.09	10	6
PO32	8	8	1	5.5	2.39	5	2
RU32	5	7	2	5.2	1.72	3	2
MA2R	1	1	1	1	0	1	1
CH32	3	8	5	6.6	1.24	2	1
HE32	1	5	5	5	0	1	1

TABLE 2.7 – Caractéristiques des matières du train 8

2.3.5 Description des devoirs

Nous trouvons au CNED plusieurs types de devoirs et, d'une manière générale, d'activités d'apprentissage. Seuls certains types de devoirs sont pris en compte pour caractériser la réussite scolaire. Ce sont les suivants :

— Devoir de type I : devoir numérique écrit,
— Devoir de type A : devoir numérique oral ou mixte écrit/oral,
— Devoir de type C : devoir papier.

Dans les données de l'application GAEL, nous trouvons des devoirs de type E. Ces derniers ne sont pas en réalité des devoirs mais simplement un mot d'excuses à la place d'un devoir.

2.3.6 Description des traces numériques

Dans cette section, nous décrivons les traces numériques des actions effectuées par les apprenants et fournies par le LMS du CNED.
La Figure 2.3 est une capture d'écran de la table des traces numériques générées par le LMS du CNED. Chaque ligne de cette figure représente une action, plus connue sous le nom de *log*.
Chaque action est définie par plusieurs attributs, notamment : son identifiant (*id*), le schéma complet de l'action donné par l'attribut *eventname*, la composante sur laquelle l'action a été effectuée (*component*),

le type de l'action donné par l'attribut *action*, l'identifiant de l'apprenant qui a effectué l'action *userid*, l'identifiant du cours sur lequel l'action a été effectuée (*courseid*) et le temps auquel l'action a été effectuée connu sous le nom du *timestamp* et donné par l'attribut *timecreated*.
D'après cette table, nous avons en total **125 191 574** actions dont les statistiques sont les suivantes :

— le nombre de clics maximal effectués par un apprenant est 1123056
— le nombre de clics minimal effectués par un apprenant est 1
— le nombre de clics moyen effectués par un apprenant est ≈1328

En effet, ces actions sont de types différents et effectuées sur des composantes différentes :

— 50 différentes composantes telles que *forum*, *discussion*, *core*...
— 52 différents types d'actions de : soumission, téléchargement, consultation, connexion, déconnexion, écriture

#	id	eventname	component	action	target	objecttabl	objectid	contextins	userid	courseid	timecreated
1	196703437	\booktool_print\...	booktool_print	printed	book	book	61	16497	[redacted]	1333	1501545617
2	196703438	\booktool_print\...	booktool_print	printed	book	book	62	16498	[redacted]	1333	1501545664
3	196703439	\core\event\cou...	core	viewed	course	NULL	NULL	1333	[redacted]	1333	1501545693
4	196703440	\mod_resource\...	mod_resource	viewed	course_...	resource	37009	127110	[redacted]	1333	1501545705
5	196703441	\core\event\cou...	core	viewed	course	NULL	NULL	1333	[redacted]	1333	1501545771
6	196703442	\core\event\cou...	core	viewed	course	NULL	NULL	448	[redacted]	448	1501545780
7	196703443	\core\event\use...	core	loggedin	user	user	255710	0	[redacted]	0	1501545783

FIGURE 2.3 – Les sources de données du CNED

2.4 Analyse diagnostic : étude de corrélation

L'objectif de cette section est de repérer et visualiser préalablement les corrélations multiples entre les variables et/ou les ressemblances entre les individus.

2.4.1 Définitions de base

Dans cette thèse, nous nous intéressons à l'échec/la réussite scolaire par matière.
Dans cette section, nous définissons les notions permettant d'établir les corrélations entre les différentes variables du jeu de données et la réussite par matière.

2.4.1.1 Matrice de corrélation

La corrélation entre plusieurs variables aléatoires ou statistiques est une notion de liaison (ce qui veut dire qu'elles évoluent ensemble à une vitesse constante) qui mesure leur indépendance [7]. Une matrice de corrélation est utilisée pour évaluer la dépendance entre plusieurs variables en même temps. Le résultat est une table contenant les coefficients de corrélation entre chaque variable et les autres ; les valeurs de ces derniers sont comprises entre -1 et 1. Il existe plusieurs mesures de corrélation, notamment :

— **le coefficient r de corrélation de Pearson :** il mesure une corrélation linéaire entre deux variables.
— **le coefficients de corrélation de τ de kendall et ρ de Spearman :** ils sont respectivement utilisées pour estimer le coefficient de corrélation basé sur le rang. Ce sont des tests statistiques dits robustes car ils ne dépendent pas de la distribution des données. Le test de corrélation de Kendall

7. https ://www.jmp.com/fr_fr/statistics-knowledge-portal.html

et celui de Spearman sont recommandés lorsque les variables ne suivent pas une loi normale.

Dans notre contexte, il s'agit d'une étude de corrélations linéaires entre deux variables. De ce fait, nous présentons les matrices de corrélations relatives aux résultats donnés par le coefficient de Pearson.

2.4.1.2 Test de significativité (p-values)

Le seuil de signification détermine si nous sommes en droit d'interpréter l'existence d'une relation entre deux variables. Ce test vient dire si le coefficient de corrélation est significativement différent de 0. Le résultat de ce test de significativité est un nombre p compris entre 0 et 1. Le seuil de signification est, par convention, 5% ou 0.05 ; Donc, les valeurs de p inférieures à 0.05 sont significativement importantes à étudier.

2.4.2 Définitions des groupes d'élèves

Dans cette partie, nous définissons formellement plusieurs groupes d'élèves selon les moyennes obtenues dans la matière à étudier à la fin d'année.
La variable que nous cherchons à observer est liée à la moyenne d'un élève.
Ensuite, nous étudions les corrélations entre les moyennes obtenues par ces groupes d'élèves inscrits à la matière en question et les variables suivantes :

— variable 1 : moyenne des notes dans toutes les matières
— variable 2 : nombre de devoirs passés dans la matière en question
— variable 3 : nombre de devoirs de type I passés dans la matière en question

La matière que nous choisissons d'étudier est obligatoire et scientifique : c'est la matière physique-chimie de nomenclature SP31.

Dans la suite de cette section, nous définissons plusieurs groupes d'élèves selon différents seuils.

2.4.2.1 Groupes d'élèves selon le seuil 12

La 1ère définition des groupes d'élèves est établie selon le seuil d'une moyenne égale à 12 et se compose de trois catégories d'élèves :

— catégorie 1 : les élèves ayant une moyenne supérieure strictement à 12.
— catégorie 2 : les élèves ayant une moyenne entre 8 et 12.
— catégorie 3 : les élèves ayant une moyenne inférieure strictement à 8.

Cette catégorisation permet d'avoir une idée sur le comportement des apprenants de la zone d'incertitude qui correspond aux apprenants ayant des moyennes entre 8 et 12.

	Train 7	**Train 8**
Catégorie 1	839	827
Catégorie 2	273	201
Catégorie 3	199	151

TABLE 2.8 – Nombre d'inscrits de SP31 appartenant à chaque catégorie de la première définition d'un groupe d'élèves

La table 2.8 présente, pour chaque train scolaire, le nombre d'élèves appartenant à chaque catégorie de la première définition des groupes d'élèves inscrits à la matière SP31.

Cas de la première catégorie : les résultats de corrélation de cette catégorie sont donnés par la Table 2.9. Les résultats du coefficient de corrélation R ainsi que du test de significativité P de la corrélation entre une moyenne supérieure strictement à 12 et la moyenne de toutes les notes obtenues dans une année scolaire montrent l'intérêt que nous devons porter à cette étude. De plus, pour le train 7, la corrélation entre la moyenne et la 2ème variable (voir 2.4.2) donne $P = 0.0033$ ce qui est inférieur au seuil fixé à 0.05.

	Train 7	**Train 8**
Variable 1	R = 0.46 P = 0.0	R = 0.36 P = 0.0
Variable 2	R = - 0.07 P = 0.033	R = -0.03 P = 0.32
Variable 3	R = 0.04 P = 0.25	R = 0.06 P = 0.11

TABLE 2.9 – Résultats des corrélations entre une moyenne supérieure strictement à 12 et toutes les variables

Cas de la deuxième catégorie : les résultats de corrélation de cette catégorie sont donnés par la Table 2.10. Ces résultats sont assez pertinents et montrent l'intérêt d'étudier les corrélations entre la 2ème catégorie de cette définition des groupes d'élèves et les variables définies dans la partie 2.4.2. Pour le train 7, le coefficient de corrélation entre une moyenne comprise entre 8 et 12 et la moyenne de toutes les notes obtenues (définie comme variable 1) est R=0.27 : c'est une valeur positive qui tend vers 1. Et le test de significativité P=0.0 est inférieur au seuil défini par 0.05.

	Train 7	**Train 8**
Variable 1	R = 0.27 P = 0.0	R = 0.25 P = 0.0003
Variable 2	R = 0.09 P = 0.1368	R = 0.12 P = 0.1010
Variable 3	R = 0.24 P = 0.0	R = 0.12 P = 0.0128

TABLE 2.10 – Résultats des corrélations entre une moyenne comprise entre 8 et 12 et toutes les variables

Cas de la troisième catégorie : La table 2.11 présente les résultats de la corrélation de cette catégorie d'élèves. À travers les valeurs de coefficient R et de test de significativité présentées, ce tableau montre l'interêt à apporter aux études des corrélation entre les moyennes obtenues par cette catégorie d'élèves et les variables mentionnées dans 2.11. Par exemple, la corrélation étudiée pour le cas du train 8 entre une moyenne strictement inférieure à 8 dans la matière SP31 et le nombre de devoirs passés dans cette dernière donne un coefficient de *Pearson* $R = 0.21$ et un test de significativité $P = 0.0091$ qui est inférieur au seuil 0.05.

2.4.2.2 Groupes d'élèves selon le seuil 10

La 2ème définition est établie selon le seuil d'une moyenne égale à 10 et se compose de deux catégories d'élèves :

	Train 7	**Train 8**
Variable 1	R = 0.27 P = 0.0001	R = 0.15 P = 0.066
Variable 2	R = 0.09 P = 0.20	R = 0.21 P = 0.0091
Variable 3	R = 0.23 P = 0.0013	R = 0.25 P = 0.0017

TABLE 2.11 – Résultats des corrélations entre une moyenne inférieure strictement à 8 et toutes les variables

— catégorie 1 : les élèves ayant une moyenne supérieure ou égale à 10
— catégorie 2 : les élèves ayant une moyenne inférieure strictement à 10

	Train 7	**Train 8**
Catégorie 1	1009	971
Catégorie 2	302	208

TABLE 2.12 – Nombre d'inscrits de SP31 appartenant à chaque catégorie de la deuxième définition des groupes d'élève

Le tableau 2.12 présente, pour chaque train scolaire, le nombre d'élèves appartenant à chaque catégorie de la deuxième définition des groupes d'élèves inscrits à la matière SP31.

Cas de la première catégorie : Les résultats des corrélations entre une moyenne supérieure ou égale à 10 et les variables, mentionnées dans 2.4.2, présentés dans la table 2.13 sont intéressants.
Par exemple, la corrélation entre la moyenne obtenue par cette catégorie d'élèves et la variable 3 donne des tests de significativité $P = 0.0025$ et $P = 0.0$ (inférieurs au seuil 0.05) pour le train 7 et 8 respectivement.

	Train 7	**Train 8**
Variable 1	R = 0.52 P = 0.0	R = 0.43 P = 0.0
Variable 2	R = -0.04 P = 0.197	R = 0.01 P = 0.68
Variable 3	R = 0.09 P = 0.0025	R = 0.14 P = 0.0

TABLE 2.13 – Résultats des corrélations entre une moyenne inférieure ou égale à 10 et toutes les variables

Cas de la deuxième catégorie : Dans la table 2.14, nous obtenons aussi des résultats des corrélations entre une moyenne inférieure à 10 et les variables mentionnées dans 2.4.2.
Pour le cas du train 7, la corrélation entre les moyennes de cette catégorie d'élèves et le nombre de devoirs passés dans SP31 donne un coefficient de Pearson $R = 0.12$ qui est positif et tend vers 1 et un test de significativité $P = 0.03$.

Ces études diagnostics montrent les corrélations qui existent entre les variables définies et la réussite des apprenants. En d'autre terme, cette étude prouve l'impact de la variation des valeurs des variables

	Train 7	Train 8
Variable 1	R = 0.41 P = 0.0	R = 0.21 P = 0.0028
Variable 2	R = 0.12 P = 0.03	R = 0.11 P = 0.09
Variable 3	R = 0.29 P = 0.0	R = 0.25 P = 0.0

TABLE 2.14 – Résultats des corrélations entre une moyenne inférieure à 10 et toutes les variables

définies sur le résultat final des apprenants.
De plus, les groupes d'élèves identifiés selon différents seuils permettent de donner une définition à la réussite scolaire.

2.5 Construction de la base de données des expérimentations

Les étapes de l'analyse descriptive et diagnostic servent à comprendre le contexte métier et étudier les corrélations entre les variables extraites à partir des données brutes. Dans les résultats de cette partie, nous avons pris en compte les données disponibles de toute la population de la classe 3ème collège des années 2017-2018 et 2018-2019 du site du Rouen. Ceci a permis une compréhension plus vaste du contexte et des comportements existants. Or, l'apprentissage au CNED est assez particulier (comme décrit dans la section 2.2.2) car les apprenants inscrits dans une même classe ne suivent pas forcément le même rythme académique pour terminer le programme de leur formation. En d'autres termes, afin de garantir l'accès à l'école à un maximum d'élèves, le CNED propose la flexibilité d'inscription et d'apprentissage. Nous remarquons dans la table 2.2 que l'inscription reste ouverte jusqu'à la fin du mois de mai de chaque année. Cette flexibilité d'inscription conduit à passer par des adaptations de programme pour que les apprenants aient les mêmes chances d'assimilation du contenu et de réussite scolaire. Une inscription est considérée tardive lorsqu'elle est faite après le 31 octobre de chaque année scolaire. Le CNED considère qu'un apprenant qui rejoint une cohorte après cette date ne pourrait plus rattraper et suivre le rythme de ses pairs. Cet apprenant passe alors par un ajustement de programme lui permettant un meilleur apprentissage et des meilleurs résultats.
Compte-tenu de ces informations fournies par le CNED, nous construisons notre jeu de données d'expérimentations qui se compose des apprenants qui se sont inscrits avant le 31 octobre des années 2017-2018 et 2018-2019. Nous justifions ce choix par la disponibilité des données et traces d'apprentissage de ces apprenants pour effectuer un large spectre d'expérimentations et avoir des résultats interprétables.
Par conséquent le jeu de données construit se compose des traces d'apprentissage de :

— 725 apprenants qui se sont inscrits avant le 31/10/2017 (Train 7)
— 762 apprenants qui se sont inscrits avant le 31/10/2018 (Train 8)

Dans tous les chapitres suivants, la matière principale de test et des expérimentations sera celle de la Physique-Chimie de nomenclature SP31. Cette matière est classifiée comme matière obligatoire (voir Table 2.5). De ce fait, nous avons plus d'inscriptions et de traces numériques d'apprentissage dans cette matière. Pour des objectifs de validation des approches, nous testons les résultats obtenus sur deux autres matières qui sont les mathématiques, obligatoire et de discipline scientifique, et histoire-géographie, obligatoire et de discipline littéraire. Le choix sur les matières obligatoires se fait toujours dans le but d'avoir plus de données d'expérimentations.

2.6 Conclusion

Les institutions d'enseignement en ligne se caractérisent très souvent par la multitude de plateformes dont elles disposent pour divers objectifs d'apprentissage, de gestion administrative, de notations et bien d'autres. Le CNED aussi, étant le plus large centre français d'enseignement à distance, ne fait pas l'exception et possède plusieurs applications par lesquelles le contenu pédagogique est proposé et la communication enseignant-élève est assurée.

Dans ce travail, nous nous intéressons à l'étude des données et traces d'apprentissage issues de deux applications du CNED qui s'avèrent les plus intéressantes pour nos objectifs.

Dans tout projet de LA, le travail sur la compréhension du mécanisme du contexte métier et l'exploration des données : compréhension, nettoyage, analyse, fusion dans un même entrepôt est la partie la plus délicate et coûteuse en temps. Le travail effectué sur les données du CNED dans ce projet de thèse ne fait pas l'exception surtout que l'apprentissage au CNED est assez particulier et que sa base de données brutes est volumineuse.

Étant donné les caractéristiques du CNED et le nombre d'inscriptions qui augmente au quotidien, les enseignants du CNED ne sont plus en mesure de fournir un suivi régulier à leurs apprenants et les assister pendant toutes les étapes de leur scolarité ; ceci est l'un des facteurs expliquant les taux élevés d'échec au CNED. L'objectif est donc de fournir un système permettant la prédiction précoce et précise des apprenants à risque.

En nous basant sur les résultats des analyses descriptive et diagnostic réalisées, nous proposons dans les chapitres suivants des méthodologies, approches et des méthodes d'évaluation pour apporter les meilleurs résultats à cet objectif.

Chapitre 3

Une méthodologie en profondeur

Synthèse
Ce présent chapitre s'appuie sur les notions des sciences de l'éducation et les approches de LA de l'état de l'art pour proposer une méthodologie qui traite la problématique de la prédiction précoce et précise du risque.
La méthodologie proposée met l'accent sur les processus d'extraction et de sélection des caractéristiques et représente l'activité des apprenants en ligne à l'aide de quatre indicateurs d'apprentissage : performance, engagement, régularité et réactivité. De plus, l'étape de l'évaluation a permis d'identifier le modèle qui prédit au plus tôt et avec précision les apprenants à risque.
À travers la mise en valeur des données temporelles, ce chapitre prépare la contribution suivante liée à l'importance de l'évaluation temporelle dans le LA.

3.1 Introduction

Dans le contexte d'un établissement d'enseignement entièrement à distance, les données sont généralement multi-sources vu la disponibilité de plusieurs applications d'apprentissage et d'administration. Ces applications génèrent des données hétérogènes et de différents types. L'hétérogénéité des données est expliquée par le fait de disposer des données administratives décrivant la démographie des apprenants, des traces numériques des activités des apprenants et aussi des données de performance académique. D'un côté, ces données peuvent être indépendantes du temps et stables ; leurs valeurs correspondantes sont fixes. De l'autre côté, nous trouvons des données dépendantes du temps, dont les valeurs évoluent et changent progressivement pendant la période d'apprentissage.
Compte-tenu du volume des données et leur diversité et du nombre d'apprenants à suivre par chaque formateur, les enseignants ne sont plus en mesure d'accompagner tous leurs apprenants en même temps et de les assister avec un suivi pédagogique adapté à la situation de chacun d'entre eux ; d'où l'intérêt de concevoir un système dédié aux enseignants répondant à leur besoins d'accompagnement et de suivi régulier. Dans ce contexte, la question de recherche que nous nous sommes posées est la suivante : **comment représenter au mieux l'activité d'un apprenant inscrit dans un établissement d'apprentissage en ligne ?**

Dans ce chapitre, nous proposons une méthodologie en profondeur dont l'objectif est la prédiction précise et au plus tôt des apprenants à risque d'échec en se basant sur l'exploitation des traces numériques et l'utilisation des modèles d'apprentissage automatique (ML) supervisé. Après l'étape de la collecte et de nettoyage des données, cette méthodologie commence par l'identification des indicateurs d'apprentissage les plus pertinents parmi la performance, l'engagement, la régularité et la réactivité. Ensuite, et sur la base de ces indicateurs, les deux étapes suivantes correspondent respectivement à l'extraction et

la sélection des caractéristiques représentant au mieux l'activité de l'apprenant en ligne. Les dernières étapes correspondent à la modélisation et l'évaluation des performances des modèles de ML utilisés. En utilisant la mesure des taux de faux positifs (FPR), nous concluons sur le meilleur modèle ML parmi ceux testés qui prédit correctement les apprenants à risque. La profondeur de cette méthodologie est expliquée par la multitude d'étapes qui la composent et surtout le passage répété de certaines étapes, particulièrement la modélisation et l'évaluation des modèles de ML. L'objectif de ces étapes est de conclure sur la pertinence de certains indicateurs d'apprentissage et d'identifier les modèles de ML avec les meilleures capacités à détecter le plus précisément les apprenants à risque.

À travers ce chapitre, nous mettons aussi en valeur l'importance de l'intégration des caractéristiques de nature temporelle pour améliorer les résultats prédictifs des modèles de ML et accélérer l'obtention des pics de performance.

À ces fins, nous utilisons un échantillon de données réelles des apprenants de classe de 3ème suivant la même matière en ligne qui est celle de la physique-chimie (voir Chapitre 2). Nous identifions les indicateurs d'apprentissage et extrayons les caractéristiques qui les définissent à partir de deux applications disponibles (MOODLE et GAEL). Ensuite, nous suivons une approche de prédiction hebdomadaire et formalisons le problème en un problème de classification à 3 classes : réussite, risque d'échec moyen et risque d'échec élevé. Les modèles ML entraînés et testés dans les différentes expérimentations sont : les k-plus proches voisins (*k-nearest neighbours (k-nn)*), la machine à vecteurs de support (*Support Vector Machine (SVM)*), la forêt aléatoire (*Random Forest (RF)*) et l'arbre de décision (*Decision Tree (DT)*). Ces modèles sont peu complexes, très utilisés dans la littérature et montrent des bonnes performances de prédiction. Plusieurs techniques de sélection de caractéristiques sont appliquées pour identifier les caractéristiques les plus représentatives de l'activité d'un apprenant. Les techniques des méthodes filtrantes et enveloppantes [8] donnent des résultats très prometteurs. De plus, l'évolution de la mesure de FPR prouve que le modèle DT montre une bonne capacité à prédire les apprenants à risque lors des premières semaines de prédiction. Contrairement à la majorité des travaux de l'état de l'art, l'évaluation des performances des modèles, à l'aide de l'exactitude et le *FPR*, sera basée sur des courbes évolutives dans le temps. Comme l'apprentissage est un processus dans le temps, ces courbes permettent d'observer l'évolution des performances d'un système prédictif dans le temps.

Pour des objectifs de validation des résultats trouvés, nous utilisons les traces numériques collectées de deux autres matières et appliquons la même démarche de la méthodologie proposée.

3.2 Description de la méthodologie proposée

Dans cette section, nous présentons une description de la méthodologie proposée. Cette méthodologie est valable pour tout projet de LA dont l'objectif est la prédiction précoce et précise des apprenants à risque. L'objectif et le(s) type(s) de(s) résultat(s) attendu(s) suite à l'application de chacune des étapes qui la composent sont aussi définis et décrits.

Les données sont essentielles pour tout projet de LA et constituent le point de départ de la méthodologie que nous proposons. Comme le montre la Figure 3.1, la 1ère étape d'un projet de LA consiste à collecter les différentes traces numériques de toutes les sources de données disponibles, étudier les relations entre toutes les tables de données pour les fusionner et les placer dans un même entrepôt. Il faut ensuite nettoyer ces données de tout bruit pouvant influencer la qualité de la prédiction. Le passage par cette première phase est très coûteux en temps mais nécessaire pour avoir des données brutes prêtes à être utilisées, exploitées et analysées dans la 2ème phase.

8. Les méthodes filtrantes et enveloppantes sont deux méthodes de sélection de caractéristiques (voir la section 3.2.2)

L'étape suivante est l'extraction des caractéristiques qui se base, en premier lieu, sur l'ingénierie des caractéristiques (en anglais *feature engineering)*. C'est une tâche centrale dans la préparation des données pour le ML. Il s'agit de la construction des caractéristiques appropriées à partir des données brutes conduisant à des meilleures performances prédictives. Ce processus implique l'application de fonctions de transformation telles que des opérateurs arithmétiques et d'agrégation sur des données primaires pour en générer des nouvelles [125]. Les transformations permettent de mettre à l'échelle une caractéristique ou de convertir une relation non linéaire entre une caractéristique et une classe cible en une relation linéaire, plus facile à apprendre [125]. L'objectif de cette deuxième phase est de préparer un nouveau jeu de données composé d'un ensemble de caractéristiques représentant les indicateurs d'apprentissage exprimant l'activité d'apprentissage des apprenants.

La 3ème étape est l'exploitation d'un ensemble d'algorithmes de classification de ML pour des objectifs de *benchmark* et de comparaison de performance. Pour cette étape, il faut commencer par distinguer les variables indépendantes de la variable cible. Cette variable cible est de type catégorie ; il s'agit d'une étiquette de classe. Ensuite, chaque modèle prend en entrée l'ensemble des caractéristiques calculées précédemment (dans la phase 2) et donne en sortie comme résultat de prédiction la classe prédite pour chaque apprenant.

Les résultats de cette prédiction passent ensuite à l'étape de l'évaluation. Cette première évaluation se fait suivant une mesure de performance générale illustrant la performance globale du modèle : c'est la mesure de l'exactitude (voir section 1.7.1.2). Cette étape sert à comparer les performances des algorithmes de ML sélectionnés pour suivre l'évolution des valeurs de leur exactitude au cours du temps. Sur la base des résultats de cette phase, nous identifions et sélectionnons pour les étapes suivantes les algorithmes avec des valeurs d'exactitude croissantes dans le temps.

L'étape suivante concerne le processus de sélection des caractéristiques afin d'identifier les plus intéressantes et pertinentes pour la prédiction. Cette étape possède deux objectifs. D'une part, elle permet l'amélioration de la performance des algorithmes de prédiction en minimisant le bruit et enlevant les caractéristiques peu importantes pour le modèle en question. D'autre part, la sélection des caractéristiques permet surtout de conclure sur les indicateurs d'apprentissage les plus représentatifs de l'activité de l'apprenant.
Les caractéristiques sélectionnées constituent la nouvelle base de données de la seconde phase de modélisation. Le passage de nouveau par cette phase permet de vérifier la pertinence des caractéristiques sélectionnées en suivant l'évolution et les changements que pourraient avoir les valeurs d'exactitude.

Les résultats de prédiction de cette dernière modélisation passent encore une fois par une étape d'évaluation. L'objectif étant la prédiction des apprenants à risque, nous avons besoin d'une mesure plus spécifique évaluant les performances des algorithmes à bien prédire cette classe d'apprenants. La mesure des taux des faux positifs (*False Positive Rate FPR*) sert à évaluer la capacité d'un modèle à détecter les apprenants à risque d'échec.

Cette méthodologie se caractérise par sa profondeur et ses phases multiples pour pouvoir identifier à la fois les caractéristiques et indicateurs d'apprentissage les plus représentatifs de l'activité de l'apprenant et le(s) modèle(s) qui prédisent le mieux la classe d'apprenants à risque. Pour résumer, la profondeur de cette méthodologie est expliquée par :

- le **benchmark** des algorithmes de classification
- le **passage par les deux processus d'extraction et de sélection** des caractéristiques pour identifier celles les plus pertinentes pour la prédiction et définir les indicateurs d'apprentissage les plus représentatifs de l'activité d'un apprenant
- la **répétitivité** de la modélisation et de l'évaluation (en comparant les résultats de plusieurs métriques de performance) des algorithmes pour identifier celui ou ceux avec la meilleure capacité à prédire la classe d'apprenants à risque.

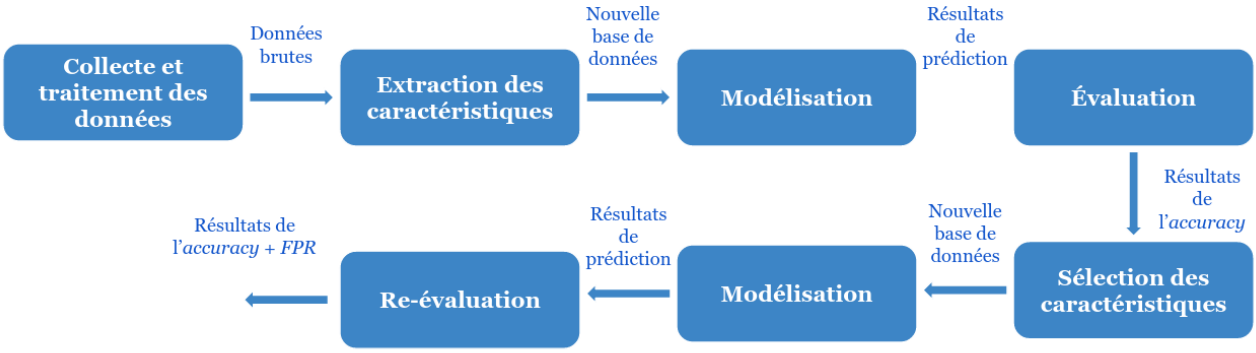


FIGURE 3.1 – Les étapes de la méthodologie proposée

Dans la suite de cette section, nous détaillons davantage toutes les phases de cette méthodologie.

3.2.1 Extraction des caractéristiques

Par définition, une caractéristique est une représentation des données brutes. En d'autre terme, une caractéristique est calculée suite à un ensemble des opérations réalisées sur une ou plusieurs données brutes. Les caractéristiques d'entrée jouent un rôle très important dans l'apprentissage des modèles de ML. Par conséquent, l'extraction des caractéristiques est une tâche centrale dans tout projet de ML, comme illustré par la Figure 3.1.
L'extraction des caractéristiques dépend fortement des indicateurs d'apprentissage. De ce fait, l'idée est de commencer par définir l'activité d'un apprenant à l'aide des indicateurs sur lesquels se base le travail de processus d'extraction des caractéristiques.

Iksal[126] définit un indicateur ainsi : "un indicateur est un observable signifiant sur le plan pédagogique, calculé ou établi à l'aide d'observés, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un Environnement Informatique pour l'Apprentissage Humain (EIAH). Il est défini en fonction d'un objectif d'observation et motivé par un objectif pédagogique". Un indicateur d'apprentissage répond donc à cette définition et est défini par un sous-ensemble de caractéristiques.

L'identification des indicateurs d'apprentissage ci-dessous a été établie sur la base d'une étude approfondie du profil comportemental d'un apprenant à risque donné par les sciences de l'éducation ainsi que sur les résultats de multiples projets de LA [12] [51] [75]. De plus, nous avons pris en considération la spécificité des outils et données disponibles dans le contexte d'une formation en ligne ainsi que la particularité du profil d'un apprenant en totale autonomie :

- **performance :** c'est l'indicateur d'apprentissage le plus utilisé et le plus répandu dans les travaux de l'état de l'art. Il est représenté par les caractéristiques liées aux notes et examens qui sont fortement corrélées avec le résultat d'un apprenant. De plus, les enseignants utilisent toujours, en premier lieu, les notes pour évaluer leurs apprenants et les établissements s'appuient sur les résultats des examens pour évaluer l'apprentissage [51] [12]
- **engagement :** c'est l'indicateur qui traduit l'implication d'un apprenant dans son travail. Dans le contexte d'un apprentissage en ligne, cette implication est définie par la présence sur la plateforme d'apprentissage et l'activité en ligne. En effet, les caractéristiques calculées à partir de la participation des apprenants aux plateformes en ligne sont fréquemment utilisées dans la littérature pour exprimer leur engagement [70].
- **régularité :** les résultats de l'état de l'art montrent la corrélation entre la prédiction des résultats finaux des apprenants et leur régularité dans l'apprentissage. Cet indicateur possède plus d'une

définition selon le domaine de son application : les actions, le temps, ou une combinaison de deux. La régularité dans les actions correspond à des motifs répétitifs dans la séquence d'activité de l'apprenant. La régularité dans le temps correspond à des motifs répétitifs dans le calendrier des sessions d'étude. La régularité dans les deux domaines combinés est reflétée par les dépendances entre les types d'actions et leur temps d'occurrence [67]. Dans un contexte d'un apprentissage en ligne où les apprenants sont en autonomie totale, nous avons besoin d'observer l'évolution de l'activité d'un apprenant d'un temps de prédiction à un autre. C'est pour cette raison que nous donnons une nouvelle définition à la régularité comme suit : la régularité de la progression d'un temps de prédiction à un autre. Cette régularité de progression pourrait être mesurée en terme d'actions effectuées sur la plateforme ou en terme de devoirs déposés.

— **réactivité :** à notre connaissance, la réactivité est un indicateur d'apprentissage qui n'a pas été introduit dans la littérature des LA. En effet, contrairement à l'enseignement traditionnel où les calendriers de scolarité sont fixés et généralement imposés, chaque apprenant en ligne a son propre rythme d'apprentissage et est en totale autonomie. La réactivité dans le cadre d'un apprentissage en ligne correspond au temps nécessaire pour devenir actif dans les plateformes d'apprentissage et au respect des délais de soumission des devoirs. Nous définissons la réactivité comme la réponse à temps pour les différents événements liés au cours. Cet indicateur sert à analyser le comportement de l'apprenant et à le comparer à celui de ses pairs.

Pour chacun de ces indicateurs, nous extrayons des caractéristiques à partir des données brutes. Pour obtenir de telles caractéristiques, nous passons par de multiples calculs de données brutes tels que la composition et les combinaisons.

3.2.2 Sélection des caractéristiques

La qualité des caractéristiques est importante pour que les modèles de ML retournent des prédictions précises. Cependant, un ensemble de caractéristiques en entrée de grande dimension peut contenir des données et informations bruyantes, redondantes et non pertinentes pour la prédiction. La caractéristique qui n'est pas nécessaire à la précision du modèle est connue sous le nom de caractéristique non pertinente [127]. Des telles données affaiblissent les performances prédictives, provoquent un sur-apprentissage [9] et augmentent le taux d'erreur. La sélection des caractéristiques (*Feature selection (FS)*) est un processus dont l'objectif est la gestion de ce problème. C'est le processus qui sélectionne un sous-ensemble de caractéristiques pertinentes à partir de l'ensemble initial en fonction de la redondance et la pertinence [127].
À cette fin, plusieurs techniques sont utilisées dans les problèmes de classification qui se répartissent en trois catégories :

— **méthodes filtrantes :** c'est l'une des premières méthodes utilisées en sélection de caractéristiques. En se basant sur des tests statistiques, le modèle sélectionne de l'ensemble initial un sous-ensemble de k caractéristiques à définir. Ce sont les caractéristiques les plus corrélées avec la variable cible [127]. L'avantage de cette méthode est que l'évaluation des caractéristiques se fait indépendamment du classificateur.
— **méthodes enveloppantes :** à la différence des méthodes filtrantes, ces méthodes intègrent un classificateur au cœur de leur processus. Le but est de prendre en compte les performances de classification au cours de la sélection en utilisant un algorithme d'apprentissage [127].
— **méthodes intégrées :** ces méthodes ont l'objectif de sélectionner les caractéristiques les plus pertinentes pendant la phase de l'apprentissage. Ces méthodes sont connues par leur plus grande

9. C'est un phénomène très fréquent dans le ML et qui apparaît quand le modèle essaie de trop s'adapter dans les données d'entraînement. IL se produit donc lorsque l'écart entre l'erreur d'entraînement et l'erreur de test est trop large [128].

rapidité par rapport aux autres méthodes. La sélection de caractéristiques intégrées peut utiliser deux techniques : la régularisation et les méthodes basées sur l'arborescence [127].

3.2.3 Approche de modélisation

La prédiction précoce des apprenants à risque est généralement formalisée en un problème de classification [12][90][44] à m étiquettes de classe $Y = \{C_1, C_2, .., C_m\}$. Le nombre m de classes est différent d'un projet à un autre selon le contexte et les objectifs. Chaque classe $C_j \in Y$ est une variable catégorique. La définition de ces classes de Y est établie en se basant sur le système d'évaluation fixé par l'établissement pour lequel l'approche est appliquée.

En fonction des besoins de chaque projet ainsi que de la fréquence du suivi des activités requises par les enseignants, la granularité temporelle t_i pour l'application de l'approche change ; il s'agit de la période du temps à laquelle la prédiction se fait et se répète. Cette période peut être un jour, une semaine, un mois, un an …. Le choix de cette granularité temporelle entre t_i et t_{i+1} dépend de la nature des résultats attendue de son application. L'ensemble de temps de prédiction est $T = \{t_1, t_2, .., t_d\}$.

Pour représenter l'activité d'un apprenant $S_q \in S = \{S_1, S_2, .., S_p\}$ à un temps de prédiction t_i, toutes les caractéristiques des indicateurs d'apprentissage sont regroupées dans un même vecteur $X_{q_{t_i}}$. Ainsi, à chaque $t_i \in T$, un apprenant S_q est représenté par un vecteur $X_{q_{t_i}}$ composé de toutes les caractéristiques f_1 jusqu'à f_n suivies de l'étiquette de classe C_j à laquelle il appartient.

$$X_{q_{t_i}} =< f_1, f_2, .., f_n, C_j >_{q_{t_i}}$$

Chacune de ces caractéristiques $f_k \in \mathbb{R}$ représente une activité d'apprentissage jusqu'au temps de prédiction t_i. À chaque temps de prédiction t_i, la valeur d'une caractéristique est ajoutée à celle du temps de prédiction t_{i-1} : nous procédons à une accumulation de valeurs.

3.2.4 Intégration d'une caractéristique temporelle dans le vecteur de l'apprenant

L'apprentissage est un processus qui se caractérise par la continuité et l'évolution au cours du temps. Le comportement présent de l'apprenant dépend de son passé et de l'historique de ses résultats. Ceci dit, la prédiction des classes des apprenants est aussi un processus évolutif dans le temps. En d'autres termes, la prédiction actuelle dépend du comportement de l'apprenant ainsi que du résultat de la classification qui la précède. Il est donc important d'intégrer la dimension temporelle de la prédiction dans le vecteur d'activité $X_{q_{t_i}}$ de l'apprenant au temps de prédiction t_i. Nous proposons d'illustrer cette continuité de prédiction à travers l'ajout d'une autre caractéristique temporelle qui est celle de la classe prédite au temps t_{i-1}. Plus explicitement, à chaque temps de prédiction t_i, le vecteur d'un apprenant $X_{q_{t_i}}$ se compose désormais, en plus des caractéristiques d'activité (performance, engagement, régularité et réactivité), de la classe $C_{p_{t_{(i-1)}}}$: c'est la classe prédite au temps t_{i-1}. Par conséquent, à chaque temps de prédiction t_i, chaque apprenant est représenté par son vecteur $X_{q_{t_i}}$ ainsi :

$$X_{q_{t_i}} =< f_1, f_2, .., f_n, C_{p_{t_{(i-1)}}}, C_j >_{q_{t_i}}$$

Où chacune des caractéristiques $f_k \in \mathbb{R}$ représente une activité d'apprentissage jusqu'au temps de prédiction t_i et $C_{p_{t_{(i-1)}}} \in Y$ est le résultat de la prédiction au temps t_{i-1}. C_j est l'étiquette de la classe réelle à laquelle un apprenant appartient et que nous cherchons à prédire.

D'un côté, cette intégration de la classe prédite au temps (t_{i-1}) sert à illustrer la continuité de la prédiction lorsqu'il s'agit d'un processus évolutif dans le temps comme celui de l'apprentissage. De l'autre côté, elle participe à l'amélioration des résultats de prédiction (voir Section 1.5.2 du chapitre 1).

3.3 Application de la méthodologie sur le contexte du CNED

Dans cette section, nous appliquons les étapes de la méthodologie décrite dans la Section 3.2 sur le contexte du CNED.
L'étape de la collecte et de traitement des données a été présenté et décrite dans le Chapitre 2.
Pour rappel, nous étudions les traces numériques de la classe 3ème du collège pendant les deux années scolaires 2017-2018 et 2018-2019 durant 37 et 35 semaines respectivement. Les apprenants de 2017-2018 et 2018-2019 sont respectivement au nombre 663 et 698.

3.3.1 Extraction des caractéristiques de la base de données CNED

Dans cette partie, nous procédons à la 2ème étape de la méthodologie proposée qui est celle de l'extraction des caractéristiques. Pour ce faire, nous nous basons sur l'ensemble de définitions et notions de la section 3.2.1 donnant une description générale de tout le processus.

Dans le cadre du CNED, en plus des données démographiques fournies par l'application GAEL, l'activité d'un apprenant est représentée par les quatre indicateurs pédagogiques introduits dans la section 3.2.1. Sur la base des caractéristiques disponibles et extraites de deux sources de données, ces indicateurs sont définis comme suit :

- **performance :** actuellement, les enseignants du CNED se basent sur les notes obtenues aux examens pour évaluer leurs apprenants. C'est pour cette raison que la performance d'un apprenant garde cette définition et est calculée à partir des caractéristiques liées à l'évaluation académiques et aux notes : nombre de devoirs soumis et moyenne de notes obtenues jusqu'à la date de prédiction. Ces caractéristiques ont tendance à changer de valeurs d'un temps de prédiction à un autre ; elles sont dites évolutives dans le temps.
- **engagement :** les apprenants du CNED étant en totale autonomie, le seul moyen de suivre leur engagement et implication dans les études est le taux de leur participation et présence sur la plateforme en ligne. De plus, bien que l'apprentissage au CNED reste encore multi-modal, les enseignants adoptent l'idée de l'apprentissage via les plateformes en ligne et incitent leurs apprenants surtout les collégiens à être plus actifs sur le LMS. Par conséquent, l'engagement est un indicateur d'apprentissage très important pour le suivi de l'activité des apprenants et représenté par 36 caractéristiques extraites de la base du CNED. Ces caractéristiques sont aussi dites évolutives comme elles peuvent changer de valeur d'un temps de prédiction à un autre.
- **régularité :** comme défini dans 3.2.1, la régularité est l'avancement réalisé par un apprenant d'un temps de prédiction à un autre. Comme le CNED se caractérise par la totale autonomie de ses apprenants, cet avancement est défini en termes de nombre d'actions réalisées au sein du LMS et nombre d'examens soumis. De ce fait, la régularité est représentée par deux caractéristiques dont les valeurs ne sont pas fixes et changent d'un temps de prédiction à un autre.
- **réactivité :** ce nouvel indicateur est introduit dans la Section 3.2.1 et il sert à comparer l'apprenant avec ses pairs. La réactivité est représentée par des caractéristiques exprimant l'aptitude d'un apprenant à soumettre un examen à temps ou à se connecter assez rapidement à la plateforme en ligne. La réactivité est donc représentée en total avec 7 caractéristiques dans le contexte du CNED.

Ainsi, chaque apprenant est représenté par 10 caractéristiques démographiques (constantes) et des indicateurs d'apprentissage représentés par les caractéristiques extraites à partir des données brutes. Au total, chaque apprenant est défini par 58 caractéristiques.
Les caractéristiques extraites et décrivant les indicateurs d'apprentissage sont assez diverses (certaines

sont constantes et d'autres changent dans le temps) et hétérogènes.
Toutes les caractéristiques extraites sont détaillées dans l'Annexe A.

3.3.2 Application de l'approche de modélisation au contexte du CNED

Les enseignants du CNED ont besoin d'un suivi régulier et fréquent de leurs apprenants. Par conséquent, la granularité temporelle choisie pour ce contexte est la semaine d'activité, $t_i \in T = semaine$. Ainsi, la prédiction des apprenants en difficulté scolaire se fait sur une base hebdomadaire.

Comme l'inscription au CNED reste ouverte et possible jusqu'à très tard dans l'année scolaire, les apprenants, même s'ils appartiennent à une même classe, ne commencent pas leur programme en même temps. Ainsi, la date du début d'activité t_0 pourrait être différente d'un apprenant à un autre. De ce fait, deux apprenants, ayant commencé l'année scolaire à des dates différentes, ne peuvent pas avoir les mêmes dates de prédiction. Par conséquent, pour chaque apprenant S_q, les semaines de prédiction dépendent de la date de son début d'activité t_{0_q}. Plus explicitement, la 1ère semaine (qui correspond à la 1ère date) de prédiction d'un apprenant $S_q \in S$ est $t_{1_q} = t_{0_q} + 7jours$, la 2ème semaine (qui correspond à la 2ème date) de prédiction est $t_{2_q} = t_{1_q} + 7jours$ et ainsi de suite jusqu'à ce que t_{d_q} corresponde à la date de la fin de l'année scolaire.

À l'exception des caractéristiques démographiques, dont les valeurs sont constantes et indépendantes du temps, le reste des caractéristiques représentant l'activité d'un apprenant est mis à jour hebdomadairement (à chaque date de prédiction). L'ensemble de caractéristiques démographiques et des indicateurs d'apprentissage, exprimés à travers les caractéristiques extraites de la base du CNED, est regroupé dans un même vecteur $X_{q_{t_i}}$ traduisant l'activité hebdomadaire de l'apprenant S_q au temps de prédiction t_i.

Pour le mode d'évaluation, le CNED suit les standards du système éducatif français qui permet aux enseignants correcteurs d'attribuer une note comprise entre 0 et 20 à chaque examen / test. En effet, une moyenne de notes obtenue égale à 10 dans un module donné détermine généralement la réussite ou l'échec d'un apprenant dans ce même module.
Cependant, il est d'une très grande importance d'étudier et analyser davantage les apprenants dans la zone d'incertitude. En d'autres termes, ce sont les apprenants avec des moyennes ni assez élevées qui garantissent leur réussite et ni assez basses qui déterminent assez tôt dans l'année leur échec. Ce sont les apprenants de la zone de doute et avec lesquels une intervention pédagogique assez tôt pourrait être bénéfique pour les aider à améliorer leurs résultats finaux. Cette classe d'apprenants est particulièrement intéressante à suivre et à analyser le comportement durant toute l'année scolaire.
Par conséquent, pour chaque module, les apprenants de la base de données du CNED sont répartis en trois classes en fonction de la moyenne de notes obtenues à la fin de l'année scolaire :

- **la classe de réussite étiquetée C_1 :** cette classe regroupe les apprenants ayant obtenu une moyenne finale supérieure strictement à 12.
- **la classe de risque moyen étiquetée C_2 :** c'est la classe d'incertitude et qui regroupe les apprenants ayant obtenu une moyenne finale comprise entre 8 et 12.
- **la classe de risque élevé étiquetée C_3 :** cette classe regroupe les apprenants défaillants ayant obtenu une moyenne finale strictement inférieure à 8.

Pour chacune des années scolaires étudiées, le Tableau 3.1 donne le nombre d'apprenants appartenant à chacune des trois classes définies ci-dessus. Comme la plupart des travaux de l'état de l'art, la classe de réussite C_1 est la classe majoritaire et la classe du risque élevé C_3 est la classe minoritaire. Entre ces deux classes, nous trouvons la classe C_2 qui regroupe aux alentours de 100 apprenants pendant les deux années scolaires.

Classes d'apprenants / Année scolaire	C_1	C_2	C_3
2017-2018	488	111	64
2018-2019	538	101	59

TABLE 3.1 – Nombre d'apprenants par classe

Dans l'étape de la modélisation, nous cherchons à faire le *benchmark* de plusieurs modèles de classification de ML et à suivre leurs prédictions au cours du temps. Ce travail nous permettrait d'analyser les performances des différents algorithmes et de conclure sur le modèle ML qui prédit le mieux les classes des apprenants à risque. Donc, dans la partie expérimentale, nous procédons à la comparaison des performances de prédiction des modèles d'apprentissage supervisé suivants : forêt aléatoire (*Random Forest (RF)*), arbre de décision (*Decision Tree (DT)*), k-plus proches voisins (*K-nearest neighbours (K-nn)*) et machines à support vecteurs (*Support Vector Machines (SVM)*). Ces modèles se caractérisent par leur faible complexité et la facilité de leur compréhension et adaptation aux différents contextes et problèmes de classification. De plus, ils sont largement adoptés et utilisés dans l'état de l'art de LA et montrent des résultats de prédiction assez pertinents.

3.4 Implantation de la méthodologie : expérimentations et résultats

En nous appuyant sur la description théorique de la méthodologie, les résultats du processus de l'extraction des caractéristiques et l'approche de modélisation, nous procédons dans cette section à l'implantation des étapes de la modélisation et l'évaluation des classificateurs de ML pour le contexte du CNED.

3.4.1 Protocole expérimental

Dans cette section, nous présentons le protocole expérimental que nous avons suivi pour l'implantation et l'évaluation de la méthodologie.

À chaque temps de prédiction t_i, les caractéristiques d'entrée des modèles (RF, DT, K-nn, SVM) sont celles qui composent les vecteurs $X_{q_{t_i}}$, représentant chacun l'activité d'un apprenant S_q jusqu'au temps t_i de l'année scolaire. Ces modèles sont entraînés et testés en s'appuyant sur la technique de la validation croisée [10] à 5 blocs.
Pour évaluer la performance des modèles ML à donner une prédiction C_{pred} similaire à C_{test}, nous avons suivi une méthode composée de deux étapes permettant l'identification de(s) modèle(s) ayant la meilleure exactitude :

1. **1ère étape :** entraîner chaque modèle en sélectionnant au hasard 80% des vecteurs d'activité des apprenants inscrits pendant l'année scolaire 2017-2018. Ensuite, tester la performance du modèle en question avec les 20% des vecteurs d'activité restants.
2. **2ème étape :** afin de pouvoir confirmer les résultats de la 1ère étape, nous entraînons chaque modèle avec les vecteurs d'activité de l'année $n-1$ (qui est dans ce cas l'année scolaire 2017-2018) et le testons avec ceux de l'année scolaire n (2018-2019).

Des données en entrée assez volumineuses favorisent un meilleur apprentissage des modèles de ML ; d'où l'objectif de la 2ème étape de ce protocole expérimental. En effet, il faudrait modéliser les algorithmes

10. La validation croisée est une méthode statistique d'évaluation et de comparaison des algorithmes d'apprentissage en divisant les données en deux segments : l'un utilisé pour apprendre ou former un modèle et l'autre utilisé pour valider le modèle [129].

avec un plus large nombre de données (tous les vecteurs des caractéristiques des apprenants de l'année $n-1$) et s'assurer de ne pas avoir une perturbation ou chute de performance.

3.4.2 Modélisation et évaluation de l'exactitude

Les étapes de la modélisation et l'évaluation sont liées l'une à l'autre. En d'autres termes, les modèles apprennent le contexte, étudient et analysent, à partir de la masse de vecteurs d'activité passés en entrée, les relations entre les variables indépendantes (les caractéristiques extraites représentant les indicateurs d'apprentissage) et la variable dépendante (la classe de l'apprenant) ; il s'agit de la phase d'entraînement. Ensuite, le modèle teste son apprentissage avec des nouvelles données ; il s'agit de la phase de test. Les résultats de l'exactitude servent à évaluer les prédictions et la performance générale du modèle.

3.4.2.1 1ère étape du protocole expérimental

Nous entraînons les quatre modèles : RF, DT, k-nn et SVM et nous procédons à la 1ère étape de l'évaluation de leurs valeurs d'exactitude au cours des temps de prédiction. La Figure 3.2 présente les courbes d'évolution des valeurs d'exactitude des prédictions retournées par les modèles appris.

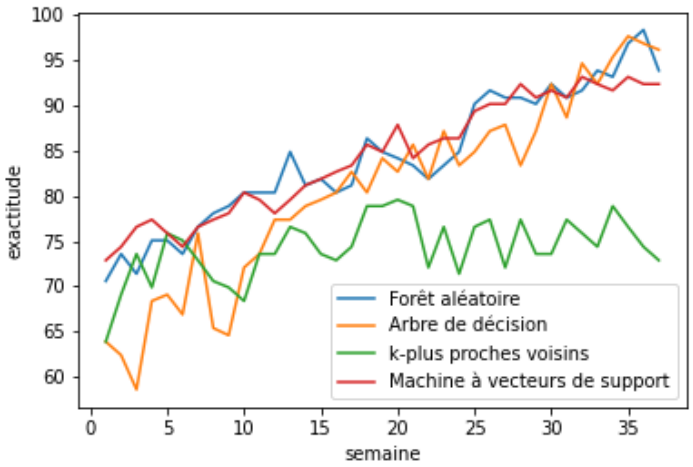


FIGURE 3.2 – Première étape de l'évaluation de l'exactitude

L'allure de la courbe des valeurs d'exactitude de k-nn montre que les prédictions de ce dernier sont instables au cours du temps. Nous remarquons que RF, DT et SVM sont les modèles les plus performants et avec des prédictions qui s'améliorent à travers les dates de prédiction. En effet, ils se caractérisent par des courbes d'exactitude croissantes au cours du temps.
À la 1ère date de prédiction, les pourcentages de la mesure d'exactitude de RF, DT et SVM respectivement sont : 70.6%, 63.9% et 72.9%. En analysant les résultats donnés par la Figure 3.2, nous remarquons qu'aux premières dates de prédiction, SVM semble être le plus performant avec les valeurs d'exactitude les plus élevées. À partir de la $semaine \approx 8$, SVM et RF retournent quasiment les mêmes valeurs d'exactitude. Cependant, d'après cette même figure, DT est le modèle le moins performant, jusqu'à la $semaine \approx 20$, en retournant les valeurs d'exactitude les plus basses en les comparant à celles retournées par RF et SVM.
Les valeurs d'exactitude les plus élevées ont été obtenues par RF (98.4%), DT (97.7%) et SVM (93.2%) aux semaines de 36, 35 et 32 respectivement. En avançant dans l'année scolaire, les apprenants laissent

plus de traces numériques, et, donc, les données sont de plus en plus disponibles. De ce fait, il est très pertinent d'obtenir ces valeurs d'exactitude à ces semaines de prédiction.

En résumé, cette première étape du protocole d'évaluation montre que les prédictions de RF, DT et SVM s'améliorent avec le temps ; ce sont les modèles que nous gardons pour la 2ème étape d'évaluation.

3.4.2.2 2ème étape du protocole expérimental

Comme expliqué ci-dessus, nous gardons, pour cette 2ème étape du protocole d'évaluation, les modèles : RF, DT et SVM et dont les résultats sont donnés par la Figure 3.2.

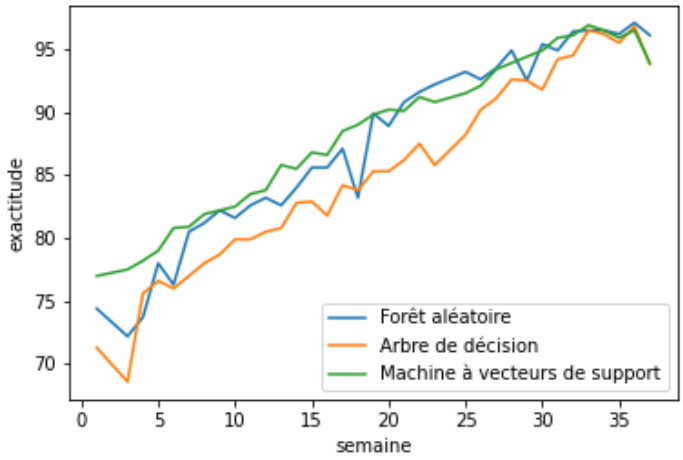


FIGURE 3.3 – Deuxième étape de l'évaluation de l'exactitude

À première vue, les courbes d'exactitude de RF, DT et SVM données par la Figure 3.3 gardent les mêmes allures croissantes que celles de la Figure 3.2. Les algorithmes sélectionnés sont pertinents. Comme pour la 1ère étape de ce protocole, les courbes d'exactitude de la Figure 3.3 montrent que, à la 1ère date de prédiction, SVM est le plus performant avec une exactitude$\approx 78\%$ et DT est le moins performant avec une exactitude$\approx 72\%$. Le modèle RF arrive au milieu de deux autres avec une exactitude$\approx 75\%$.

Le Tableau 3.2 présente les résultats de l'exactitude des prédictions des modèles RF, DT et SVM qui prennent en entrée seulement les caractéristiques démographiques des apprenants. Les première et deuxième lignes du tableau sont respectivement les résultats de l'application de la 1ère et 2ème étapes du protocole d'évaluation. Les résultats donnés par ce tableau prouvent que les performances de prédiction des algorithmes au cours des premières dates des Figures 3.2 et 3.3 sont dues aux caractéristiques démographiques. Ceci est expliqué par les traces numériques peu disponibles et l'inactivité des apprenants pendant les premières semaines de l'année scolaire.

Modèle / Étape de l'évaluation	RF	DT	SVM
1ère étape	0.608	0.655	0.756
2ème étape	0.713	0.614	0.77

TABLE 3.2 – Exactitude des modèles en utilisant uniquement les caractéristiques démographiques

Afin de gagner en dimension, en temps de calcul et en précision, nous procédons dans la prochaine section à l'étape suivante de notre méthodologie qui est la sélection des caractéristiques.

3.4.3 Processus de la sélection des caractéristiques

La littérature prouve que la sélection des caractéristiques aide à améliorer les performances de prédiction, l'évolutivité et la capacité de généralisation du modèle de classification. Dans la découverte des connaissances, ce processus joue un rôle fondamental dans la réduction de la complexité de calcul, du stockage et du coût.

Pour cette étude, nous suivons les courbes d'exactitude données par la Figure 3.2 et appliquons, pour chaque modèle, les techniques de sélection des caractéristiques sur les vecteurs d'apprentissage qui donnent la valeur maximale d'exactitude. Pour les prochaines expériences, nous entraînons et testons les modèles (RF, DT et SVM) avec les vecteurs des apprenants inscrits pendant l'année scolaire 2017-2018. Dans les parties suivantes, nous appliquons les différentes méthodes de sélection des caractéristiques et présentons leurs résultats.

3.4.3.1 Application des méthodes filtrantes

Dans cette partie, nous présentons les résultats de la sélection des caractéristiques données par deux techniques de la méthode filtrante. Il s'agit de l'application de deux tests statistiques les plus utilisés et connus qui sont : khi-deux et ANOVA [127].

Sélection des caractéristiques L'application de ces deux tests statistiques nécessite de fixer, au préalable, le nombre k de caractéristiques à sélectionner. Le choix de k se fait sur la base de la qualité des prédictions obtenues ; en d'autres termes, le nombre k optimal correspond aux caractéristiques avec les meilleures valeurs d'exactitude.
Dans cette étude, nous fixons le nombre de caractéristiques à sélectionner à $k = 20$ qui est la valeur optimale pour ce contexte du CNED. En effet, nous avons appliqué les deux tests, khi-deux et ANOVA, avec d'autres valeurs de k (supérieures et inférieures à 20) et qui n'ont pas donné des meilleurs résultats d'exactitude. Ainsi, à chaque semaine de prédiction, chaque apprenant S_q est désormais représenté par un vecteur $X_{q_{t_i}}$ de 20 caractéristiques d'apprentissage, en plus de son étiquette de classe (C_1, C_2, C_3).
La majorité des caractéristiques sélectionnées par les deux tests (khi-deux et ANOVA) représente l'indicateur d'*engagement*. En effet, ces caractéristiques concernent directement les taux d'activité de l'apprenant sur le LMS. De plus, les caractéristiques liées à l'indicateur de *performance*, telles que le nombre d'examens passés et la moyenne de notes obtenues jusqu'à la date de prédiction t_i, sont aussi sélectionnées par les deux tests.
Le test ANOVA sélectionne des caractéristiques qui définissent l'indicateur de *régularité* telles que la progression en nombre d'actions et en examens effectués par un apprenant par rapport à la date de prédiction précédente. Les caractéristiques démographiques sélectionnées par le test du khi-deux sont la ville et le pays de résidence et celles sélectionnées par ANOVA sont les caractéristiques concernant la bourse et redoublant ou non. À première vue, les résultats retournés par les modèles utilisant les caractéristiques sélectionnées par le test ANOVA sont meilleurs.

Suivi de l'exactitude Après avoir sélectionné les $k = 20$ caractéristiques par les deux tests, nous procédons à l'entraînement des modèles (RF, DT et SVM) et testons leur performance prédictive.
Nous avons suivi la prédiction des valeurs d'exactitude retournées par les trois modèles. Les résultats des prédictions au cours du temps suite à l'adoption des caractéristiques sélectionnées par khi-deux et ANOVA sont représentés respectivement par les courbes des figures 3.4 et 3.5.

D'après ces deux figures, nous pouvons constater qu'il n' y a pas une chute ou dégradation de l'exactitude des modèles. En effet, les courbes des figures 3.4 et 3.5 gardent les mêmes allures que celles de

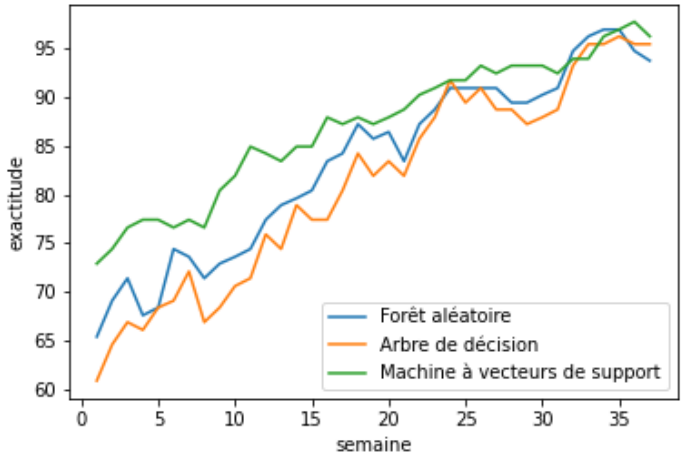


FIGURE 3.4 – L'évolution des valeurs d'*exactitude* suite à la sélection par le test khi-deux

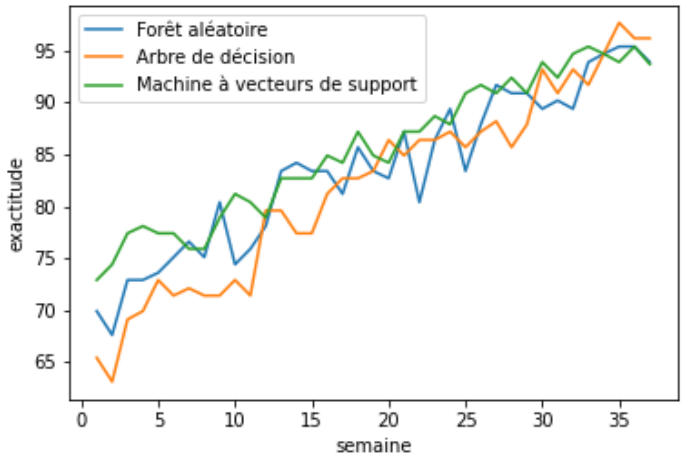


FIGURE 3.5 – L'évolution des valeurs d'*exactitude* suite à la sélection par le test ANOVA

3.2 et 3.3. Comme pour les résultats des sections 3.4.2.1 et 3.4.2.2, SVM est toujours le modèle avec les meilleures valeurs d'exactitude aux premières dates de prédiction et DT est celui qui montre les plus faibles valeurs d'exactitude.
Cependant, en entraînant et testant les modèles avec les caractéristiques sélectionnées par khi-deux et ANOVA, nous constatons une augmentation plus rapide des valeurs d'exactitude surtout pendant les premières dates de prédiction. De plus, pour certaines dates de prédiction, les caractéristiques d'entrée sélectionnées avec le test ANOVA donnent une exactitude de prédiction plus élevée que celles sélectionnées par le test khi-deux.

L'avantage de la méthode filtrante notamment la sélection basée sur les tests khi-deux et ANOVA est l'indépendance des modèles de ML. En effet, il s'agit d'une étude statistique des corrélations entre les variables indépendantes et la variable dépendante. D'après les résultats d'exactitude que nous venons d'obtenir, les caractéristiques sélectionnées par le test ANOVA, représentant les indicateurs d'engage-

ment, de performance et de régularité, sont pertinentes et retournent des valeurs d'exactitude élevées.

3.4.3.2 Application des méthodes enveloppantes

Dans cette partie, nous présentons les résultats de la sélection des caractéristiques données par une technique de la méthode enveloppante. Il s'agit de la technique de l'élimination récursive des caractéristiques avec validation croisée (en anglais : *The Recursive Feature Elimination with Cross Validation (RFECV)*).
La sélection des caractéristiques par la méthode enveloppante, notamment la technique de RFECV, se fait comme dans une boîte noire ; c'est-à-dire qu'il n'y a aucune connaissance de l'algorithme sous-jacent [127]. Des ensembles des caractéristiques sont sélectionnés sur la base d'algorithmes inductifs. Chaque ensemble de caractéristiques choisi considère plutôt l'exactitude du modèle pendant la phase d'entraînement. En fonction de l'exactitude mesurée à l'étape précédente, la méthode décidera d'ajouter ou de supprimer une caractéristique du sous-ensemble sélectionné. Pour cette raison, les méthodes enveloppantes sont plus complexes en termes de calcul [127].

Le nombre de caractéristiques sélectionnées par la techniques de RFECV diffère d'un modèle à un autre. En effet le nombre k de caractéristiques sélectionnées par cette technique avec RF, DT et SVM est respectivement 10, 29 et 13. Le nombre de caractéristiques sélectionnées le plus élevé a été repéré avec le modèle DT.
Avec les trois modèles, RFECV sélectionne des caractéristiques indiquant l'engagement d'apprenant, telles que le nombre d'actions effectuées sur le LMS jusqu'à la date de prédiction. Les caractéristiques de la performance de l'apprenant données par le nombre d'examens passés et la moyenne de notes obtenues jusqu'à la date de prédiction sont aussi sélectionnées. La réactivité de l'apprenant exprimée par le nombre de jours entre la date de début de l'activité et la date de la première connexion sur le LMS est sélectionnée par les trois modèles. Ceci confirme la forte corrélation entre cet indicateur pédagogique et le résultat final d'un apprenant. SVM et DT sélectionnent les caractéristiques de progression en nombre d'actions effectuées et examens passés, exprimant la régularité de l'apprentissage de l'apprenant.
Quant aux caractéristiques démographiques, le lieu de naissance et la ville de résidence sont sélectionnés par les trois modèles. RF et DT sélectionnent la caractéristique informant sur la bourse. SVM et DT sélectionnent les caractéristiques sur l'âge et le sexe.

Suivi de l'exactitude : les résultats de l'application de cette technique RFECV avec les trois modèles RF, DT et SVM sont donnés par la Figure 3.6.
À première vue, les courbes de cette figure gardent les mêmes allures que celles des figures 3.2 et 3.3. En effet, SVM est toujours le modèle avec les valeurs d'exactitude les plus élevées aux premières dates de prédiction. Alors que, DT est l'algorithme avec les plus faibles valeurs d'exactitude pendant les premières semaines de prédiction. Pendant toutes les dates de prédiction, il n'y pas une dégradation dans les valeurs d'exactitude par rapport aux résultats des sections 3.4.2.1 et 3.4.2.2 qui évaluent ces mêmes modèles appris avec toutes les caractéristiques. De plus, nous constatons que les caractéristiques d'entrée sélectionnées par la technique RFECV semblent avoir un meilleur impact sur l'exactitude de la prédiction des trois modèles. Autrement dit, les courbes de la Figure 3.6 montrent une augmentation plus rapide des valeurs d'exactitude aux premières dates de prédiction.
La méthode enveloppante à travers l'application de la technique de RFECV montre des bons résultats de prédiction. En effet, les caractéristiques sélectionnées par cette technique montrent, à la fois, leur pertinence et la pertinence des indicateurs d'apprentissage définis.

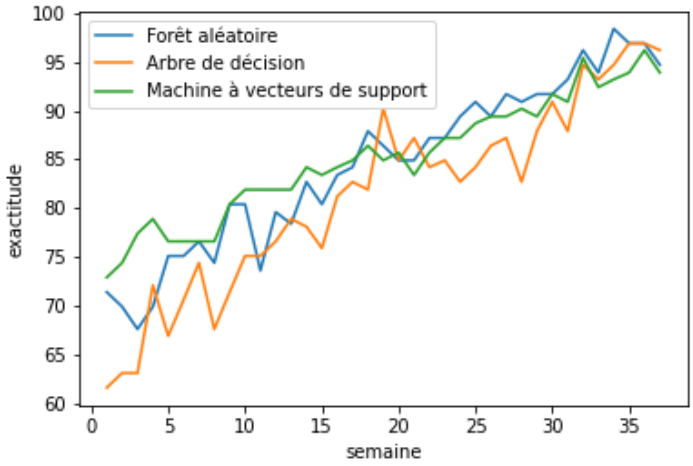


FIGURE 3.6 – L'évolution des valeurs d'exactitude suite à la sélection par la technique de RFECV

3.4.3.3 Application des méthodes intégrées

Jusqu'à présent, les méthodes de sélection de caractéristiques que nous avons testées ci-dessus sont utilisées au niveau de la phase du pré-traitement. Dans cette partie, nous appliquons une technique de la méthode intégrée qui fonctionne de manière à ce que les caractéristiques les plus pertinentes soient sélectionnées au cours de la phase d'apprentissage des modèles. Ceci présente les avantages d'améliorer le coût de calcul, l'exactitude de la classification et d'éviter également d'entraîner le modèle à chaque fois qu'une nouvelle caractéristique est ajoutée [127].

Grâce à leur structure puissante, les algorithmes basés sur les arbres se caractérisent par l'hyper-paramètre d'importance des caractéristiques qui sert à sélectionner les caractéristiques les plus importantes pour une prédiction précise. C'est pour cette raison, nous utilisons le RF (comme algorithme basé sur la notion des arbres de décision) pour la sélection des caractéristiques.
Afin d'entraîner RF, DT et SVM et tester leur exactitude de prédiction, l'algorithme basé sur la notion des arbres sélectionne 8 caractéristiques pertinentes. Les principales caractéristiques sélectionnées appartiennent aux indicateurs de performance, données par les notes et moyennes obtenues jusqu'à la date de prédiction, et de réactivité, exprimée par le temps que l'apprenant met pour se connecter à la plate-forme. Les caractéristiques démographiques retenues sont la ville de résidence, le lieu de naissance et l'âge.

Suivi de l'exactitude : les résultats d'exactitude de l'application de cette technique avec les trois modèles sont donnés par la Figure 3.7. Les courbes de suivi des valeurs d'exactitude au cours du temps gardent une même allure que celles des Figures 3.4.2.1 et 3.4.2.2 et ne montrent pas une dégradation de performance. Les caractéristiques sélectionnées avec cette technique dépendent du modèle arborescent et montrent une meilleure performance avec les modèles de la même catégorie.

3.4.3.4 Analyse des résultats de la première phase d'évaluation de la méthodologie

Dans les parties précédentes, nous avons procédé au processus de sélection des caractéristiques et appliqué ses différentes méthodes et techniques. Les caractéristiques qui définissent les indicateurs de performance et d'engagement sont sélectionnées par toutes les techniques testées ; d'où leur pertinence. Les caractéristiques exprimant la régularité de l'avancement d'un apprenant sont sélectionnées par les

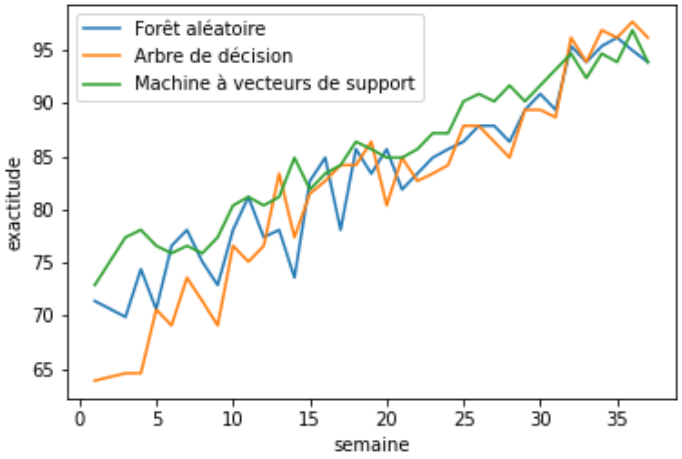


FIGURE 3.7 – L'évolution des valeurs d'exactitude suite à la sélection par la technique intégrée

techniques ANOVA (technique de méthode filtrante) et RFECV (technique de méthode enveloppante). Les caractéristiques de l'indicateur de réactivité sont sélectionnées par les techniques des méthodes enveloppantes et intégrées.

Aux premières dates de prédiction, en considérant les valeurs d'exactitude, SVM est le modèle avec les meilleures performances prédictives. À ces mêmes dates, DT est le modèle ML le moins performant, comme il présente les plus faibles valeurs d'exactitude. C'est le modèle RF qui arrive au juste milieu des deux autres modèles avec une performance pertinente et des valeurs d'exactitude élevées.
En entraînant et testant RF, DT et SVM avec les caractéristiques sélectionnées par les différentes techniques, nous constatons une amélioration dans la performance prédictive générale de ces modèles ; et c'est particulièrement pendant les premières dates de prédiction. En conclusion, passer par le processus de la sélection de caractéristiques permet de réduire la dimension des données en enlevant les données bruitées et de garder voire d'améliorer les prédictions des modèles.

Certaines techniques de sélection sont liées aux algorithmes de ML et d'autres sont des tests statistiques indépendants des algorithmes. Par exemple, le test ANOVA (de la méthode filtrante) sélectionne les caractéristiques qui sont corrélées avec la variable cible indépendamment des modèles. L'utilisation des caractéristiques sélectionnées par cette technique montre des valeurs d'exactitude élevées tout au long des dates de prédiction.

De plus, à travers les résultats d'exactitude suite à la sélection de caractéristiques, nous confirmons la pertinence des indicateurs d'apprentissage définis et identifiés dans la Section3.2.1. L'indicateur d'engagement est très pertinent et ses caractéristiques sont sélectionnées par toutes les techniques de sélection testées. Les nouvelles définitions données aux indicateurs de régularité et réactivité montrent leur pertinence. Leurs caractéristiques sont, en effet, sélectionnées par la majorité des techniques. Ces indicateurs d'apprentissage sont très importants pour suivre les résultats des apprenants et servent aussi dans les diagnostics des problèmes d'apprentissage chez les apprenants à risque moyen et élevé.

3.4.4 Ré-évaluation : résultats de la mesure FPR

Évaluer la performance de la classification en ne se focalisant pas sur une classe cible est un appui très général pour comparer les modèles de ML. La mesure d'exactitude ne précise pas le nombre d'individus de chaque classe ; elle retourne la performance globale du modèle. Pour un traitement optimal de

certaines problématiques comme celle de la prédiction précise des apprenants à risque, il est nécessaire d'aller en profondeur dans l'évaluation des performances prédictives des modèles ML en s'appuyant sur des métriques de performance plus spécifiques. Dans notre étude de cas, ces métriques doivent évaluer la performance des modèles ML à correctement prédire les apprenants à risque moyen et élevé (C_2 et C_3 respectivement). Ce sont ces classes d'apprenants qui nous intéressent le plus et c'est auprès d'eux que les interventions pédagogiques auront lieu.
Dans notre contexte, un modèle performant et précis est celui qui attribue à chaque apprenant l'étiquette de sa classe réelle et qui prédit surtout avec un minimum d'erreur les apprenants à risque d'échec. En effet, l'objectif est de minimiser la classification des apprenants à risque en classe de réussite. À cette fin, nous proposons de suivre au cours de toutes les dates de prédiction l'évolution de la mesure *FPR* donnée par :

$$FPR = \frac{FP}{FP+VN}$$

D'après sa formule, *FPR* s'intéresse essentiellement aux classes d'apprenants à risque en retournant la proportion de cas d'apprenants à risque incorrectement identifiés comme apprenants réussis. Plus la valeur du FPR est faible, plus le modèle est qualifié pour avoir une capacité significative à prédire les apprenants à risque.

3.4.4.1 Suivi des résultats de la mesure FPR

Dans cette partie, nous présentons le suivi de la mesure de FPR au cours du temps en modélisant RF, DT et SVM avec les caractéristiques d'entrée sélectionnées par toutes les techniques testées dans la section précédente.

Les résultats du suivi hebdomadaire des valeurs FPR sont donnés par les Figures 3.8, 3.9, 3.10 et 3.11.

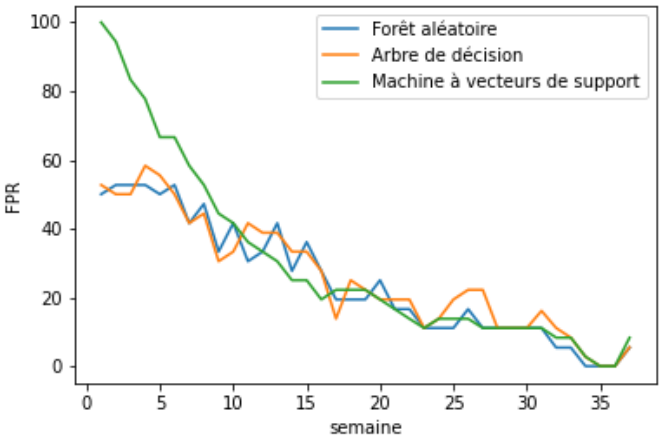


FIGURE 3.8 – L'évolution des valeurs de FPR suite à la sélection par la technique de khi-deux

Ces figures montrent que SVM est l'algorithme avec le FPR le plus élevé pendant les premières dates de prédiction. En se basant seulement sur les résultats de l'exactitude, toutes les expérimentations effectuées dans la Section 3.4.3 montrent que SVM est le modèle le plus performant. Cependant, ce dernier ne prédit pas correctement les apprenants de deux classes de risque (C_2 et C_3) aux premières dates de prédiction. Ces taux d'exactitude élevés sont dus à sa capacité à bien prédire la classe des apprenants qui

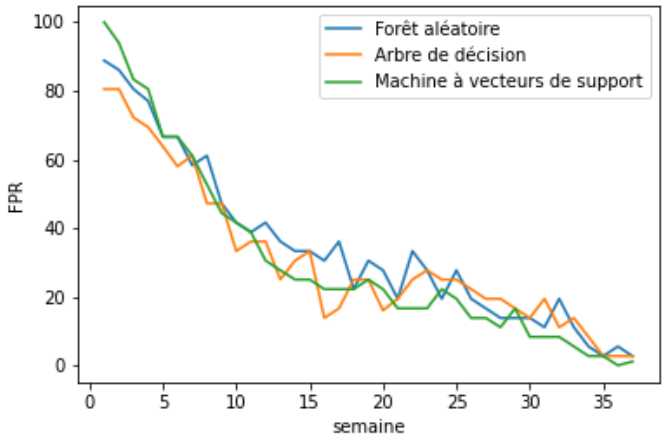


FIGURE 3.9 – L'évolution des valeurs de FPR suite à la sélection par la technique de ANOVA

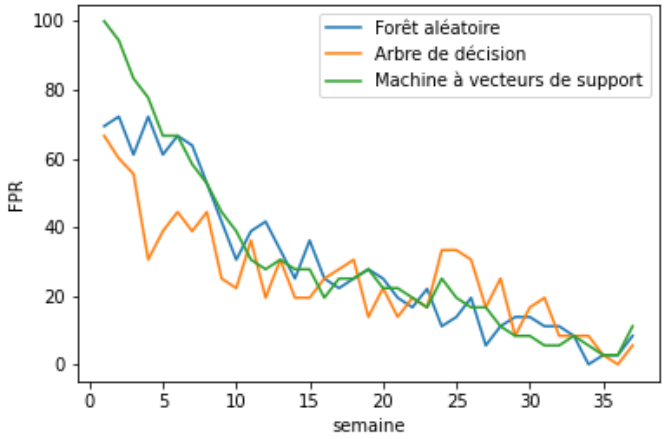


FIGURE 3.10 – L'évolution des valeurs de *FPR* suite à la sélection par la technique de RFECV

réussissent et qui constituent une majorité dans la base de données ; c'est la classe qui nous intéresse le moins.
D'un autre côté, les Figures 3.8, 3.9, 3.10 et 3.11 montrent que, aux premiers temps de prédiction, DT est celui avec les plus faibles valeurs de FPR. Ces taux de FPR montrent la capacité de ce modèle à correctement détecter les apprenants des classes du risque (C_2 et C_3). Cependant, les courbes d'exactitude de la Section 3.4.3 ont montré que DT est le moins performant aux premiers temps de prédiction comme ses valeurs d'exactitude sont les plus faibles.
RF est le modèle qui arrive encore une fois au juste milieu des deux autres et ses valeurs FPR présentent une légère différence par rapport à celles de DT.

3.4.4.2 Analyse des résultats de la 2ème phase d'évaluation de la méthodologie

Aux premiers temps de prédiction, SVM est le modèle avec les valeurs d'exactitude les plus élevées et les valeurs de FPR les plus élevées aussi. Ces résultats traduisent la capacité de ce modèle à correcte-

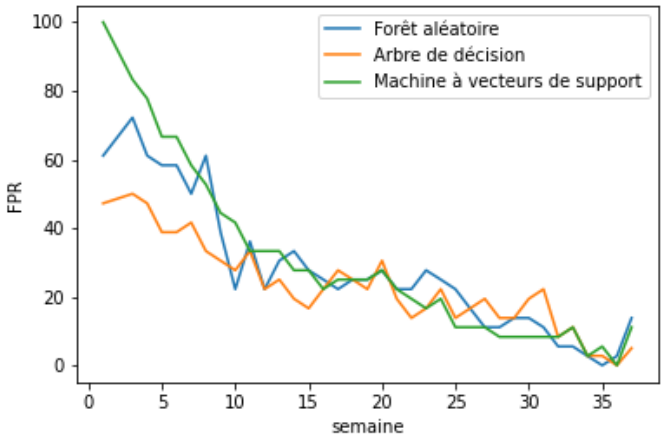


FIGURE 3.11 – L'évolution des valeurs de *FPR* suite à la sélection par la technique intégrée

ment détecter les apprenants qui réussissent.
Pendant les premières semaines de prédiction, DT est l'algorithme avec les valeurs d'exactitude et de FPR les plus faibles. Les résultats de la mesure de FPR montrent que ce modèle est le meilleur à prédire correctement et avec précision les apprenants à risque.
Les résultats de la mesure de FPR montrent que le RF est légèrement moins performant que l'arbre de décision pour prédire les apprenants à risque, mais reste bien meilleure que SVM.
À partir de la semaine 10, tous les algorithmes montrent presque les mêmes valeurs d'exactitude et de FPR. À partir de la semaine 20, nous prédisons avec le minimum d'erreur les apprenants à risque.

La sélection des caractéristiques par les techniques de la méthode filtrante est à la fois simple, peu complexe et surtout indépendante du modèle de ML utilisé. De plus, à travers toutes les expérimentations ci-dessus, nous remarquons que les modèles entraînés avec les caractéristiques de ANOVA sont légèrement plus performants que ceux entraînés avec les caractéristiques sélectionnées par le test khi-deux.
Dans la suite de la partie expérimentale, nous utilisons les $k = 20$ caractéristiques sélectionnées par le test ANOVA.

3.4.5 Résultats de l'intégration de la caractéristique temporelle dans le vecteur d'activité de l'apprenant

Dans cette partie, à chaque temps de prédiction t_i, un apprenant S_q est représenté par ses caractéristiques sélectionnées par le test *ANOVA* et sa classe $C_{p_{t_{(i-1)}}}$ prédite au temps t_{i-1}. Toutes ces caractéristiques servent à modéliser le modèle RF au cours de toutes les semaines de prédiction. RF montre des performances et des valeurs d'exactitude assez élevées. De plus selon les résultats de FPR, la capacité de détection des apprenants à risque de ce modèle est bien meilleure que celle du modèle SVM et s'approche du modèle DT. RF prédit très bien toutes les classes d'apprenants ; d'où le choix de ce modèle.

La Figure 3.12 présente les résultats du suivi des valeurs d'exactitude et de FPR en modélisant RF avec les caractéristiques sélectionnées par le test ANOVA et la classe $C_{p_{t_{(i-1)}}}$. À chaque temps de prédiction t_i, un apprenant est maintenant représenté par 21 caractéristiques : les 20 caractéristiques sélectionnées par le test *ANOVA* et la caractéristique correspondante à la classe $C_{p_{t_{(i-1)}}}$ prédite au temps t_{i-1}.

En comparant les résultats et l'allure de la courbe d'exactitude de cette figure avec ceux de la Figure 3.5, nous constatons une évolution plus rapide des valeurs d'exactitude depuis la 2ème semaine de prédiction. L'exactitude augmente plus rapidement au cours du temps et atteint des valeurs maximales élevées. À partir de la 20ème date de prédiction, le modèle montre des valeurs d'exactitude supérieures à 90% et atteint une exactitude de 100% à un point donné de prédiction (la semaine 34).
En comparant l'allure de la courbe de FPR de la Figure 3.12 avec celles de la Figure 3.9, nous remarquons une baisse de valeur plus rapide au cours des premières dates de prédiction. De plus les plus faibles valeurs sont atteintes plus rapidement au cours du temps. À la 14ème date de prédiction, le modèle donne un FPR $\approx 8.5\%$. À partir de la 20ème semaine, le modèle prédit les classes d'apprenants avec les plus faibles valeurs de FPR qui atteint le 0% à la 25ème semaine. En rajoutant la caractéristique temporelle de la classe prédite au temps précédent, le modèle fait moins d'erreur pour détecter les apprenants à risque.

Synthèse : l'ajout de la classe prédite au temps t_{i-1} améliore les performances prédictives du modèle et accélère surtout l'atteinte des valeurs maximales d'exactitude et minimales de FPR. La caractéristique temporelle ajoutée au vecteur d'activité d'apprentissage hebdomadaire joue un rôle important dans la détection du modèle des apprenants à risque. Elle permet surtout d'accélérer l'atteinte des bonnes performances et montre l'importance de l'intégration de la dimension temporelle lorsqu'il s'agit d'un processus continu dans le temps, comme celui de l'apprentissage.

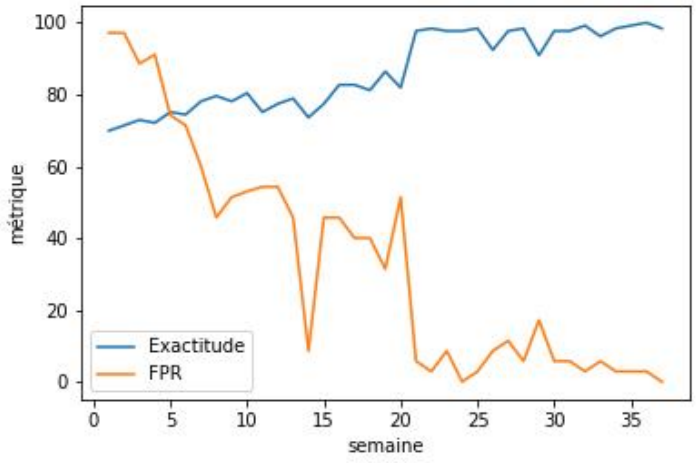


FIGURE 3.12 – L'évolution des valeurs d'exactitude et de FPR avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant

3.5 Validation : Application des résultats de la méthodologie sur d'autres matières

Dans cette partie, nous modélisons l'apprentissage de deux nouvelles matières qui sont les mathématiques et l'histoire-géographie. Les mathématiques sont une matière scientifique de même type que les physiques-chimie et l'histoire-géographie fait partie des matières littéraires ; d'où le choix de ces matières. En effet, l'objectif est de vérifier la validité de la méthodologie et des résultats trouvés sur différents types de matières.
Nous modélisons les modèles avec les caractéristiques extraites par le test *ANOVA* pour leur pertinence.

De plus, le test *ANOVA* se caractérise par son indépendance des modèles d'apprentissage.
À travers ces expérimentations, nous ne nous attendons pas à des performances exactement égales à celles obtenues avec les traces numériques de la physique-chimie. En effet, même si les natures des caractéristiques utilisées sont les mêmes, leurs valeurs changent. Par conséquent, une légère diminution avec les traces numériques des nouvelles matières est attendue. L'objectif est de ne pas avoir une chute de performance et de conserver les mêmes allures des courbes des métriques de ML. L'obtention de tels résultats permet la généralisation de la pertinence des indicateurs d'apprentissage sélectionnés et modèles de ML choisis.

Les tableaux 3.3 et 3.4 présentent le nombre total et par étiquette de classe (C_1, C_2 et C_3) d'apprenants inscrits, pendant les trains 7 et 8, respectivement dans les matières des mathématiques et de l'histoire-géographie.

	Train 7	Train 8
Nombre d'apprenants total	679	719
Nombre d'apprenants de C_1	530	587
Nombre d'apprenants de C_2	99	94
Nombre d'apprenants de C_3	50	38

TABLE 3.3 – Nombre d'apprenants total et par étiquette de classe inscrits dans la matière des Mathématiques

	Train 7	Train 8
Nombre d'apprenants total	684	729
Nombre d'apprenants de C_1	551	615
Nombre d'apprenants de C_2	117	99
Nombre d'apprenants de C_3	16	15

TABLE 3.4 – Nombre d'apprenants total et par étiquette de classe inscrits dans la matière Histoire-Géographie-

3.5.1 1ère matière : les mathématiques

En utilisant les traces numériques effectuées sur les mathématiques et les caractéristiques sélectionnées par le test ANOVA, nous présentons l'évolution de l'exactitude des prédictions des algorithmes RF, DT et SVM.
En suivant le protocole expérimental expliqué dans la Section 3.4.1, les Figures 3.13 et 3.14 présentent l'évolution de l'exactitude de trois modèles ML selon respectivement les 1ère et 2ème étapes de ce protocole.

Analyse des courbes de la Figure 3.13 : la Figure 3.13 présente les résultats d'exactitude de la 1ère phase du protocole expérimental suivi. Nous remarquons que les courbes d'exactitude de RF, DT et SVM gardent les mêmes allures ainsi que la même évolutivité que celles de la Figure 3.2. À partir de la semaine 3, SVM est le modèle avec les valeurs d'exactitude les plus élevées. Tandis que, DT est le modèle avec les plus faibles valeurs d'exactitude aux premières dates de prédiction. RF est le modèle avec des valeurs d'exactitude assez élevées et qui arrivent au milieu de celles retournées par DT et SVM.
Les courbes de la Figure 3.13 conservent les mêmes caractéristiques que celles de la Figure 3.2.

Analyse des courbes de la Figure 3.14 : la Figure 3.14 présente les résultats de la 2ème étape du

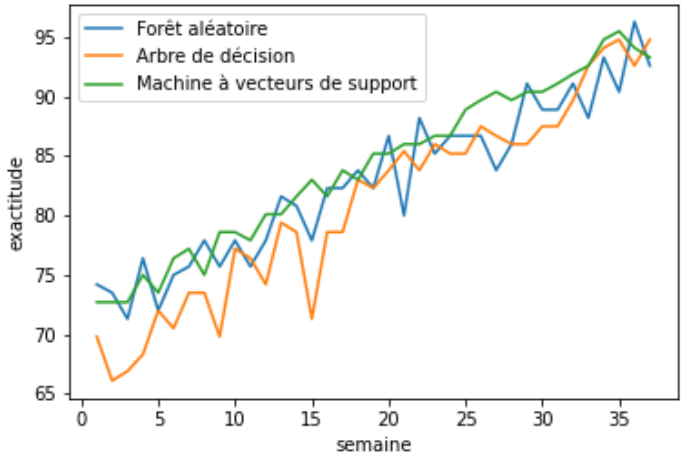


FIGURE 3.13 – L'évolution de l'exactitude en utilisant les traces numériques des mathématiques - 1ère étape du protocole expérimental

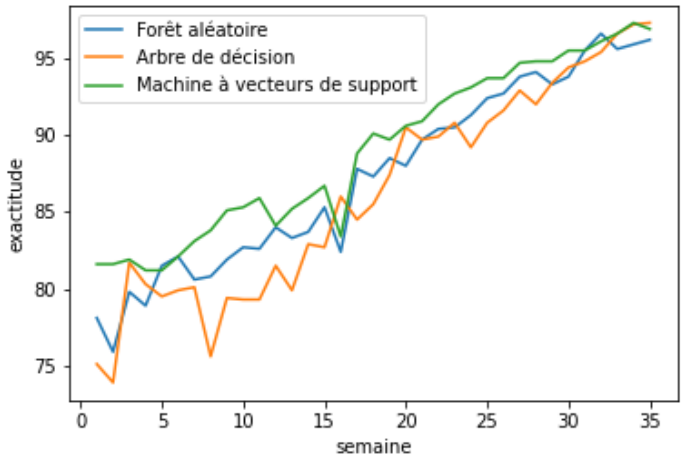


FIGURE 3.14 – L'évolution de l'exactitude en utilisant les traces numériques des mathématiques - 2ème étape du protocole expérimental

protocole expérimental. Nous avons entraîné les modèles RF, DT et SVM avec les données d'apprentissage des mathématiques des apprenants inscrits pendant le train 7 et les avons testés avec celles des apprenants inscrits dans le train 8. Les courbes de cette figure conservent les mêmes allures que celles de la Figure 3.3 et gardent l'aspect évolutif des valeurs d'exactitude des prédictions retournées. Comme pour le cas des Figures 3.2 et 3.3, les modèles montrent des prédictions plus précises, traduites par des valeurs d'exactitude plus élevées, dès les premières semaines lorsqu'ils sont entraînés avec les vecteurs des traces numériques de toute une population du train 7. Entraîner les modèles avec une plus grande quantité de données permet d'obtenir des prédictions plus exactes et précises. Pendant les premières dates de prédiction, le modèle SVM est celui avec les valeurs d'exactitude les plus élevées et DT est le modèle avec les valeurs d'exactitude les plus faibles.

Analyse des courbes de la Figure 3.15 : la Figure 3.15 montre l'évolution des valeurs de la métrique de FPR retournées par les modèles RF, DT et SVM respectivement. Nous constatons que les courbes de cette Figure conservent des allures décroissantes comme celles données par la Figure 3.9. Nous rappelons que plus la valeur de FPR est faible, plus le modèle se caractérise par une meilleure capacité à bien prédire les classes du risque.
Pendant les premières semaines, SVM possède les valeurs de FPR les plus élevées et DT montre des valeurs de FPR les plus faibles. Ces résultats confirment qu'avec les traces numériques des mathématiques, DT est le modèle qui prédit le mieux les apprenants à risque. SVM est celui qui prédit le moins précisément les apprenants à risque. RF retourne des valeurs de FPR entre celles de DT et SVM.
Ces résultats sont valables jusqu'à la 10ème semaine de prédiction. À partir de cette date, les performances des modèles se croisent et montrent des oscillations. DT atteint la valeur minimale de FPR 0% en premier lieu. SVM n'atteint jamais une valeur de FPR égale à 0%.

Après avoir vérifié la validité de la méthodologie suivie, les indicateurs d'apprentissage identifiés ainsi que les caractéristiques d'apprentissage sélectionnées, nous passons à la validation de l'impact de l'intégration d'une caractéristique temporelle dans le vecteur d'activité en ligne d'un apprenant sur les performances des modèles de ML.
Comme expliqué dans la Section 3.2.4, à chaque semaine de prédiction t_i, l'apprenant est désormais représenté par, en plus de ses caractéristiques d'apprentissage, la classe résultat de la prédiction à la semaine t_{i-1}. Pour les raisons citées au niveau de la Section 3.4.5, nous entraînons le modèle RF et testons ses performances.

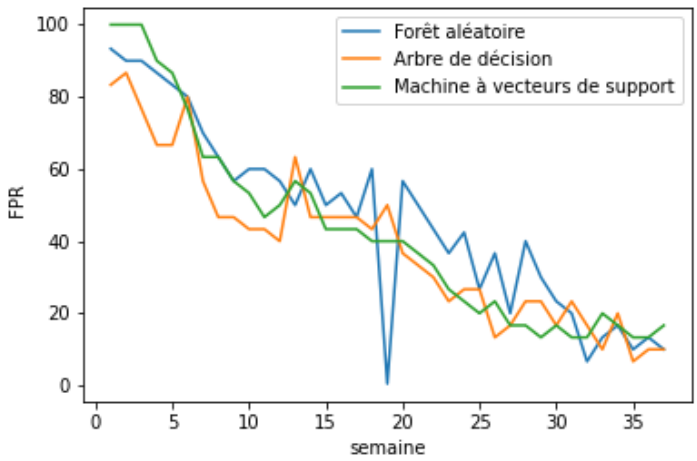


FIGURE 3.15 – L'évolution de FPR en utilisant les traces numériques des mathématiques

Analyse des résultats de la Figure 3.16 :la Figure 3.16 montre l'évolution de deux métriques exactitude et FPR pour évaluer les performances de RF suite à l'ajout de la caractéristique temporelle dans les vecteurs de la population d'apprenants. En comparant les deux courbes de cette Figure avec les courbes correspondantes à l'évolution de l'exactitude et le FPR du modèle RF respectivement des Figures 3.13 et 3.15, nous constatons que la performance de ce modèle s'améliore plus rapidement dans le temps. Plus explicitement, RF retourne des valeurs d'exactitude maximales et de FPR minimales plus tôt. La Figure 3.13 montre que RF arrive à atteindre une exactitude égale à 100% vers la fin de la prédiction.

De plus, l'ajout de la caractéristique temporelle montre son impact dans la précision de prédiction des classes des apprenants à risque à travers des valeurs de FPR plus faibles et plus rapide à obtenir.

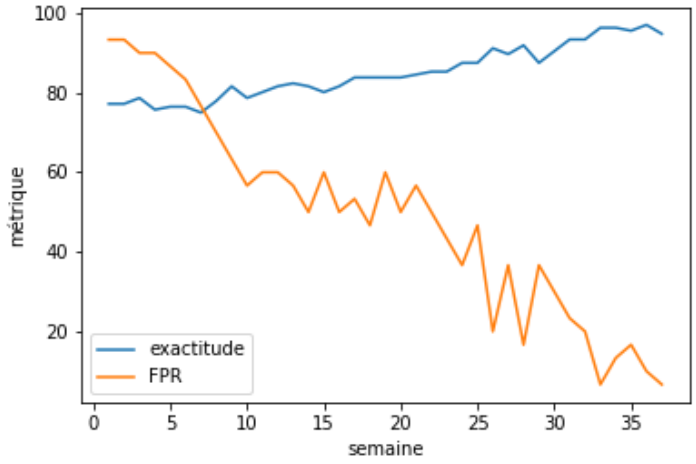


FIGURE 3.16 – L'évolution de l'exactitude et FPR avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant- Les mathématiques-

Synthèse : la méthodologie suivie regroupant la sélection des modèles, l'identification et la définition des indicateurs d'apprentissage, la sélection des caractéristiques d'apprentissage et finalement l'intégration d'une caractéristique temporelle prouvent leur validité avec l'utilisation des traces numériques d'une autre matière : les mathématiques. Toutes les courbes présentées dans cette section gardent les mêmes allures que celles de leurs correspondantes dans les sections précédentes. Ces résultats montrent encore une fois que la nature des caractéristiques extraites pour définir les indicateurs d'apprentissage de la performance, l'engagement, la régularité et la réactivité sont pertinentes et représentatives de l'activité académique de l'apprenant. De plus, DT, malgré son architecture simple, est le plus performant quand il s'agit de la précision de la prédiction des apprenants à risque. Nous concluons sur la pertinence de l'intégration de la caractéristique temporelle et son impact pour obtenir des prédictions plus précises et plus tôt dans la période d'apprentissage.
Dans la section suivante, nous appliquons la même démarche sur les traces numériques d'un autre type de matière qui est celle de l'histoire-géographie.

3.5.2 2ème matière : Histoire-géographie

En utilisant les traces numériques de la matière Histoire-Géographie et les caractéristiques sélectionnées par le test *ANOVA*, nous présentons, dans cette partie, les résultats de toute la démarche d'évaluation des prédictions des modèles RF, DT et SVM.
Les Figures 3.17 et 3.18 présentent les courbes résultantes des étapes du protocole expérimental introduit dans la Section 3.4.1.

Analyse des courbes de la Figure 3.17 : la Figure 3.17 présente les courbes d'évolution des valeurs d'exactitude des modèles RF, DT et SVM suite à l'application de la 1ère étape du protocole. Nous constatons que ces courbes gardent les mêmes allures que celles de la Figure 3.2 ; elles sont évolutives et croissantes dans le temps. Similairement aux résultats précédents, DT est le modèle qui montre les plus

faibles valeurs d'exactitude aux premières semaines de prédiction. SVM est le modèle qui commence avec l'exactitude la plus élevée et qui reste inchangée jusqu'à la 19ème date de prédiction. Pour le cas de cette expérimentation, RF montre une évolution très rapide des valeurs d'exactitude. À partir de la semaine 20, les performances des trois modèles convergent vers quasiment les mêmes valeurs.

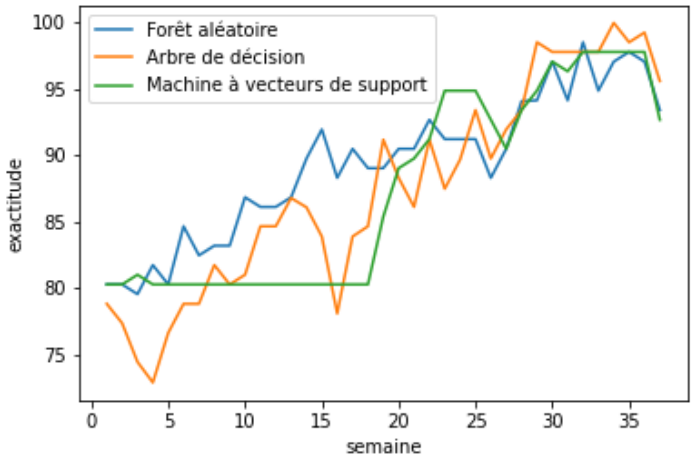


FIGURE 3.17 – L'évolution de l'*exactitude* en utilisant les traces numériques effectuées sur la matière histoire-géographie - 1ère étape du protocole expérimental

Analyse des courbes de la Figure 3.18 : la Figure 3.18 présente les résultats des prédictions des trois modèles en suivant la 2ème étape du protocole expérimental. Comme ils sont entraînés avec une plus grande quantité de données, les modèles RF, DT et SVM montrent des valeurs d'exactitude encore plus élevées dès la première semaine de prédiction. En considérant les courbes de cette Figure, SVM est le modèle le plus performant aux premières dates de prédiction avec une valeur d'exactitude$\approx 85\%$ et qui reste inchangée jusqu'à la 20ème semaine de prédiction. À partir de cette date, SVM montre des valeurs d'exactitude décroissantes qui s'améliorent vers la fin de la période d'apprentissage. DT est toujours le modèle avec les valeurs d'exactitude les moins élevées aux premières dates de prédiction. Sa courbe montre une évolution continue de l'exactitude de ses résultats de prédiction. RF est le modèle avec des valeurs d'exactitude élevées et qui arrivent entre celles de SVM et DT particulièrement aux premières dates de prédiction. À partir de la 15ème semaine de prédiction, les performances de RF et de DT convergent et se croisent au niveau d'un certain nombre de temps de prédiction, comme à la semaine 22 et à la semaine 25.

Pour valider les performances des modèles à bien prédire les classes d'apprenants à risque, nous passons au suivi de l'évolution des valeurs de la métrique de FPR retournées par RF, DT et SVM.

Analyse des courbes de la Figure 3.19 : la Figure 3.19 montrent les courbes de FPR de RF, DT et SVM. Ces courbes montrent des allures décroissantes au long de toutes les dates de prédiction. SVM est le modèle avec les valeurs de FPR les plus élevées pendant toutes les premières dates de prédiction. De la 1ère date jusqu'à la 16ème date de prédiction, SVM montre une valeur de *FPR* égale à 100% : à ces dates, ce modèle ne détecte aucun apprenant à risque et ses valeurs élevées d'exactitude revient au fait qu'il prédit très bien les apprenants qui réussissent. Dès la 1ère semaine et jusqu'à la fin de l'apprentissage, DT retourne des valeurs de FPR les plus faibles : ce modèle montre la meilleure capacité à détecter les

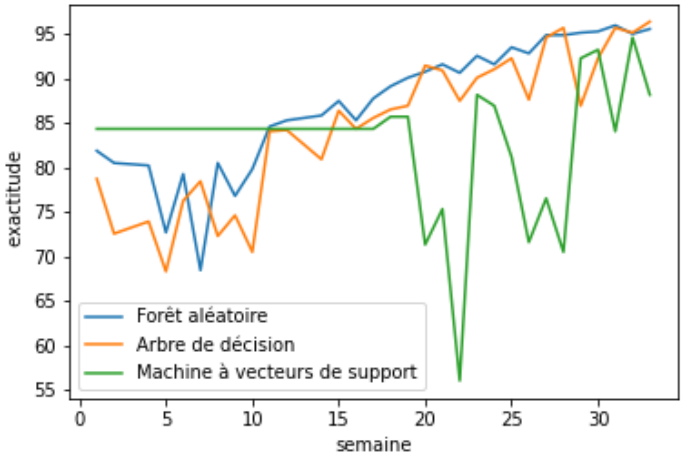


FIGURE 3.18 – L'évolution de l'*exactitude* en utilisant les traces numériques effectuées sur la matière histoire-géographie - 2ème étape du protocole expérimental

apprenants à risque. Les performances de RF à identifier les classes d'apprenants à risque, traduites par des valeurs de FPR descendantes, s'approchent de celles de DT. Les taux de FPR de ces deux derniers modèles convergent au niveau de certains points de prédiction.

Les caractéristiques sélectionnées par le test *ANOVA* pour définir les indicateurs d'apprentissage de la performance, l'engagement, la régularité et la réactivité ainsi que les modèles RF, DT et SVM confirment les résultats de l'application de la méthodologie suivie sur les traces numériques de la matière Histoire-géographie.
Nous analysons les résultats de l'intégration de la caractéristique temporelle de la classe prédite au temps de prédiction t_{i-1} dans le vecteur d'activité de l'apprenant de la matière Histoire-Géographie. Les résultats des prédictions retournées par le modèle RF sont évalués à l'aide des métriques de l'exactitude et le FPR et données par les courbes de la Figure 3.20.

Analyse des courbes de la Figure 3.20 : les deux courbes de la Figure 3.20 montrent que RF prédit les classes des apprenants avec une exactitude très élevée et qui s'améliore très rapidement dans le temps. La courbe d'exactitude garde une allure croissante tout au long des dates de prédiction et atteint vite des valeurs supérieures à 90%. Les taux de FPR commencent par une valeur $\approx 100\%$ mais ils diminuent très rapidement et atteignent des valeurs minimales qui s'approchent de 0%. En plus d'améliorer les performances du modèle, l'intégration de la caractéristique temporelle accélère l'obtention des valeurs maximales de performance.

Synthèse : à travers les résultats des expérimentations effectuées dans ces deux Sections 3.5.1 et 3.5.2, nous vérifions la validité de la méthodologie proposée en utilisant des traces numériques d'autres matières. Les différentes courbes et figures obtenues confirment :

— la pertinence de la nature des caractéristiques d'apprentissage sélectionnées par le test *ANOVA* Ces caractéristiques décrivent la démographie des apprenants et définissent leur performance, engagement, régularité et réactivité.
— la pertinence des modèles de ML sélectionnés et leur capacité à prédire les classes d'apprenants, particulièrement la classe d'apprenants à risque (avec le modèle DT et RF).

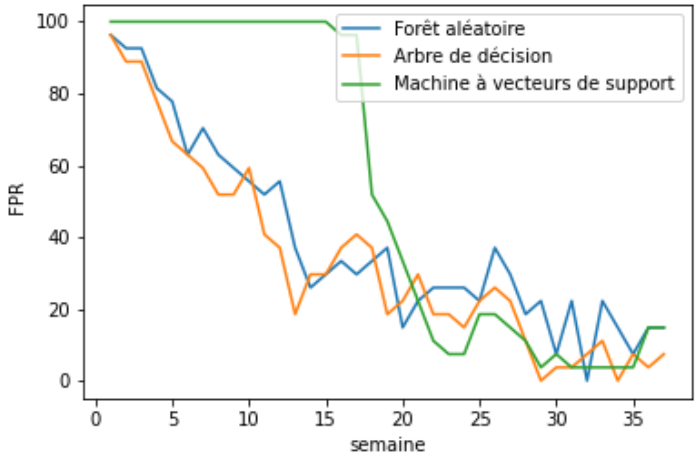


FIGURE 3.19 – L'évolution de *FPR* en utilisant les traces numériques effectuées sur la matière histoire-géographie -

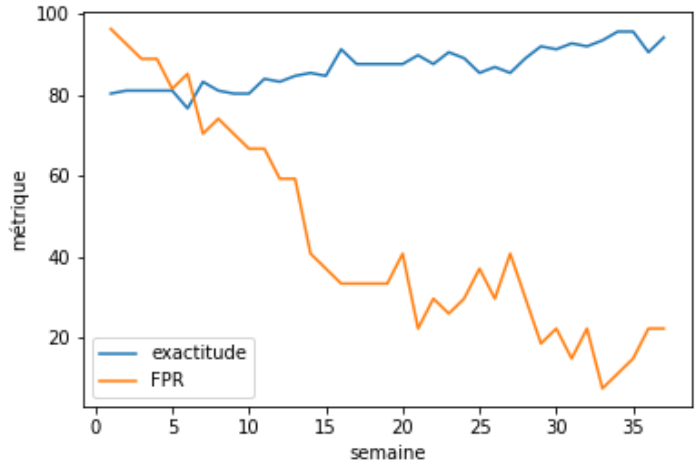


FIGURE 3.20 – L'évolution de l'*exactitude* et *FPR* avec l'ajout de la classe prédite au temps t_{i-1} au vecteur d'activité de l'apprenant-Histoire géographie-

— l'importance de la profondeur de l'évaluation pour identifier le modèle qui prédit et détecte le mieux les apprenants à risque. Toutes les expérimentations du suivi de la métrique de FPR montrent que DT est le modèle qui a la meilleure capacité à prédire les apprenants à risque moyen et élevé.
— la pertinence de l'intégration de la caractéristique temporelle dans le vecteur d'activité des apprenants pour améliorer ainsi qu'accélérer les performances du modèle de ML. Ces expérimentations confirment l'importance des données temporelles.

3.6 Conclusion

La prédiction précoce des apprenants à risque inscrits dans des établissements en ligne est un sujet populaire dans la littérature. Cependant, cette problématique est moins discutée lorsqu'il s'agit des collégiens en totale autonomie. Le CNED ne fait pas l'exception et enregistre chaque année un taux élevé d'échec parmi ses apprenants. Ainsi, à travers ce travail, nous visons à fournir aux enseignants du CNED un système capable d'identifier au plus tôt at avec précision les apprenants à risque. Ce système de LA se base sur une méthodologie profonde qui s'intéresse aux données employées, la pertinence des modèles de ML utilisés et leur capacité à bien identifier les apprenants à risque. Dans ce chapitre, nous proposons une méthodologie qui se base sur les étapes classiques de ML. L'objectif de cette méthodologie est l'évaluation de la pertinence des indicateurs et caractéristiques d'apprentissage sélectionnées ainsi que des modèles ML employés.
Au long de cette méthodologie, en croisant les définitions issues des sciences de l'éducation, les résultats de l'état de l'art et les traces numériques disponibles, nous avons commencé par identifier les indicateurs d'apprentissage définissant l'activité de tout apprenant en ligne : la performance, l'engagement, la régularité et la réactivité. L'identification de ces indicateurs est pertinente pour plusieurs objectifs comme la description de l'activité de l'apprenant et le diagnostic de sa situation d'apprentissage.... Ensuite, à partir des définitions données à chacun de ces indicateurs, nous sommes passés à l'extraction des caractéristiques à partir des données brutes disponibles. Chaque indicateur est représenté par un ensemble de caractéristiques extraites. Avant de passer à la modélisation et afin d'avoir un suivi régulier et détaillé des apprenants, nous avons proposé une approche de prédiction hebdomadaire et formalisé le problème en un problème de classification à 3 classes selon le niveau du risque. Les premiers résultats de l'évaluation montrent que RF, DT et SVM sont les modèles avec les valeurs d'*exactitude* les plus élevées. Dans l'étape suivante, il s'agit de la sélection des caractéristiques les plus corrélées avec le résultat final de l'apprenant. Le passage par ce processus permet d'améliorer les résultats de prédiction et surtout éliminer toutes les données bruitées qui pourraient dégrader la prédiction. L'application de plusieurs techniques de sélection de caractéristiques montrent la pertinence des caractéristiques définissant la performance, l'engagement, la régularité dans sa nouvelle définition ainsi que la réactivité.
En plus de l'identification des caractéristiques pertinentes, la phase de l'évaluation permet aussi l'identification du modèle avec la meilleure capacité à détecter les apprenants à risque moyen et élevé. Le suivi de la métrique de FPR a permis de conclure que le modèle DT identifie le mieux les apprenants à risque. Une dernière contribution de ce chapitre est la mise en valeur de l'ajout d'une caractéristique temporelle dans le vecteur d'activité d'un apprenant. En effet, les résultats des mesures d'exactitude et de FPR montrent que l'intégration de la caractéristique temporelle améliore la qualité de prédiction et surtout accélère l'obtention des pics de performance.
Tous ces résultats ont été validés par l'application de la même démarche en utilisant les traces numériques d'autres matières : mathématiques et histoire-géographie.

Les résultats trouvés dans ce chapitre ouvrent les perspectives sur d'autres questions. D'un côté, les résultats de la phase d'évaluation montrent l'importance du choix de la métrique à suivre. De l'autre côté, le suivi des métriques au cours du temps et leurs oscillations ainsi que l'intégration de la caractéristique temporelle sont pertinents dans la performance des modèles. Par conséquent, une évaluation basée sur des métriques temporelles serait intéressante pour que le système proposé fournisse les meilleurs résultats de prédiction et pour pouvoir prédire au plus tôt et avec une grande précision les apprenants à risque. De plus, l'objectif de ce système est toujours l'intervention auprès des apprenants avec des difficultés et donc commencer par générer des alertes bien précises. C'est l'évaluation basée sur des métriques temporelles qui permettrait cette fonctionnalité dans le système final à proposer. Tout cela sera l'objet du chapitre suivant.

Chapitre 4

Évaluation temporelle des systèmes prédictifs

Synthèse
Ce présent chapitre interprète les résultats du chapitre précédent et conclut sur l'importance de l'évaluation temporelle dans un contexte d'un processus continu dans le temps.
En nous basant sur les techniques d'extraction d'erreur (error mining [11]) et de suivi régulier des résultats de prédiction, nous proposons des métriques d'évaluation temporelles qui permettent de quantifier et qualifier la précocité et la stabilité des systèmes prédictifs. De plus, étant donné l'importance des métriques de précision habituelles utilisées dans l'état de l'art, nous étudions les différents compromis qui existent entre ces dernières et les métriques temporelles.
Ces différentes métriques sont ensuite utilisées dans les chapitres suivants pour évaluer les profils types d'apprenants et proposer un algorithme d'alerte précoce.

4.1 Introduction

L'apprentissage rencontre plusieurs défis et problèmes, dont le plus répandu est les taux élevés d'échec parmi les apprenants. Afin de répondre à cette problématique, les systèmes d'alerte précoce *(Early Warning Systems (EWS))* basés sur les modèles de ML ont été proposés. L'objectif des EWS est la prédiction précise et au plus tôt des apprenants à risque d'échec pour alerter les enseignants des cas de ces derniers. Ces systèmes sont basés à la fois sur des flux de données d'apprentissage en temps réel et sur des modèles de ML.

Les données générées par les plateformes d'apprentissage en ligne sont disponibles progressivement dans le temps, car elles dépendent fortement du moment auquel les apprenants interagissent avec le contenu éducatif. Ainsi, ces données sont appelées séries chronologiques, car elles consistent en un ensemble de séquences d'événements qui se produisent au fil du temps. La référence temporelle peut être un horodatage (*timestamp*) ou tout autre intervalle du temps à grain fini (par exemple, une semaine). Ces données alimentent les EWSs et contribuent à leur performance.

Afin d'évaluer l'efficacité des prédictions fournies par les EWS, certains types de métriques sont communément utilisés. D'une part, des mesures de performance sont utilisées pour évaluer la complexité temporelle et spatiale des systèmes par rapport aux données utilisées. D'autre part, les mesures de ML

11. L'extraction d'erreurs est une technique entièrement indépendante du système et de la langue. Cette technique pourrait être appliqué sans aucun changement sur les résultats d'analyse produits par n'importe quel système travaillant sur n'importe quelle langue. La seule information nécessaire est une valeur booléenne pour chaque phrase qui indique si elle a été analysée avec succès ou non [130].

telles que la précision, le rappel, la sensibilité...sont utilisées pour qualifier la capacité du système à prédire correctement les résultats finaux des apprenants du contexte étudié. Malgré la diversité de métriques d'évaluation existantes, aucune d'entre elles n'est dédiée à évaluer la *précocité*, la *continuité* et l'*évolution* des prédictions dans le temps. Cependant, le temps est une dimension importante qui doit être prise en compte lors de l'évaluation d'un EWS à mesure que l'apprentissage et la prédiction sont deux processus évolutifs dans le temps. De plus, comme leur nom l'indique, les EWS se caractérisent à la fois par la précision ainsi que la *précocité* de leurs prédictions.

Dans ce chapitre, nous nous concentrons sur la question de recherche suivante : ***comment évaluer l'évolution temporelle des systèmes de prédiction ?***
Afin de répondre à cette question, nous proposons des nouvelles métriques dépendantes du temps, dites ***métriques temporelles***, et qui sont utilisées pour évaluer ces systèmes au fil de temps et mettre en valeur l'importance de la dimension temporelle et son impact sur les prédictions. Ces métriques temporelles sont la ***précocité*** et la ***stabilité***.

Dans un premier temps, la précocité a été définie comme la première date de prédiction à laquelle le système prédit correctement la classe réelle de l'apprenant. L'application de cette définition sur un contexte réel donne des résultats promoteurs. Mais, à des temps de prédiction proches, le modèle pourrait être inexact et le premier temps de prédiction correct illustre peu la performance du système. C'est pour cette raison que nous avons, dans un second temps, adapté et reformulé la définition de la précocité pour que son algorithme retourne une liste de valeurs de temps qui correspondent toutes à de bonnes prédictions.

La stabilité des prédictions a été définie comme la capacité d'un EWS à fournir les plus longues séquences de prédictions successives correctes sur les temps de prédiction.

Un EWS est dit efficace lorsqu'il fournit des prédictions précises et au plus tôt dans l'année scolaire. En effet, il ne suffit pas d'avoir une prédiction précoce sans pouvoir assurer l'exactitude du système prédictif. À cette fin, nous proposons la mesure de Précocité-Exactitude (**Earliness-exactitude Score (EAS)**) pour étudier le compromis entre les métriques de précision de ML (notamment l'exactitude) et celle de précocité puisqu'elles sont proportionnellement inversées.
De plus, nous proposons d'utiliser une mesure de précocité-stabilité (**Earliness-Stability Score (ESS)**) pour étudier la relation entre la précocité d'un classifieur et sa stabilité. La mesure ESS est utilisée pour examiner le compromis entre les métriques dépendantes du temps. L'objectif est de comparer l'utilisation de la mesure ESS à celle de la mesure EAS.

Dans la partie expérimentale, nous validons l'importance et la pertinence des métriques temporelles proposées et l'intérêt des études de compromis entre les différents types de métriques (temporelle et de précision). Pour ce faire, nous appliquons les algorithmes de précocité, la fonction du calcul de la stabilité ainsi que les formules de EAS et de ESS pour évaluer la performance de quatre EWS réels [9] [131] ayant comme objectif la prédiction des apprenants à risque d'échec. Ces systèmes utilisent les données réelles des collégiens inscrits au module de physique-chimie en ligne au sein du CNED.

4.2 Description du problème

Dans cette section, nous commençons par une présentation formelle du problème de la classification des séries chronologiques. Nous introduisons par la suite la technique du suivi régulier que nous avons adoptée pour proposer les métriques d'évaluation temporelle dans les prochaines sections.

4.2.1 Formalisation du problème

Supposons que $Y = \{C_1, C_2, .., C_m\}$ est l'ensemble de classes prédéfinies et qui est déterminé à l'aide de données d'apprentissage existantes.
Soit $S = (S_1, S_2, ..., S_k)$ l'ensemble d'apprenants dans l'ensemble de données de test D_{test} et $T = \{t_1, t_2, .., t_c\}$ l'ensemble des temps de prédiction.

Chaque apprenant $S_p \in S$, à un temps de prédiction $t_i \in T$, est présenté par un vecteur $X_{t_i} =< f_1, f_2, ..., f_n, C_j >$ où chaque $f_k \in \mathbb{R}$ présente la k^{ieme} caractéristique extraite des données brutes, et $C_j \in Y$ est l'étiquette de la j^{ieme} classe.

Un EWS efficace a la capacité de déterminer à chaque temps de prédiction t_i la bonne classe pour chaque apprenant de S. Cette prédiction doit être à la fois précoce et précise et garantir la stabilité de tout le système. L'objectif de ce travail sur l'évaluation des systèmes prédictifs à l'aide des métriques temporelles est d'identifier le premier point dans le temps où l'étiquette de la classe est correctement prédite et le classificateur maintient une prédiction stable et précise dans le temps.

Comme le montre la Figure 4.1, à chaque temps de prédiction $t_i \in T$, chaque apprenant $S_p \in S$ est prédit appartenir à une classe $C_j \in Y$. La Figure 4.1 présente un exemple de suivi régulier de prédictions d'un ensemble d'apprenants $S = (S_1, S_2, S_3)$ sur un intervalle de temps $T = \{t_1, t_2, t_3, t_4, t_5\}$. À chacune des date de prédiction de T, une seule étiquette parmi $Y = \{C_1, C_2, C_3\}$ est attribuée à chaque apprenant.

4.2.2 Suivi régulier des prédictions

La qualité des prédictions des EWS basés sur des classifieurs est évaluée en s'appuyant sur diverses métriques et mesures de performance algorithmique [112] [113] [114]. Mais aucune d'entre elles ne considère la dimension temporelle pour déterminer l'efficacité de ces systèmes. En effet, le suivi régulier des résultats de prédiction est à l'origine de l'identification de nouvelles métriques temporelles comme la précocité et la stabilité.
Le principe du suivi régulier consiste à sauvegarder à chaque temps de prédiction les classifications obtenues. Ce travail sert à comparer les prédictions du temps t_i par rapport à celles du temps t_{i-1} ainsi que par rapport à l'étiquette réelle. Ceci permet d'avoir un aperçu sur l'exactitude de la prédiction de chaque étiquette de classe et de découvrir de près les fluctuations et les variations possibles dans la performance de tout le système prédictif. La Figure 4.1 illustre bien un exemple d'application de ce principe et montre les variations dans les résultats du modèle.

De plus, en appliquant ce principe sur les résultats d'un contexte réel (celui des résultats des prédictions du chapitre précédent), nous nous rendons compte que l'évaluation de la performance d'un EWS à l'aide d'une métrique de précision non évolutive dans le temps (telle que l'exactitude) est insuffisante pour détecter les défaillances de stabilité du système et de juger sur sa capacité à fournir des prédictions précoces et précises.

4.3 1ère métrique temporelle : La précocité

L'objectif d'un EWS est la prédiction précoce des apprenants les moins performants afin que les tuteurs puissent intervenir efficacement auprès d'eux. La prédiction précoce est communément définie comme le temps adéquat et pertinent permettant à la fois une prédiction précise des performances des apprenants et des interventions efficaces du tuteur auprès des apprenants à risque [132]. Une intervention pédagogique a un impact positif sur le parcours académique d'un apprenant lorsqu'elle est dispensée au bon moment.

	t_1		t_2		t_3		t_4		t_5	
	l_1	p_1	l_2	p_2	l_3	p_3	l_4	p_4	l_5	p_5
S_1	C_2	C_3	C_2	C_2	C_2	C_2	C_1	C_2	C_1	C_1
S_2	C_1	C_1	C_1	C_1	C_2	C_2	C_2	C_2	C_1	C_1
S_3	C_3	C_1	C_3	C_1	C_3	C_2	C_3	C_3	C_3	C_3

FIGURE 4.1 – Un exemple de suivi régulier des prédictions pendant un intervalle de temps T

Dans cette section, nous présentons les définitions proposées pour cette nouvelle métrique temporelle et les mises à jour qui y sont faites pour qu'elle soit appliquée aux cas les plus généraux. Nous décrivons aussi les algorithmes qui permettent le calcul de cette métrique temporelle.

4.3.1 La précocité : première date d'une bonne prédiction

La précocité est une métrique temporelle que nous proposons pour évaluer la performance des algorithmes de classification de ML lorsque la prédiction est un processus évolutif dans le temps ; un processus qui se fait à plusieurs points de temps d'un intervalle fixé.
Nous définissons la précocité comme la première date à laquelle la prédiction est correcte. L'Algorithme 1 présente la suite des instructions que nous proposons pour retourner ce premier point du temps de la bonne prédiction pour chacune des étiquettes des classes de l'ensemble Y.

4.3.1.1 Description de l'algorithme

L'Algorithme 1 prend comme paramètres d'**entrée** : la liste d'élèves (S), l'ensemble d'étiquettes de classes (Y) et les données de la base de test (D_{test}). Cet algorithme retourne en **sortie** les valeurs de la métrique de précocité pour chacune des étiquettes de classes (E_Y).
L'algorithme commence par itérer sur l'étiquette de la classe ($Cj \in Y$) (Ligne 1). Pour chaque classe C_j, l'Algorithme initialise deux variables ($Precoce_{tot}$ et $compte_S$), qui contiendront respectivement la somme de dates des premières prédictions correctes et le nombre d'élèves qui ont été affectés au moins une fois à la classe C_j en question (Lignes 2-3). Ensuite, l'Algorithme itère sur l'ensemble d'élèves ($S_k \in S$) (Ligne 4). Il vérifie d'abord si S_k a été affecté au moins une fois à la classe C_j (ligne 5). Si c'est le cas, l'algorithme cherche la première fois que S_k a été dans C_j (Ligne 6) ainsi que la première fois que S_k est prédit correctement dans C_j (Ligne 7). Le premier temps de prédiction correcte est ensuite calculé (Ligne 8) et les deux variables ($Precoce_{tot}$ et $compte_S$) sont mises à jour (Lignes 9-10). La précocité (donnée par la variable $Early_{C_j}$) qui correspond à l'étiquette de classe C_j est déterminée à la Ligne 13. Puis avant d'itérer sur l'étiquette de classe suivante, la précocité mesurée pour la classe C_j en question est sauvegardée dans E_y (Ligne 14).

En bref, cet Algorithme retourne le temps de la première prédiction correcte pour chacune des étiquettes des classes de l'ensemble Y. Pour chacune des étiquettes C_j, ce temps est calculé comme la moyenne de la somme de tous les premiers temps de prédictions correctes des apprenants ayant déjà été étiquetés comme appartenant à C_j.

4.3.1.2 Exemple de calcul de la précocité

Pour l'exemple du calcul de la précocité, nous prenons les résultats des prédictions du système illustré par la Figure 4.1.
L'ensemble S est composé de trois élèves $\{S_1, S_2, S_3\}$. À chaque t_i, un apprenant appartient à une étiquette de classe (l_i) et a une classe prédite (p_i). Supposons que t_i représente une semaine et que nous devons calculer la précocité par rapport à la classe C_1. La première bonne prédiction pour chacun des trois apprenants est représentée par la case verte. L'apprenant S_3 n'est pas pris en compte car il n'a jamais été étiqueté comme appartenant à C_1. En appliquant l'Algorithme 1, les valeurs de la mesure de précocité pour S_1 et S_2 sont égales respectivement à 2 et 1. Ainsi, la précocité pour l'étiquette de classe $C1 = 1,5(3/2)$. En d'autres termes, C_1 est correctement prédite à partir de (1.5) semaine de l'apprentissage.

Algorithm 1 Algorithme de précocité

Entrée: S, Y, D_{test}
Sortie: E_Y

pour chaque C_j dans Y **faire**
  $Precoce_{tot} \leftarrow 0$
  $compte_S \leftarrow 0$
  pour chaque S_k dans S **faire**
    si (attribué$(S_k, C_j) ==$**Vrai) alors**
      $t_0 \leftarrow$ première_Étiquette(S_k, C_j)
      $t_1 \leftarrow$ première_Prediction(S_k, C_j)
      $Precoce_{S_kC_j} \leftarrow t_1 - t_0 + 1$
      $Precoce_{tot} \leftarrow Prcoce_{tot} + Precoce_{S_kC_j}$
      $compte_S \leftarrow compte_S + 1$
    fin si
  fin pour
  $Precoce_{C_j} \leftarrow Prcoce_{tot}/compte_S$
  $E_Y \leftarrow mettre(C_j, Precoce_{C_j})$
fin pour

4.3.2 La précocité : un ensemble de dates de bonnes prédictions

Dans la section 4.3.1, la précocité est définie comme le premier point dans le temps auquel le modèle de classification prédit correctement l'étiquette de classe. Il s'agit, en effet, d'une forte hypothèse, car le modèle pourrait être imprécis aux prochains points de prédiction.
Comme nous le savons, le comportement d'apprentissage des apprenants est dynamique dans le temps, et aussi le sont les prédictions. Compte tenu de l'oscillation qui peut caractériser les prédictions, il est important d'étudier plusieurs points de précocité dans le temps afin d'identifier avec précision la bonne valeur de précocité.
Dans la suite de cette section, nous introduisons un algorithme qui extrait tous les points temporels de précocité des bonnes prédictions pour chacune des étiquettes des classes de Y.
Pour des raisons de compréhension et de facilité d'interprétation et de calcul, nous divisons l'algorithme en deux sous-algorithmes présentés et décrits dans les sous-sections suivantes.

4.3.2.1 Description du 1er sous-algorithme

Le premier sous-Algorithme 2, appelé *Précocité par apprenant*, retourne pour un apprenant S_p une liste E_C de x valeurs de temps, qui correspondent aux premières prédictions correctes successives pour une classe C donnée. Cet Algorithme prend en entrée l'apprenant S_p, l'étiquette de classe C, x qui est le nombre de points de précocité à retourner et le jeu de données de test D_{test}. L'Algorithme 2 commence par affecter l'ensemble vide à la variable E (Ligne 1), qui est la séquence de x premiers points temporels de bonne prédiction. Au niveau de la Ligne 2, l'Algorithme initialise la variable j, qui est un compteur d'itérations. Si l'apprenant S_p a été affecté au moins une fois à l'étiquette de classe C (Ligne 3), l'algorithme détermine le temps t_0 qui correspond à la première date à laquelle S_p a été étiqueté C (Ligne 4). Sinon, l'algorithme retourne une liste vide (Ligne 14 - Ligne 15). L'algorithme utilise une variable t_s pour désigner le point de départ dans le temps pour rechercher la j^{eme} date d'une prédiction correcte. t_s est initialisée à t_0 (Ligne 5). Tant que le nombre calculé de points de précocité est différent de x (Ligne 6), l'algorithme continue sa recherche. Il détermine le temps t_j où S_p est correctement prédit dans C, à partir de t_s (Ligne 7). Ensuite, le temps de prédiction correcte donné par e_j est calculé (Ligne 8). L'algorithme met à jour la liste E qui contient les premières dates de prédictions correctes (Ligne 9). Les variables j et t_s sont mises à jour respectivement au niveau des (Ligne 10) et (Ligne 11). En fait, la recherche suivante commence avec le point temporel de prédiction actuellement trouvé (t_j). Lorsque la recherche est terminée, la liste de x points temporels de prédiction les plus tôt est sauvegardée dans E_C (Ligne 13).

En bref, cet Algorithme retourne, pour un apprenant S_p et une étiquette de classe C donnés et passés en paramètre d'entrée, une liste de mesures de points de temps précoces et successifs de bonne prédiction. Cette liste est de taille x à définir selon le contexte et les besoins.

Algorithm 2 Précocité par apprenant - Précocité(S_p, C, x, D_{test})

Entrée: S_p, C, x, D_{test}
Sortie: E_C
1: $E \leftarrow \emptyset$
2: $j \leftarrow 0$
3: **si (attribue**$(S_p, C) ==$**Vrai) alors**
4: $t_0 \leftarrow$ premiere_Etiquette(S_p, C)
5: $t_s \leftarrow t_0$
6: **tant que** $j < x$ **faire**
7: $t_j \leftarrow$ premiere_Prediction(S_p, C, t_s)
8: $e_j \leftarrow t_j - t_0 + 1$
9: $E \leftarrow mettre_a_jour(E, e_j)$
10: $j \leftarrow j + 1$
11: $t_s \leftarrow t_j + 1$
12: **fin tant que**
13: $E_C \leftarrow mettre(C, E)$
14: **sinon**
15: apprenant n'a jamais été dans la classe C
16: **fin si**

4.3.2.2 Description du 2ème sous-algorithme

Le second sous-Algorithme 3, nommé *Précocité Totale(S, Y , x, D_{test})* retourne une liste de x premières dates de prédictions correctes, pour l'ensemble S et pour chacune des étiquettes de classe C_j dans

Y. Chaque valeur du temps i de cette liste est la moyenne de toutes les i valeurs du temps retournées par chaque élève dans S.
L'Algorithme 3 prend en entrée la liste d'élèves (S), l'ensemble d'étiquettes de classe (Y), le nombre de points de précocité à mesurer (x) et l'ensemble de données de test (D_{test}). Cet algorithme commence par itérer sur les étiquettes de classe (C_j dans Y) (Ligne 1). À l'intérieur de cette boucle itérative, l'algorithme affecte l'ensemble vide à la variable L (Ligne 2). Cette variable contient toutes les listes de x premières prédictions correctes de tous les apprenants de S. Au niveau de la (Ligne 3), l'Algorithme initialise la variable $compte_L$ qui fait référence à la taille de la variable L. Ensuite, l'Algorithme itère sur les apprenants de S (Ligne 4). Nous appellons l'Algorithme Précocité(S_p, C_j , x, D_{test}) (décrit par l'Algorithme 2) et dont le résultat est affecté à la variable l_p (Ligne 5). Ensuite, L est mise à jour (Ligne 6) et $compte_L$ est incrémentée de 1 (Ligne 7). Une fois que L est construite, l'Algorithme procède à une itération dont le compteur i est compris entre 1 et x (Ligne 9). L'Algorithme initialise la variable $precoce_{i_{c_j}}$ qui contient la somme de toutes les valeurs de temps (Ligne 10). Au niveau de la (Ligne 11), l'Algorithme itère sur les listes dans L. Ensuite, la valeur de chaque l dans L d'indice i est ajoutée à $precoce_{iC_j}$ (Ligne 12). La valeur du point de précocité numéro i de l'étiquette de classe C_j donnée par la variable $precoce_{iC_jtot}$ est déterminée au niveau de la (Ligne 14). Le résultat de $precoce_{iC_jtot}$ est sauvegardé dans la variable E_Y avec l'étiquette de classe C_j (Ligne 15).

En bref, cet Algorithme, en s'appuyant sur les résultats de l'Algorithme 2, retourne pour chaque apprenant S_p de la base de données de test la liste de mesures de points de temps précoces et successifs de bonne prédiction pour chacune des étiquettes de classe de l'ensemble Y. Cette liste est de taille x à déterminer selon les besoins de chaque contexte.

Algorithm 3 Algorithme de précocité - Précocité_Totale(S, Y, x, D_{test})

Entrée: S, Y, x, D_{test}
Sortie: E_Y

1: **pour chaque** C_j in Y **faire**
2: $L \leftarrow \emptyset$
3: $compte_L \leftarrow 0$
4: **pour chaque** S_p dans S **faire**
5: $l_p \leftarrow Precocite(S_p, C_j, x, D_{test})$
6: $L \leftarrow mettre_a_jour(L, l_p)$
7: $compte_L \leftarrow compte_L + 1$
8: **fin pour**
9: **pour** ($i = 1$ et $i \leq x$) **faire**
10: $precoce_{iC_j} \leftarrow 0$
11: **pour chaque** l dans L **faire**
12: $precoce_{iC_j} \leftarrow precoce_{iC_j} + l[i]$
13: **fin pour**
14: $precoce_{iC_jtot} \leftarrow precoce_{iC_j}/compte_L$
15: $E_Y \leftarrow mettre(C_j, precoce_{iC_jtot})$
16: **fin pour**
17: **fin pour**

4.3.2.3 Exemple de calcul de la précocité

Pour l'exemple de calcul de la précocité, nous prenons les résultats du système de la Figure 4.1. L'ensemble S est composé de trois apprenants {S_1, S_2, S_3}. À chaque t_i, un apprenant appartient à une

étiquette de classe (l_i) et à une classe prédite (p_i). Supposons que t_i représente une semaine et que nous devons calculer $x = 2$ points de précocité par rapport à la classe C_2. Les prédictions correctes sont cadrées en rouge. L'apprenant S_3 n'est pas pris en compte car il n'a jamais été affecté à l'étiquette de classe C_2.

Application du sous-Algorithme 2 :

— la liste de deux premiers précoces pour l'apprenant S_1 est : {2, 3}

— la liste de deux premiers précoces pour l'apprenant S_2 est : {3, 4}

Application du sous-Algorithme 3 : Pour la classe C_2, les moyennes de deux premiers temps précoces pour une prédiction correcte sont respectivement : $(2+3)/2 = 2.5$, $(3+4)/2 = 3.5$.

L'Algorithme 2 et 3 retournent : (C_2, 2.5, 3.5) ce qui signifie que la classe C_2 est correctement prédite à partir de la semaine ≈ 2.5. De plus, le 2ème temps précoce pour prédire correctement cette classe C_2 est la semaine ≈ 3.5.

4.4 2ème métrique temporelle : la stabilité

L'évaluation de la performance des EWS à l'aide des métriques temporelles plutôt que se contenter des métriques de précision de ML est un axe de recherche très intéressant. En effet, ces mesures temporelles mettent l'accent sur l'évolution et la continuité de deux processus d'apprentissage et de prédiction dans le temps.

Comme le comportement d'apprentissage et les performances des apprenants ont tendance à changer au fil du temps, les performances d'un modèle de classification suivent ces changements et peuvent montrer des instabilités et oscillations de prédiction. Pour l'exemple de la Figure 4.1, les apprenants S_1 et S_2 ont été affectés à au moins deux différentes étiquettes de classes pendant les 5 temps de prédiction ; et c'est dû au changement dans les comportements d'apprentissage de ces apprenants.

La stabilité temporelle, notamment dans le domaine de l'éducation, est rarement mentionnée dans l'état de l'art comme une mesure caractérisant la capacité d'un EWS à maintenir les mêmes bonnes prédictions pendant un intervalle du temps prédéfini.

Nous définissons la stabilité temporelle comme **la moyenne des plus longues séquences de prédictions successives correctes** sur un intervalle du temps $[t_1, t_2]$ où $t_1 < t_2$. La stabilité est donc calculée à l'aide de l'équation suivante :

$$Stabilite = \frac{\sum_{p=1}^{n} |h(s_p)|}{|D|} \tag{4.1}$$

où $h : S \rightarrow T^n$ est une fonction qui associe à chaque apprenant de $D \subseteq D_{test}$ la plus longue séquence de prédictions correctes successives à partir de t_1 jusqu'à un temps t_2. Cette équation permet de calculer la stabilité sur chaque étiquette de classe parmi Y ou sur l'ensemble du jeu de données $D \subseteq D_{test}$. Cependant, nous pensons que la stabilité est plus pertinente et interprétable lorsqu'elle est calculée sur le jeu de test ($D = D_{test}$).

4.5 Étude des compromis

Dans cette partie, nous présentons une étude théorique des différents compromis qui existent entre les métriques de précision de ML, particulièrement l'exactitude, et les métriques temporelles que nous proposons.

Dans un premier temps, nous étudions le compromis entre l'exactitude et la précocité en utilisant la mesure Précocité-Exactitude (***Earliness-Accuracy Score (EAS)***). Ensuite, nous proposons une mesure

qui permet d'étudier le compromis entre la précocité et la stabilité temporelle : c'est la Précocité-Stabilité ***(Earliness-Stability Score (ESS))***.

4.5.1 Le compromis Précocité-Exactitude : EAS

Pour un EWS, plus tôt les prédictions sont correctes, meilleur est le système. Cependant en améliorant les résultats de la précocité, les résultats des métriques de précision de ML, y compris l'exactitude, doivent rester suffisamment élevés pour fournir aux utilisateurs de ces systèmes des prédictions aussi exactes et précises que possible. Pour cette raison, nous proposons d'étudier le compromis entre la précocité et l'exactitude des prédictions fournies par un EWS. Afin d'atteindre cet objectif, nous nous basons sur le calcul de la moyenne harmonique entre ces deux métriques. La moyenne harmonique permet de mesurer la tendance centrale et elle est utilisée lorsqu'un rapport moyen est demandé. C'est un moyen pour mettre en évidence la proportionnalité inverse entre deux variables.
[107] a déjà utilisé cette mesure pour étudier la relation entre l'exactitude et la précocité pour la classification des signaux électroniques. Nous adaptons ce travail et utilisons cette même mesure pour déterminer la relation entre la précocité et l'exactitude dans un domaine complètement différent qui est celui de l'éducation. Nous proposons ainsi la mesure de EAS. À notre connaissance, la moyenne harmonique entre une métrique temporelle et une métrique de précision n'a jamais été utilisée dans le domaine de l'éducation pour exprimer la capacité d'un EWS à fournir au plus tôt des prédictions exactes. La mesure de EAS est calculée comme suit :

$$EAS = \frac{2 * (1 - precocite) * exactitude}{(1 - precocite) + exactitude} \tag{4.2}$$

D'après sa définition et son équation, plus la mesure de EAS est élevée, plus le système est qualifié d'avoir une capacité à fournir des prédictions précoces et précises.

4.5.2 Le compromis Précocité-Stabilité : ESS

L'apprentissage et la prédiction des résultats des apprenants sont deux processus évolutifs dans le temps ; d'où l'importance des métriques de précocité et de stabilité pour évaluer la performance des EWS dans le temps.

Comme les données d'apprentissage sont plus disponibles en avançant dans l'année scolaire, la précision des prédictions devrait aussi suivre et augmenter ; le système prédictif gagne en stabilité. En d'autres termes, l'exactitude des prédictions et leur stabilité augmentent côte à côte avec le temps. Cependant, [133] prouve que cette hypothèse n'est pas toujours évidente et montre que les courbes d'exactitude et de stabilité peuvent légèrement diverger l'une de l'autre. Cela peut être expliqué par les principales définitions de l'exactitude et de la stabilité. En effet, l'exactitude correspond à la fraction de prédictions correctes faites sur l'ensemble des prédictions. Cette métrique ne tient pas compte de la succession de ces bonnes prédictions dans le temps. La métrique de stabilité, quant à elle, décrit la capacité du système à maintenir la plus longue séquence de prédictions correctes sur un intervalle du temps donné. Ainsi, une valeur fixe de l'exactitude dans le temps, peut traduire une faible stabilité du système. Par conséquent, comme les prédictions peuvent osciller dans le temps, il est insuffisant de mesurer la précocité de la prédiction correcte indépendamment de la stabilité du système.

L'équilibre entre les mesures de la précocité et la stabilité d'un EWS permet d'identifier le point de précocité approprié qui garantit à la fois des prédictions précises et la stabilité du système prédictif.

Pour atteindre cet objectif, nous proposons une mesure de Précocité-Stabilité (*Earliness-Stability Score (ESS)*) qui illustre le compromis entre la précocité des prédictions et la stabilité temporelle du

EWS. Cette mesure se base aussi sur la moyenne harmonique entre la précocité et la stabilité et elle est calculée comme suit :

$$ESS = \frac{2 * (1 - precocite) * stabilite}{(1 - precocite) + stabilite} \tag{4.3}$$

Cette mesure cherche à identifier le temps le plus précoce de prédictions correctes et à partir duquel les performances du système restent stables pendant les temps de prédiction suivants. La stabilité ici est la plus longue séquence de prédictions correctes successives à partir de t_1 (temps de la première prédiction) jusqu'à un temps de précocité t_e.

4.6 Description des systèmes

Afin de prouver la pertinence des métriques temporelles proposées, nous les avons utilisé pour comparer quatre EWS existants. Le même jeu de donnée a été utilisé pour les quatre systèmes. Ce jeu de données correspond à celui du CNED présenté et décrit dans le Chapitre 2 de ce manuscrit. Cette section décrit ces systèmes ainsi que leurs approches.

L'objectif de ces systèmes est de suivre les performances des apprenants d'une manière hebdomadaire afin d'identifier au plus tôt ceux qui risquent d'échouer. Ainsi, le temps de prédiction $t_i \in T$ correspond à une semaine. Nous rappelons que notre jeu de données est composé de traces numériques de 647 apprenants ayant suivi le module de physique-chimie pendant 37 semaines. Les apprenants sont classés en trois classes en fonction de la moyenne des notes : $Y = \{C_1, C_2, C_3\}$:

— **réussite** C_1 **:** lorsque la moyenne des notes est strictement supérieure à 12.
— **risque d'échec moyen** C_2 **:** lorsque la moyenne des notes est comprise entre 8 et 12.
— **risque d'échec élevé** C_3 **:** lorsque la moyenne des notes est strictement inférieure à 8.

À chaque semaine, un apprenant est représenté par un ensemble de caractéristiques d'apprentissage et une étiquette de classe. Ces caractéristiques d'apprentissage incluent la performance, l'engagement, la régularité et réactivité (voir Chapitre 3).

Pour donner un exemple d'application des métriques temporelles proposées, nous les avons utilisées pour comparer quatre EWS différents (EWS_1, EWS_2 [9], EWS_3, EWS_4 [131]).
Les deux premiers systèmes (EWS_1, EWS_2) utilisent le modèle (RF), tandis que les deux derniers (EWS_3, EWS_4) utilisent le modèle des réseaux de neurones artificiels (ANN).
Pour une prédiction hebdomadaire, EWS_1 et EWS_3 utilisent toutes les caractéristiques d'apprentissage extraites de la base du CNED et exprimant les indicateurs d'apprentissage (introduits dans le Chapitre 3), y compris la performance, l'engagement, la régularité et la réactivité, en plus des caractéristiques démographiques. Tandis que, EWS_2 et EWS_4 se servent uniquement de l'engagement, défini par les caractéristiques d'actions sur la plateforme en ligne, pour identifier les apprenants à risque d'échec.

Une autre différence entre les systèmes est présentée par la manière par laquelle l'étiquette de classe (C_1, C_2, C_3) est attribuée dans le temps. En effet, pour les deux premiers systèmes (EWS_1, EWS_2), les prédictions sont réalisées par rapport à la classe finale de l'apprenant en fin d'année [9]. En d'autres termes, chaque apprenant appartient à une seule et unique étiquette de classe pendant toutes les semaines de prédiction et le modèle tente de la prédire correctement le plus tôt possible. Pour les deux derniers systèmes (EWS_3, EWS_4), l'attribution de l'étiquette de classe est dynamique et a tendance à changer en fonction de la performance des apprenants [131]. Par exemple, un apprenant peut être dans la classe de réussite pendant 3 semaines consécutives, mais à la 4ème semaine, il peut être affecté à une étiquette de

classe différente en raison de fluctuations de performance. Le modèle doit donc capter ces changements d'une semaine à une autre pour prédire correctement de classe de l'apprenant.

La Figure 4.2 montre l'évolution de l'exactitude de quatre systèmes (EWS_1, EWS_2, EWS_3, EWS_4) sur l'ensemble de dates de prédiction. Comme le montre la Figure 4.2, EWS_1 et EWS_3, qui utilisent toutes les caractéristiques d'apprentissage, montrent une bonne performance et un meilleur fontionnement en termes d'exactitude hebdomadaire que les deux systèmes EWS_2 et EWS_4 qui n'utilisent que les caractéristiques d'engagement.
Pour EWS_2, l'exactitude ne présente pas de grandes variations et elle reste quasiment stable au fil des semaines. ; elle atteint une valeur maximale de $\approx 73\%$.
EWS_4 montre une exactitude$\approx 91\%$ au temps de prédiction t_1. Cette valeur est, en effet, due à l'attribution imposée de l'étiquette C_1 à tous les apprenants de la base de données. En avançant dans le temps, les classes de ce système se forment et les valeurs d'exactitude changent. De la semaine 1 à 6, l'exactitude de ce système connaît une diminution, puis, elle augmente de nouveau à partir de la semaine 7 pour atteindre une valeur $\approx 72\%$ à la dernière date de prédiction.
À travers leurs courbes d'exactitude, ces deux systèmes (EWS_2 et EWS_4) restent assez performants en général avec des exactitudes qui ne se dégradent pas au cours du temps.

Cependant, cette métrique d'exactitude est encore une fois insuffisante pour mettre l'accent sur la précocité des systèmes par rapport à toutes les classes et notamment celles du risque élevé et moyen. De plus, l'utilisation de cette métrique n'est pas un appui pertinent pour conclure sur la stabilité des prédictions de ces systèmes.

Ainsi, dans la partie expérimentale, les quatre systèmes en question seront évalués en se basant sur les métriques temporelles de la précocité et stabilité. Cette évaluation sera aussi suivie des résultats de l'étude des différents compromis entre les métriques proposées et celles déjà existantes.

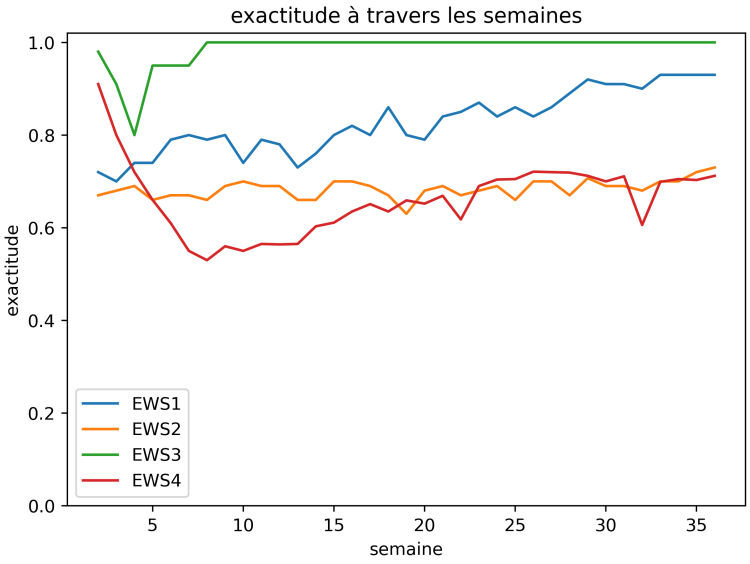


FIGURE 4.2 – EWS_1 VS EWS_2 EWS_3 EWS_4 en terme d'exactitude

4.7 1ère partie expérimentale

Dans cette partie, nous présentons les expérimentations conduites pour valider la première définition proposée pour la précocité et la définition de la stabilité. Nous proposons aussi de calculer la mesure de EAS selon cette définition de précocité. Pour ce faire, nous procédons à l'évaluation de la performance de quatre EWS introduits ci-dessus.

4.7.1 Résultats des métriques temporelles sur les 12 premières semaines

Selon les travaux réalisés sur l'importance de la précocité dans les systèmes prédictifs [45], la précocité est toujours ciblée aux premières semaines d'apprentissage pour que l'intervention soit efficace. C'est pour cette raison que nous commençons par présenter les résultats des métriques temporelles et de précision en considérant seulement les prédictions des 12 premières semaines. Dans le système français, un trimestre est composé de 12 semaines ; ce qui explique le choix de 12 semaines pour la première partie de l'expérimentation. En effet, il serait pertinent de pouvoir détecter les apprenants à risque au premier trimestre de l'apprentissage pour intervenir auprès d'eux et leur laisser le temps suffisant pour se rattraper.

4.7.1.1 Résultats de précocité, d'exactitude et de EAS sur 12 semaines

Les Tableaux 4.1 et 4.2 présentent les valeurs de précocité, d'exactitude et de EAS pour les systèmes EWS_1, EWS_2, EWS_3 et EWS_4. La précocité est pertinente, surtout pour détecter les apprenants appartenant aux classes du risque moyen et élevé (C_2 et C_3). Ainsi, nous présentons les résultats de précocité et EAS par étiquette de classe.

Comparaison entre EWS_1 et EWS_2 : pour ces deux systèmes, la classe dominante est C_1 (la classe de la réussite). Sur l'intervalle de prédiction fixé à 12 semaines, la précocité de prédiction de cette classe pour EWS_1 et EWS_2 correspond respectivement à 9.5% ($\approx 1.14 semaines$) et 8.83% ($\approx 1 semaine$). Les valeurs de précocité et d'exactitude de cette classe diffèrent légèrement entre les deux systèmes. Par contre, en nous référant aux valeurs de la mesure de EAS, nous remarquons que EWS_2 possède une EAS plus élevée et donc il est qualifié meilleur pour la prédiction de cette classe C_1.
En regardant les résultats de précocité, d'exactitude et de EAS des classes C_2 et C_3, nous constatons que le système EWS_2 est le pire lorsqu'il s'agit de détecter ces deux classes. En effet, sur les 12 premières semaines, ce système montre une mesure de EAS égale à 0% pour C_2 et C_3.

Comparaison entre EWS_3 et EWS_4 : pour ces deux systèmes EWS_3 et EWS_4, la classe dominante est C_3. Sur l'intervalle de prédiction de 12 semaines, les temps de précocité de prédiction de cette classe pour EWS_3 et EWS_4 sont respectivement à 9.33% ($\approx 1.11 semaines$) et 8.5% ($\approx 1 semaine$). Ces deux valeurs de précocité sont légèrement différentes, mais EWS_3 montre une plus grande exactitude (qui est égale à 100%) et donc une valeur de EAS plus élevée de $\approx 5\%$.
Par convention, un bon modèle prédit avec une plus haute précision en avançant dans le temps. Cependant, EWS_4 prédit la classe C_1 plus tard que EWS_3 mais moins précisément. De plus, bien qu'il ait prédit la classe du risque moyen C_3 avec précision et assez tôt, EWS_4 n'a jamais prédit les apprenants de la classe C_2 au cours des 12 premières semaines.
En nous basant sur les résultats de EAS, EWS_3 présente des bonnes performances pour la prédiction des trois classes et montre surtout une meilleure capacité à détecter les apprenants des classes du risque C_2 et C_3.

	EWS_1			EWS_2		
	Précocité	exactitude	EAS	Précocité	exactitude	EAS
C_1	9.5%	92.63%	90.95%	8.83%	91%	91.29%
C_2	26.6%	9%	16.03%	0%	0%	0%
C_3	44.1%	46%	50.46%	0%	0%	0%

TABLE 4.1 – Mesures de précocité, exactitude et EAS-EWS_1 VS EWS_2

	EWS_3			EWS_4		
	Précocité	exactitude	EAS	Précocité	exactitude	EAS
C_1	10.75%	78.94%	83.83%	37.85%	21.62%	25%
C_2	17.66%	55.75%	66.52%	0%	0%	0%
C_3	9.33%	100%	95.12%	8.5%	92.37%	90.65%

TABLE 4.2 – Mesures de précocité, exactitude et EASA-EWS_3 VS EWS_4

Synthèse : pour les quatre systèmes, la classe dominante est toujours prédite en premier et au plus tôt par rapport au reste des classes.
Si nous interprétons le taux de précocité de prédiction de la classe C_2 pour EWS_2 et EWS_4, nous pensons que ces derniers prédisent mieux cette classe que EWS_1 et EWS_3 respectivement. En effet, théoriquement, 0% est la meilleure valeur de précocité accompagnée d'une exactitude de valeur égale à 100%. Cependant, ceci n'est pas vrai puisque les valeurs d'exactitude et de EAS sont égales à 0. Ainsi, 0% n'explique rien à moins d'étudier l'exactitude et par conséquent la mesure de EAS. Ces résultats prouvent encore une fois l'importance de la sélection des caractéristiques pour prédire au plus tôt les apprenants à risque d'échec. En effet, EWS_2 et EWS_4, qui n'utilisent que des caractéristiques d'engagement, sont les meilleurs exemples, où les classes du risque peuvent ne pas être détectées. Ces résultats prouvent également que l'évaluation d'un EWS basée soit sur l'exactitude soit sur la précocité n'est pas assez pertinente pour conclure sur la performance d'un modèle de classification. Le compromis entre les deux types de métriques à travers la mesure de EAS donne un aperçu plus précis sur la performance d'un EWS. La précocité et le EAS peuvent également être utilisés pour évaluer la pertinence des approches adaptées par les systèmes prédictifs (exemple : EWS_1 et EWS_3). Le premier système (EWS_1) effectue les prédictions par rapport à la classe finale à la fin de l'année scolaire. Tandis que le second (EWS_3) effectue les prédictions à un instant t par rapport à l'étiquette de classe à l'instant $t+1$. Comme le prouve les Tableaux 4.1 et 4.2, les mesures EAS montrent que EWS_1 est moins précoce et moins exacte que EWS_3 pour prédire C_2 et C_3, sachant qu'ils utilisent la même base de test.

4.7.1.2 Résultats de la stabilité sur 12 semaines

La stabilité est complémentaire à la mesure de EAS. Le Tableau 4.3 présente les mesures de stabilité des quatre systèmes sur les 12 premières semaines de prédiction. En nous référant à la définition de la stabilité (Section 4.4), $t_1 = t_e$ où t_e est un point de temps de précocité et $t_2 = 12^{\text{ème}}$ semaine d'apprentissage.
Selon le résultat de ce tableau, nous constatons que la stabilité par classe est plus pertinente pour les deux systèmes EWS_1 et EWS_2. Ce premier résultat est expliqué par la manière dont ces deux systèmes attribuent les étiquettes de classe aux apprenants. En effet, EWS_1 et EWS_2 prédisent par rapport à une classe finale. Tandis que, EWS_3 et EWS_4 prédisent par rapport aux étiquettes de classe au temps $t+1$. Pour cette raison, nous envisageons de suivre également la stabilité des prédictions sur l'ensemble du jeu de données de test D_{test}.

Les résultats de la stabilité de D_{test} sont plus cohérents avec ceux de EAS et la stabilité globale (de tout le système) est plus pertinente ; elle reflète la vraie stabilité du système.
Comme le montre le tableau 4.3, EWS_1 est plus stable que EWS_2 en terme de stabilité par classe et globale. Donc, EWS_1 réussit à prédire correctement les étiquettes de classe et à construire des séquences plus longues de prédictions correctes.
Alors que, en ce qui concerne EWS_3 et EWS_4, la stabilité globale est plus pertinente, et cela montre que EWS_3 est plus stable.
Contrairement à la précocité qui est plus intéressante lorsqu'elle est calculée par étiquette de classe, la stabilité est plus pertinente lorsqu'elle est calculée par rapport aux résultats des prédictions de toutes les classes du système. La stabilité globale traduit l'évolution de l'ensemble du système au fil de temps. Dans le reste des expérimentations, nous considérons la stabilité globale du système.

	EWS_1	EWS_2	EWS_3	EWS_4
C_1	92.26%	89.12%	58.79%	18.24%
C_2	11.36%	0%	35.85%	0%
C_3	27.56%	0%	47.70%	39.86%
D_{test}	72.72%	65.12%	88.10%	48.48%

TABLE 4.3 – Mesure de stabilité

4.7.2 Résultats des métriques temporelles sur toute l'année

Dans cette 2ème partie, nous appliquons la 1ère définition de la précocité, la stabilité et la mesure de EAS en considérant la totalité des semaines de prédiction. Ceci sert à une comparaison des résultats et une analyse des performances des systèmes.

4.7.2.1 Résultats de précocité, d'exactitude et de EAS sur toute l'année

Les Tableaux 4.4 et 4.5 présentent les résultats des mesures de précocité, exactitude et de EAS calculés sur toute l'année pour respectivement les systèmes EWS_1, EWS_2, EWS_3 et EWS_4. Ces mesures sont calculées par étiquette de classe.

Comparaison entre les systèmes EWS_1 et EWS_2 : pour ces deux systèmes, la classe dominante est toujours la classe C_1. Sur 37 semaines de prédiction, EWS_1 et EWS_2 prédisent cette classe avec une précocité égale à 3.24% qui correspond à ≈ 1.2 semaines. Les valeurs d'exactitude et de EAS correspondantes à C_1 montrent que EWS_1 prédit mieux les apprenants de cette classe.
EWS_1 montre les meilleures performances pour la prédiction de la classe C_2 que le système EWS_2. En effet, les mesures de précocité de cette classe de EWS_1 et EWS_2 correspondent respectivement à 72.37% ($\approx 12.27 semaines$) et 71.95% ($\approx 26.62 semaines$). EWS_1 prédit plus tôt les apprenants de C_2. De plus, la mesure de EAS de cette classe par EWS_1, qui est égale à 72.37%, est beaucoup plus élevée que celle donnée par EWS_2 et qui est égale à 39.54%.
En ce qui concerne la prédiction de la classe C_3, EWS_1 est aussi plus performant que EWS_2 qui ne prédit jamais cette classe. 0% ne signifie pas encore une fois une précocité très élevée ; mais plutôt, l'incapacité de détecter la classe par le système.

Comparaison entre les systèmes EWS_3 et EWS_4 : pour ces deux systèmes, la classe dominante est C_3. Sur toute la période de l'apprentissage, EWS_3 et EWS_4 prédisent cette classe C_3 avec une précocité égale 2.77% qui correspond à $\approx 1 semaine$. Les résultats de EAS montrent que EWS_3 prédit

mieux cette classe d'apprenants.
EWS_3 prédit aussi très bien la classe C_1 avec une précocité égale à 2.77%. Par contre, EWS_4 prédit cette classe avec une précocité plus élevée et égale à 8.33%. Les résultats de EAS de cette classe montre que EWS_3 prédit mieux C_1 que EWS_4.
Cependant, EWS_4 ne détecte pas les apprenants de la classe C_2. 0% de précocité signifie ici l'incapacité du système à prédire cette classe. EWS_3 est plus performant que EWS_4 pour la prédiction de C_2

	EWS_1			EWS_2		
	Précocité	exactitude	EAS	Précocité	exactitude	EAS
C_1	3.24%	69.9%	81.16%	3.24%	67%	79.17%
C_2	33.16%	78.9%	72.37%	71.95%	67%	39.54%
C_3	22.02%	79.6%	78.78%	0%	0%	0%

TABLE 4.4 – Mesures de précocité et EAS -EWS_1 VS EWS_2

	EWS_3			EWS_4		
	Précocité	exactitude	EAS	Précocité	exactitude	EAS
C_1	2.77%	98%	97.61%	8.33%	78%	84.28%
C_2	5.55%	96%	95.21%	0%	0%	0%
C_3	2.77%	98%	97.61%	2.77%	91%	94.01%

TABLE 4.5 – Mesures de précocité et EAS -EWS_3 VS EWS_4

Comparaison des tableaux de précocité et EAS (Tableaux 4.1 et 4.4) : Les deux Tableaux 4.1 et 4.4 présentent les résultats des mesures de précocité, exactitude et EAS de deux systèmes EWS_1 et EWS_2 sur respectivement les 12 premières semaines de prédiction et la totalité des semaines de prédiction.
Selon le résultat de EAS, EWS_1 et EWS_2 du Tableau 4.1 montrent des meilleurs résultats pour la prédiction de la classe C_1.
Pour la prédiction de C_2, EWS_1 du Tableau 4.1 montre une meilleure valeur de précocité que celle du Tableau 4.4. Alors que, la mesure de EAS de EWS_1 du Tableau 4.4 est plus élevée que celle du Tableau 4.1. De plus, en considérant seulement les résultats des 12 premières semaines, EWS_2 ne prédit jamais les apprenants de C_2. Cependant, en tenant compte de la totalité des semaines de prédiction, la précocité de cette classe avec ce même système s'élève à 71.95% et donc aussi sa mesure de EAS. Les résultats de prédiction de cette classe sont expliqués par la disponibilité des données au cours du temps. En d'autres termes, en avançant dans le temps, EWS_2 apprend mieux les caractéristiques des apprenants de C_2 et donc les prédictions s'améliorent.
En tenant en considération l'intervalle de toutes les semaines de prédiction, la précocité de prédiction de C_3 par le système EWS_1 est meilleure que celle lorsque l'intervalle d'étude est fixé à 12 semaines donné par le Tableau 4.1. C_3 est prédit plus tôt avec EWS_1 du Tableau 4.4. La mesure de EAS est donc aussi plus élevée. Cependant, EWS_2 n'arrive pas à détecter les apprenants de cette classe C_3 pendant toutes les semaines de prédiction ce qui explique le pourcentage de 0% de précocité donné par le Tableau 4.4.

Comparaison des tableaux de précocité et EAS (Tableaux 4.2 et 4.5) : les Tableaux 4.2 et 4.5 présentent les résultats des mesures de précocité, exactitude et EAS de deux systèmes EWS_3 et EWS_4 sur respectivement les 2 premières semaines de prédiction et la totalité de temps de prédiction.
Selon les résultats de EAS, EWS_3 et EWS_4 du Tableau 4.5 montrent des meilleures performances pour la prédiction de la classe C_1.

La classe C_2 n'est pas détectée par le système EWS_4 dans les deux cas des Tableaux 4.2 et 4.5. Cependant, les résultats de EWS_3 du Tableau 4.5 montrent des meilleures performances de prédiction de C_2 que ceux du Tableau 4.2 avec un EAS égal à 66.52%.
Pour la classe C_3, les résultats du Tableau 4.5, notamment de EAS, montrent que EWS_3 et EWS_4 prédisent mieux cette classe sur toutes les semaines de prédiction.

Synthèse : les différences des résultats de précocité et EAS entre les Sections 4.7.1.1 et 4.7.2.1 montrent que le système a tendance à avoir des changements de performance dus aux changements des comportements d'apprentissage des apprenants. Ceci met en question la pertinence de la précocité lorsqu'elle est définie comme la première date de prédiction correcte. En effet un apprenant peut être bien prédit correctement à un temps t_i et est mal prédit pendant une succession de dates de prédiction. Ce sont les résultats qui nous laissent nous poser des questions sur la représentativité de la première date de bonne prédiction de la performance du système pour prédire une classe : est-ce que la précocité telle qu'elle est définit permet vraiment d'évaluer le système ? C'est à partir de toutes ces idées et questions que la définition de la précocité a été mise à jour dans la Section 4.3.2.

4.7.2.2 Résultats de la stabilité sur toute l'année

Un EWS doit maintenir une performance stable tout au long la période d'apprentissage. De plus, la Section 4.7.1.2 montre qu'il est plus pertinent de suivre la stabilité générale du système que la stabilité par classe.
La figure 4.3 montre l'évolution de la stabilité des jeux de données D_{test} des quatre systèmes (EWS_1, EWS_2, EWS_3 et EWS_4) tout au long de l'année scolaire.
La stabilité de EWS_1 augmente légèrement avec le temps ; il enregistre des taux de stabilité qui sont à $\approx 70\%$ et $\approx 76\%$ respectivement à la première et dernière semaine de prédiction. Tandis que, la stabilité de EWS_2 diminue légèrement au fil des semaines. En effet, ce système enregistre des taux de stabilité de $\approx 64\%$ et $\approx 59\%$ respectivement à la première et dernière date de prédiction.
Les deux systèmes EWS_3 et EWS_4 commencent avec des taux de stabilité assez élevés ($\approx 100\%$ et $\approx 92\%$ respectivement). Cela s'explique par la dominance de la classe C_3, puisqu'au départ tous les apprenants sont affectés par défaut à cette classe. Cependant, ces valeurs élevées diminuent rapidement au cours des semaines suivantes. Vers la 4ème semaine, le système EWS_3 commence à affecter correctement chaque apprenant à une classe parmi C_1, C_2 et C_3. À partir de la 5ème semaine, la stabilité du système EWS_3 recommence son augmentation continûment et atteint un taux de $\approx 96\%$ à la dernière date de prédiction.
L'allure de la courbe de stabilité du système EWS_4 est identique à celle du EWS_3 avec un décalage des valeurs. Jusqu'à la semaine 13, la stabilité globale de EWS_3 montre une diminution et commence à ré-augmenter à partir de la semaine 14 pour atteindre un taux de $\approx 57\%$ à la dernière date de prédiction.

À partir des courbes de la Figure 4.3, nous constatons une baisse partielle ou totale de stabilité des systèmes EWS_2, EWS_3 et EWS_4. Bien que la courbe de stabilité de EWS_1 ne montre pas une diminution au cours du temps, les taux de stabilité enregistrés par EWS_3 confirment que ce dernier reste plus stable.

L'exactitude et la stabilité sont proportionnelles ; plus le système montre des valeurs d'exactitude élevées, plus il est stable. Cependant, à partir des courbes d'exactitude (Figure 4.2) et celles de stabilité (Figure 4.3) des systèmes EWS_2 et EWS_4, nous ne pouvons pas conclure sur l'efficacité de ces derniers à maintenir des prédictions correctes dans le temps. Ce résultat montre qu'il faut aller encore en profondeur avec l'évaluation des systèmes prédictifs et étudier les différents compromis qui peuvent exister entre les métriques temporelles et de précision.

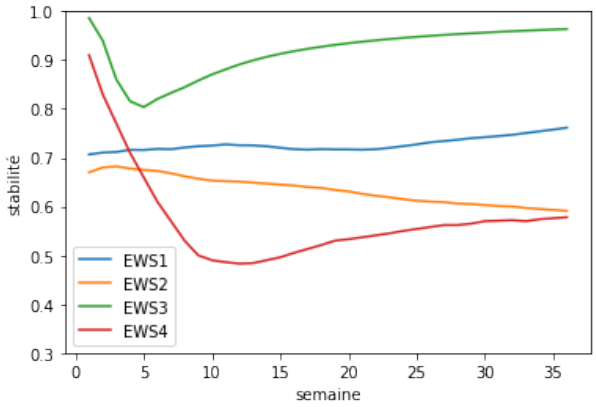


FIGURE 4.3 – Comparer EWS_1, EWS_2, EWS_3 and EWS_4 en terme de stabilité temporelle

4.8 2ème partie expérimentale

Dans cette partie, nous utilisons la mise à jour effectuée sur l'algorithme de la précocité pour calculer les deux mesures de EAS et ESS. Nous comparons ensuite leur pertinence.
Dans cette partie, et nous référant à l'équation de la stabilité 4.1, t_1 correspond au temps de la 1ère prédiction et t_e correspond à un point du temps de précocité.

4.8.1 Analyse des mesures de compromis

Les Figures 4.4 et 4.5 présentent respectivement les résultats des mesures EAS et de ESS pour les quatre systèmes EWS_1, EWS_2, EWS_3 et EWS_4 et pour chaque étiquette de classe. L'axe des abscisses de ces figures correspond aux points de précocité. Chaque point de précocité contient une valeur temporelle qui peut être différente d'une étiquette de classe à une autre. Par exemple, pour EWS_3, pour le point d'axe $x = 1$, les valeurs du temps de précocité pour C_1, C_2 et C_3 sont respectivement, semaine 1, semaine 2 et semaine 1. L'axe des ordonnées correspond aux taux EAS et ESS.
Comme le montrent les Figures 4.4 et 4.5, les étiquettes de classe C_3 et C_2 n'ont jamais été détectées respectivement par EWS_2 et EWS_4 au cours des semaines de prédiction.

Selon les définitions de deux mesures EAS et ESS (Sections 4.5.1 et 4.5.2), plus les taux de ces métriques sont élevés, meilleur est le système.

Atteindre un EAS ou un ESS de 100% n'est pas toujours possible car cela correspond à un EWS idéal. Pour cette raison, nous devons établir une valeur du seuil pour identifier le point de précocité le plus approprié du système. Pour ce travail, nous avons défini 70% comme une valeur acceptable de EAS et ESS.

Pour chacune de ces métriques, la définition de ce seuil permet d'identifier le premier point de précocité à partir duquel le système est précis et stable et les prédictions correctes sont précoces. Ce point est ensuite utilisé comme point initial pour surveiller l'évolution de EAS et ESS sur le reste des points de précocité. En d'autres termes, nous examinons si les valeurs des EAS et ESS restent au-dessus du seuil prédéfini au fil du temps.
En surveillant les valeurs de EAS et ESS restantes, nous distinguons trois situations différentes :

1. 1ère situation (S1) : les taux des métriques restent supérieurs ou égaux au seuil fixé sur le reste des points de précocité. Dans ce cas, le point de précocité du système correspond au premier point de précocité.

2. 2ème situation (S2) : les taux des métriques diminuent par rapport au seuil fixé sur le reste des points de précocité. Dans ce cas, le EWS n'est pas assez exact/stable. Et, nous avons deux observations concernant l'exactitude/la stabilité du système : si la valeur de EAS/ESS chute directement après les points de précocité initiale, le EWS n'est pas précis/stable et nous ne pouvons pas conclure sur sa précocité. Sinon, si les valeurs de EAS/ESS restent supérieures ou égales au seuil sur des points de précocité successifs puis diminuent, alors le système est partiellement exact/stable.
3. 3ème situation (S3) : les valeurs de EAS/ESS n'atteignent jamais le seuil. Dans ce cas, le système n'est pas exact/stable et ne fournit pas de prédictions correctes au début comme requis.

Dans les sections suivantes, nous analysons davantage les résultats de EAS et ESS obtenus avec les quatre systèmes de prédiction.

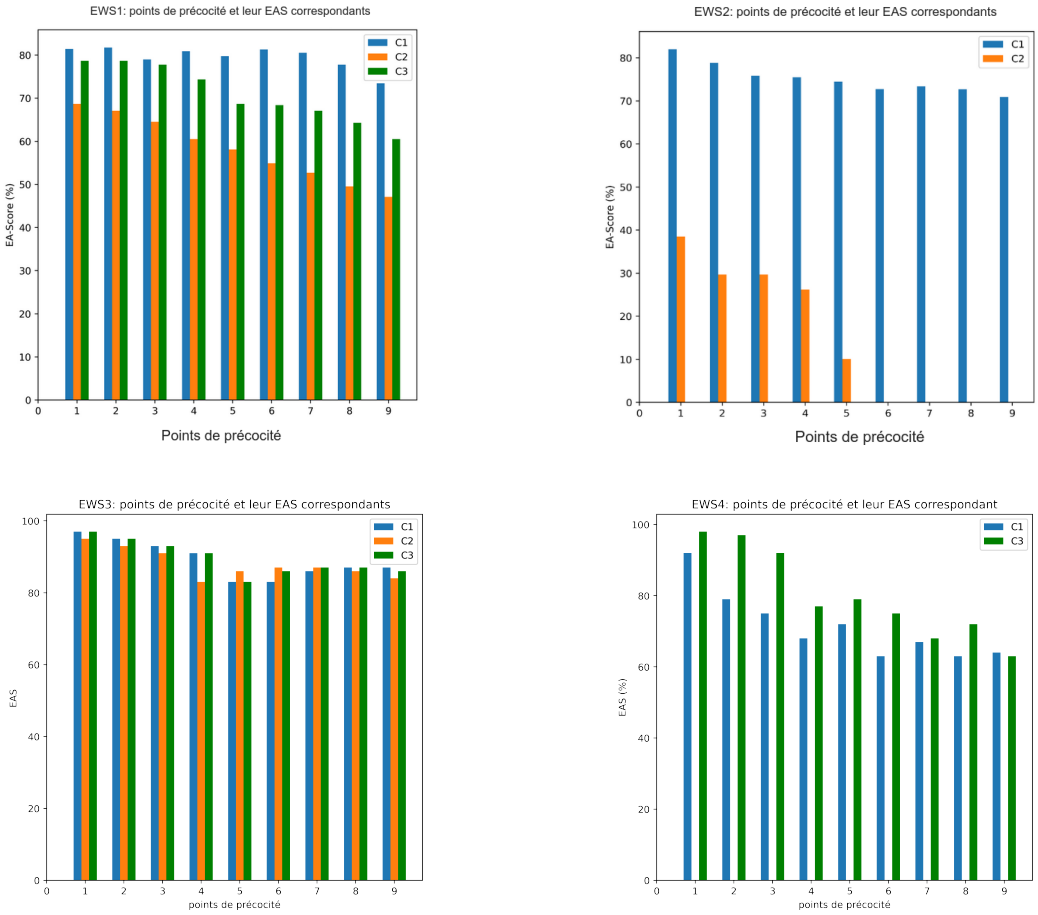


FIGURE 4.4 – Taux de la mesure de EAS par étiquette de classe -tous les systèmes-

4.8.1.1 Analyse des résultats de la mesure EAS

Dans cette section, nous présentons les résultats de EAS selon chacune des situations évoquées ci-dessus (Section 4.8.1) :

1. (S1) : comme le montre la Figure 4.4, pour le système EWS_3, pour chacune des étiquettes de classe C_1, C_2 et C_3, le premier point de précocité qui correspond à un EAS $\geq$70% est le point

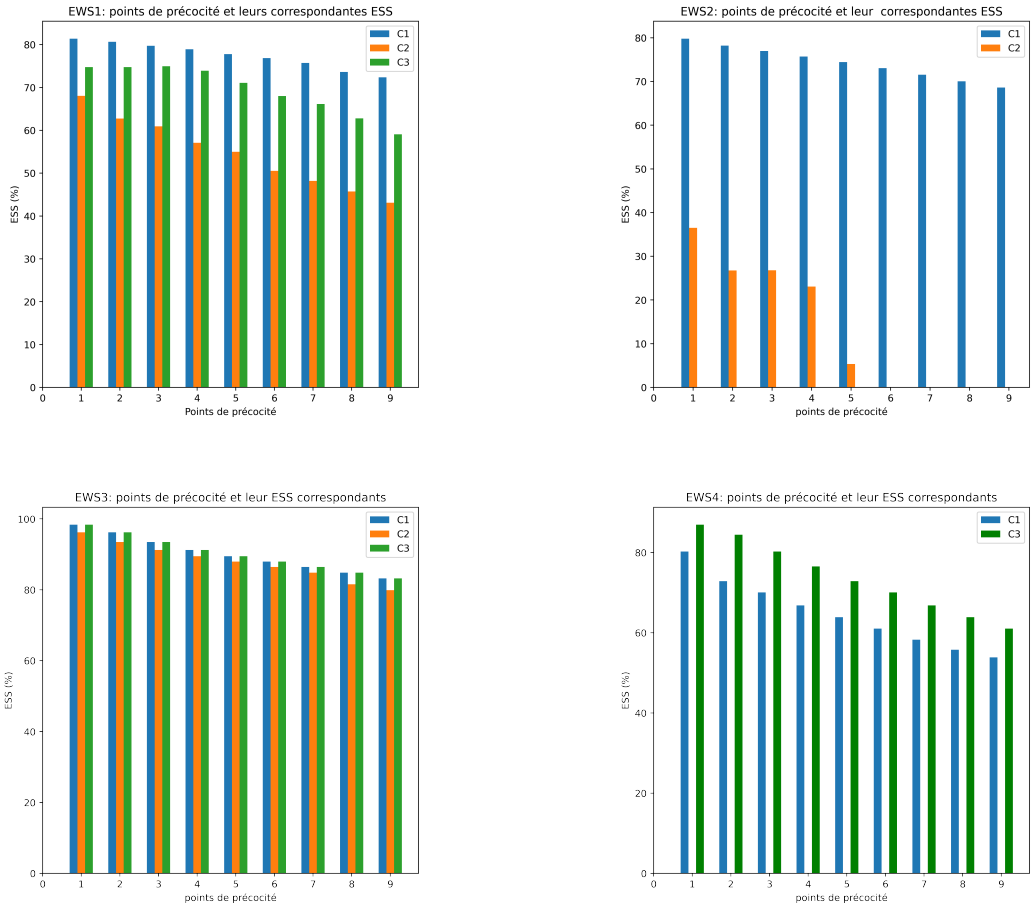


FIGURE 4.5 – Taux de la mesure de ESS par chaque étiquette de classe -tous les systèmes-

d'axe $x = 1$. Le reste des valeurs de EAS restent supérieures à 70%. Ainsi, ce premier point (x=1) représente le point de précocité pour les trois étiquettes de classe.

2. (S2) : pour le système EWS_1, le premier point de précocité pour l'étiquette de classe C_3 où la valeur de EAS est supérieure à 70% est le point d'axe x=1. Les valeurs EAS restent au-dessus de 70% jusqu'à x=4, temps auquel elles commencent à diminuer sur le reste des points de précocité restants. Pour EWS_4, pour le premier point de précocité pour l'étiquette de classe C_1 où l'EAS est supérieur à 70% est le point avec l'axe x=1. Les valeurs EAS restent supérieures à 70% jusqu'à x=3. Dans ce cas, la conclusion sur les points de précocité du système n'est pas évidente et nous devons vérifier les résultats de ESS (voir Section 4.8.1.2).
3. (S3) : pour EWS_2, les valeurs de EAS de l'étiquette de classe C_2 sont toutes inférieures à 70%. Dans ce cas, ce résultat peut être confirmé davantage avec les résultats de ESS (voir Section 4.8.1.2).

4.8.1.2 Analyse des résultats de la mesure ESS

Dans cette section, nous présentons les résultats de la mesure ESS (donnés par la Figure 4.5) correspondants à chacune des situations citées dans la Section 4.8.1 :

1. (S1) : par exemple, comme le montre la Figure 4.5, pour le système EWS_3, pour chacune des étiquettes de classe C_1, C_2 et C_3, le premier point de précocité qui correspond à un ESS $\geq$70% est le point d'axe x=1. Les valeurs de ESS pour chacune des étiquettes de classe restent supérieures au seuil sur le reste des points de précocité. Ainsi, nous pouvons conclure que le point avec x=1, est le point de précocité approprié. Pour EWS_3, les valeurs de temps de précocité pour C_1, C_2 et C_3 sont respectivement la semaine 1, la semaine 2 et la semaine 1. Cette observation confirme celle fournie avec EAS.
2. (S2) : Par exemple, pour EWS_4, le premier point de précocité pour l'étiquette de classe C_1, où la valeur de ESS est supérieure à 70% est le point d'axe x=1. Les valeurs ESS restent supérieures à 70% jusqu'à x=3, temps auquel elles commencent à diminuer sur le reste des points de précocité. De même pour EWS_1, les valeurs de ESS restent supérieures au seuil jusqu'au point x=5. Ensuite, elles diminuent. Dans ce cas, nous pensons qu'il serait intéressant d'étudier la métrique de ESS par intervalle du temps (par exemple un trimestre). En effet, les données deviennent progressivement disponibles au fil du temps et peuvent présenter quelques différences en termes de représentation des étiquettes de classe. Ainsi, cela peut affecter la stabilité du système et par conséquent les valeurs de ESS.
3. (S3) : par exemple, pour EWS_2, les valeurs de ESS de l'étiquette de classe C_2 sont toutes inférieures à 70%.

Pour résumer, selon le seuil fixé pour EAS et ESS, EWS_1 et EWS_3 sont les plus précis et les plus stables car ils fournissent des prédictions correctes précoces. EWS_3 est d'autant plus efficace que toutes les valeurs EAS et ESS pour toutes les étiquettes de classe sont supérieures au seuil.

4.8.1.3 Comparaison entre EAS et ESS :

EAS et ESS sont deux mesures complémentaires et leur utilisation permet d'identifier le point de précocité le plus approprié. Cependant, l'utilisation de ESS est plus pertinente. En effet, même si les deux mesures se croisent au premier point de précocité, le suivi des résultats des points de précocité qui suivent peut montrer que les valeurs de EAS sont probablement supérieures à celles de ESS. Cette observation est plus nette avec les systèmes les moins performants tels que EWS_2 et EWS_4. Par exemple, pour EWS_4, pour l'étiquette de classe C_3, les 6 premiers points de précocité ont des valeurs de EAS supérieures à 70%. Cependant, pour cette même classe, seuls les cinq premiers points de précocité ont des valeurs de ESS supérieures à 70%.

4.8.2 Discussion

En résumé, cette étude expérimentale a permis de tirer les résultats suivants :

— la stabilité et l'exactitude sont proportionnelles lorsque l'exactitude du système augmente ou diminue continûment avec le temps. Cependant, lorsque l'exactitude stagne ou varie légèrement, la stabilité du système diminue. De ce fait, l'utilisation de deux mesures de EAS et ESS est complémentaire et permet de mieux définir le point de précocité en étudiant la relation entre la précocité et l'exactitude d'une part et la précocité et la stabilité d'autre part.

— en général, une bonne valeur de EAS ou ESS tend vers un taux de 100%, ce qui, selon les définitions de ces deux métriques, correspond à une précocité de 0 (le système est capable de prédire correctement dès la première fois), une exactitude de 100% et une stabilité de 100%. Cependant, atteindre cette valeur pour un EWS est presque impossible. Par conséquent, afin d'identifier le point du temps précoce, nous avons établi un seuil pour EAS et ESS. Il n'y a

pas de règle pour déterminer le seuil approprié, mais il peut être défini en fonction des besoins, concernant la précocité de la prédiction à effectuer, de l'utilisateur final.

— lorsque l'exactitude varie légèrement ou stagne dans le temps et la stabilité diminue, l'utilisation de EAS pour identifier un point du temps précoce est insuffisante. Dans une telle situation, l'utilisation de ESS a permis de conclure sur le temps de la précocité. En général, nous avons constaté que ESS et EAS peuvent identifier le même point de précocité initial dans le temps. Cependant, lors du suivi de l'évolution de leurs valeurs sur le reste des points de précocité, nous avons constaté que pour certains des systèmes étudiés, les valeurs de EAS sont supérieures au seuil à certains points de précocité. Et, les valeurs de ESS sur ces mêmes points de précocité sont inférieures au seuil. Comme l'exactitude ne reflète pas les performances réelles du système lorsqu'il s'agit de sa capacité à maintenir des prédictions successives correctes dans le temps, nous concluons ainsi que l'ESS est une métrique plus pertinente que l'EAS pour caractériser le EWS en termes de précocité et de stabilité.

4.9 Conclusion

L'objectif de ce chapitre est l'évaluation de la performance des systèmes de prédiction en tenant en considération l'importance de la dimension temporelle lorsqu'il s'agit d'un contexte d'application évolutif dans le temps. Bien que l'apprentissage et la prédiction soient deux processus évolutifs dans le temps, l'évaluation temporelle est peu présente dans les LA.

Afin de surmonter ces limites de l'existant, nous avons introduit deux nouvelles métriques temporelles, à savoir la précocité et la stabilité pour évaluer des systèmes de prédiction éducatifs. De plus, un compromis entre la précocité et l'exactitude est important à étudier pour pouvoir évaluer la capacité d'un EWS à prédire précisemment et au plus tôt les apprenants à risque d'échec. L'utilisation de la mesure de EAS sert à illustrer cet équilibre. En outre, nous avons étudié la relation entre les métriques de la précocité et la stabilité à l'aide de la mesure de ESS permettant d'identifier le meilleur temps de précocité pour une prédiction d'étiquette de classe précoce et un système de prédiction stable.

Les différentes expérimentations réalisées montrent des résultats prometteurs et prouve l'importance des métriques temporelles proposées et les études de compromis effectuées pour garantir la précocité et l'exactitude du système, à travers le calcul de leurs mesures de EAS et ESS. Les résultats obtenus ont montré que le premier point de précocité n'est pas toujours le plus approprié pour prédire les apprenants à risque et maintenir un système stable. De plus, la fixation d'un seuil minimum pour la valeur de ESS et EAS permet d'avoir un moyen de comparaison et d'identifier l'intervalle de recherche du ou des meilleurs points de précocité. Les résultats ont aussi permis de conclure que la mesure de ESS est plus utile que celle de EAS pour identifier le point de précocité et suivre la stabilité du système sur un ensemble de points de précocité dans le temps.

En perspective de ce travail, ces métriques temporelles pourraient être utilisées de manière inverse pour déterminer la granularité temporelle de notre approche de prédiction. De plus, pour des utilisations plus génériques de ces métriques temporelles, nous les adaptons afin de pouvoir les appliquer dans l'évaluation des systèmes basés sur les modèles de régression.

Chapitre 5

Identification des profils types d'apprentissage pour évaluer les systèmes de prédiction

Synthèse :
Ce chapitre utilise les indicateurs d'apprentissage et les algorithmes des métriques temporelles pour une évaluation approfondie des systèmes prédictifs.
Étant donné l'hétérogénéité des apprenants existants dans une même cohorte en ligne et l'importance de fournir à chacun d'entre eux un suivi efficace et personnalisé, l'objectif de ce chapitre est de proposer une méthodologie permettant une évaluation plus fine des performances du système de prédiction. Tout d'abord, cette méthodologie identifie les profils d'apprenants existants et associe une description à chacun d'eux. Ensuite, elle évalue les prédictions du système par profil d'apprenants.

Ce travail permettra par la suite de générer des alertes par profil d'apprenants.

5.1 Introduction

Le phénomène des taux élevés d'échec et d'abandon dans les établissements en ligne est dû à de nombreuses raisons telles que la grande diversité de profils d'apprenants exprimant chacun des besoins et attentes différents et nécessitant un accompagnement personnalisé [134].

Les systèmes de LA proposés aux enseignants doivent tenir compte des divers comportements d'apprentissage existants. En effet, tous les apprenants doivent être bien accompagnés et soutenus dans leurs tâches d'apprentissage et aucun ne doit être privilégié, ou au contraire défavorisé [5]. Par conséquent, une question que nous pouvons nous poser est la suivante : ***les systèmes de prédiction sont-ils aussi performants sur l'ensemble de la population ?***
Afin de répondre à cette question, nous proposons une méthodologie, dont l'objectif est une évaluation plus détaillée d'un système de prédiction éducatif.

Pour ce faire, cette méthodologie repose sur l'identification des *personas* existants dans un contexte d'apprentissage et sur l'évaluation des performances des modèles sur chacun de ces personas. Les personas sont définis comme des représentations des profils d'apprenants [135] [136]. Nous illustrons cette méthodologie sur une étude de cas : modèle de prédiction utilisant des traces numériques réelles.

En bref, cette contribution consiste en une évaluation plus précise des systèmes éducatifs, prenant en compte les différents profils d'apprentissage et se basant sur un large éventail de métriques de précision et temporelles.
Nous procédons selon les étapes suivantes :

— compte-tenu de la disparité des traces numériques disponibles, nous définissons plusieurs indicateurs d'apprentissage caractérisant le comportement d'un apprenant. Ensuite, nous identifions des groupes homogènes d'apprenants partageant des comportements similaires selon ces indicateurs. Ces groupes d'apprenants sont finalement caractérisés en personas (travail fait dans [5]).
— nous examinons les métriques d'évaluation existantes (notamment l'exactitude) et avons recours aux métriques temporelles déjà introduites (voir Chapitre 4).
— nous effectuons une évaluation précise sur un contexte réel tout en nous appuyant sur une approche de prédiction hebdomadaire.

La partie expérimentale est menée à partir des données réelles des collégiens inscrits au CNED. L'apprentissage de la cohorte sélectionnée illustre la diversité des profils d'apprenants et la bi-modalité de l'apprentissage au CNED.

5.2 La méthodologie d'évaluation

Dans cette section, nous commençons par une description générale de la méthodologie proposée et ses différentes étapes. Ensuite, nous introduisons l'étude de cas de la partie expérimentale.

5.2.1 Description de la méthodologie

Afin d'atteindre nos objectifs d'évaluation, la méthodologie, présentée par la Figure 5.1, se déroule autour de trois étapes principales qui sont les suivantes :

1. ***identification des profils types d'apprenants*** issus du contexte d'apprentissage. Les profils types sont caractérisés par des personas, contenant des informations clés sur les comportements des apprenants [5]. Chaque apprenant est associé à un persona spécifique qui correspond au mieux à son comportement d'apprentissage. Afin d'atteindre les objectifs de cette étape, nous devons d'abord passer par l'analyse de la base de données et l'identification des indicateurs d'apprentissage. Ensuite, nous passons à l'extraction des caractéristiques en appliquant les opérations d'accumulation et/ou d'agrégation nécessaires.
2. ***exécution du système de prédiction*** sur les données et mesures d'un ensemble de métriques, contenant à la fois des métriques de précision de ML, comme l'exactitude, et métriques temporelles : la précocité et la stabilité. Afin d'atteindre ces objectifs d'évaluation, nous commençons par définir l'approche de prédiction et passer à la modélisation. Ensuite, nous appliquons les méthodes d'extraction d'erreur notamment le suivi régulier des prédictions (voir Chapitre 4) pour pouvoir calculer les métriques temporelles.
3. ***évaluation plus approfondie du système*** en fonction des profils types d'apprenants identifiés et des différentes mesures de performance. L'objectif de cette étape est d'évaluer les performances du système de prédiction selon les profils types d'apprenants identifiés à l'étape 1 et en utilisant les métriques sélectionnées à l'étape 2.

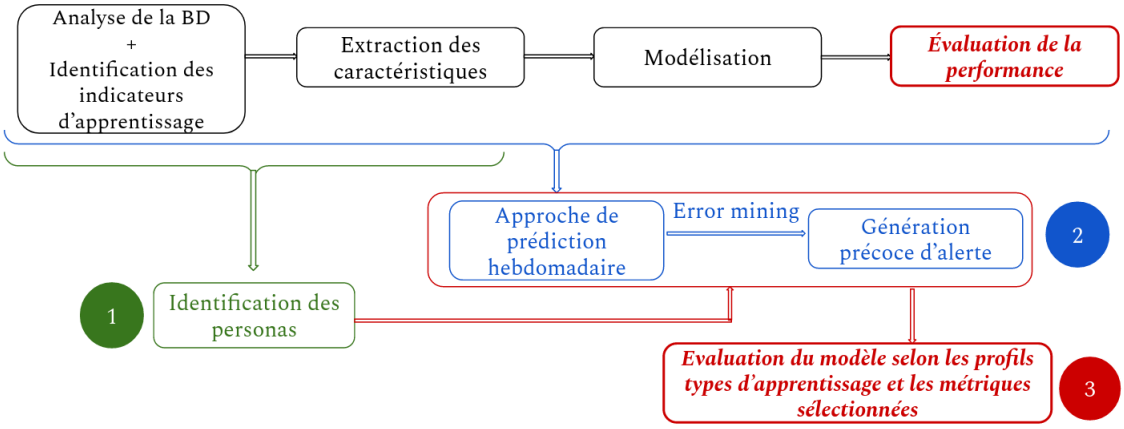


FIGURE 5.1 – Les étapes de la méthodologie d'évaluation

5.2.2 Description des données

Notre étude de cas concerne les collégiens du CNED inscrits en ligne dans le module de physique-chimie pendant l'année scolaire 2017-2018. Comme pour tous les modules d'apprentissage au CNED (voir Chapitre 2), les supports du cours de cette matière ainsi que les exercices d'application sont disponibles en ligne et en version papier. Ces différentes options donnent à l'apprenant la liberté de choisir le mode d'apprentissage qui lui correspond le mieux. Étant donné le grand nombre d'apprenants et la particularité de l'apprentissage au CNED, il est de plus en plus chronophage et difficile pour les enseignants de fournir à chacun de leurs apprenants un suivi pédagogique efficace et personnalisé ; d'où l'importance de l'identification des profils d'apprenants similaires.

Nous sélectionnons les traces numériques de la matière physique chimie effectuées par 639 apprenants ayant été inscrits entre le 1er et le 31 Octobre 2017. Ces apprenants partagent entre eux le même parcours et les mêmes caractéristiques d'apprentissage.
La durée d'apprentissage de cette matière est de 300 jours. Durant cette période, les apprenants peuvent soumettre jusqu'à 6 examens. En moyenne, les apprenants soumettent $4,51$ examens. La note moyenne obtenue par chaque apprenant dans les examens soumis est de $13,73$. Cependant, si nous attribuons 0 à chaque devoir non rendu, cette moyenne baisse et est égale à $10,21$.

La bi-modalité de l'apprentissage (numérique ou papier) rend difficile l'étude de ce jeu de données. En effet, les apprenants qui suivent le cours exclusivement au format papier peuvent ne pas produire de traces et il n'est donc pas pertinent de les comparer avec des apprenants actifs sur la plateforme d'apprentissage en ligne. Dans notre jeu de données, nous avons remarqué que $37,25\%$ de la population ne s'est jamais connectée tout au long de l'année. Afin de gérer cette particularité, le jeu de données a été divisé en deux sous-ensembles :

1. ***traces :*** ce sous-ensemble regroupe les apprenants ayant fait au moins une action (log) sur l'environement d'apprentissage en ligne.
2. ***pas_traces :*** ce sous-ensemble regroupe les apprenants qui ne se sont jamais connectés à a plateforme d'apprentissage en ligne.

Ensuite, nous définissons 4 classes d'apprenants de la cohorte selon leur performance :
— **réussite** (C_1) : la moyenne de notes finale obtenue par l'apprenant est strictement supérieure à 12.

— **risque d'échec moyen** (C_2) : la moyenne de notes finale obtenue est comprise entre 8 et 12.
— **risque d'échec élevé** (C_3) : la moyenne de notes finale est strictement inférieure à 8.
— **abandon** (C_4) : au moins les deux derniers examens ne sont pas rendus.

Le Tableau 5.1 résume le nombre d'apprenants appartenant à chaque sous-ensemble de *traces* et *pas_traces* et par chacune des classes C_1, C_2, C_3 et C_4. Nous remarquons que le nombre d'apprenants appartenant au sous-ensemble de *log* reste bien supérieur à celui des apprenants qui appartiennent au sous-ensemble de *pas_traces*. Les apprenants étiquetées comme appartenant aux classes C_1 et C_4 sont ceux qui se connectent le plus. Les apprenants à risque élevé d'échec (la classe C_3) sont ceux qui se connectent le moins. En effet, nous avons plus d'apprenants de C_3 qui ne se sont jamais connectés (28 apprenants) que d'apprenants de cette même classe qui se sont connectés au moins une fois pendant toute l'année scolaire (17 apprenants). Pour les apprenants de la classe du risque moyen C_2, la majorité d'entre eux appartient au sous-ensemble de *traces* (53 apprenants parmi 82).

	traces	***pas_traces***
C_1	178	64
C_2	53	29
C_3	17	28
C_4	153	117
Total	401	238

TABLE 5.1 – Nombre d'apprenants dans chaque classe de chaque sous-ensemble de données

5.3 1ère étape de la méthodologie : identification des profils types d'apprenants

L'objectif de cette partie est la présentation du processus d'identification des profils types d'apprenants existants dans une cohorte. Pour ce faire, nous commençons par une brève étude bibliographique sur les personas : leur définition et utilité. Ensuite, nous passons à la sélection des indicateurs d'apprentissage. Cette dernière sous-étape conduit à l'identification des profils types d'apprenants et leur représentation sous forme de personas.

5.3.1 État de l'art

Dans un contexte d'apprentissage en ligne, les comportements des apprenants sont observés à travers leurs traces numériques effectuées au sein des plateformes en ligne. Plusieurs projets de LA exploitent ces données pour comparer les apprenants en fonction de divers indicateurs, tels que l'engagement, la performance ou la régularité. Dans notre contexte, les comportements d'apprentissage sont décrits selon un ensemble d'indicateurs, permettant une caractérisation plus fine des apprenants [136] [137]. Pour cette raison, nous définissons les personas des apprenants correspondants aux apprenants typiques identifiés par les processus de classification du ML [135]. D'une part, l'identification de tels personas permet une description plus précise du corpus, notamment en termes de représentation des profils d'apprenants. D'autre part, ces personas répondent au besoin d'une mise en œuvre de l'analytique de l'apprentissage éthique [138], assurent un soutien équitable entre les apprenants et fournissent des outils utiles aux acteurs de terrain qui ont besoin d'aider leurs apprenants avec un soutien égal [139]. Cependant, la diversité des données disponibles rend la tâche délicate : la variété des données sauvegardées ne permet pas de

calculer tout le temps les mêmes indicateurs d'apprentissage. Les indicateurs que nous calculons pour notre contexte sont décrits dans la partie suivante.

5.3.2 Identification des indicateurs d'apprentissage

Les traces numériques disponibles dans l'ensemble de données du CNED sont diverses et contiennent à la fois des traces et des données de performance. Ces données ont d'abord été utilisées pour définir cinq indicateurs absolus (calculés pour chaque apprenant) :

- **engagement :** reflète l'activité de l'apprenant sur l'environnement d'apprentissage en ligne (traces).
- **régularité :** traduit la constance de connexion de l'apprenant entre le début et la fin du cours (fréquence de connexion).
- **curiosité :** exprime la motivation intrinsèque de l'apprenant à consulter diverses ressources pédagogiques (variété de contenus consultés) [140].
- **performance :** correspond aux notes des examens.
- **réactivité :** fournit des informations sur l'aptitude de l'apprenant à répondre aux différents événements liés au cours (rapidité de la passation évaluations).

Pour aller plus loin, nous complétons ces indicateurs absolus par un jeu d'indicateurs relatifs. La moyenne de chaque indicateur est calculée pour tous les apprenants, et l'indicateur relatif associé renseigne sur le comportement d'un profil d'apprenant spécifique par rapport à ses pairs (différence négative ou positive par rapport au reste du groupe).
Les deux types d'indicateurs ont servi de base pour l'identification des profils d'apprenants, décrits dans la section suivante. De toute évidence, les indicateurs d'engagement, de curiosité et de régularité, basés sur des caractéristiques extraites à partir des actions (traces) sur la plateforme en ligne, n'ont pas été calculés pour le sous-ensemble de *no_traces* car les apprenants associés ne se sont jamais connectés.

5.3.3 Identification des profils types d'apprenants

Pour chaque sous-ensemble de données, l'étude des indicateurs d'apprentissage permet d'identifier les profils types d'apprenants. Ces profils types correspondent à des sous-ensembles homogènes d'apprenants, partageant des comportements similaires. Ces profils d'apprenants sont identifiés suivant différentes étapes :

- **pré-traitement des données :**
  - *normalisation des données* : utilisation de la méthode RobustScaler [12] pour améliorer les performances du modèle.
  - *identification des valeurs aberrantes (outliers)* : il s'agit des apprenants avec des comportements particuliers. Pour ce faire, nous utilisons l'algorithme IsolationForest [13] pour distinguer les données atypiques et augmenter les performances du modèle. Cette étape est cruciale car l'atypicité des outliers ne leur permet pas d'être associés à d'autres élèves.
- **identification des groupes homogènes d'apprenants :** une fois les *inliers* et les valeurs aberrantes sont identifiés et séparés dans des sous-ensembles de données distincts, l'objectif est d'identifier des groupes homogènes d'apprenants parmi les *inliers* et selon les indicateurs d'apprentissage sélectionnés. L'algorithme k-means est utilisé pour identifier des groupes homogènes d'apprenants (Ce travail est fait dans [5]). Les résultats sont évalués avec :

12. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html
13. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html

- *l'analyse de silhouette* : mesure la distance entre chaque point d'un groupe avec les points des autres groupes. L'intervalle de valeurs de cette mesure est $[-1, 1]$: plus les valeurs sont proches de 1 plus la classification est meilleure [141].
- *le critère de Davies-Bouldin* : mesure le rapport des distances intra-groupe aux distances inter-groupe de chaque groupe avec son groupe le plus similaire. Ses valeurs appartiennent à l'intervalle $[0, +\infty]$: les valeurs qui s'approchent de 0 indiquent une meilleure classification [142].

— **description des profiles types d'apprenants :** les groupes homogènes d'étudiants identifiés donnent des informations utiles sur les différents comportements d'apprentissage en ligne, auxquels les systèmes LA doivent s'adapter. Cependant, les enseignants ont besoin de comprendre ces comportements et leur signification selon les différents indicateurs d'apprentissage. Afin d'aider les enseignants dans cette tâche, les groupes identifiés sont traduits en personas d'apprenants dans ([5]. Ces personas sont définis comme des descriptions narratives des apprenants typiques qui peuvent être identifiés grâce aux centroïdes des processus de classification de ML [135]. Les personas sont couramment utilisés pendant la phase de développement numérique, en particulier dans la conception UX [143]. Dans notre cas, l'objectif des personas est différent : ils permettent d'améliorer l'explicabilité des résultats des systèmes de LA aux enseignants, et aussi d'étudier la représentativité d'un corpus. De ce fait, chaque groupe identifié est traduit par un persona représentant une description narrative d'un apprenant fictif. Le persona contient quelques données démographiques, un texte montrant des points clés du comportement d'apprentissage(travail fait [5]). Les valeurs aberrantes ne sont pas écartées mais sont étudiées individuellement.

Nous avons appliqué ces étapes sur notre jeu de données composé de 639 apprenants. Les résultats obtenus sont donnés par le Tableau 5.2. D'après ce Tableau, nous avons obtenus 12 valeurs aberrantes dont 7 appartiennent au sous-ensemble de *traces* et 5 appartiennent au sous-ensemble de *pas_traces*. Le reste des apprenants sont des *inliers* dont chacun d'eux est associé à l'un des 21 profils d'apprenants identifiés. La valeur de k groupes homogènes optimale pour chacune des classes des sous-ensembles de *traces* et *pas_traces* est donnée par la 3ème ligne du Tableau 5.2. Les valeurs d'analyse de silhouette ainsi que l'indexe de Davies-Bouldin sont aussi donnés par ce Tableau.

	traces				**pas_traces**			
	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
Nombre d'*inliers*	176	52	16	151	63	28	27	115
Nombre de valeurs aberrantes	2	1	2	2	1	1	1	2
Valeur optimale de k	2	2	2	2	3	3	3	4
Analyse de silhouette	0.44	0.30	0.85	0.28	0.43	0.36	0.36	0.34
Index de Davies-Bouldin	1.00	1.45	0.07	1.32	0.85	0.96	0.92	1.04

TABLE 5.2 – Résultats du *clustering* par sous-ensembles de données et par classes

5.3.4 Personas : des exemples

Chaque persona contient une large variété d'informations : description narrative du comportement d'apprentissage, sa proportion dans l'ensemble de données auquel il appartient, indicateurs visuels du risque d'échec et modalité d'apprentissage. La Figure 5.2 illustre un exemple de deux personas du jeu de données étudié.

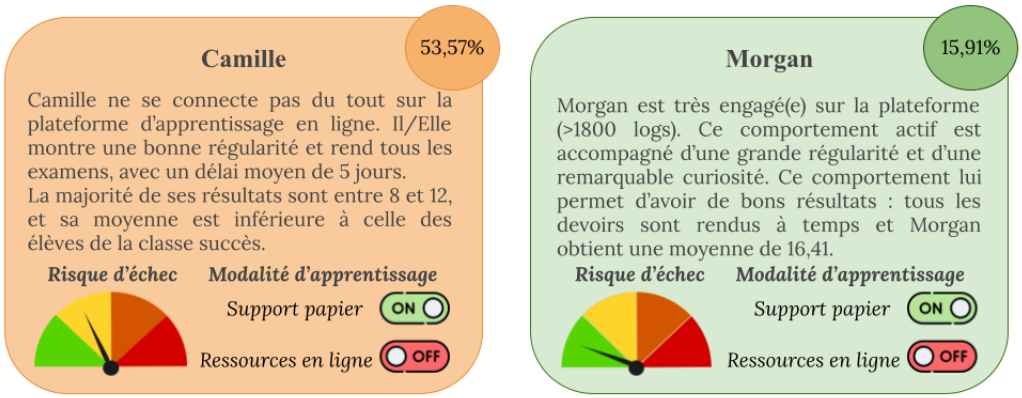


FIGURE 5.2 – Exemple de deux personas [5]

L'utilité de tels personas est triple. En plus de fournir des informations valorisantes sur les comportements des apprenants, ils contribuent à l'amélioration de l'interprétation des résultats d'un système de LA. Enfin, ils sont particulièrement intéressants pour notre étude car ils peuvent être utilisés pour affiner l'évaluation d'un système éducatif. Les résultats présentés dans la Section 5.5 confirment ce dernier point.

5.4 2ème étape de la méthodologie : Métriques d'évaluation

L'évaluation est une étape clef pour montrer la performance du système de prédiction avec les différents profils d'apprenants. Cette étape est considérée plus complète lorsqu'elle considère l'importance de l'exactitude des prédictions et leur précocité, continuité et stabilité dans le temps.
C'est pour ces raisons, nous proposons de suivre deux types de métriques pour évaluer les performances du système de prédiction avec les différents sous-ensembles d'apprenants et personas :

— métriques de précision : la métrique d'exactitude proposé par le ML
— métriques temporelles : la précocité et la stabilité temporelle introduites dans le Chapitre 4.

5.5 3ème étape de la méthodologie : une évaluation en profondeur d'un système prédictif

Cette section présente l'ensemble de la méthodologie : de la description du système de prédiction aux étapes de modélisation et d'évaluation. Elle se termine par une étude comparative basée sur les résultats obtenus.

5.5.1 Une brève description du modèle de prédiction

Notre système est basé sur un modèle de prédiction hebdomadaire des apprenants à risque d'échec ou de décrochage. Comme expliqué, les apprenants de la cohorte sont classés en quatre classes. Tout d'abord, nous sommes passés par les deux processus d'extraction et de sélection de caractéristiques (voir Chapitre 3). Passer par ces processus est important pour sélectionner les caractéristiques de l'activité les plus corrélées au résultat final de l'apprenant ainsi que pour minimiser le bruit dans le modèle. Ainsi, à chaque semaine t_i, un apprenant $S_q \in S$ (ensemble de tous les apprenants de la cohorte) est représenté par un vecteur $X_{q_{t_i}}$ composé des caractéristiques allant de f_1 à f_n et de la classe $C_j \in Y$ (l'ensemble

d'étiquettes de classe) à laquelle il appartient. Chaque apprenant appartient à une seule et unique classe sur l'année.

$$X_{q_{t_i}} = < f_1, f_2, ..., f_n, y >_{q_{t_i}}$$

Chaque caractéristique f_1 à f_n représente une activité d'apprentissage jusqu'au temps de prédiction t_i. Pour chaque temps de prédiction t_i, la valeur d'une caractéristique est ajoutée à celle du temps de prédiction t_{i-1} : nous procédons à une accumulation de valeurs. Sur la base des résultats d'exactitude de [9], nous utilisons le modèle RF comme modèle ML pour notre système.

5.5.2 Résultats

Dans la première phase d'évaluation, nous avons divisé la population de l'ensemble de données de test en deux sous-ensembles (*traces*, *pas_ traces*). Dans cette partie expérimentale, nous rapportons les résultats de trois métriques : exactitude, précocité et stabilité.

5.5.2.1 Résultats de la mesure d'exactitude

Dans cette partie, nous exposons les résultats de l'évaluation du système LA proposé en terme de la mesure d'exactitude de ML.

Analyse : Les courbes de la Figure 5.3 montrent les évolutions de la mesure d'exactitude au cours des semaines de prédiction de toute la population de tout le jeu de données, du sous-ensemble *log* et du sous-ensemble *pas_traces*. À travers ces courbes, nous constatons une différence des valeurs d'exactitude entre l'ensemble de données de test de la population totale et les sous-ensembles de données de *traces* et *pas_traces*.
En effet, nous remarquons que jusqu'à la semaine 15, les classes d'apprenants qui appartiennent au sous-ensemble de *pas_traces* sont largement mieux prédites. Étant donnée sa forte présence dans ce sous-ensemble, la classe d'abandon C_4 est la plus prédictible. Jusqu'à la semaine 4, le sous-ensemble de *log* est celui avec les meilleures valeurs d'exactitude ($\approx 85\%$). Entre les semaines 6 et 15, la performance en terme d'exactitude du jeu de données de données initiales de toute la population dépasse le sous-ensemble de *traces* et atteint une valeur de $\approx 87\%$.
Cependant, plus nous avançons dans l'année scolaire, plus les résultats de prédiction de deux sous-ensembles de *traces* et *pas_traces* convergent vers quasiment les mêmes valeurs, particulièrement entre les semaines 16 et 24. Ces résultats sont donc insuffisants pour observer les performances des différents profils d'apprenants existants dans les différents sous-ensembles. Il est donc pertinent de passer à l'évaluation du système par profil d'apprenant.

La Figure 5.4 montre que les courbes d'évolution de la mesure d'exactitude au cours du temps des personas identifiés pour chacun de deux sous-ensembles de données *traces* (A) et *pas_traces* (B). Afin d'assurer la lisibilité de la figure, nous ne présentons que les résultats d'un persona par classe et par sous-ensemble de données.

Les personas des numéros 1, 4, 5 et 8 sont sélectionnés pour le sous-ensemble de *traces* et ils représentent respectivement les classes C_1, C_2, C_3 et C_4. Les personas des numéros 10, 13, 16 et 13 sont sélectionnés pour le sous-ensemble de *pas_traces* et ils représentent respectivement les classes C_1, C_2, C_3 et C_4.
Les courbes d'évolution de la mesure d'exactitude de la Figure 5.4 illustre bien les différences de performance entre les personas appartenant au même sous-ensemble. Pour le sous-ensemble de *traces*, le persona 8 est celui avec les meilleures performances et atteint une exactitude de 100% à partir de la

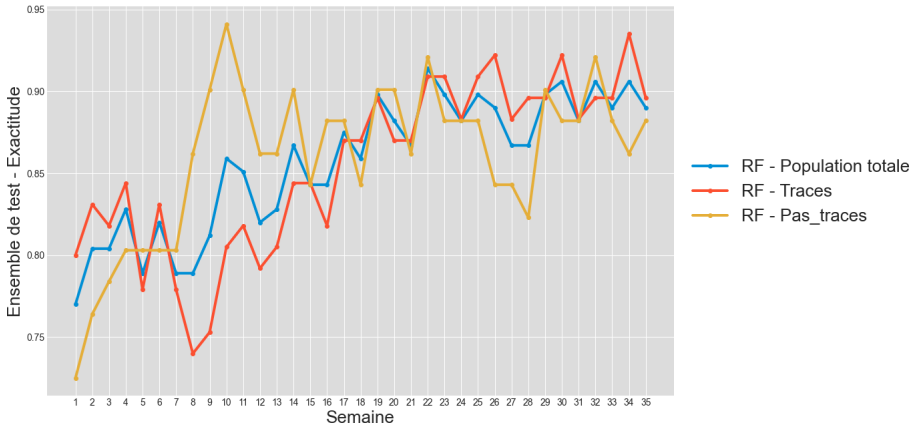


FIGURE 5.3 – Évolution de la mesure d'exactitude avec toute la population et les deux sous-ensembles de données

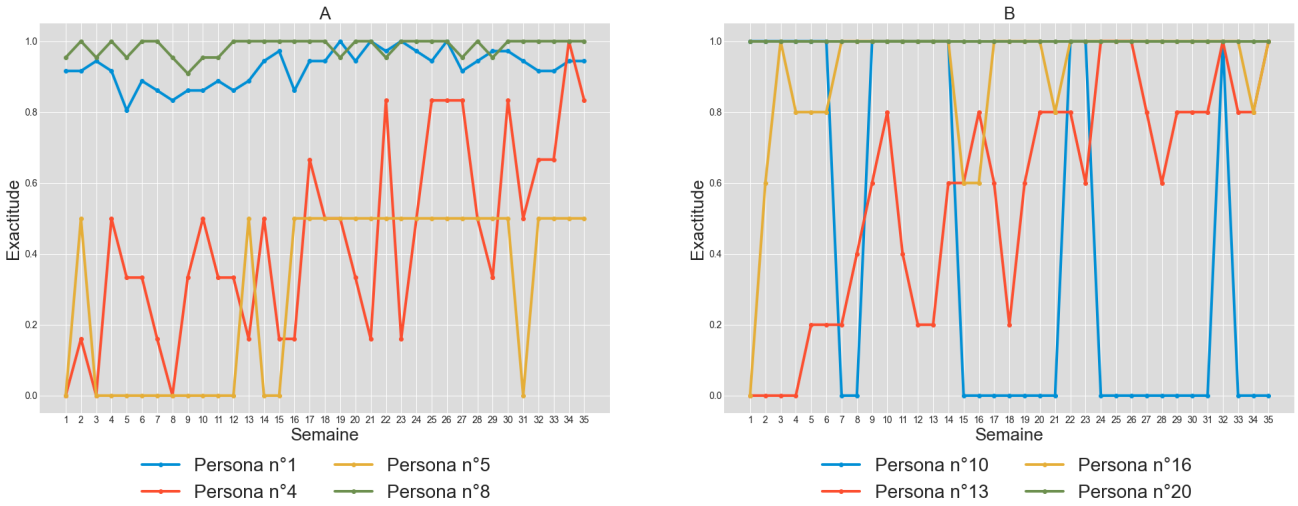


FIGURE 5.4 – Évolution de la mesure d'exactitude des profils types d'apprenants des sous-ensembles de données *traces (A)* et *pas_traces (B)*

semaine 13. Le persona 1 connaît des bonnes performances pendant toutes les dates de prédiction. Les valeurs d'exactitude de la prédiction de ce dernier basculent entre 95% et 100%. Tandis que, le persona 4, bien qu'il ait d'assez bonnes performances à la semaine ($\approx 55\%$), sa performance connaît une chute de la semaine 3 jusqu'à la semaine 12. Sur le reste de la période de l'apprentissage, les valeurs d'exactitude de la prédiction de ce persona connaissent des fortes oscillations et ne dépassent toujours pas 55%. Les prédictions du persona 4 s'améliorent considérablement à partir de la semaine 17 et la mesure d'exactitude atteint des valeurs $\approx 81\%$ et 100% entre les semaines 22 et 34.

Synthèse : à partir de l'analyse des résultats de la Figure 5.4, nous remarquons les différences des valeurs d'exactitude entre les personas d'un même sous-ensemble de données. Ces différences sont parfois très larges comme entre le persona 8 et le persona 5. De plus, contrairement aux résultats de la Figure 5.3 qui s'avèrent être généraux et peu représentatifs des sous-ensembles d'apprenants, les courbes d'exactitude ne convergent pas vers une même valeur ou des valeurs approximatives à des points de temps de

prédiction.
Les premiers résultats donnés avec la mesure d'exactitude de ML montrent l'importance de l'identification des profils types d'apprenants pour évaluer en profondeur le fonctionnement du système LA avec chacun de ces derniers. Nous passons dans la section suivante à une évaluation temporelle des performances des sous-ensembles de données et des personas existants.

5.5.2.2 Résultats de la précocité et la stabilité temporelle

Dans cette partie, nous exposons les résultats des métriques temporelles de la précocité et la stabilité sur les différents sous-ensembles ainsi que les groupes homogènes d'apprenants identifiés. Nous appliquons le premier algorithme de précocité 1 qui retourne le premier point du temps d'une prédiction correcte.

Le Tableau 5.3 présente les résultats des métriques temporelles de précocité et de stabilité évaluant les prédictions de la totalité de l'ensemble de test et de deux sous-ensembles de *traces* et *pas_traces*. Les mesures de ce tableau sont présentées en semaine. La couleur verte de ce Tableau met l'accent sur les meilleurs résultats. Tandis que la couleur orange représente les résultats les plus faibles.

Analyse : une première constatation du Tableau 5.3 est la suivante : les résultats de deux métriques temporelles sont différents avec les différents ensembles de données. Cette différence du résultat est nettement saisie avec les deux classes du risque moyen et élevé. Par exemple, la précocité de la classe du risque élevé (C_3) avec la population totale, le sous-ensemble *traces* et le sous-ensemble *pas_traces* est égale respectivement aux semaines 3,1 et 3,57. En effet, nous remarquons que les différences des valeurs de précocité sont plus concrètes entre les sous-ensembles de *traces* et *pas_traces*. Tandis que, les différences des valeurs de la mesure de stabilité de deux classes du risque sont larges entre l'ensemble de toute la population et les deux sous-ensembles de *traces* et *pas_traces*. Par exemple, pour la classe du risque élevé, les mesures de stabilité sont égales à 16,55, 7,5 et 19,14 avec respectivement l'ensemble de la population totale, le sous-ensemble de *traces* et le sous-ensemble de *pas_traces*.
Ce Tableau 5.3 montre une caractéristique commune entre les différents ensembles : la classe d'abandon a toujours les meilleures valeurs de précocité et de stabilité. Cependant, la classe du risque moyen est toujours prédite le plus tardivement avec les faibles valeurs de stabilité.
Ce tableau montre déjà les différences de performance d'un système de LA entre deux grands sous-ensembles d'apprenants de la même population. De ce fait, il serait intéressant de suivre en profondeur cette évaluation avec les groupes homogènes d'apprenants existants dans chacun de ces sous-ensembles.

Le Tableau 5.4 présente les résultats des métriques temporelles de précocité et stabilité avec chaque groupe de profil ou persona existant dans les deux sous-ensembles de *traces* et *pas_traces*. Les mesures de ce tableau sont aussi présentées en semaine. Le code des couleurs désigne l'appartenance de chaque persona à chacune des classes C_1, C_2, C_3 et C_4. La couleur verte de ce Tableau met l'accent sur les meilleurs résultats. Tandis que la couleur orange représente les plus faibles résultats.
Le Tableau 5.4 montre les différences de performance entre les différents personas appartenant au même sous-ensemble. Par exemple, les mesures de la précocité de deux personas 2 et 3 du sous-ensemble de *traces* sont respectivement égales à 1 et 12,66 : la différence est très large. Un autre exemple du sous-ensemble de *pas_traces* est le suivant : les mesures de stabilité de deux personas 9 et 10 sont respectivement 35 et 6. Il s'agit encore une fois d'une large différence entre les mesures. En effet, ces deux personas 9 et 10 appartiennent à la même classe d'apprenants qui est la classe de réussite (C_1). De ce fait, nous pouvons avoir des différences de performance entre des personas qui appartiennent à la même classe. Les résultats de précocité de deux personas 3 et 4 de la classe C_2 du sous-ensemble *traces* confirment aussi cette constatation. En effet, cette différence dans les résultats de précocité et stabilité entre les personas d'une même classe sont plus tangibles avec ceux de classe du risque moyen et élevé.

Ce Tableau montre que les personas des classes de réussite (C_1) et d'abandon (C_4) ont les meilleures valeurs de précocité (ils sont généralement correctement prédits à partir de la semaine 1) et de stabilité.

Synthèse : les résultats de ces Tableaux 5.3 et 5.4 montrent les différences dans les résultats des métriques de précocité et stabilité entre les sous-ensembles existants dans une population ainsi qu'entre les profils d'apprenants de chacun de ces sous-ensembles. En effet, le Tableau 5.3 représente peu les différences de performance qui peuvent caractériser un système de LA au cours du temps et ne montre pas l'écart entre les résultats des métriques temporelles correspondants à chacun des profils d'apprenants existant dans un sous-ensemble. De ce fait, il est d'un grand intérêt de tenir compte des divers profils d'apprenants dans un groupe d'élèves pour évaluer les performances d'un système de LA et les améliorer pour un fonctionnement équitable avec les différents personas.

Compte-tenu des résultats de deux sections 5.5.2.1 et 5.5.2.2, le suivi de l'évolution de deux types de métriques de précision et temporelles a permis de mettre en évidence les différences de performance d'un système de LA entre les différents profils d'apprenants existants dans un même groupe. En appliquant ces métriques sur les sous-ensembles d'apprenants, ces différences de performance sont parfois peu constatées. Cependant, les écarts de performance sont plus nets quand les métriques de précision et temporelle sont appliquées sur les profils d'apprenants. De plus, une large différence des résultats de ces métriques est saisie entre les personas d'une même étiquette de classe.

Par conséquent, il est important d'évaluer les performances prédictives d'un système de LA selon la diversité de profils d'apprenants existants dans une cohorte. De plus, pour que le système de LA fonctionne équitablement avec tous les apprenants, il est important de tenir compte de leur hétérogénéité et de la diversité de leurs profils d'apprentissage lors de sa conception.

	Population totale		***Sous-ensemble de traces***		***Sous-ensemble de pas_traces***	
	Précocité	**Stabilité**	**Précocité**	**Stabilité**	**Précocité**	**Stabilité**
Abandon	1.0	31.4	1.0	31.0	1.0	31.8
Risque élevé	3.0	16.5	1.0	7.5	3.5	19.1
Risque moyen	8.0	6.6	6.7	4.3	8.2	9.5
Réussite	1.1	28.3	1.1	28.6	1.0	28.6
Total	2.0	25.3	1.7	26.1	2.3	26.4

TABLE 5.3 – Mesure de la précocité et de la stabilité par chaque classe d'un groupe de profils

Sous-ensemble	**Persona**	**Précocité**	**Stabilité**
traces	1	1.1	28.6
	2	1.0	26.0
	3	12.6	1.6
	4	5.5	5.6
	5	1.0	7.5
	7	1.0	25.2
	8	1.0	32.3

C1
C2
C3
C4

Sous-ensemble	**Persona**	**Précocité**	**Stabilité**
pas_traces	9	1.0	35.0
	10	1.0	6.0
	11	1.0	35.0
	12	5.0	12.0
	13	8.2	10.8
	14	12.0	1.0
	15	6.5	8.5
	16	24.0	23.2
	18	1.0	31.6
	19	1.0	35.0
	20	1.0	35.0
	21	1.0	33.1

TABLE 5.4 – Les mesures de la précocité et stabilité pour chaque persona

5.5.3 Discussions

En considérant toute la population, la Figure 5.3 et le Tableau 5.3 montrent que le modèle de prédiction est globalement performant et atteint une exactitude $\approx 93\%$, une précocité totale de $2,06 semaines$ et une stabilité de $25,35 semaines$. Cependant, en tenant compte de la présence de deux sous-ensembles d'apprenants se caractérisant chacun par l'existence de plusieurs profils d'apprenants, le modèle de prédiction ne montre pas les mêmes performances avec les différents personas. Par exemple, dans notre contexte, les apprenants qui appartiennent au sous-ensemble *traces* sont prédits plus précisemment et au plus tôt que ceux qui appartiennent au sous-ensemble *pas_traces*. De plus, les métriques temporelles et de précision montrent des meilleurs résultats avec les apprenants de (C_1) et (C_4) qu'avec les apprenants de deux classes (C_2) et (C_3). Ajoutons à cela que nous avons parfois des larges écarts des résultats de performance entre les personas de la même classe d'apprenants appartenant au même sous-ensemble de population.

En bref, les résultats de deux différents types de métriques d'évaluation montrent que la performance du modèle prédictif est étroitement dépendante des profils types d'apprenants. Par conséquent, afin de fournir aux acteurs de l'éducation, notamment les enseignants, des prédictions précises, fiables, précoces et stables au cours du temps, il est important que le système de LA considère les différents profils d'apprentissage existant au sein d'une cohorte.

5.6 Conclusion

Les profils d'apprenants identifiés, caractérisés par des personas, au sein de notre jeu de données sont diversifiés et confirment que les apprenants adoptent des comportements différents et doivent recevoir un accompagnement adapté. De plus, l'étape de l'évaluation révèle que les performances du modèle n'étaient pas les mêmes avec tous les personas. Les résultats obtenus répondent à notre question de recherche liée au bon fonctionnement d'un système de LA avec l'ensemble d'une population. Ces résultats confirment l'intérêt d'une identification au préalable des profils d'apprenants pour une évaluation approfondie de ces systèmes.

De plus, les métriques temporelles telles que la précocité et la stabilité renseignent sur la confiance qu'un utilisateur peut avoir dans le système. En effet, l'utilisation de ces nouvelles métriques temporelles montrent encore une fois l'insuffisance des métriques de précision usuelles pour évaluer les résultats d'un système de prédiction au cours du temps. C'est pour cette raison que nous prévoyons de creuser encore plus dans cet axe de recherche en proposant de nouveaux raffinements d'évaluation des systèmes prédictifs reposant à la fois sur l'identification des profils d'apprenants et l'importance des métriques temporelles et des mesures de compromis entre les métriques de précision et temporelles (Chapitre 4).

De ce fait, comme perspectives à ce travail, nous estimons appliquer le 2ème algorithme de la précocité (chapitre 4) qui retourne plusieurs points de temps précoce dans l'année scolaire. Les résultats retournés permettront de calculer les différentes mesures de compromis qui permettront d'identifier, pour chaque profil d'apprenant, le point du temps avec la meilleure précocité, exactitude et stabilité du système de prédiction. Ce contexte de recherche mérite plus d'attention car il permettrait de fournir des retours rapides aux enseignants sur la situation de leurs apprenants. Cette information précoce et précise pourrait les aider à élaborer rapidement des solutions efficaces et personnalisées pour les apprenants à risque.

Ensuite, nous prévoyons de travailler sur la séparation du jeu de données initial selon les modalités d'apprentissage (*traces* et *pas_traces*) qui peuvent aussi influencer les performance du système de prédiction. Par conséquent, il semble intéressant de comparer les résultats avec différentes partitions du jeu de données. D'une part, cela pourrait permettre de mettre en évidence les caractéristiques clés qui sont essentielles au bon fonctionnement du modèle. D'autre part, cela améliorerait encore plus l'explicabilité en permettant aux enseignants et aux universitaires de sélectionner la partition appropriée en fonction de

leurs objectifs pédagogiques.

Chapitre 6

Génération précoce d'alerte

Synthèse

Ce chapitre exploite les résultats et les interprétations de l'évaluation temporelle des systèmes prédictifs pour répondre aux limites des systèmes d'alerte précoce (Early Warning Systems (EWS)) de l'état de l'art.
En nous basant sur la notion de la précocité de prédiction ainsi que sur des règles d'alerte, nous proposons un algorithme qui permet la génération automatique d'alertes précoces et précises pour les apprenants à risque d'échec. La validation de cette approche sur des systèmes prédictifs réels met en valeur l'importance, d'une part, du calcul du point du temps précoce pour pouvoir alerter efficacement, et d'autre part, de la stabilité temporelle des EWS pour choisir la bonne règle d'alerte.
Les alertes générées sont très utiles pour l'étape de la prescription des actions correctives. Elles permettent aux enseignants, utilisateurs des EWS, d'intervenir efficacement et au bon moment auprès des apprenants à risque.

6.1 Introduction

Les systèmes d'alerte précoce (en anglais *Early Warning Systems (EWS)*) sont utilisés dans plusieurs domaines [46] et pour différents objectifs : alerter sur les catastrophes naturelles, alerter sur les ralentissements financiers et économiques des sociétés, de la bourse. . ., alerter les patients de la détérioration de leurs signes cliniques . . .[46]. En effet, il s'agit toujours d'alerter la partie concernée par des éventuels risques futurs. Afin de faire face aux taux élevés d'échec et de décrochage, les LA ont aussi adopté l'utilisation de ces EWS pour la prédiction précoce et précise des apprenants à risque dans l'enseignement en ligne. Afin de répondre au mieux aux attentes de leurs utilisateurs (enseignants, apprenants, institutions. . .), les EWS se basent à la fois sur les données historiques des apprenants et la puissance des modèles de ML.

Lorsqu'un EWS est dédié aux enseignants, il permet l'identification précoce des apprenants avec des difficultés d'apprentissage [50]. Comme son nom l'indique, en utilisant un EWS, l'enseignant s'attend à être alerté en cas du risque. En d'autres termes, la génération d'alerte suite à l'identification des apprenants à risque devrait être la finalité d'un EWS.
Cependant, les travaux de l'état de l'art des LA, s'intéressant à la conception et au développement des EWS, ne proposent pas des méthodes et / ou algorithmes particuliers pour une génération automatique d'alertes. En effet, les EWS de la littérature ne vont pas au delà de la prédiction du risque et l'identification des apprenants en situation d'apprentissage défaillante.

La génération d'alerte dans le domaine des LA est un axe de recherche peu développé à notre connais-

sance et les méthodes d'alerte n'ont pas été largement élaborées. Étant donné l'importance de l'alerte précoce des enseignants des apprenants à risque pour une intervention efficace auprès de ces derniers, nous cherchons à répondre à la question suivante : ***comment alerter au plus tôt les enseignants des apprenants à risque ?***.

Dans ce chapitre, nous proposons un algorithme pour la génération automatique d'alertes précoces et précises pour les enseignants, utilisateurs des EWS. Ces alertes ciblent essentiellement les apprenants à risque d'échec et qui rencontrent des difficultés d'apprentissage.
Pour atteindre cet objectif, l'algorithme utilise, d'une part, les métriques temporelles, notamment la précocité et la mesure du compromis précocité-stabilité ESS (voir Chapitre 4). De plus, cet algorithme utilise la notion de *la règle d'alerte (R)* que nous définissons comme une succession de i prédictions comme à risque. Afin de choisir la bonne valeur de i, toute règle d'alerte R_i est évaluée à l'aide de deux métriques inspirées de la matrice de confusion habituelle de ML et adaptées à la problématique d'alerte : la *précision* et le *taux de fausses positives* (FPR) des alertes générées suite à l'application de R_i.

Afin de valider cette approche, nous testons cet algorithme d'alerte sur quatre différents EWS réels et qui utilisent les traces numériques d'une population de collégiens inscrits en ligne.

6.2 Formalisation du problème d'alerte

L'objectif est d'alerter les enseignants des apprenants qui sont en situation d'apprentissage défaillante. Cette génération d'alerte doit être précoce et précise et cibler un nombre maximal d'apprenants à risque. Pour ce faire, nous utilisons les résultats donnés par une classification où les apprenants d'une cohorte sont affectés à différentes classes en fonction de leurs performances scolaires.

Soit $Y = \{C_1, C_2, .., C_m\}$ l'ensemble des étiquettes de classe et soit $y_{risque} \subseteq Y$ un sous-ensemble de Y qui regroupe les étiquettes des classes du risque.
Soit $S = \{S_1, S_2, .., S_q\}$ l'ensemble d'apprenants de la base de données de test D_{test} et soit $T = \{t_1, t_2, .., t_k\}$ l'ensemble de temps de prédiction.
À chaque $t_k \in T$, chaque apprenant $S_p \in S$ est représenté par un vecteur $X_{p_k} = \{f_1, f_2, .., f_z, C_j\}_{p_k}$ où $f_n \in \mathbb{R}$ sont les activités d'apprentissage de S_p et $C_j \in Y$ est la classe à laquelle l'apprenant appartient.

L'objectif de cette étude est d'alerter l'enseignant au plus tôt des apprenants identifiés comme à risque. En d'autres termes, la prédiction de l'apprenant $S_p \in S$ dans une classe $C_j \in y_{risque}$ est à l'origine de la génération d'alerte auprès de l'enseignant.

La Figure 6.1 est une illustration d'un système composé de $S = \{S_1, S_2, S_3\}$. L'ensemble d'étiquettes de classe de ce système est $Y = \{C_1, C_2, C_3\}$ avec $y_{risque} = \{C_2, C_3\}$; Seule C_1 n'est pas une classe du risque. Pour chaque apprenant de S, $\{l_1, l_2, l_3, l_4, l_5\}$ et $\{p_1, p_2, p_3, p_4, p_5\}$ représentent respectivement les étiquettes réelles de classe et les prédictions obtenues aux temps $T = \{t_1, t_2, t_3, t_4, t_5\}$. Une case est colorée en rouge lorsqu'un apprenant S_p est prédit dans une classe $C_j \in y_{risque} = \{C_2, C_3\}$; c'est ainsi que l'alerte des enseignants est produite.

6.3 Règle d'alerte

La génération d'alerte est un processus qui suit celui de la prédiction. En s'appuyant sur la définition des EWS et les attentes des enseignants derrière l'utilisation de ces systèmes, la génération d'alerte vise particulièrement les apprenants en classes du risque. Par conséquent, une première hypothèse que nous pouvons établir est la suivante : une alerte est produite dès qu'un apprenant est prédit comme à risque. En d'autres termes, la première prédiction d'un apprenant dans une classe du risque est le temps auquel l'alerte est générée et l'enseignant est donc alerté de la situation académique de ce dernier.
En appliquant cette hypothèse, nous pouvons constater deux types d'alerte :

	t_1		t_2		t_3		t_4		t_5	
	l_1	p_1	l_2	p_2	l_3	p_3	l_4	p_4	l_5	p_5
S_1	C_1	C_3	C_1	C_1	C_1	C_1	C_1	C_1	C_1	C_3
S_2	C_1	C_1	C_1	C_1	C_2	C_2	C_2	C_2	C_1	C_1
S_3	C_3	C_1	C_3	C_2	C_3	C_2	C_3	C_3	C_3	C_3

FIGURE 6.1 – Tableau de génération d'alerte

— **bonne alerte :** alerter un enseignant pour avertir un apprenant dont le comportement d'apprentissage est défaillant.
— **fausse alerte :** alerter un enseignant pour avertir un apprenant réussi.

Pour l'exemple de la Figure 6.1, au temps t_1, l'apprenant S_1 appartient à la classe d'étiquette C_1 alors qu'il a été prédit dans C_2 (c'est une classe du risque). L'alerte produite pour S_1 à t_1 est une fausse alerte. Au temps de prédiction t_2, S_1 est correctement prédit dans la classe de réussite C_1.
Au temps, t_2, l'apprenant S_3 appartient à la classe C_3 (une classe du risque) et il a été prédit pour la première fois dans C_2 (classe du risque). L'alerte produite pour S_3 à t_2 est une bonne alerte.

Notre objectif étant d'alerter les enseignants des apprenants à risque, nous avons l'intention de maximiser le taux de bonnes alertes et minimiser les fausses alertes. Ceci dit, produire une fausse alerte reste moins grave que passer à côté d'une bonne alerte. En effet, rater un certain nombre de bonnes alertes réduit la *confiance* de l'enseignant envers la performance d'EWS. Pour l'exemple de la Figure 6.1, à t_1, l'apprenant S_3 est prédit dans C_1 alors qu'il appartient à la classe du risque C_3. En appliquant la première hypothèse d'alerte, aucune alerte n'est produite pour S_3 à t_1 : il s'agit d'une bonne alerte ratée. Ainsi, l'application de la première hypothèse définie ci-dessus peut engendrer un taux élevé de fausses alertes et surtout un taux élevé de bonnes alertes ratées. En effet, il ne suffit pas de produire des alertes précoce sans avoir vérifié leur précision. Autrement dit, le processus de la génération d'alerte ne doit pas privilégier la précocité de l'alerte à sa précision et à la stabilité du système.
Par conséquent, il serait pertinent de proposer une méthode de génération d'alerte qui considère plusieurs paramètres :
— la précocité de l'alerte : alerter les enseignants au plus tôt des apprenants à risque pour une intervention efficace.
— la précision de l'alerte : maximiser le taux de bonnes alertes par rapport au nombre total d'alertes générées.
— la stabilité du système : des prédictions stables et successivement correctes dans le temps.

Ainsi, nous proposons de suivre la succession des prédictions et définissons une règle d'alerte R_i comme suit :

règle$_i$ (R_i) : alerter un enseignant lorsqu'un apprenant est prédit i fois de suite en classe du risque.

Le chiffre i est à définir et représente le nombre de fois où l'apprenant a été successivement prédit comme à risque. La définition de la règle R_i a pour objectif d'assurer *la confiance* de l'enseignant en

maximisant essentiellement les bonnes alertes et aussi minimisant les fausses alertes.
Pour une illustration de cette notion, nous prenons plusieurs valeurs de i et appliquons les règles R_i sur les résultats du système de la Figure 6.1. Si $i = 2$, l'apprenant S_2 est prédit dans C_2 (classe du risque) aux temps t_3 et t_4. Donc, S_2 est classé **2** fois de suite en classe du risque : une alerte est produite au temps t_4. Si $i = 4$, l'apprenant S_3 est prédit dans C_2 et C_3 (deux classes du risque) aux temps t_2, t_3, t_4 et t_5. S_3 est classé **4** fois de suite en classe du risque : une alerte est donc générée au temps t_5.

Afin d'évaluer la pertinence d'une règle d'alerte R_i, nous adaptons la matrice de confusion habituelle de ML à la problématique d'alerte comme le montre la Figure 6.2. Ainsi, afin de fixer la valeur optimale de i, nous proposons d'évaluer R_i en suivant deux métriques de la matrice de confusion : la précision et le taux de faux positifs (*FPR*). Ces deux métriques sont désormais adaptées à la problématique de la génération d'alerte comme suit :

$$\textit{précision} = \text{taux de bonnes alertes} \ / \ \text{nombre total d'alertes}$$

$$\textit{FPR} = \text{taux de fausses alertes / nombre total d'alertes}$$

Le nombre total d'alertes = $\sum$ (Bonne alerte) + $\sum$ (Fausse alerte).

Une alerte est dite ratée lorsqu'un apprenant appartient réellement à la classe du risque mais la règle d'alerte ne le détecte pas. De ces faits, la valeur de i de R_i est optimale lorsque la précision est élevée, le FPR est minimal et le nombre d'alertes ratées est aussi minimal.

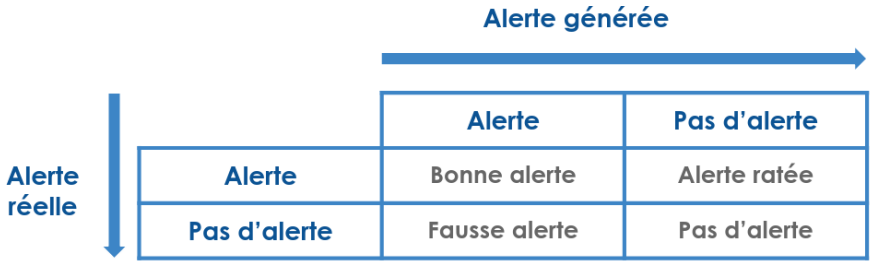


FIGURE 6.2 – Matrice de la confusion adaptée à l'évaluation des règles d'alerte

6.4 Algorithme d'alerte

Dans cette section, nous proposons un algorithme pour une génération automatique d'alerte précoce et précise. Dans un premier temps, nous exposons les pré-requis de l'algorithme d'alerte. Ensuite, nous présentons et décrivons l'algorithme de l'alerte.

6.4.1 Les pré-requis de l'alerte : la précocité de prédiction et la stabilité

La génération d'alerte se sert des résultats de prédiction et de l'évaluation statique et temporelle des performances des modèles analytiques utilisés par le EWS. Ainsi, afin de proposer un algorithme pour une génération d'alerte **efficace**, **précoce** et **précise**, il faut considérer un certain nombre de paramètres permettant de répondre aux questions de *quand* et *comment* générer des alertes.

Quand alerter ? : la génération d'alerte est étroitement liée à la dimension temporelle : il faut alerter au temps dit *adéquat*. Ce temps doit à la fois :

— être suffisamment précoce pour que l'apprenant candidat à l'alerte ait le temps d'améliorer ses résultats.
— garantir la stabilité des prédictions correctes dans le temps

Ce temps *adéquat* est le résultat de l'étude du compromis entre la précocité et la stabilité donné par la mesure ESS (voir le Chapitre 4). Autrement dit, le point du temps assurant l'équilibre précocité-stabilité mesurée à travers le calcul de ESS est le temps à partir duquel le processus de la génération d'alerte peut être déclenché.
$\Longrightarrow$**L'algorithme de la génération d'alerte utilise les métriques de performance temporelles pour identifier le temps *adéquat* pour déclencher l'alerte.**

Comment alerter ? : dès que le temps dit *adéquat* pour le déclenchement de la génération d'alerte est fixé, nous définissons la méthode d'alerte. C'est à ce niveau que la notion de règle d'alerte introduite dans la Section 6.3 intervient.
Plus explicitement, à partir du temps *adéquat* pour l'alerte, les prédictions de chaque apprenant suivent la définition de la règle d'alerte R_i où la valeur de i est à définir selon le contexte et les besoins. En effet, afin de générer une alerte, il faut que, à partir du temps *adéquat* fixé, l'apprenant ait été prédit i fois de suite comme à risque.
$\Longrightarrow$ **L'algorithme de la génération d'alerte utilise les règles d'alerte pour identifier la méthode d'alerte.**

En bref, les pré-requis de l'algorithme de la génération d'alerte sont les suivants :

1. la précocité de la prédiction des classes du risque.
2. la stabilité du système prédictif.
3. le point du temps assurant le compromis entre la précocité des prédictions et la stabilité du système.
4. la règle d'alerte R_i à appliquer comme méthode d'alerte.

Dans la section suivante, nous présentons l'algorithme d'alerte et le décrivons en détails.

6.4.2 Description de l'algorithme d'alerte

L'Algorithme 4 propose une méthode automatique pour générer des alertes en se basant sur l'Algorithme 3 de la précocité et l'Equation 4.3 de calcul de *ESS* (voir Chapitre 4).
L'Algorithme 4 prend en entrée la liste des apprenants S, le sous-ensemble y_{risque} de Y contenant seulement les étiquettes de classe du risque, le jeu de données de test D_{test}, le nombre de temps de prédiction donné par T et la règle d'alerte à appliquer donnée par le chiffre i. Cet Algorithme retourne A_S qui est l'ensemble d'apprenants à risque candidats à l'alerte.
Cet Algorithme commence par affecter l'ensemble vide à la variable $E_{precocite}$ qui contiendra les valeurs des points de précocité pour chaque classe du risque de y_{risque} (Ligne 1). Afin d'identifier le point du temps adéquat à partir duquel la génération d'alerte est possible, l'Algorithme 4 itère sur l'ensemble d'étiquettes de classe de risque ($C_k \in y_{risque}$) (Ligne 2). Pour chaque classe C_k, la variable $precoce_{C_k}$ de type liste reçoit le résultat de l'appel de l'Algorithme 3 nommé Précocité_Totale (Ligne 3) (voir Section 4.3.2.2). Au niveau de la Ligne 4, l'Algorithme affecte la liste vide à la variable L_{C_k}. Ensuite, cet Algorithme itère sur chaque point de précocité de la liste $precoce_{C_k}$ (Ligne 5). Pour chaque point de précocité j, l'Algorithme procède au calcul de sa mesure ESS correspondant donné par la variable ess_{j_k} (Ligne 6). Le couple $< j_k, ess_{j_k} >$ est ensuite ajouté à la liste L_{C_k} (Ligne 7). Pour chaque classe C_k, l'Algorithme sélectionne le meilleur couple $< precocite_{C_k}, ess_{C_k} >$ assurant la précocité de la prédiction de C_k et la stabilité du système (Ligne 9). Par la suite, la valeur de temps $precocite_{C_k}$ est

ajoutée à l'ensemble $E_{precocite}$ (Ligne 10). Une fois que l'ensemble $E_{precocite}$ est établi, l'Algorithme en extrait la valeur de precocité minimale et l'affecte à la variable $precocite$ (Ligne 12). Dans sa 2ème partie, l'Algorithme itère sur l'ensemble des apprenants (S_p dans S) (Ligne 13). Pour chaque S_p, cet Algorithme initialise les variables du temps de prédiction t à $precocite$ (Ligne 14), de la règle à appliquer r à 0 (Ligne 15) et d'alerte à faux (Ligne 16). Tant que le nombre total de temps de prédiction T n'est pas atteint et que la valeur de r est inférieure à celle de i, l'Algorithme continue à parcourir les prédictions (Ligne 17). Si S_p est prédit dans une classe du risque parmi y_{risque} (Ligne 18) alors la valeur de r est incrémentée de 1 (Ligne 19). Sinon, la valeur de r est encore une fois remise à 0 (Ligne 21). Ensuite, la valeur de t est incrémentée de 1 (Ligne 23). Pour chaque S_p, en sortant de la boucle tant que, si r est égal à i (Ligne 25) alors la variable booléenne alerte reçoit Vrai (Ligne 26) et S_p est ajouté à la liste A_S des apprenants candidats à l'alerte (Ligne 27). Sinon, l'apprenant S_p n'est pas candidat à l'alerte.

Algorithm 4 Algorithme d'alerte - Alerte(S, y_{risque}, D_{test}, T, i)

Entrée: $S, y_{risque}, D_{test}, T, i$
Sortie: A_S

1: $E_{precocite} \leftarrow \emptyset$
2: **pour chaque** C_k dans y_{risque} **faire**
3: $precoce_{C_k} \leftarrow$ Precocité_Totale($S, y_{risque}, x, D_{test}$)
4: $L_{C_k} \leftarrow \emptyset$
5: **pour chaque** j dans $precoce_{C_k}$ **faire**
6: $ess_{j_k} \leftarrow \frac{2*(1-j)*stabilite}{(1-j)+stabilite}$
7: $L_{C_k} \leftarrow mettre(L_{C_k}, < j_k, ess_{j_k} >)$
8: **fin pour**
9: $< precocite_{C_k}, ess_{C_k} > \leftarrow meilleur(L_{C_k})$
10: $E_{precocite} \leftarrow mettre(E_{precocite}, precocite_{C_k})$
11: **fin pour**
12: $precocite \leftarrow minimum(E_{precocite})$
13: **pour chaque** S_p dans S **faire**
14: $t \leftarrow precocite$
15: $r \leftarrow 0$
16: $alerte \leftarrow faux$
17: **tant que** ($t \leq T$ et $r < i$) **faire**
18: **si (prediction**(S_p, y_{risque}) **== Vrai) alors**
19: $r \leftarrow r + 1$
20: **sinon**
21: $r \leftarrow 0$
22: **fin si**
23: $t \leftarrow t + 1$
24: **fin tant que**
25: **si** $r == i$ **alors**
26: $alerte \leftarrow Vrai$
27: $A_S \leftarrow mettre(S_p)$
28: **sinon**
29: S_p ne doit pas être alerté
30: **fin si**
31: **fin pour**

6.5 Description des EWS

Afin de valider l'algorithme d'alerte de la Section 6.4.2, nous proposons de l'appliquer sur quatre EWS différents. Ces systèmes utilisent différents modèles de ML et données qui appartiennent au jeu de données du CNED. Nous décrivons ainsi ces systèmes :

1. EWS_1 : le jeu de données de ce système est composé des traces numériques de 647 apprenants ayant suivi le module de physique-chimie pendant 35 semaines durant l'année scolaire 2017-2018. EWS_1 utilise le modèle RF [9].
2. EWS_2 : le jeu de données de ce système est composé des traces numériques de 647 apprenants ayant suivi le module de physique-chimie pendant 35 durant l'année scolaire 2018-2019. EWS_2 utilise le modèle de DT.
3. EWS_3 : le jeu de données de ce système est composé des traces numériques de 679 apprenants ayant suivi le module des mathématiques pendant 35 semaines durant l'année scolaire 2017-2018. EWS_3 utilise le modèle RF.
4. EWS_4 : ce système utilise le jeu de données composé des traces numériques de 729 apprenants ayant suivi le module de l'histoire-géographie pendant 35 semaines durant l'année scolaire 2018-2019. EWS_4 utilise le modèle DT.

Ces quatre systèmes utilisent toutes les caractéristiques extraites du jeu de données du CNED et liées aux différents indicateurs d'apprentissage notamment : la performance, l'engagement, la régularité et la réactivité.

L'objectif de ces quatre systèmes est de suivre les performances des apprenants d'une manière hebdomadaire pour pouvoir générer des alertes du risque d'échec au plus tôt. Ainsi, la granularité temporelle pour la prédiction, $t_k \in T$, est la semaine.
Les apprenants de chacun de ces systèmes sont classifiés en trois classes selon la moyenne de notes obtenue à la fin d'année : $Y = \{C_1, C_2, C_3\}$

— réussite C_1 : lorsque la moyenne est supérieure strictement à 12.
— risque moyen C_2 : lorsque la moyenne est entre 8 et 12.
— risque élevé C_3 : lorsque la moyenne est strictement inférieure à 8.

Pour ces quatre systèmes, $y_{risque} = \{C_2, C_3\}$. À chaque semaine de prédiction $t_k \in T$, tout apprenant est représenté par son vecteur composé d'un ensemble de caractéristiques et une étiquette de classe.

La Figure 6.3 montre l'évolution de la stabilité temporelle des quatre systèmes tout au long des semaines de prédiction. La stabilité de EWS_1 augmente légèrement avec le temps. Elle est de $\approx 70\%$ et $\approx 76\%$ respectivement à la première et à la dernière semaine de prédiction. La stabilité de EWS_2 diminue dans le temps. Elle commence avec une valeur $\approx 70\%$ et est égale à $\approx 63\%$ à la dernière semaine de prédiction. Les deux systèmes EWS_3 et EWS_4 commencent par une stabilité assez élevées qui est respectivement égale à $\approx 78\%$ et $\approx 80\%$. Cependant, ces valeurs diminuent rapidement au cours des semaines suivantes. Les stabilités de EWS_3 et EWS_4 à la dernière semaine de prédiction sont égales à $\approx 75\%$ et $\approx 73\%$.

6.6 Expérimentations et résultats

Dans cette partie, nous présentons les résultats de l'application de l'Algorithme de génération d'alerte sur les quatre systèmes EWS_1, EWS_2, EWS_3 et EWS_4. Nous commençons à travers la Section 6.6.1 par exposer les résultats des différents points de précocité de chacune des classes des systèmes et leurs mesures correspondantes de ESS. Ces premières expérimentations permettent d'identifier, pour chacun des systèmes, le point du temps précoce à partir duquel l'application des règles d'alerte peut être déclenchée. Nous appliquons cinq différentes règles R_1, R_2, R_3, R_4 et R_5 et nous comparons leurs résultats.

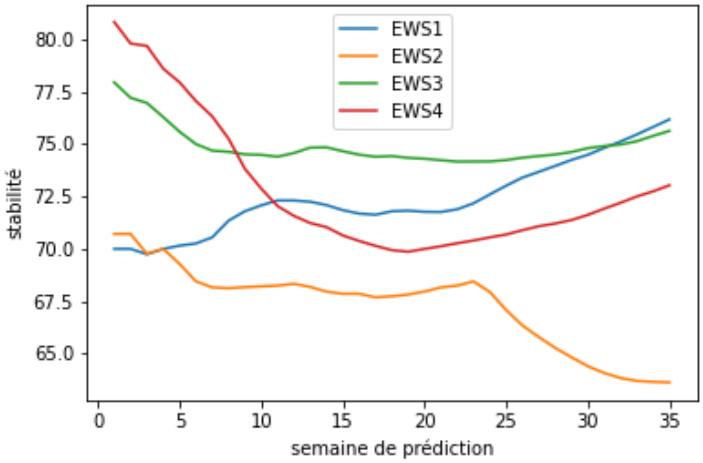


FIGURE 6.3 – Comparaison des EWS_1, EWS_2, EWS_3 et EWS_4 en terme de stabilité temporelle

6.6.1 Les résultats de ESS des quatre systèmes

Dans cette partie, nous commençons par appliquer l'Algorithme de précocité introduit dans la Section 4.3.2. En se basant sur les résultats de la mesure ESS, l'objectif de cette partie est de définir, pour chacun des EWS, le point du temps précoce à partir duquel l'application des règles d'alerte peut être déclenchée.
L'Algorithme d'alerte s'effectue essentiellement pour les apprenants à risque. C'est pour cette raison que la sélection du point de précocité tient compte des résultats de ESS correspondants aux classes du risque. Les Figures 6.4, 6.5, 6.6 et 6.7 présentent les résultats de calcul de ESS pour respectivement les systèmes EWS_1, EWS_2, EWS_3 et EWS_4. Pour chaque système, ces figures illustrent, pour chacune des classes C_1, C_2 et C_3, les 9 premiers points de précocité et leurs taux de ESS correspondants. L'axe des abscisses de ces figures correspond aux points de précocité. Chaque point de précocité contient une valeur temporelle qui peut être différente d'une étiquette de classe à l'autre. Par exemple, pour EWS_1, pour le point d'axe $x = 1$, les valeurs de temps de précocité pour C_1, C_2 et C_3 sont respectivement, semaine 1, semaine 13 et semaine 8. L'axe des ordonnées correspond aux taux de ESS.

Afin de sélectionner le point de précocité le plus approprié pour une prédiction précise, précoce et permettant la stabilité du système, nous fixons aussi un seuil minimal pour la valeur du ESS. Comme dans le Chapitre 4, il n'y a pas de règle pour la définition de la valeur de ce seuil. Nous fixons ce seuil selon les besoins et les résultats de chaque système.

Analyse des taux de ESS et fixation du seuil minimal : les taux ESS des trois classes des apprenants suivent une allure décroissante au cours des points de précocité. Nous constatons aussi que les valeurs des taux ESS correspondantes aux différents points de précocité de C_2 et C_3 sont différents d'un système à un autre. De plus, pour tous les EWS, la classe du risque moyen C_2 se caractérise toujours par les plus faibles taux de ESS ; ceci revient au fait que les apprenants de cette classe sont correctement identifiés plus tard dans l'année que ceux des autres classes. De ce fait, pour chacun des EWS de notre étude, la définition du seuil minimal dépend de la valeur de ESS correspondant au 1er point de précocité de la classe C_2.

Pour EWS_1 : au premier point de précocité, C_2 et C_3 commencent respectivement avec $ESS = 68.04\%$ et $ESS = 74.73\%$. Au 9ème point de précocité, C_2 et C_3 atteignent respectivement des taux

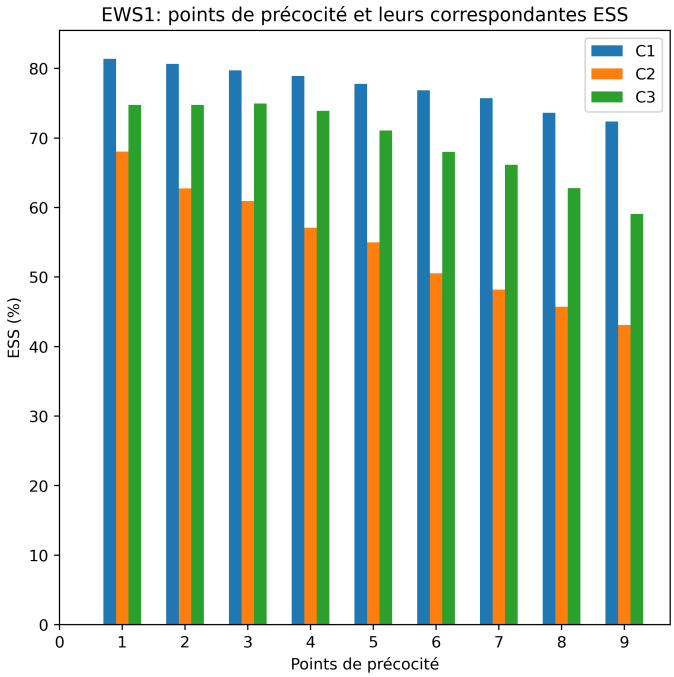


FIGURE 6.4 – Les points de précocité et leurs ESS correspondants - EWS_1

$ESS = 43.09\%$ et $ESS = 59\%$. Pour ce système, le seuil minimal de ESS correspond à un taux de 65%.

Pour EWS_2 : au premier temps précoce, les classes C_2 et C_3 montrent respectivement des $ESS = 65.79\%$ et $ESS = 73.25\%$. Au 9ème point précoce, C_2 et C_3 ont un taux $ESS = 45.68\%$. Pour ce système, le seuil minimal de ESS est égal à un taux de 65%.

Pour EWS_3 : au premier temps de précocité, C_2 et C_3 commencent respectivement avec des $ESS = 62.88\%$ et $ESS = 76.33\%$. À son 7ème et dernier point précoce, le taux ESS de C_2 est égale à 37.19%. Au 9ème point précoce, le taux ESS de C_3 est égale à 46.8%. Pour ce système, le seuil minimal de ESS correspond à un taux de 60%.

Pour EWS_4 : Au premier temps précoce, les deux classes C_2 et C_3 ont respectivement des taux $ESS = 75.18\%$ et $ESS = 77.31\%$. À son 7ème et dernier point, le taux de ESS de C_2 est égal à 53.17%. Au 9ème point précoce, le taux de ESS de C_3 est égal à 46.06%. Pour ce système, le seuil minimal de ESS correspond à un taux de 75%.

Identification du temps précoce pour appliquer les règles d'alerte pour chaque système : après avoir fixé le seuil minimal de ESS, il est maintenant opportun d'identifier pour chaque système le temps précoce pour pouvoir appliquer les règles d'alerte. Le Tableau 6.1 présente, pour les deux classes C_2 et C_3 (les classes du risque) de chaque système, le 1er taux de ESS supérieur ou égal au seuil prédéfini et sa valeur de précocité correspondante. La précocité est donnée en semaine. La valeur de précocité à retenir est le temps à partir duquel l'application des règles d'alerte peut être déclenchée (voir Algorithme 4). Cette précocité retenue correspond à la valeur du temps minimale entre les précocités des classes C_2 et C_3.

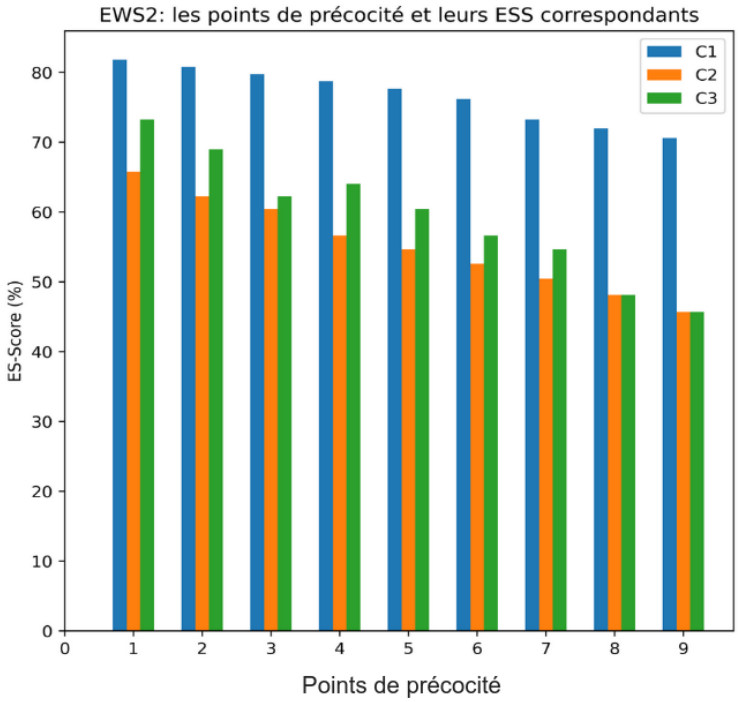


FIGURE 6.5 – Les points de précocité et leurs ESS correspondants - EWS_2

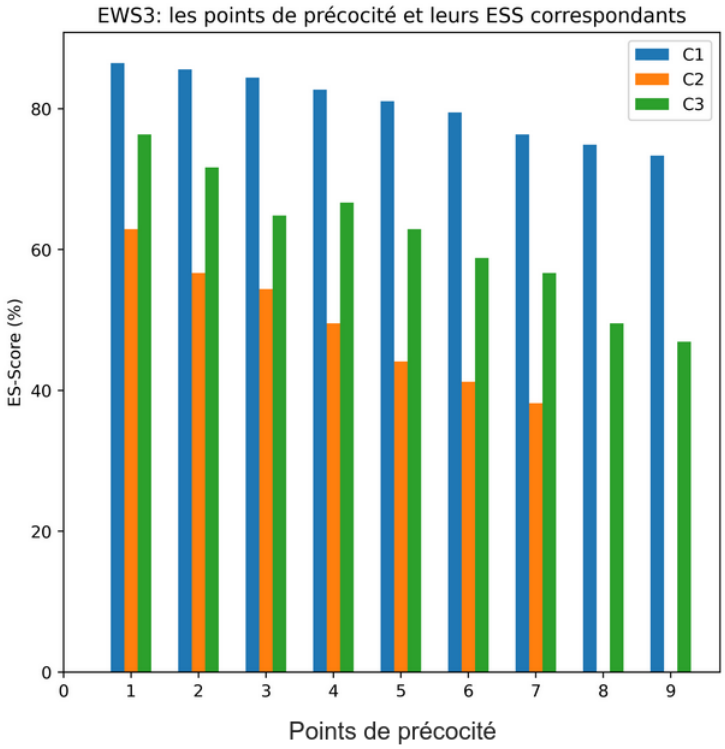


FIGURE 6.6 – Les points de précocité et leurs ESS correspondants - EWS_3

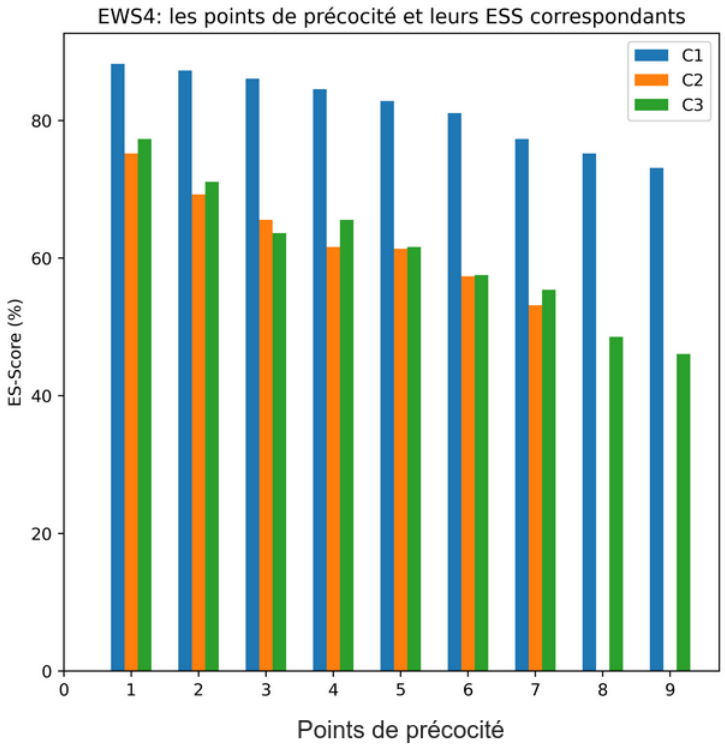


FIGURE 6.7 – Les points de précocité et leurs ESS correspondants - EWS_4

D'après le Tableau 6.1, la précocité retenue pour tous les systèmes est la semaine 8. Autrement dit, à partir de la 8ème semaine de prédiction, nous pouvons suivre les prédictions des apprenants pour appliquer les règles d'alerte.
Pour d'autres systèmes utilisant d'autres données et modèles et développés pour d'autres contextes, ce point du temps précoce pourrait probablement être différent de la semaine 8.

	EWS_1		EWS_2		EWS_3		EWS_4	
	C_2	C_3	C_2	C_3	C_2	C_3	C_2	C_3
ESS (%)	68.04	74.73	65.79	73.25	62.88	76.33	75.18	77.31
Précocité (semaine)	13	8	13	8	16	8	9	8
Précocité retenue (semaine)	8		8		8		8	

TABLE 6.1 – ESS des quatre EWS et leurs précocités correspondantes

6.6.2 Application des règles d'alerte et résultats

La 2ème partie de l'Algorithme d'alerte 4 consiste à appliquer des règles d'alerte sur les prédictions de chacun des systèmes EWS_1, EWS_2, EWS_3 et EWS_4.
Dans toute sa partie, nous étudions les résultats de l'application des règles suivantes :

- R_1 : alerter un enseignant dès qu'un apprenant est prédit en classe du risque
- R_2 : alerter un enseignant lorsqu'un apprenant est prédit 2 fois de suite en classe du risque
- R_3 : alerter un enseignant lorsqu'un apprenant est prédit 3 fois de suite en classe du risque
- R_4 : alerter un enseignant lorsqu'un apprenant est prédit 4 fois de suite en classe du risque

— R_5 : alerter un enseignant lorsqu'un apprenant est prédit 5 fois de suite en classe du risque

Ces cinq règles sont appliquées à partir de la semaine 8. Cette date a été fixée comme la précocité de tous les systèmes pour identifier les apprenants candidats à l'alerte.
Le Tableau 6.2 présente les résultats de l'application des règles R_1, R_2, R_3, R_4 et R_5 sur les systèmes EWS_1, EWS_2, EWS_3 et EWS_4. Pour chacune de ces règles, ce tableau montre le nombre total d'alertes générées, la précision (en pourcentage) de ces alertes ainsi que leurs valeurs de FPR (en pourcentage).

	EWS_1			EWS_2			EWS_3			EWS_4		
	Nombre total d'alerte	Précision (%)	FPR (%)	Nombre total d'alerte	Précision (%)	FPR (%)	Nombre total d'alerte	Précision (%)	FPR (%)	Nombre total d'alerte	Précision (%)	FPR (%)
R_1	53	66.96	33.96	89	35.95	64.04	48	62.5	37.5	72	31.94	68.05
R_2	45	77.77	22.22	62	51.61	48.38	42	69.04	30.95	57	38.59	61.40
R3	43	81.39	18.60	54	59.25	40.74	34	79.41	20.58	43	51.16	48.83
R4	41	85.36	14.63	51	62.74	37.25	28	85.71	14.28	32	65.625	34.375
R5	39	87.17	12.82	46	67.39	32.60	24	87.5	12.5	32	65.625	34.375

TABLE 6.2 – Application des cinq règles d'alerte sur les quatre systèmes

Analyse : d'après ce tableau, l'application de la règle R_1 engendre un nombre élevé d'alertes. La grande majorité de ces alertes est fausse. Par exemple, en appliquant R_1 sur les prédictions de EWS_2, 89 alertes sont générées dont seulement 35.95% sont bonnes. L'application de R_1 sur EWS_4 génère de 72 alertes dont plus la moitié (68.95%) est fausse. Pour EWS_1 et EWS_3, R_1 a donné un nombre considérablement important d'alertes dont seulement un peu plus que la moitié est une bonne alerte (66.96% et 62.6% pour respectivement EWS_1 et EWS_3). Étant donnée l'instabilité des résultats des EWS (en particulier aux premières dates de prédiction), l'application de R_1 ne peut pas assurer des alertes précises en diminuant au maximum le nombre d'alertes pour les apprenants qui réussissent. Un taux très élevé de fausses alertes engendre la diminution de la *confiance* de l'enseignant dans l'EWS.

L'application de R_2 sur les EWS permet de générer un nombre moins important d'alertes qui se caractérise par une meilleure précision en la comparant avec les alertes de R_1. Cette amélioration est constatée surtout avec les résultats de EWS_2 : 62 alertes sont générées dont 51.61% sont bonnes ; il s'agit donc d'une amélioration de 16 points par rapport à la précision de R_1 avec ce même système. Les résultats de EWS_4 sont légèrement meilleurs avec R_2 : 68 alertes sont générées dont seulement 38.59% sont bonnes. En appliquant R_2 sur EWS_3, 42 alertes sont générées dont 69.04% sont bonnes. R_2 donne les meilleurs résultats avec EWS_1 : 45 alertes sont générées dont 77.77% sont bonnes.

L'application de R_3 sur les EWS étudiés permet d'améliorer nettement la précision des alertes générées. Les améliorations les plus importantes sont observées avec EWS_3 et EWS_4 avec respectivement des taux de précision de 79.41% et 51.16%. Il s'agit d'une augmentation de précision de $\approx 10\%$ et $\approx 13\%$ par rapport à R_2 avec respectivement EWS_3 et EWS_4. R_3 permet une légère amélioration par rapport à R_2 avec EWS_2 et EWS_1. La règle R_3 considère plus la stabilité temporelle des prédictions ce qui explique la nette amélioration des valeurs de précision et la diminution des taux de FPR.

L'application de R_4 sur les résultats des différents systèmes permet l'amélioration des précisions des alertes par rapport aux règles précédentes. Cette amélioration est surtout observée avec EWS_3 et EWS_4 dont les précisions sont respectivement égales à 85.71% et 65.62%. Une légère amélioration de 4% et $\approx 3\%$ par rapport à R_3 est observée avec respectivement EWS_1 et EWS_2. Plus la règle tient compte de l'importance de la stabilité temporelle des prédictions, moins d'alertes sont générées par le système. Cependant, ces alertes se caractérisent par une précision élevée.

L'application de R_5 améliore légèrement les précisions des alertes par rapport à R_4. Par exemple, R_4 et R_5 avec EWS_4 ont les mêmes valeurs du nombre total d'alertes générées, de précision et de FPR.

À partir d'une certaine valeur de i de la règle R_i, la différence des valeurs de précision et de FPR n'est plus remarquable avec celles de la règle R_{i-1}. Il s'agit d'un compromis à établir entre la précocité des alertes et leurs précisions. En effet, certes plus la valeur de i est élevée, plus la précision du nombre total d'alertes est élevée et moins nombreuses sont les fausses alertes. Mais, il est aussi probable d'avoir raté la génération plus tôt des bonnes alertes avec une valeur de i plus faible.

Synthèse : d'après les résultats de l'application des cinq règles d'alerte sur les différents EWS, nous pouvons constater que :

- le nombre élevé d'alertes générées ne reflète pas forcément leurs précisions. Comme les apprenants à risque sont souvent moins nombreux que ceux qui réussissent, il n'est pas pertinent que l'EWS génère un nombre très élevé d'alertes. Il s'agit sinon de fausses alertes.
- la règle R_1 est insuffisante pour une génération d'alerte précise. L'application de cette règle avec la plupart des EWS engendre un nombre élevé d'alertes qui sont majoritairement fausses. En effet, particulièrement aux premiers temps de prédiction, les EWS sont susceptibles de fournir des prédictions fausses en identifiant des apprenants qui réussissent comme appartenant à des classes du risque et vice versa. Et comme cette règle ne tient pas compte de la stabilité de prédiction, un taux élevé de fausses alertes peut être observé ce qui diminue la *confiance* des enseignants en l'utilisation des EWS et leur pertinence.
- à partir de l'application de la règle R_2, il y a une amélioration des précisions et une diminution des valeurs de FPR des alertes générées avec la plupart des EWS. Cette règle donne les meilleurs résultats avec EWS_1.
- les plus faibles précisions des alertes générées sont observées avec les systèmes les moins stables dans le temps. Par exemple, la stabilité temporelle de EWS_2 et EWS_4 diminue dans le temps (voir Figure 6.3). Toutes les règles appliquées sur ces deux systèmes se caractérisent par une faible précision. Cependant, la stabilité de EWS_1 augmente légèrement dans le temps et nous remarquons une précision supérieure à 75% dès l'application de R_2. La stabilité temporelle de EWS_3 est moyenne dans le temps et nous remarquons qu'avec l'application de R_3 les alertes sont précises de plus de 75%. Cependant, comme EWS_2 et EWS_4 se caractérisent par des stabilités décroissantes dans le temps, les précisions des règles R_1, R_2 et R_3 restent faibles. C'est seulement avec R_5 que ces deux systèmes atteignent des précisions égales respectivement à $\approx 67\%$ et $\approx 65\%$.
- la valeur de i de la règle R_i à appliquer est différente d'un système à un autre et dépend clairement de la stabilité temporelle de ce dernier. Plus le système est stable, plus la valeur de i est faible. De plus, la valeur de i doit assurer le compromis entre la précocité des alertes (pour qu'elles soient plus efficaces) et leurs précisions. Autrement dit, si la valeur de i est élevée et comme l'alerte n'est générée que lorsque le nombre de prédictions en classe du risque est égale à i, nous pouvons probablement rater une génération plus tôt de bonnes alertes. Ajoutons à cela que certains systèmes ne vérifient pas la condition de i prédictions successives en classe du risque lorsque i est élevée ; d'où la possibilité de rater carrément des bonnes alertes.
- à partir d'une certaine valeur de i, la précision de R_i est légèrement supérieure à R_{i-1}. Il est peut être plus pertinent dans ce cas de favoriser la précocité à la succession d'un nombre i élevé de prédictions en classe du risque ; c'est-à-dire, opter pour la règle R_{i-1}. Par exemple, pour le système EWS_1, la précision de R_2 est déjà de $\approx 78\%$ et légèrement inférieure à celles de R_3, R_4 et R_5. Dans ce cas, appliquer R_2 avec EWS_1 permet de générer des alertes plus tôt pour les apprenants à risque ; ce qui permet une intervention plus efficace de l'enseignant. De plus, la différence des taux de FPR entre l'application de R_2 et R_3 est minimale.

6.7 Conclusion

En utilisant un système d'alerte précoce, et au delà de l'identification des apprenants à risque d'échec/ abandon, les enseignants s'attendent à être alertés de leur présence. Cependant, la littérature de LA ne présente pas de méthodes particulières pour une génération d'alerte précoce et précise. L'objectif de ce chapitre était donc de répondre aux limites des EWS de l'état de l'art.

Dans ce chapitre, nous avons proposé un algorithme pour une génération d'alerte précoce et précise. Cet algorithme se sert des métriques temporelles et de la notion des règles d'alerte R_i. En calculant la mesure ESS, l'algorithme fixe le point du temps précoce à partir duquel l'application d'une règle d'alerte peut être déclenchée. Comme l'alerte concerne essentiellement les apprenants à risque, ce point de précocité est identifié selon les résultats de précocité des classes du risque.

Cette approche a été testée sur quatre différents EWS éducatifs utilisant les traces numériques d'une population de collégiens. Les expérimentations conduites ont permis de déduire qu'un nombre élevé d'alertes générées ne reflète pas leurs précisions. Les résultats de l'alerte sont évalués à l'aide de deux métriques : précision et FPR. En effet, plus la valeur de i de la règle R_i est élevée, plus les alertes sont précises. Cette valeur optimale de i est différente d'un système à un autre et étroitement dépendante de la stabilité temporelle du EWS. En d'autres termes, plus le système est stable, plus i est faible et R_i est précise. De plus, afin d'augmenter *la confiance* des enseignants dans l'EWS, il est d'une grande importance de fournir des alertes les plus précises. Cependant, plus i est élevée, plus l'alerte est générée tard dans l'année. Il s'agit d'un compromis à établir entre la précocité de l'alerte et sa précision élevée.

En perspectives à ce travail, nous avons l'intention d'améliorer encore plus l'algorithme d'alerte pour qu'il tienne compte des oscillations qui peuvent caractériser les prédictions d'un EWS. Pour ce faire, la notion des règles d'alerte devrait être mise pour détecter au mieux les apprenants à risque. De plus, il serait pertinent de proposer un algorithme pour une génération d'alerte par profil d'apprenants.

Conclusion générale

1 Rappel du cadre et des objectifs de la thèse

L'éducation est importante pour l'épanouissement des peuples. C'est pour cette raison que nombreux sont les efforts fournis pour garantir un meilleur apprentissage aux différentes catégories et tranches d'âge d'une société. Parmi ces efforts, nous trouvons l'apprentissage en ligne. Ce mode d'apprentissage nécessite la disposition d'infrastructures numériques puissantes et efficaces (plateformes, applications, ...) permettant un accès régulier et aisé aux différentes ressources pédagogiques et académiques.

Certes l'apprentissage en ligne permet la continuité d'accès aux contenus pédagogiques à tout moment et de n'importe où dans le monde, et la situation de la pandémie de la COVID-19 en est le meilleur exemple. Mais ce mode d'apprentissage connaît plusieurs difficultés et défis. Le problème le plus souvent rapporté par les différentes institutions académiques en ligne est les taux élevés d'échec enregistrés annuellement. Ce problème est dû à plusieurs raisons, notamment l'hétérogénéité des apprenants, la diversité de leur comportement d'apprentissage, le manque et /ou l'inefficacité du suivi des tuteurs et des enseignants, des interventions tardives. ...

Le CNED, le plus grand centre français pour l'enseignement à distance, ne fait pas l'exception et ses statistiques annuelles font preuve aussi des pourcentages d'échec élevés parmi ses inscrits.

Afin de faire face à cette problématique, plusieurs projets de la littérature s'intéressent à l'identification précoce des apprenants à risque pour pouvoir intervenir efficacement auprès d'eux.

Dans cette thèse, nous nous sommes aussi intéressés à l'analyse des traces numériques des apprenants en ligne pour apporter une solution au problème d'échec à travers la prédiction précoce et précise des apprenants à risque.

De ce fait, cette thèse s'inscrit dans le domaine des LA et a recours aux domaines de l'intelligence artificielle et l'apprentissage automatique pour concevoir un système prédictif dédié aux enseignants, leur permettant d'être alertés au plus tôt des apprenants à risque.

Un tel système a pour objectif d'assurer une intervention pertinente et au bon moment auprès des apprenants avec des difficultés d'apprentissage, ce qui permettrait une éventuelle amélioration des résultats scolaires finaux.

Le contexte applicatif de cette thèse est le CNED. En d'autres termes, nous avons exploité les traces numériques des collégiens du CNED pour aborder la problématique d'échec. Rares sont les projets de l'état de l'art qui s'intéressent à l'enseignement secondaire, notamment au cas des collégiens. La majorité des travaux est dédiée à l'enseignement supérieur ou aux MOOCs.

Afin d'atteindre les objectifs de cette thèse, nous nous sommes trouvés face à plusieurs défis :

- la complexité de la compréhension des traces numériques, la multitude de sources de données et leur hétérogénéité, le grand volume de données à gérer et exploiter.
- la diversité des comportements des apprenants et la variété de leurs profils
- les caractéristiques et indicateurs d'apprentissage les plus significatifs.
- le(s) modèle(s) de ML qui correspond(ent) le mieux au résultat attendu.

— la nature évolutive et temporelle de l'apprentissage et donc du système à fournir.
— l'importance de l'alerte précoce et précise pour une intervention efficace.

2 Principales contributions

Pour répondre aux différents défis mentionnés, nous proposons les contributions suivantes :

2.1 Une méthodologie en profondeur

Cette première contribution propose une méthodologie qui se base sur les étapes classiques de l'apprentissage automatique. L'objectif de cette méthodologie est la sélection des caractéristiques d'apprentissage les plus corrélées avec le comportement d'un apprenant et du modèle de ML qui permet une prédiction précise des apprenants à risque.
En croisant les notions des sciences de l'éducation, les résultats des travaux de l'état de l'art ainsi que les données disponibles, nous définissons quatre indicateurs d'apprentissage qui sont : la performance, l'engagement, la réactivité et la régularité. De plus, cette méthodologie met l'accent sur l'importance des données temporelles dans l'amélioration des résultats de prédiction.
Afin d'évaluer la capacité des modèles de ML à prédire les apprenants à risque, qui sont la population cible, cette méthodologie propose de suivre l'évolution temporelle des mesures d'exactitude et de FPR.

2.2 Évaluation temporelle des systèmes prédictifs

L'objectif de cette contribution est l'évaluation de la performance des systèmes de prédiction en considérant l'importance de la dimension temporelle de l'apprentissage. Cette contribution propose deux métriques temporelles qui sont : la précocité et la stabilité. Ces deux métriques permettent d'évaluer les systèmes prédictifs dans le temps. La précocité retourne le temps à partir duquel la prédiction est correcte, tandis que la stabilité sert à suivre l'évolution des résultats de prédiction au cours du temps.
Cette contribution propose deux mesures permettant d'étudier les différents compromis existants entre les métriques de précision de ML, notamment l'exactitude, et les métriques temporelles. Ces deux mesures permettent d'identifier le point du temps à partir duquel la prédiction est précoce et le système est à la fois exact et stable.

2.3 Identification des profils d'apprenants pour l'évaluation des systèmes prédictifs

Les établissements d'apprentissage se caractérisent par la diversité de profils d'apprenants. Afin de concevoir un système qui fonctionne efficacement avec tous les apprenants, nous proposons une méthodologie d'évaluation qui repose sur l'identification des profils d'apprenants existants dans une cohorte et sur un large spectre de métriques de précision et temporelles. L'objectif est d'assurer le bon fonctionnement du système de prédiction avec les différents profils d'apprentissage d'une cohorte.

2.4 Génération précoce d'alerte

Toutes les contributions précédentes servent à la conception d'un système d'alerte précoce dédié aux enseignants et dont le résultat final est la génération d'alerte précoce et précise des apprenants à risque. Cette contribution propose un algorithme qui permet d'alerter les enseignants au plus tôt des apprenants à risque d'échec.
Afin de déterminer le temps adéquat pour la génération d'alerte, cet algorithme utilise les résultats des

métriques temporelles. Ensuite, il introduit la notion des règles d'alerte pour définir la méthode d'alerte. Afin d'évaluer la pertinence d'une alerte, deux métriques inspirées de la matrice de confusion habituelle et adaptées à la problématique d'alerte sont utilisées : la précision et le FPR.

3 Limites des travaux effectués

— Les indicateurs d'apprentissage identifiés par la méthodologie proposée dans le Chapitre 3 correspondent aux apprenants qui n'ont pas été soumis à des ajustements de parcours en cours d'année. Ainsi, le cas des apprenants avec des inscriptions très tardives dans l'année n'a pas été étudié. Cette catégorie d'apprenants pourrait avoir d'autres indicateurs et caractéristiques d'apprentissage qui caractérisent mieux leurs parcours académiques particuliers. De plus, la partie modélisation de l'activité de ces apprenants pourrait être différente de ceux qui ont été étudiés. De ce fait, toute l'approche prédictive pourrait être différente de celle proposée dans le Chapitre 3.

— La contribution du Chapitre 4 propose des métriques temporelles pour évaluer les systèmes basés sur des modèles de classification. Cependant, même si ces métriques gardent les mêmes définitions avec les systèmes de régression, leurs algorithmes seront différents de ce qui a été proposé.

— Dans le Chapitre 5, l'évaluation temporelle des personas identifiés a eu recours à la stabilité et seulement l'algorithme correspondant à la première définition de la précocité. Cette contribution se restreint à l'identification des profils d'apprenants existants à la fin de la période de l'apprentissage. Il est aussi d'un grand intérêt d'observer l'évolution des profils d'apprenants dans le temps. En effet, nous pouvons avoir des apprenants qui basculent entre plusieurs profils tout au long de la période d'apprentissage. Une étude pareille met encore plus en avant la nature évolutive et continue de l'apprentissage.

— Nous dégageons deux principales limites de l'algorithme d'alerte proposé. Ce dernier ne tient pas compte des importantes oscillations d'un système de prédiction. En effet, la notion d'une règle d'alerte exige que les prédictions en classes du risque soient successives. Cet algorithme ne détecte pas les apprenants à risque pour lesquels les règles d'alerte ne sont pas satisfaites, surtout lorsqu'il s'agit d'un i de R_i élevé. La 2ème limite consiste en la génération individuelle d'alerte. Plus explicitement, l'algorithme ne tient pas compte de la notion des profils types d'apprenants. L'objectif d'un tel algorithme est d'assurer les mêmes performances avec les différents profils d'apprentissage.

4 Perspectives

Étant donné les limites évoqués et soucieuses de la continuité de nos travaux, nous avons dégagé un ensemble de perspectives. Ces perspectives appartiennent à deux catégories qui sont : des perspectives à court terme et d'autres à long terme.

1. **Perspectives à court terme :** les perspectives à court terme sont les suivantes :
   — nous prévoyons appliquer le 2ème algorithme de précocité sur les prédictions des différents profils d'apprenants identifiés. Cet algorithme considère les oscillations des comportements des personas et retourne une liste de temps précoces de prédiction correcte. À partir de cette liste, des études de compromis sont réalisées à l'aide des mesures de ESS et EAS et qui permettront de fixer le temps précoce pour une prédiction correcte de chaque persona.

— nous envisageons de combiner l'algorithme de la génération d'alerte avec la notion de profils types d'apprenants. Une telle expérimentation aura pour but de générer des alertes par profil d'apprenants.
— nous prévoyons d'utiliser les métriques temporelles et les mesures de compromis établies pour développer un indicateur de confiance. Il s'agit de mesurer la confiance que pourrait avoir un enseignant envers les résultats d'un EWS. Cette notion de confiance serait aussi importante à intégrer dans la génération d'alerte.
— nous avons l'intention d'améliorer la notion des règles d'alerte pour qu'elles tiennent compte des oscillations des prédictions par lesquelles se caractérisent les systèmes de classification.

2. **Perspectives à long terme :** les perspectives à long termes sont les suivantes :
— nous envisageons d'appliquer tous les travaux et les résultats trouvés aux systèmes basés sur les modèles de régression. Il est certain que ces systèmes sont différents de ceux de la classification. Il serait pertinent d'établir des adaptations sur quelques contributions, notamment les métriques temporelles.
— dans cette thèse, nous avons considéré les cas des apprenants qui ne passent pas par des ajustements de programmes académiques. Nous prévoyons faire une étude de cas sur les apprenants avec des ajustements de parcours. Cette étude de cas aura pour but d'identifier d'autres indicateurs d'apprentissage qui caractérisent le comportement et les activités académiques de cette catégorie d'apprenants.
— Comme l'apprentissage est une suite d'événements qui se produisent dans le temps, il serait pertinent d'utiliser les modèles séquentiels. Une telle modélisation permet une illustration du scénario réel de l'apprentissage.

Bibliographie

[1] Mohammad Khalil and Martin Ebner. Learning analytics : principles and constraints. In *EdMedia+ Innovate Learning*, pages 1789–1799. Association for the Advancement of Computing in Education (AACE), 2015.

[2] Javier López Zambrano, Juan Alfonso Lara Torralbo, Cristóbal Romero Morales, et al. Early prediction of student learning performance through data mining : A systematic review. *Psicothema*, 2021.

[3] Alhothali Areej, Albsisi Maram, Assalahi Hussein, and Aldosemani Tahani. Predicting student outcomes in online courses using machine learning techniques : A review. *Journal of Learning Analytics*, 2022.

[4] Rüdiger Wirth and Jochen Hipp. Crisp-dm : Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester, 2000.

[5] Célina Treuillier and Anne Boyer. A new way to characterize learning datasets. *Proceedings of the 14th International Conference on Computer Supported Education*, 2022.

[6] Yves Soulabail. Histoire de l'humanité. volume i : De la préhistoire aux débuts de la civilisation. *La Revue des Sciences de Gestion : Direction et Gestion*, 45(242) :126, 2010.

[7] L'Enseignement ouvert et à distance : tendances, considérations politiques et stratégiques - UNESCO Bibliothèque Numérique.

[8] Jean-Charles Marty and Alain Mille. *Analyse de traces et personnalisation des environnements informatiques pour l'apprentissage humain*. Hermes Sciences Publications, 2009.

[9] Amal Ben Soussia, Azim Roussanaly, and Anne Boyer. An in-depth methodology to predict at-risk learners. In *European Conference on Technology Enhanced Learning*, pages 193–206. Springer, 2021.

[10] Anonymus AC08926112. *Equity and quality in education : Supporting disadvantaged students and schools*. OECD, 2012.

[11] Papia Bawa. Retention in online courses : Exploring issues and solutions—a literature review. *Sage Open*, 6(1) :2158244015621777, 2016.

[12] Jakub Kuzilek, Martin Hlosta, Drahomira Herrmannova, Zdenek Zdrahal, Jonas Vaclavek, and Annika Wolff. Ou analyse : analysing at-risk students at the open university. *Learning Analytics Review*, pages 1–16, 2015.

[13] George Siemens and Phil Long. Penetrating the fog : Analytics in learning and education. *EDUCAUSE review*, 46(5) :30, 2011.

[14] Mark Van Harmelen and David Workman. Analytics for learning and teaching. *CETIS Analytics Series*, 1(3) :1–40, 2012.

[15] Yeonjeong Park and I-H Jo. Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 21(1) :110, 2015.

[16] David Baneres, Abdulkadir Karadeniz, Ana-Elena Guerrero-Roldán, M Elena Rodríguez-Gonzalez, and Montse Serra. Analysis of the accuracy of an early warning system for learners at-risk : A case study. In *Proceedings of the 11th International Conference on Education and New Learning Technologies—EDULEARN19, Palma, Spain*, pages 1–3, 2019.

[17] Balqis Albreiki, Nazar Zaki, and Hany Alashwal. A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, 11(9) :552, 2021.

[18] Gary James. Advantages and disadvantages of online learning. *Retrieved July*, 1 :2006, 2002.

[19] Melody W Alexander, Allen D Truell, and Jensen J Zhao. Expected advantages and disadvantages of online learning : Perceptions from college students who have not taken online courses. *Issues in Information Systems*, 13(2) :193–200, 2012.

[20] Potvin Pierre. La réussite éducative : Texte pour le cadre de référence du ctreq. 30-04-10.

[21] Feyfant Annie. Réussite éducative, réussite scolaire ? *Note de Veille de l'IFÉ : Revue de littérature de recherche*, Février 2014.

[22] Glasman Dominique. "il n'y a pas de réussite scolaire !" "le sens du programme de "réussite éducative". *Caisse nationale d'allocations familiales | « Informations sociales »*, page n° 141 | pages 74 à 85, 2007.

[23] Lise Saint-Laurent. Programme d'intervention auprès des élèves à risque : une nouvelle option éducative, Éditions gaëtan morin, boucherville. 1995.

[24] Référentielles : Les élèves à risque et hdaa). *Fédération des syndicats de l'enseignement*, Avril 2013.

[25] Godard L. Notes de cours sur les difficultés en lecture et en écriture. *UQM, Montréal*, 2007.

[26] Goupil G. Les élèves en difficulté d'adaptation et d'apprentissage, Éditions gaëtan morin, boucherville, 3ème édition. 2007.

[27] Potvin P., Fortin L., Marcotte D., and collab. Y'a une place pour toi !,guide de prévention du décrochage scolaire, ctreq,québec. 2007.

[28] Poliquin-Verville H. and Royer É. Les troubles du comportement : état des connaissances et perspectives d'intervention, meq, québec. 1992.

[29] Kew Si Na and Zaidatun Tasir. Identifying at-risk students in online learning by analysing learning behaviour : A systematic review. In *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, pages 118–123. IEEE, 2017.

[30] JoAnn T Funk. At-risk online learners : Reducing barriers to success. *eLearn*, 2005(8) :3, 2005.

[31] Hsiang-Yu Chien, Oi-Man Kwok, Yu-Chen Yeh, Noelle Wall Sweany, Eunkyeng Baek, and William McIntosh. Identifying at-risk online learners by psychological variables using machine learning techniques. *Online Learning*, 24(4) :131–146, 2020.

[32] Carlos J Asarta and James R Schmidt. Access patterns of online materials in a blended course. *Decision Sciences Journal of Innovative Education*, 11(1) :107–123, 2013.

[33] Bruce W Tuckman. Relations of academic procrastination, rationalizations, and performance in a web course with deadlines. *Psychological reports*, 96(3_suppl) :1015–1021, 2005.

[34] Viola Osborn. Identifying at-risk students in videoconferencing and web-based distance education. *American Journal of Distance Education*, 15(1) :41–54, 2001.

[35] Mark Taormino. Student preparation for distance education. *Distance Learning*, 7(3) :55, 2010.

[36] Filippo Sciarrone. Machine learning and learning analytics : Integrating data with learning. In *2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–5. IEEE, 2018.

[37] Srećko Joksimović, Vitomir Kovanović, and Shane Dawson. The journey of learning analytics. *HERDSA Review of Higher Education*, 6 :27–63, 2019.

[38] Peter Reimann. Connecting learning analytics with learning research : The role of design-based research. *Learning : Research and Practice*, 2(2) :130–142, 2016.

[39] Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, and Hendrik Thüs. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6) :318–331, 2012.

[40] Cristóbal Romero and Sebastián Ventura. Educational data mining : a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6) :601–618, 2010.

[41] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. Early detection of students at risk–predicting student dropouts using administrative student data and machine learning methods. *Available at SSRN 3275433*, 2018.

[42] Cristóbal Romero and Sebastian Ventura. Guest editorial : Special issue on early prediction and supporting of learning performance. *IEEE Transactions on Learning Technologies*, 12(2) :145–147, 2019.

[43] Martin Hlosta, Zdenek Zdrahal, and Jaroslav Zendulka. Ouroboros : early identification of at-risk students without models based on legacy data. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 6–15, 2017.

[44] Muhammad Adnan, Asad Habib, Jawad Ashraf, Shafaq Mussadiq, Arsalan Ali Raza, Muhammad Abid, Maryam Bashir, and Sana Ullah Khan. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access*, 9 :7519–7539, 2021.

[45] Josep Figueroa-Cañas and Teresa Sancho-Vinuesa. Early prediction of dropout and final exam performance in an online statistics course. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 15(2) :86–94, 2020.

[46] Martín Liz-Domínguez, Manuel Caeiro Rodríguez, Martín Llamas Nistal, and Fernando A Mikic-Fonte. Predictors and early warning systems in higher education-a systematic literature review. *LASI-SPAIN*, pages 84–99, 2019.

[47] Reid Basher. Global early warning systems for natural hazards : systematic and people-centred. *Philosophical transactions of the royal society a : mathematical, physical and engineering sciences*, 364(1845) :2167–2182, 2006.

[48] Matthieu Bussiere and Marcel Fratzscher. Towards a new early warning system of financial crises. *journal of International Money and Finance*, 25(6) :953–973, 2006.

[49] ME Beth Smith, Joseph C Chiovaro, Maya O'Neil, Devan Kansagara, Ana R Quiñones, Michele Freeman, Makalapua L Motu'apuaka, and Christopher G Slatore. Early warning system scores for clinical deterioration in hospitalized patients : a systematic review. *Annals of the American Thoracic Society*, 11(9) :1454–1465, 2014.

[50] Anjeela Jokhan, Bibhya Sharma, and Shaveen Singh. Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 44(11) :1900–1911, 2019.

[51] Kimberly E Arnold and Matthew D Pistilli. Course signals at purdue : Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 267–270, 2012.

[52] Andrew E Krumm, R Joseph Waddington, Stephanie D Teasley, and Steven Lonn. A learning management system-based early warning system for academic advising in undergraduate engineering. In *Learning analytics*, pages 103–119. Springer, 2014.

[53] Zhuping Wang, Chenjing Zhu, Zelin Ying, Ying Zhang, Ben Wang, Xinyu Jin, and Huansong Yang. Design and implementation of early warning system based on educational big data. In *2018 5th International Conference on Systems and Informatics (icsai)*, pages 549–553. IEEE, 2018.

[54] Gökhan Akçapınar, Mohammad Nehal Hasnine, Rwitajit Majumdar, Brendan Flanagan, and Hiroaki Ogata. Developing an early-warning system for spotting at-risk students by using ebook interaction logs. *Smart Learning Environments*, 6(1) :1–15, 2019.

[55] Anat Cohen. Analysis of student activity in web-supported courses as a tool for predicting dropout. *Educational Technology Research and Development*, 65(5) :1285–1304, 2017.

[56] Ya-Han Hu, Chia-Lun Lo, and Sheng-Pao Shih. Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, 36 :469–478, 2014.

[57] Jeanette Samuelsen, Weiqin Chen, and Barbara Wasson. Integrating multiple data sources for learning analytics—review of literature. *Research and Practice in Technology Enhanced Learning*, 14(1) :1–20, 2019.

[58] William Villegas-Ch, Milton Roman-Cañizares, Angel Jaramillo-Alcázar, and Xavier Palacios-Pacheco. Data analysis as a tool for the application of adaptive learning in a university environment. *Applied Sciences*, 10(20) :7016, 2020.

[59] Beat A Schwendimann, Maria Jesus Rodriguez-Triana, Andrii Vozniuk, Luis P Prieto, Mina Shirvani Boroujeni, Adrian Holzer, Denis Gillet, and Pierre Dillenbourg. Perceiving learning at a glance : A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1) :30–41, 2016.

[60] Simon Knight, Alyssa Friend Wise, and Bodong Chen. Time for change : Why learning analytics needs temporal analysis. *Journal of Learning Analytics*, 4(3) :7–17, 2017.

[61] Bodong Chen, Simon Knight, and Alyssa Wise. Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, 2018.

[62] Alireza Ahadi, Raymond Lister, Heikki Haapala, and Arto Vihavainen. Exploring machine learning methods to automatically identify students in need of assistance. In *Proceedings of the eleventh annual international conference on international computing education research*, pages 121–130, 2015.

[63] Zoe YR Chen, Anna YQ Huang, Owen HT Lu, and Stephen JH Yang. Considering temporal features in early prediction of at-risk students. In *2021 International Conference on Advanced Learning Technologies (ICALT)*, pages 350–351. IEEE, 2021.

[64] IKSAL Sébastien. Ingénierie de l'observation basée sur la prescription en eiah. Septembre 2012.

[65] Su Mu, Meng Cui, and Xiaodi Huang. Multimodal data fusion in learning analytics : A systematic review. *Sensors*, 20(23) :6856, 2020.

[66] Fedia Hlioui, Nadia Aloui, and Faiez Gargouri. A withdrawal prediction model of at-risk learners based on behavioural indicators. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 16(2) :32–53, 2021.

[67] Mina Shirvani Boroujeni, Kshitij Sharma, Łukasz Kidziński, Lorenzo Lucignano, and Pierre Dillenbourg. How to quantify student's regularity ? In *European conference on technology enhanced learning*, pages 277–291. Springer, 2016.

[68] N Sclater, A Peasgood, and J Mullan. Case study : A traffic lights and interventions : Signals at purdue university, 2016.

[69] Min Hu and Hao Li. Student engagement in online learning : A review. In *2017 International Symposium on Educational Technology (ISET)*, pages 39–43. IEEE, 2017.

[70] Mushtaq Hussain, Wenhao Zhu, Wu Zhang, and Syed Muhammad Raza Abidi. Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational intelligence and neuroscience*, 2018, 2018.

[71] Kerri-Lee Krause and Hamish Coates. Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, 33(5) :493–505, 2008.

[72] Alexander W Astin. Student involvement : A developmental theory for higher education. *Journal of college student personnel*, 25(4) :297–308, 1984.

[73] Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. Learning latent engagement patterns of students in online courses. In *Twenty-eighth AAAI conference on artificial intelligence*, 2014.

[74] Hajra Waheed, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman, Salem Alelyani, and Raheel Nawaz. Predicting academic performance of students from vle big data using deep learning models. *Computers in Human behavior*, 104 :106189, 2020.

[75] Naif Radi Aljohani, Ayman Fayoumi, and Saeed-Ul Hassan. Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability*, 11(24) :7238, 2019.

[76] Fu Chen and Ying Cui. Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, 7(2) :1–17, 2020.

[77] Barry J Zimmerman. Investigating self-regulation and motivation : Historical background, methodological developments, and future prospects. *American educational research journal*, 45(1) :166–183, 2008.

[78] Randy Moore. Attendance and performance. *Journal of College Science Teaching*, 32(6) :367, 2003.

[79] DV Eastmond. Adult learning of distance students through computer conferencing. *Unpublished doctoral dissertation, New York : Syracuse University*, 1993.

[80] Ji Won You. Identifying significant indicators using lms data to predict course achievement in online learning. *The Internet and Higher Education*, 29 :23–30, 2016.

[81] Jelena Jovanović, Mohammed Saqr, Srećko Joksimović, and Dragan Gašević. Students matter the most in learning analytics : The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, 172 :104251, 2021.

[82] TM Mitchell. Machine learning mcgraw-hill international. 1997.

[83] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9 :381–386, 2020.

[84] Veronika Gamper and Stefan Nothelfer. *The Future of Education. Trend Report Fall 2015*. CDTM, Center for Digital Technology and Management, 2015.

[85] Kangwook Lee, Jichan Chung, and Changho Suh. Large-scale and interpretable collaborative filtering for educational data. *ML4ED KDD Work*, pages 1–7, 2017.

[86] Neema Mduma, Khamisi Kalegele, and Dina Machuve. A survey of machine learning approaches and techniques for student dropout prediction. 2019.

[87] Geraldine Gray, Colm McGuinness, and Philip Owende. An application of classification models to predict learner progression in tertiary education. In *2014 IEEE International Advance Computing Conference (IACC)*, pages 549–554. IEEE, 2014.

[88] Srečko Natek and Moti Zwilling. Student data mining solution–knowledge management system related to higher education institutions. *Expert systems with applications*, 41(14) :6400–6407, 2014.

[89] Mushtaq Hussain, Wenhao Zhu, Wu Zhang, Syed Muhammad Raza Abidi, and Sadaqat Ali. Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, 52(1) :381–407, 2019.

[90] Annika Wolff, Zdenek Zdrahal, Drahomira Herrmannova, Jakub Kuzilek, and Martin Hlosta. Developing predictive models for early detection of at-risk students on distance learning modules. 2014.

[91] Wanli Xing, Xin Chen, Jared Stein, and Michael Marcinkowski. Temporal predication of dropouts in moocs : Reaching the low hanging fruit through stacking generalization. *Computers in human behavior*, 58 :119–129, 2016.

[92] Jiazhen He, James Bailey, Benjamin Rubinstein, and Rui Zhang. Identifying at-risk students in massive open online courses. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[93] Samuel PM Choi, Sze Sing Lam, Kam Cheong Li, and Billy TM Wong. Learning analytics at low cost : At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, 21(2) :273–290, 2018.

[94] Gökhan Akçapınar, Arif Altun, and Petek Aşkar. Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1) :1–20, 2019.

[95] Lin Qiu, Yanshen Liu, and Yi Liu. An integrated framework with feature selection for dropout prediction in massive open online courses. *IEEE Access*, 6 :71474–71484, 2018.

[96] David Monllaó Olivé, Du Q Huynh, Mark Reynolds, Martin Dougiamas, and Damyon Wiese. A supervised learning framework : Using assessment to identify students at risk of dropping out of a mooc. *Journal of Computing in Higher Education*, 32(1) :9–26, 2020.

[97] Edmund Evangelista. A hybrid machine learning framework for predicting students' performance in virtual learning environment. *International Journal of Emerging Technologies in Learning (iJET)*, 16(24), 2021.

[98] B Loganathan, Indrajit Patra, Vipul Garchar, Harikumar Pallathadka, Mohd Naved, and Sanjeev Gour. Development of machine learning based framework for classification and prediction of students in virtual classroom environment. In *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, pages 1–5. IEEE, 2022.

[99] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification : an overview. *arXiv preprint arXiv :2008.05756*, 2020.

[100] David Bañeres, M Elena Rodríguez, Ana Elena Guerrero-Roldán, and Abdulkadir Karadeniz. An early warning system to detect at-risk students in online higher education. *Applied Sciences*, 10(13) :4427, 2020.

[101] Sunbok Lee and Jae Young Chung. The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences*, 9(15) :3093, 2019.

[102] Halit Karalar, Ceyhun Kapucu, and Hüseyin Gürüler. Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education*, 18(1) :1–18, 2021.

[103] Raza Hasan, Sellappan Palaniappan, Salman Mahmood, Ali Abbas, Kamal Uddin Sarker, and Mian Usman Sattar. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, 10(11) :3894, 2020.

[104] Tiago Santos and Roman Kern. A literature survey of early time series classification and deep learning. *Sami@ iknow*, 2016.

[105] Jiaojiao Wang, Dongjin Yu, Chengfei Liu, and Xiaoxiao Sun. Outcome-oriented predictive process monitoring with attention-based bidirectional lstm neural networks. In *2019 IEEE International Conference on Web Services (ICWS)*, pages 360–367. IEEE, 2019.

[106] Aneesh G Nath, Anshul Sharma, Sandeep S Udmale, and Sanjay Kumar Singh. An early classification approach for improving structural rotor fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 70 :1–13, 2020.

[107] Patrick Schäfer and Ulf Leser. Teaser : early and accurate time series classification. *Data mining and knowledge discovery*, 34(5) :1336–1362, 2020.

[108] Michel Philipp, Thomas Rusch, Kurt Hornik, and Carolin Strobl. Measuring the stability of results from supervised statistical learning. *Journal of Computational and Graphical Statistics*, 27(4) :685–700, 2018.

[109] Riccardo Guidotti and Salvatore Ruggieri. Assessing the stability of interpretable models. *arXiv preprint arXiv :1810.09352*, 2018.

[110] Irene Teinemaa, Marlon Dumas, Anna Leontjeva, and Fabrizio Maria Maggi. Temporal stability in predictive process monitoring. *Data Mining and Knowledge Discovery*, 32(5) :1306–1338, 2018.

[111] Abdallah Namoun and Abdullah Alshanqiti. Predicting student performance using data mining and learning analytics techniques : A systematic literature review. *Applied Sciences*, 11(1) :237, 2020.

[112] Jay Bainbridge, James Melitski, Anne Zahradnik, Eitel JM Lauría, Sandeep Jayaprakash, and Josh Baron. Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, 21(2) :247–262, 2015.

[113] Samira Khouljiand OuahiMariame et al. Feature engineering, mining for predicting student success based on interaction with the virtual learning environment using artificial neural network. *Annals of the Romanian Society for Cell Biology*, 25(6) :12734–12746, 2021.

[114] Gábor Kőrösi and Richard Farkas. Mooc performance prediction by deep learning from raw clickstream data. In *International Conference on Advances in Computing and Data Sciences*, pages 474–485. Springer, 2020.

[115] Viet Anh Nguyen, Quang Bach Nguyen, and Vuong Thinh Nguyen. A model to forecast learning outcomes for students in blended learning courses based on learning analytics. In *Proceedings of the 2nd International Conference on E-Society, E-Education and E-Technology*, pages 35–41, 2018.

[116] David Baneres, M Elena Rodríguez-Gonzalez, and Montse Serra. An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Transactions on Learning Technologies*, 12(2) :249–263, 2019.

[117] M Srilekshmi, S Sindhumol, Shiffon Chatterjee, and Kamal Bijlani. Learning analytics to identify students at-risk in moocs. In *2016 IEEE Eighth international conference on technology for education (T4E)*, pages 194–199. IEEE, 2016.

[118] Chen-Hsiang Yu, Jungpin Wu, and An-Chi Liu. Predicting learning outcomes with mooc clickstreams. *Education sciences*, 9(2) :104, 2019.

[119] Jui-Long Hung, Morgan C Wang, Shuyan Wang, Maha Abdelrasoul, Yaohang Li, and Wu He. Identifying at-risk students for early interventions—a time-series clustering approach. *IEEE Transactions on Emerging Topics in Computing*, 5(1) :45–55, 2015.

[120] Imani Mwalumbwe and Joel S Mtebe. Using learning analytics to predict students' performance in moodle learning management system : A case of mbeya university of science and technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1) :1–13, 2017.

[121] Raghad Al-Shabandar, Abir Jaafar Hussain, Panos Liatsis, and Robert Keight. Detecting at-risk students with early interventions using machine learning techniques. *IEEE Access*, 7 :149464–149478, 2019.

[122] Colin Shearer. The crisp-dm model : the new blueprint for data mining. *Journal of data warehousing*, 5(4) :13–22, 2000.

[123] Fu Tak-chung. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, 2011.

[124] CNED. L'histoire du cned depuis 1939.

[125] Fatemeh Nargesian, Horst Samulowitz, Udayan Khurana, Elias B Khalil, and Deepak S Turaga. Learning feature engineering for classification. In *Ijcai*, pages 2529–2535, 2017.

[126] Sébastien Iksal. *Ingénierie de l'observation basée sur la prescription en EIAH*. PhD thesis, Université du Maine, 2012.

[127] B Venkatesh and J Anuradha. A review of feature selection and its methods. *Cybernetics and information technologies*, 19(1) :3–26, 2019.

[128] Haotian Zhang, Lin Zhang, and Yuan Jiang. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2019.

[129] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, 5 :532–538, 2009.

[130] Benoît Sagot and Eric Villemonte de La Clergerie. Error mining in parsing results. In *The 21st International Conference of the Association for Computational Linguistics (ACL 2006)*, pages 329–336, 2006.

[131] Chahrazed Labba and Anne Boyer. When and how to update online analytical models for predicting students performance ? In *European Conference on Technology Enhanced Learning*, pages 173–186. Springer, 2022.

[132] David Bañeres, M Elena Rodríguez, Ana Elena Guerrero-Roldán, and Abdulkadir Karadeniz. An early warning system to detect at-risk students in online higher education. *Applied Sciences*, 10(13) :4427, 2020.

[133] Amal Ben soussia, Chahrazed Labba, Azim Roussanaly, and Anne Boyer. Assess performance prediction systems : Beyond precision indicators. *Proceedings of the 14th International Conference on Computer Supported Education*, 1 :489–496, 2022.

[134] Di Xu and Shanna S Jaggars. Performance gaps between online and face-to-face courses : Differences across types of students and academic subject areas. *The Journal of Higher Education*, 85(5) :633–659, 2014.

[135] Christopher Brooks and Jim Greer. Explaining predictive models to learning specialists using personas. In *Proceedings of the fourth international conference on learning analytics and knowledge*, pages 26–30, 2014.

[136] Celina Treuillier and Anne Boyer. Identification of class-representative learner personas. In *LA4SLE 2021-Learning Analytics for Smart Learning Environments*, volume 3024, pages 38–45, 2021.

[137] Ningyu Zhang, Gautam Biswas, and Yi Dong. Characterizing students' learning behaviors using unsupervised learning methods. In *International conference on artificial intelligence in education*, pages 430–441. Springer, 2017.

[138] Sharon Slade and Paul Prinsloo. Learning analytics : Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10) :1510–1529, 2013.

[139] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova, Ig Ibert Bittencourt, et al. Ethics of ai in education : Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, pages 1–23, 2021.

[140] Graham Pluck and HL Johnson. Stimulating curiosity to enhance learning. *GESJ : Education Sciences and Psychology*, 2, 2011.

[141] Silhouettes Rousseeuw. A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, (20) :53.

[142] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2) :224–227, 1979.

[143] Carine Lallemand and Guillaume Gronier. *Méthodes de design UX : 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles, 2015.

Annexe A

Les caractéristiques extraites de la base de données CNED

Dans cette annexe, nous présentons toutes les caractéristiques extraites de la base de données du CNED.

A.1 Les caractéristiques démographiques

Le Tableau A.1 présente les caractéristiques démographiques des apprenants du CNED.

Caractéristique	Signification
bourse	avoir une bourse ou pas
redoublant	refaire l'année ou pas
qualité	le genre (femme ou homme)
cp	code postale
paysExp	pays exportateur
paysRes	le pays de résidence
paysNat	le pays natal
depNaiss	le départemenet de naissance
paysNaiss	le pays de naissance
age	l'âge de l'élève

TABLE A.1 – Tableau des caractéristiques démographiques

A.2 Les caractéristiques de l'indicateur de performance

Dans cette partie, nous présentons les caractéristiques extraites de la base de données du CNED pour illustrer la performance des apprenants. Elles sont les suivantes :

— $nExamen_{t_i}$: nombre d'examens soumis pour la matière en question jusqu'à la semaine de prédiction t_i.
— $moyNote_{t_i}$: moyenne de notes de la matière en question jusqu'à la semaine de prédiction t_i.
— $totalExamen$: nombre total d'examens effectués pour la matière en question.

A.3 Les caractéristiques de l'indicateur d'engagement

Le Tableau A.2 présente les caractéristiques extraites de la base de données du CNED pour illustrer l'engagement des apprenants.

Caractéristique	**Signification**
$actionLundi_{t_i}$	nombre d'actions en ligne effectuées sur la matière en question tous les Lundis jusqu'à la semaine de prédiction t_i
$actionMardi_{t_i}$	nombre d'actions en ligne effectuées la matière en question tous les Mardis jusqu'à la semaine de prédiction t_i
$actionMercredi_{t_i}$	nombre d'actions en ligne effectuées sur la matière en question tous les Mercredis jusqu'à la semaine de prédiction t_i
$actionJeudi_{t_i}$	nombre d'actions en ligne effectuées sur la matière en question tous les Jeudis jusqu'à la semaine de prédiction t_i
$actionVendredi_ti$	nombre d'actions en ligne effectuées la matière en question tous les Vendredis jusqu'à la semaine de prédiction t_i
$actionSamedi_{t_i}$	nombre d'actions en ligne effectuées sur la matière en question tous les Samedis jusqu'à la semaine de prédiction t_i
$actionDimanche_{t_i}$	nombre d'actions en ligne effectuées sur la matière en question tous les Dimanches jusqu'à la semaine de prédiction t_i
$actionResourcet_i$	nombre d'actions effectuées sur la composante de type Resource de la matière en question jusqu'à la semaine de prédiction t_i
$actionDialogue_{t_i}$	nombre d'actions effectuées sur la composante de type Dialogue de la matière en question jusqu'à la semaine de prédiction t_i
$actionForum_{t_i}$	nombre d'actions effectuées sur la composante de type Forum de la matière en question jusqu'à la semaine de prédiction t_i
$actionFolder_{t_i}$	nombre d'actions effectuées sur la composante de type Folder de la matière en question jusqu'à la semaine de prédiction t_i
$actionCore_{t_i}$	nombre d'actions effectuées sur la composante de type Core de la matière en question jusqu'à la semaine de prédiction t_i
$actionScorm_{t_i}$	nombre d'actions effectuées sur la composante de type Scorm de la matière en question jusqu'à la semaine de prédiction t_i
$actionCreated_{t_i}$	nombre d'actions de type Created effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionViewed_{t_i}$	nombre d'actions de type Viewed effectuées sur la matière en question jusqu'à la semaine de prédiction t_i

$actionDownloaded_{t_i}$	nombre d'actions de type Downloaded effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionGraded_{t_i}$	nombre d'actions de type Graded effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionLaunched_{t_i}$	nombre d'actions de type Launched effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionSubmitted_{t_i}$	nombre d'actions de type Submitted effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionCreated_{t_i}$	nombre d'actions de type Created effectuées sur la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedCore_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Core de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedResource_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Resource de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedDialogue_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Dialogue de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedForum_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Forum de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedFolder_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Folder de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedScorm_{t_i}$	nombre d'actions de type Viewed effectués sur la composante Scorm de la matière en question jusqu'à la semaine de prédiction t_i
$actionCreatedTargetConversation_{t_i}$	nombre d'actions de type Created effectuées avec l'objet Conversation de la matière en question jusqu'à la semaine de prédiction t_i
$actionCreatedTargetReply_{t_i}$	nombre d'actions de type Created effectuées avec l'objet Reply de la matière en question jusqu'à la semaine de prédiction t_i
$actionCreatedTargetConversationt_i$	nombre d'actions de type Created effectuées avec l'objet Conversation de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedTargetCourse_{t_i}$	nombre d'action de type Viewed effectuées avec l'objet Course de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedTargetCourseModule_{t_i}$	nombre d'action de type Viewed effectuées avec l'objet Course Module de la matière en question jusqu'à la semaine de prédiction t_i

$actionViewedTargetConversation_{t_i}$	nombre d'action de type Viewed effectuées avec l'objet Conversation de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedTargetUserProfile_{t_i}$	nombre d'action de type Viewed effectuées avec l'objet User Profile Module de la matière en question jusqu'à la semaine de prédiction t_i
$actionViewedTargetUserReport_{t_i}$	nombre d'action de type Viewed effectuées avec l'objet User Report de la matière en question jusqu'à la semaine de prédiction t_i

TABLE A.2: Les caractéristiques extraites de la base du CNED correspondantes à l'indicateur d'engagement

A.4 Les caractéristiques de l'indicateur de régularité

Dans cette partie, nous présentons les caractéristiques extraites de la base de données du CNED pour illustrer l'indicateur de régularité. Elles sont comme suit :

— $progressAction_{t_i}$: c'est la caractéristique qui démontre s'il y a eu un avancement en nombre d'actions effectués sur la matière en question par rapport à la semaine t_{i-1}.
— $progressExamen_{t_i}$: c'est la caractéristique qui démontre s'il y a eu un avancement en nombre d'examens de ma matière en question soumis par rapport à la semaine t_{i-1}.
— $nbreDifferentJourConnexion_{t_i}$: c'est la caractéristique qui calcule le nombre de jours de connexion de l'élève entre la semaine t_{i-1} et t_i

A.5 Les caractéristiques de l'indicateur de réactivité

Dans cette partie, nous présentons les caractéristiques extraites de la base de données du CNED pour illustrer l'indicateur de réactivité. Elles sont comme suit :

— $premiereConnexion$: différence en nombre de jours entre la date de début d'activité et la date de la première connexion à la plate-forme.
— $soumissionDevoir1$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 1er devoir.
— $soumissionDevoir2$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 2ème devoir.
— $soumissionDevoir3$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 3ème devoir.
— $soumissionDevoir4$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 4ème devoir.
— $soumissionDevoir5$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 5ème devoir.
— $soumissionDevoir6$: la différence en nombre de jours entre la date de début d'activité et la date du dépôt du 6ème devoir.

Annexe B

Tableau de bord avec les indicateurs d'apprentissage identifiés

Cette annexe présente l'utilisation des indicateurs d'apprentissage identifiés dans la 1ère contribution pour le développement d'un tableau de bord dédié aux enseignants. Pour le calcul de ces indicateurs, ce tableau exploite les données du CNED de la classe 3ème inscrits en physique-chimie.
Ce tableau de bord se compose de deux vues principales : un vue de toute la classe et une vue individuelle pour chaque élève.

B.1 Vue de toute la classe

La Figure B.1 représente la vue globale de l'activité de toute la classe suivie par l'enseignant. Cette vue comporte 3 principaux graphes montrant chacun un type des actiités des apprenants.



FIGURE B.1 – Le tableau de bord vue de classe

La Figure B.2 montre l'évolution de trois indicateurs d'engagement, régularité et performance dans le temps. Chacun de ces indicateurs est calculé comme suit :

— engagement : c'est le nombre d'actions sur la plate-forme par semaine
— régularité : c'est le nombre de jours actifs par semaine
— performance : c'est la moyenne de notes.

À chaque semaine de prédiction, chacune de ces valeurs correspond à la moyenne de la classe.

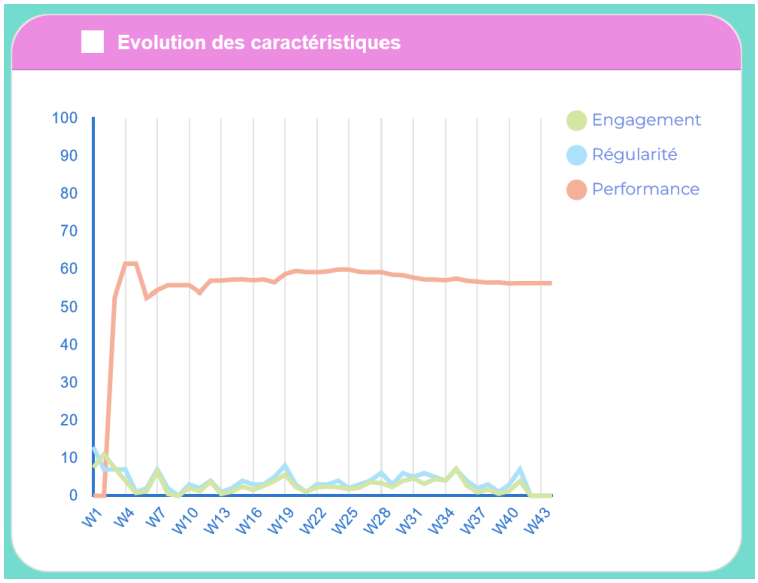


FIGURE B.2 – Évolution de trois indicateurs d'apprentissage pour toute la classe

La Figure B.3 illustre les pouls d'actions pour chaque élèves. Pour chaque élève, nous avons une représentation de son nombre d'actions entre deux périodes choisies. Toute action est représentée par un point. Dans les cas où plusieurs actions sont proches dans le temps, elles sont regroupées pour former des ellipses.

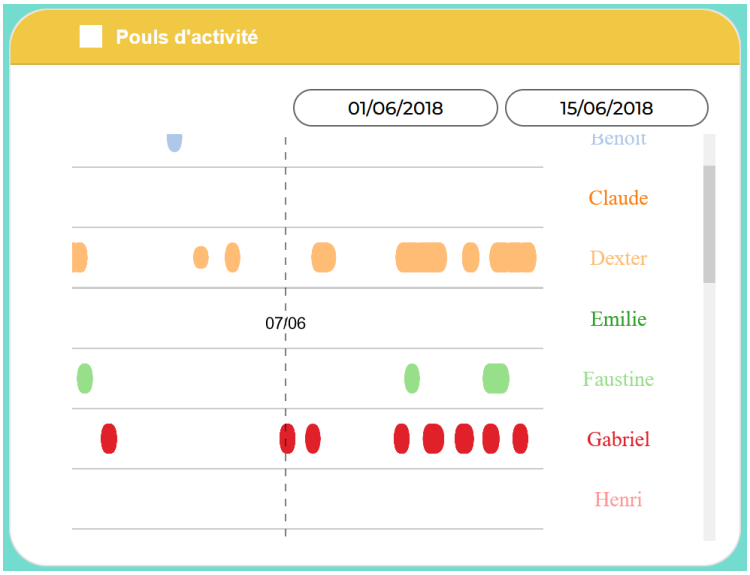


FIGURE B.3 – Pouls d'activité des élèves pendant une période choisie

La Figure B.4 illustre la proximité éducative des élèves en considérant l'indicateur d'engagement. C'est une représentation du nombre d'actions de chaque élève par rapport à la moyenne de toute la classe.

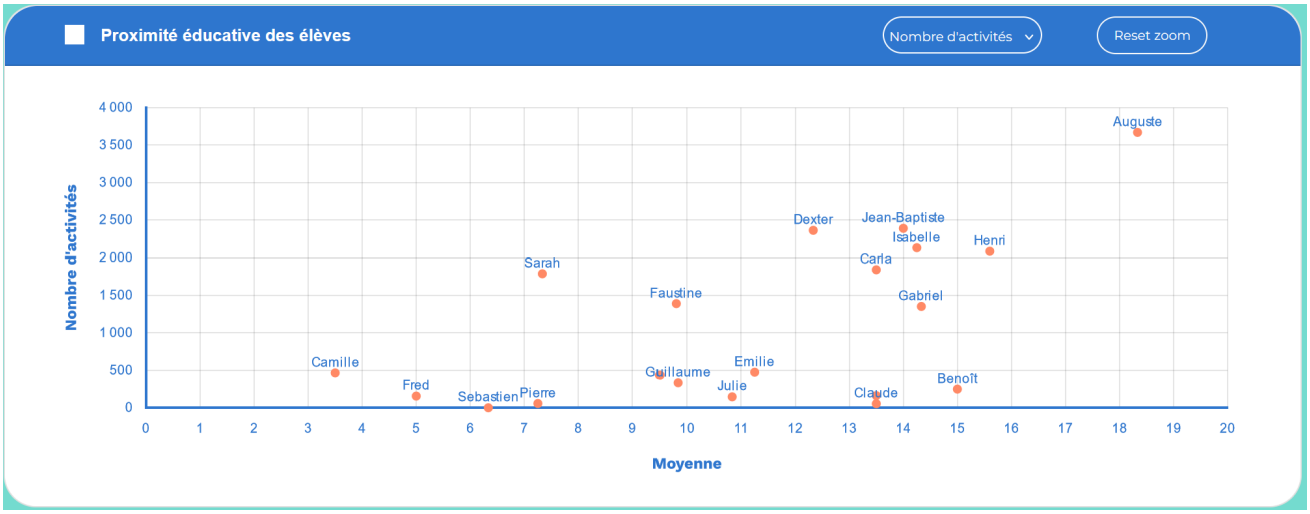


FIGURE B.4 – Proximité éducative des élèves- Indicateur d'engagement-

La Figure B.5 illustre la proximité éducative des élèves en considérant l'indicateur de la réactivité. C'est une représentation de la réactivité de chaque élève par rapport à la moyenne. La réactivité est égale au nombre de jours avant la première connexion.

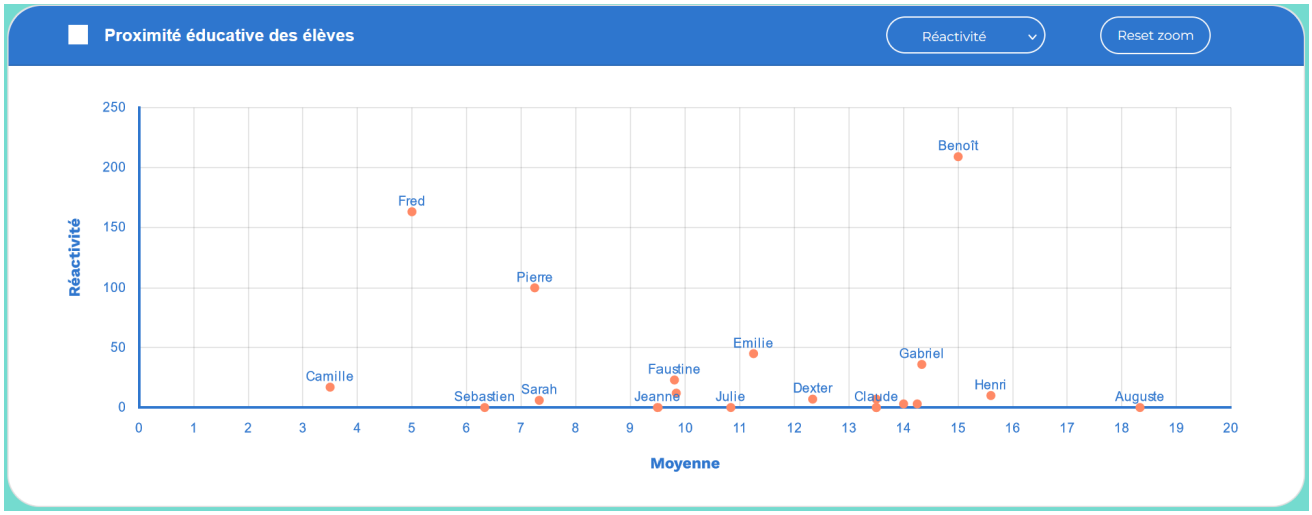


FIGURE B.5 – Proximité éducative des élèves -Indicateur de réactivité-

B.2 Vue d'un élève

La Figure B.6 représente la vue qui s'affiche à l'enseignant quand il choisit un élève donné.



FIGURE B.6 – Le tableau de bord vue de classe

La Figure B.7 montre l'évolution temporelle de trois indicateurs d'engagement, régularité et performance d'un élève. Ces indicateurs utilisent la même nature de données que la Figure B.2.

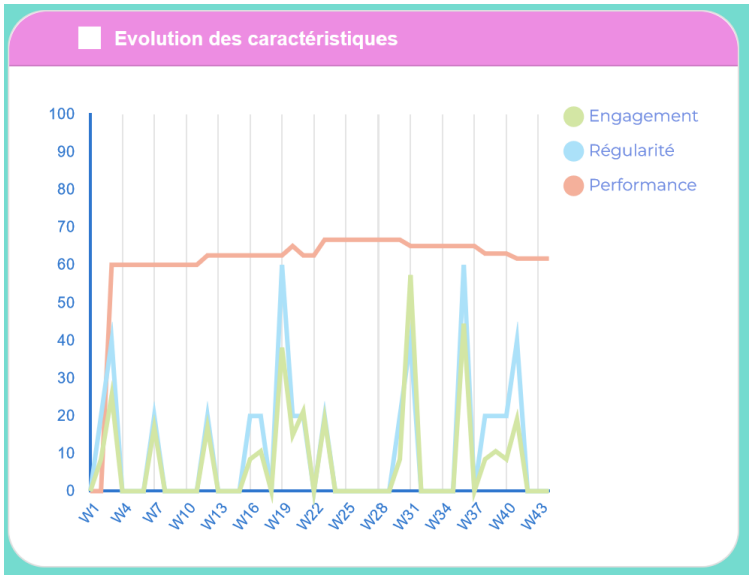


FIGURE B.7 – Évolution de trois indicateurs d'apprentissage pour toute la classe

La Figure B.8 est une représentation des prédictions hebdomadaires pour un élève.

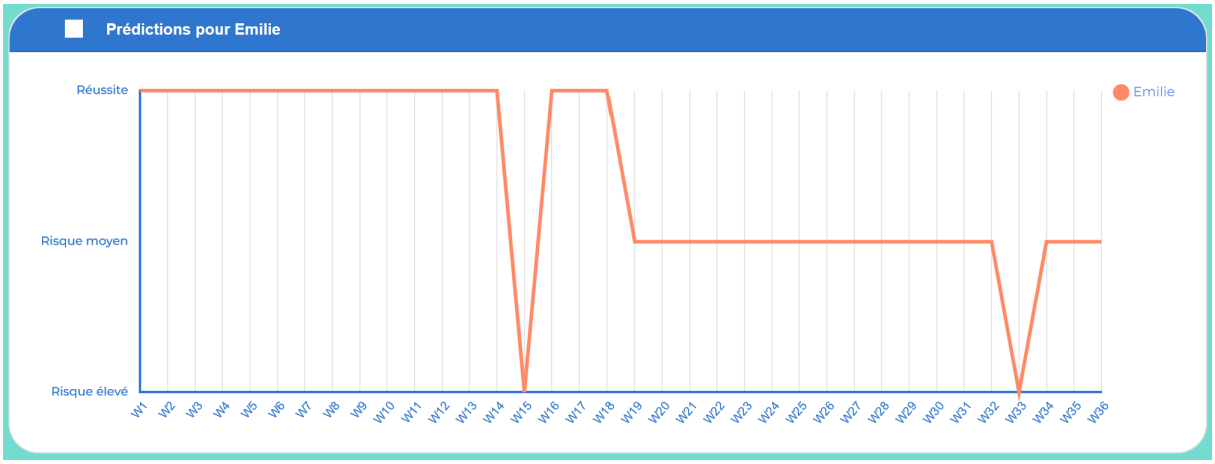


FIGURE B.8 – Évolution de trois indicateurs d'apprentissage pour toute la classe

La Figure B.9 est une représentation de l'évolution de la moyenne des indicateurs (engagement, régularité et performance) pour un élève en comparaison avec la moyenne de toute la classe avec ces indicateurs.

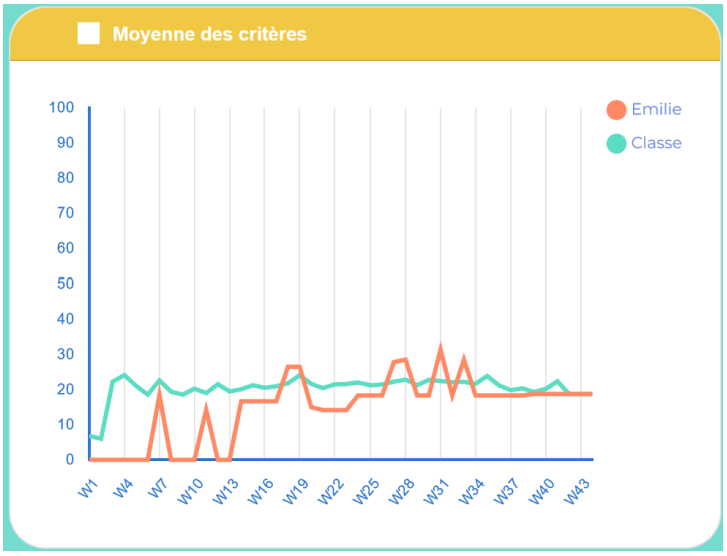


FIGURE B.9 – Évolution de trois indicateurs d'apprentissage pour toute la classe