UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

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. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

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

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Modélisation prédictive dans les environnements éducatifs : vers des approches fiables et adaptatives

THÈSE

présentée et soutenue publiquement le 05 mars 2026

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention Informatique)

par

Ikram Cherifa GAGAOUA

Composition du jury

Présidente :	Malika SMAIL-TABBONE	Professeure, Université de Lorraine
Rapporteurs :	Sébastien IKSAL	Professeur, Université du Mans
	Guillaume CLEUZIOU	Professeur, Université d'Orléans
Directrice de thèse :	Armelle BRUN	Professeure, Université de Lorraine
Invites :	Anne BOYER	Professeure retraitée, Université de Lorraine
	Chahrazed LABBA	Maîtresse de Conférences, Université de Lorraine
	François THIOUNN	Head of Product, Rise Up

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

Mis en page avec la classe thesul.

Remerciements

Cette thèse est l'aboutissement de plusieurs années de recherche, mais aussi d'un parcours humain fait de rencontres, de soutiens et de moments partagés. Je souhaite adresser ici mes plus sincères remerciements à toutes les personnes qui ont contribué, de près ou de loin, à cette aventure.

Je remercie tout d'abord les rapporteurs de cette thèse, Sébastien Iksal et Guillaume Cleuziou, pour l'attention qu'ils ont portée à ce travail et pour la richesse de leurs remarques. Je remercie également les membres du jury pour l'intérêt qu'ils ont manifesté pour mes travaux. Un remerciement particulier à Malika Smail-Tabbone, qui a accompagné cette thèse durant ces quatre années et dont les encouragements ont été précieux.

Je tiens à remercier Armelle Brun, Anne Boyer et Chahrazed Labba pour leur accompagnement, leurs conseils et les nombreux échanges scientifiques qui ont contribué à faire mûrir mon esprit de chercheuse. Je remercie également les membres de l'équipe BIRD que j'ai eu la chance de croiser au fil de ces années, pour leur accueil, leurs échanges stimulants et les cafés partagés lors de mes passages à Nancy.

Cette thèse s'est également construite dans le cadre de ma collaboration avec Domoscio et Rise Up. Je remercie tout particulièrement Benoît Praly, qui a proposé le sujet de cette thèse et l'a défendu à mes côtés dès ses débuts. Merci également à Arnaud et Guillaume Blachon, qui ont permis au projet de continuer à avancer malgré les changements liés au rachat de l'entreprise, ainsi qu'à Ségolène Carniel pour sa patience et son aide précieuse dans la gestion des nombreux sujets administratifs liés à cette thèse CIFRE. Je remercie également François Thiounn, mon manager durant ces quatre années, pour sa confiance et son accompagnement, ainsi que mes collègues Théo Bocquelet, Onur Polat et Dorian Guitel pour leur soutien et la bienveillance du cadre de travail. Un remerciement particulier à Camille Duchasseint et Louise Michel pour leur gentillesse et les moments de partage qui ont rendu cette aventure encore plus humaine.

Je souhaite également remercier ma famille, qui a toujours été un pilier tout au long de ce parcours. Merci à mes parents Farida et Mourad et à mon frère Rachid, pour leur soutien indéfectible, leur confiance et leurs encouragements constants. J'adresse également mes remerciements à Taha Aboulkacem ainsi qu'à ma belle-famille et mes belle-soeurs Khadija, Alaa et Aicha pour leur présence et leur soutien.

Je veux adresser un remerciement très spécial à Fatima Radja Khadar, qui a toujours été là, dans les moments faciles comme dans les moments plus difficiles. Ta présence, ton écoute et ton soutien ont compté bien plus que tu ne l'imagines.

Enfin, je souhaite remercier les deux personnes qui ont partagé mon quotidien pendant ces années de thèse. À Mohamed Ould Sad Saoud, mon conjoint, qui m'a accompagnée dans cette aventure avec patience, soutien et amour. Merci d'avoir été présent dans les moments de doute comme dans les moments de joie, et d'avoir toujours cru en moi.

Et à ma fille Mouna, qui a grandi en même temps que cette thèse. Merci pour ta présence pas toujours calme lors de mes longues soirées de rédaction. Ta lumière et tes sourires ont été une source infinie de motivation.

C'est grâce à vous tous que ce travail a pu voir le jour.

Résumé

Les environnements d'apprentissage numériques génèrent aujourd'hui une grande quantité de traces, permettant d'entraîner des modèles prédictifs capables d'accompagner les apprenants tout au long de leur parcours. Pourtant, ces modèles restent confrontés à plusieurs défis qui conditionnent leur utilisation en situation réelle. Cette thèse se consacre à l'étude de ces défis, et propose des approches visant à améliorer la qualité et la fiabilité des prédictions produites dans des environnements d'apprentissage virtuels.

L'objectif principal est de développer des modèles capables de suivre la dynamique réelle des traces d'apprentissage tout en renforçant la fiabilité et la confiance accordée aux prédictions. Pour y parvenir, la thèse s'appuie sur quatre dimensions complémentaire. La frugalité désigne la capacité à limiter l'usage des données et des ressources computationnelles, en ne mobilisant que les informations réellement nécessaires au fonctionnement du modèle. Elle vise à réduire la complexité tout en préservant la capacité prédictive. La dimension suivante est l'estimation de l'incertitude, qui consiste à mesurer la confiance associée à chaque prédiction. Elle permet de repérer les situations où le modèle devient fragile, où les décisions doivent être interprétées avec prudence, et où des actions pédagogiques adaptées peuvent être envisagées.

Nous nous intéressons ensuite à la sélection dynamique des caractéristiques, qui correspond au processus d'identification, à chaque instant, dans un objectif de frugalité, des variables les plus informatives pour la prédiction. Cette sélection évolue dans le temps afin de s'ajuster aux dérives comportementales, à l'arrivée de nouvelles variables et au déséquilibre naturel des classes. Enfin, nous nous penchons sur l'apprentissage incrémental déclenché, qui permet de mettre à jour le modèle uniquement lorsque les données révèlent un changement significatif. Cette approche frugale, évite les réentraînements inutiles, maintient la stabilité du modèle et contribue à l'efficacité computationnelle recherchée.

Les expérimentations menées sur deux jeux de données éducatives réelles montrent que les approches proposées permettent d'améliorer la sensibilité des modèles aux cas de décrochage tout en réduisant significativement la complexité et le volume de caractéristiques mobilisées. Ces résultats confirment la pertinence d'un apprentissage sélectif et adaptatif, capable de maintenir des performances élevées dans des environnements évolutifs tout en limitant la consommation de ressources. Cette recherche contribue ainsi à la conception de modèles prédictifs plus fiables et économes, ouvrant la voie à une intelligence artificielle éducative plus responsable, mieux calibrée et véritablement centrée sur les besoins des apprenants.

Mots-clés: intelligence artificielle en éducation, fiabilité des modèles prédictifs, incertitude des prédictions, frugalité, sélection dynamique des caractéristiques, apprentissage incrémental, dérive conceptuelle, personnalisation des apprentissages, données éducatives.

Abstract

Virtual learning environments now produce large volumes of trace data, enabling the development of predictive models that can support learners throughout their journeys. Despite this potential, such models still face several challenges that limit their effective deployment in real-world settings. This thesis examines these challenges and introduces methodological approaches designed to improve the quality and reliability of predictions in virtual learning environments.

The main objective is to build models that can follow the real temporal dynamics of learning traces while strengthening both the reliability of predictions and the trust that can be placed in them. To do so, the thesis is structured around four dimensions. Frugality refers to the ability to limit the use of data and computational resources by relying only on what that is genuinely necessary. Its goal is to reduce complexity while preserving predictive power. The following dimension is uncertainty estimation, which focuses on quantifying the confidence associated with each prediction. This makes it possible to identify situations where the model becomes fragile, where outputs should be interpreted with caution, and where tailored pedagogical actions may be required.

The thesis then explores dynamic feature selection, a process that identifies, at every moment, the most informative variables for prediction with a frugal mindset. This selection adapts over time to account for feature drift, the arrival of new variables, and the natural imbalance between classes. Finally, the work investigates triggered incremental learning, which updates the model only when the incoming data indicate a meaningful change. This frugal mechanism avoids unnecessary retraining, maintains model stability, and contributes to the computational efficiency sought.

Experiments conducted on two real-world educational datasets show that the proposed approaches increase the model's sensitivity to dropout cases while significantly reducing complexity and the number of features required. These findings confirm the value of selective and adaptive learning approaches that maintain strong predictive performance in evolving environments while limiting resource consumption. Overall, this research contributes to the development of more reliable and efficient predictive models, opening the way toward a more responsible form of educational artificial intelligence, better calibrated, sustainable, and genuinely centred on learners' needs.

Keywords: artificial intelligence in education, predictive model reliability, predictive uncertainty, frugality, dynamic feature selection, incremental learning, feature drift, personalized learning, educational data.

Table des matières

Introduction **1**

1 Contexte général 1
2 Contexte industriel 3
3 Problématique et questions de recherche 4
4 Objectifs et contributions 7
4.1 Intégrer fiabilité, frugalité et incertitude dans la modélisation des apprenants 7
4.2 Un algorithme de sélection dynamique de caractéristiques pour données déséquilibrées et temporelles 8
4.3 Un mécanisme de déclenchement dynamique d'apprentissage incrémental 9
5 Organisation de la thèse 10

2 État de l'art **11**

Introduction 11
2.1 Systèmes d'apprentissage en ligne 13
2.1.1 Définition 13
2.1.2 Traces numériques et potentiel analytique 14
2.1.3 Modélisation des apprenants à partir de traces 15
2.1.4 Prédiction dans les environnements virtuels d'apprentissage 17
2.1.5 Structuration temporelle des données 19
2.1.6 Discussion 20
2.1.7 Synthèse de la section 21
2.2 Fiabilité et incertitude des modèles d'apprentissage automatique 23
2.2.1 Définition de la fiabilité 23
2.2.2 Évaluer la fiabilité : apports de la littérature 23
2.2.3 Définition de l'incertitude 24
2.2.4 Approches probabilistes et bayésiennes 25
2.2.5 Ensembles de modèles (Deep Ensembles) 26
2.2.6 Métriques d'incertitude 27
2.2.7 Applications en éducation 27

2.2.8 Discussion . . . 28
2.3 Frugalité des modèles et des données . . . 29
2.3.1 Définition . . . 29
2.3.2 Axes de frugalité . . . 29
2.3.3 Réduction des données nécessaires . . . 29
2.3.4 Réduction des ressources d'entraînement . . . 30
2.3.5 Modèles légers et déployables . . . 31
2.3.6 Applications en éducation . . . 31
2.3.7 Discussion . . . 31
2.4 Sélection dynamiques des caractéristiques . . . 33
2.4.1 Définition . . . 33
2.4.2 Classification des caractéristiques . . . 33
2.4.3 Panorama des approches existantes . . . 34
2.4.4 Limites identifiées . . . 36
2.4.5 Applications de la sélection dynamique en éducation . . . 37
2.4.6 Discussion . . . 37
2.5 Apprentissage incrémental . . . 39
2.5.1 Définition . . . 39
2.5.2 Limites et défis actuels . . . 40
2.5.3 Applications en contexte éducatif . . . 40
2.5.4 Discussion . . . 41
Conclusion de l'état de l'art . . . 42

3 Jeux de données de la littérature en éducation 43
3.1 Jeux de données de référence en EDM et LA . . . 43
3.2 Jeux de données choisies . . . 45
3.2.1 Problématique étudiée : la prédiction du décrochage . . . 45
3.3 OULAD (Open University Learning Analytics Dataset) . . . 46
3.3.1 Présentation générale et contexte institutionnel . . . 46
3.3.2 OULAD dans la littérature . . . 47
3.3.3 Organisation et schéma relationnel . . . 47
3.3.4 Cours sélectionné . . . 47
3.3.5 Statistiques descriptives du module CCC (2014J) . . . 49
3.4 MoocCN (XuetangX) . . . 55
3.4.1 Présentation générale et origine . . . 55
3.4.2 MoocCN dans la littérature . . . 56
3.4.3 Structure et format des données . . . 57
3.4.4 Cours sélectionné . . . 57

3.4.5 Statistiques descriptives du cours sélectionné 58
3.5 Synthèse 63

4 Métrique temporelle d'incertitude et modélisation frugale pour des prédictions fiables 65
4.1 Introduction 65
4.2 Incertitude des prédictions 66
4.2.1 Interprétation en contexte éducatif 66
4.2.2 Calibration des seuils de variance et d'entropie 66
4.2.3 Une métrique temporelle : le TWPUS 68
4.2.4 Profils d'évolution de l'incertitude 69
4.3 Frugalité des données 71
4.3.1 Le modèle FEP 71
4.3.2 Frugalité guidée par la fiabilité des prédictions intermédiaires 73
4.3.3 Frugalité guidée par l'incertitude 74
4.4 Expérimentation 75
4.4.1 Expérimentation sur la frugalité et la fiabilité (modèle FEP) 75
4.4.2 Expérimentation sur l'incertitude et la fiabilité 77
4.5 Résultats 78
4.5.1 Fiabilité des prédictions 78
4.5.2 Frugalité 81
4.5.3 Incertitude et fiabilité 82
4.6 Discussion 88
4.7 Conclusion du chapitre 91

5 Sélection dynamique des caractéristiques pertinentes pour la prédiction 93
5.1 Introduction 94
5.2 Méthodologie 95
5.2.1 Définition du problème 95
5.2.2 Présentation détaillée de l'algorithme CI-DFS 96
5.2.3 Évaluation de la pertinence : *check_ relevance* 98
5.2.4 Détection de la redondance : *check_ redundancy* 100
5.2.5 Gestion du drift temporel : *manage_ drift* 101
5.2.6 Vue globale et complémentarité des composants 103
5.3 Métriques d'évaluation 104
5.3.1 Méthodes comparatives : SAOLA et baseline sans sélection 105
5.4 Évaluation sur le jeu de données OULAD 107
5.4.1 Présentation du jeu de données 107

5.4.2 Construction temporelle des caractéristiques . . . 107
5.4.3 Préparation du jeu d'entraînement et de test . . . 108
5.4.4 Modèle de prédiction utilisé . . . 108
5.4.5 Paramétrage de CI-DFS . . . 108
5.4.6 Paramétrage des modèles baseline . . . 109
5.5 Évaluation sur le jeu de données MoocCN . . . 109
5.5.1 Présentation du jeu de données MoocCN . . . 109
5.5.2 Construction temporelle des caractéristiques . . . 109
5.5.3 Spécificités du protocole expérimental . . . 110
5.5.4 Paramétrage de CI-DFS sur MoocCN . . . 110
5.5.5 Paramétrage des comparateurs . . . 110
5.6 Résultats . . . 111
5.6.1 Comparaison globale des algorithmes . . . 111
5.6.2 Ablation des composants de CI-DFS . . . 121
5.6.3 Analyse qualitative des caractéristiques . . . 124
5.7 Discussion . . . 126
5.7.1 Interprétation globale des résultats . . . 126
5.7.2 Apport de chaque composant . . . 127
5.7.3 Limites et perspectives . . . 127
5.8 Conclusion du chapitre . . . 129

6 Apprentissage incrémentale et déclenchement adaptatif 131
6.1 Introduction . . . 131
6.2 Méthodologie . . . 132
6.2.1 Vue d'ensemble du cadre proposé . . . 133
6.2.2 Rôle et gestion de la mémoire tampon . . . 133
6.2.3 Description détaillée de l'algorithme . . . 134
6.2.4 Sélection dynamique des caractéristiques . . . 135
6.2.5 Déclenchement de la mise à jour du modèle . . . 136
6.2.6 Mise à jour incrémentale du modèle . . . 137
6.2.7 Métriques et méthodes d'évaluation . . . 138
6.3 Évaluation expérimentale . . . 139
6.3.1 Jeu de données OULAD . . . 139
6.3.2 Jeu de données MoocCN (XuetangX) . . . 139
6.3.3 Modèle prédictif et paramétrage . . . 140
6.4 Résultats . . . 141
6.4.1 Résultats sur le jeu de données MoocCN . . . 141
6.4.2 Résultats sur le jeu de données OULAD . . . 142

6.4.3 Résumé comparatif 144
6.5 Discussion 144
6.5.1 Performance prédictive et sensibilité aux décrocheurs 145
6.5.2 Réduction du coût d'adaptation 145
6.5.3 Robustesse face à la variabilité temporelle 145
6.5.4 Limites et perspectives 145
6.6 Conclusion du chapitre 147

Conclusion générale **149**
1 Retour sur la problématique et les objectifs 149
2 Synthèse des contributions de la thèse 149
3 Limites de la recherche 151
4 Perspectives et pistes futures 151

Bibliographie **153**

Table des figures

2.1 Organisation de l'état de l'art 13
2.2 Composants d'un modèle d'apprenant 16
2.3 Typologie des caractéristiques selon leur pertinence et redondance [170] 34

3.1 Schéma relationnel d'OULAD 48
3.2 Visualisation de la répartition des statuts finaux dans CCC (2014J) 50
3.3 Interactions hebdomadaires dans le module CCC (2014J) 50
3.4 Répartition du nombre total de clics par étudiant selon le statut (Décrocheurs vs Non-Décrocheurs) 51
3.5 Participation hebdomadaire aux évaluations (Décrocheurs vs Non-Décrocheurs) . 52
3.6 Score moyen hebdomadaire aux évaluations (Décrocheurs vs Non-Décrocheurs) . 52
3.7 Distribution des dates d'inscription des étudiants (CCC 2014J) 53
3.8 Répartition du plus haut niveau d'éducation déclaré (CCC 2014J) 54
3.9 Heatmap des types d'activités consultées par semaine dans le module CCC (2014J) 55
3.10 Répartition des classes (Décrocheurs vs Non-Décrocheurs) pour le jeu d'entraînement (gauche) et le jeu de test (droite). 58
3.11 Distribution du nombre total de clics par étudiant pour le jeu d'entraînement (gauche) et le jeu de test (droite). 59
3.12 Clics moyens par étudiant et par semaine pour les décrocheurs et les non-décrocheurs (jeu d'entraînement à gauche, jeu de test à droite). 60
3.13 Répartition du nombre total de clics par étudiant selon la classe (train à gauche, test à droite). 61
3.14 Évolution du nombre d'étudiants ayant soumis une évaluation par semaine (train à gauche, test à droite). 61
3.15 Répartition hebdomadaire des types d'activités dans les jeux d'entraînement (gauche) et de test (droite) 62

4.1 Workflow du modèle FEP 72
4.2 Workflow des modèles SS-MS et FEP 75
4.3 Illustration du principe d'évaluation intermédiaire comme métrique de fiabilité. . 76
4.4 Exactitude des modèles au fil des semaines (OULAD). 79
4.5 Évolution de l'ESS des modèles au fil des semaines (OULAD). 79
4.6 Exactitude des modèles au fil des semaines (MOOC-CN). 80
4.7 Évolution de l'ESS des modèles au fil des semaines (MOOC-CN). 81
4.8 Évolution de l'entropie moyenne par profil d'étudiants (OULAD). 83
4.9 Évolution de la variance moyenne par profil d'étudiants (OULAD). 83
4.10 Courbe de calibration entre entropie et probabilité observée (OULAD). 84

4.11 Taux d'erreur selon l'indicateur d'incertitude au fil des semaines (OULAD). . . . 85
4.12 Évolution de l'entropie moyenne par profil d'étudiants (MOOC-CN). 85
4.13 Évolution de la variance moyenne par profil d'étudiants (MOOC-CN). 86
4.14 Courbe de calibration entre entropie et probabilité observée (MOOC-CN). 86
4.15 Taux d'erreur selon l'indicateur d'incertitude au fil des semaines (MOOC-CN). . 87
4.16 Transitions entre profils d'incertitude au fil des semaines (MOOC-CN). 88

5.1 Exactitude hebdomadaire sur OULAD. 112
5.2 Sensibilité hebdomadaire (Rappel). 113
5.3 Spécificité hebdomadaire. 113
5.4 Nombre de caractéristiques sélectionnées. 114
5.5 Temps de sélection 115
5.6 Temps d'entraînement du modèle. 116
5.7 Sensibilité sur MoocCN (MoocCN) 117
5.8 Exactitude hebdomadaire sur MoocCN (MoocCN) 117
5.9 Spécificité sur MoocCN (MoocCN) 118
5.10 Nombre de caractéristiques sélectionnées sur MoocCN 119
5.11 Temps de sélection sur MoocCN (MoocCN) 120
5.12 Temps d'entraînement sur MoocCN (MoocCN) 120
5.13 Sensibilité sur MoocCN(MoocCN) avec ablation 122
5.14 Sensibilité sur OULAD avec ablation 122
5.15 Nombre de caractéristiques sur MoocCN(MoocCN) avec ablation 123
5.16 Nombre de caractéristiques sur OULAD avec ablation 123
5.17 Temps de sélection sur OULAD avec ablation. 124
5.18 Temps de sélection sur MoocCN(MoocCN) avec ablation 125

6.1 Évolution de l'exactitude hebdomadaire (MoocCN) 141
6.2 Évolution de la sensibilité hebdomadaire (MoocCN) 142
6.3 Évolution de l'exactitude hebdomadaire (OULAD) 143
6.4 Évolution de la sensibilité hebdomadaire (OULAD) 143

Liste des tableaux

2.1 Typologie des environnements virtuels d'apprentissage 14
2.2 Typologie des approches de modélisation des apprenants 16
2.3 Comparatif critique des algorithmes supervisés pour la prédiction dans les VLEs 19
2.4 Comparaison des principales approches de sélection dynamique des caractéristiques 34

3.1 Résumé des fichiers composant OULAD et leur contenu 48
3.2 Répartition des statuts finaux dans CCC (2014J) 49
3.3 Structure des attributs contenus dans les fichiers de logs de MoocCN 57
3.4 Répartition des décrocheurs et non-décrocheurs dans les jeux d'entraînement et de test 59

4.1 Évaluation de la frugalité sur les jeux de données *OULAD* et *MOOC-CN* 82

5.1 Résumé des métriques utilisées pour l'évaluation de CI-DFS 105
5.2 Performances moyennes sur les deux jeux de données (semaines combinées) . . . 111
5.3 Impact de chaque composant de CI-DFS sur les performances 121

6.1 Variables d'entrée, de sortie intermédiaire de l'algorithme 5 134
6.2 Configuration du modèle prédictif 140
6.3 Seuils de déclenchement utilisés pour les mises à jour incrémentales 140
6.4 Paramétrage de la méthode CI-DFS 141
6.5 Résultats comparatifs finaux sur MoocCN / OULAD 144

Introduction

1 Contexte général

Apprendre est un besoin humain fondamental, une quête de compréhension, de maîtrise, parfois de transformation [118, 50]. Depuis des siècles, les dispositifs pour transmettre le savoir ont évolué au rythme des innovations culturelles et techniques. L'invention de l'imprimerie, au milieu du XVe siècle, a marqué un premier tournant majeur, en rendant possible la diffusion massive et stable du savoir écrit [59]. L'arrivée du numérique, quelques siècles plus tard, a bouleversé encore plus radicalement les conditions d'accès à l'information et les manières d'apprendre [99].

Aujourd'hui, apprendre ne se limite plus à l'école ou à l'université. On apprend au travail, en ligne, sur mobile, parfois seul, parfois en réseau [153]. L'apprentissage peut être formel (encadré par une institution), informel (dans la vie quotidienne), ou auto-organisé dans une logique d'autodidaxie. Il peut s'appuyer sur des contenus structurés comme sur des interactions sociales, des forums ou des expériences pratiques. Cette diversité des formes d'apprentissage dessine les contours d'une nouvelle culture numérique du savoir, où chacun est potentiellement acteur de son propre parcours [153, 36, 99].

Les environnements informatiques pour l'apprentissage humain (EIAH), qu'ils soient entièrement numériques ou hybrides, incarnent cette transformation. Ils recouvrent une large gamme de dispositifs allant des plateformes d'apprentissage en ligne LMS (*Learning Management System*), comme Moodle, aux MOOCs (Massive Open Online Courses), en passant par les outils mobiles, les solutions hybrides présentiel/distanciel en ligne, les assistants intelligents ou encore les environnements immersifs. Ces EIAH rendent possible une multiplicité de scénarios d'utilisation [139, 35]. Leurs fonctions dépassent la simple mise à disposition de ressources : elles permettent aussi d'organiser les parcours, de déclencher des recommandations, d'adapter les contenus, et de suivre finement les progressions [3, 76].

Ce faisant, les EIAH ne se contentent pas de diffuser du contenu : ils produisent également des données. À chaque interaction : ouverture de ressource, clic sur un lien, soumission d'un devoir, participation à une discussion, etc., une trace est générée. Ces traces sont horodatées, contextualisées (groupées par session, par module, etc.), et stockées de manière structurée dans les bases du système. On peut en extraire des indicateurs variés : nombre de connexions, durée de session, type d'activité effectuée, score obtenu, fréquence de participation, mais aussi métadonnées techniques telles que le type d'appareil utilisé, la localisation approximative, ou les configurations linguistiques. Ces traces permettent de documenter le comportement d'apprentissage de manière longitudinale et de construire des représentations de l'évolution possible d'un individu dans son parcours [139, 35, 76].

Cette production de données a conduit à la naissance d'un nouveau domaine : la modélisation des apprenants. Celui-ci consiste à construire un modèle d'apprenant explicite, dynamique de l'état de l'apprenant dans son interaction avec le système [6]. Ce modèle peut inclure des dimensions cognitives (niveau de maîtrise, vitesse d'acquisition), motivationnelles (régularité d'engagement, décrochage potentiel), ou comportementales (stratégies d'exploration, schémas d'erreur) [83]. L'objectif est de fournir une base conceptuelle pour la personnalisation de l'apprentissage, qui permet d'adapter, par exemple les contenus, les rétroactions, ou les modalités d'évaluation [156].

Les travaux dans le champ de l'intelligence artificielle en éducation ont montré l'intérêt de ces modèles, notamment pour prédire le succès ou l'échec d'un apprenant, détecter une baisse d'engagement, ou recommander une ressource adaptée [125, 81] via des modèles prédictifs. La modélisation des apprenants fournit ainsi le cadre conceptuel et les variables nécessaires à l'entraînement de ces modèles prédictifs, en traduisant les comportements observés en indicateurs exploitables pour l'apprentissage automatique [83]. Ces prédictions s'appuient généralement sur des modèles statistiques ou d'apprentissage automatique, supervisés ou non, entraînés à partir de traces. Certains modèles d'apprentissages automatiques sont statiques, c'est-à-dire qu'ils restent inchangés après un entraînement initial. D'autres, dits dynamiques ou en ligne, s'adaptent aux nouvelles données en intégrant progressivement l'évolution des données des apprenants. D'autres encore combinent règles expertes et modèles statistiques dans une approche hybride [26, 171, 150].

Mais au-delà des avancées techniques, un constat traverse l'ensemble de ces approches : la qualité de la modélisation des apprenants dépend étroitement de la qualité des données et pas uniquement de la quantité. Le principe bien connu du "Garbage In, Garbage Out" nous rappelle que même le meilleur algorithme ne saurait produire des résultats pertinents à partir de données incohérentes, incomplètes ou mal contextualisées [65]. La fiabilité du modèle prédictif est donc indissociable du modèle d'apprenant qui l'alimente : un profil mal construit peut entrainer des prédictions biaisées ou incertaines. Les systèmes prédictifs supposent souvent que les données sont abondantes, structurées et cohérentes. Or, dans les EIAH, ces conditions sont rarement réunies. Certains apprenants laissent très peu de traces, en se connectant rarement ou en interagissant peu avec les ressources, ce qui limite la fiabilité de leur profil. D'autres adoptent des comportements atypiques, en suivant des parcours d'apprentissage inhabituels ou irréguliers. Dans les deux cas, la rareté ou l'originalité des données complique l'interprétation par les modèles.

En effet, cette diversité de caractéristiques génère ce que nous appelons des modélisations d'apprenants peu fiables. Le modèle prédictif, privé d'un historique suffisant ou confronté à des comportements atypiques, peine à produire une prédiction fiable [82, 84]. Dans la plupart des systèmes actuels, cette fiabilité n'est ni mesurée, ni signalée, ni prise en compte, engendrant un risque d'erreur ou d'absence de personnalisation [7, 141].

Dans la majorité des travaux, l'amélioration des modèles est recherchée par l'ajout continu de nouvelles données, dans l'idée que plus de données conduit nécessairement à de meilleures performances [151, 130]. Or, cette logique d'accumulation n'apporte pas toujours les gains attendus et peut même entraîner une surconsommation inutile de ressources, ou même une baisse de qualité. Plutôt que de chercher à alimenter uniformément tous les modèles avec toujours plus de données, nous faisons l'hypothèse qu'une approche ciblée, tenant compte de la pertinence et du moment d'intégration des données, permettrait d'améliorer la personnalisation tout en limitant la surconsommation. Dans cette perspective, la frugalité en intelligence artificielle est envisagée comme une approche qui, selon la littérature, vise à restreindre l'usage des données, de la puis-

sance de calcul et de la mémoire, tout en maintenant des performances satisfaisantes [61, 44, 21]. Transposée à la modélisation des apprenants, elle consiste à mobiliser uniquement les données réellement utiles pour améliorer le modèle, en interrogeant leur pertinence, leur sensibilité et leur potentiel informatif, au-delà des seules contraintes de volume ou de coût.

Dans la prochaine section, nous présentons le contexte industriel de cette thèse, en détaillant les enjeux rencontrés par l'entreprise partenaire et la manière dont ils ont conduit à la proposition de cette thèse.

2 Contexte industriel

Cette thèse s'inscrit dans un cadre de recherche appliquée, menée en partenariat avec Domoscio (Rise Up) [1], une entreprise française fondée en 2013 et spécialisée dans les technologies d'apprentissage adaptatif. Située à l'intersection des sciences cognitives, de l'intelligence artificielle et de l'analyse de données éducatives, Domoscio développe des solutions qui visent à personnaliser les parcours d'apprentissage en fonction des besoins de chaque individu. Ces solutions sont conçues pour s'intégrer dans des écosystèmes de formation existants, via des connecteurs standardisés, et sont déployées aussi bien dans l'enseignement initial que dans la formation continue, au sein d'établissements éducatifs ou d'organisations privées.

Réalisée dans le cadre d'une convention CIFRE [2], qui soutient activement l'articulation entre les objectifs scientifiques du projet et ses enjeux industriels concrets. Elle vise à répondre à des problématiques réelles rencontrées par l'entreprise dans le déploiement de ses outils, en particulier autour des limites liées à la disponibilité et à l'exploitation des traces d'apprentissage.

Plusieurs contraintes limitent aujourd'hui le potentiel adaptatif des produits développés par Domoscio. D'une part, l'intégration s'effectue sans accès direct aux bases de données internes des LMS, ce qui restreint la granularité et la richesse des données exploitables. D'autre part, les traces disponibles sont souvent fragmentées entre modules ou sessions, peu contextualisées, et dépourvues de métadonnées explicites. Ces limitations freinent la construction d'une modélisation des apprenants complète et cohérente. Dans ce contexte, la qualité de la personnalisation offerte par Domoscio dépend directement de la qualité des traces disponibles. Lorsque celles-ci sont incomplètes ou peu informatives, le système peine à proposer des ajustements pertinents, ce qui peut conduire à des recommandations sous-optimales, voire à l'absence totale de réponse du modèle.

La problématique industrielle rejoint ainsi les enjeux scientifiques explorés dans cette thèse. Les limitations observées dans l'exploitation des données d'apprentissage mettent en lumière la nécessité de repenser les conditions d'amélioration des modèles prédictifs, en particulier dans les situations où les données sont insuffisantes. Les travaux que nous présentons ici s'inscrivent dans cette perspective, en proposant une approche ciblée, raisonnée et explicable de l'enrichissement des modélisations d'apprenants.

Dans la section suivante, nous détaillons la problématique de recherche ainsi que les questions scientifiques qui structurent ce travail.

1. https ://domoscio.com/fr/
2. La CIFRE (Convention Industrielle de Formation par la Recherche) est un dispositif français qui permet à un·e doctorant·e de réaliser sa thèse en collaboration avec une entreprise, tout en étant financé·e par l'ANRT.

3 Problématique et questions de recherche

Les systèmes éducatifs s'appuient de plus en plus sur des modèles d'intelligence artificielle pour personnaliser les parcours, recommander des ressources ou anticiper les risques de décrochage [12, 128]. Ces modèles sont alimentés par des traces d'interaction, souvent issues de plateformes d'apprentissage [7]. Leur efficacité repose sur une hypothèse forte : les données disponibles sont suffisamment riches, cohérentes et représentatives de l'apprenant, pour permettre une modélisation fiable des apprenants.

Or, la quantité de ces données ne garantit pas à elle seule la qualité des modèles. Même lorsque l'information est abondante et correctement structurée, les prédictions peuvent varier d'un contexte à l'autre ou manquer de stabilité au fil du temps [29, 74]. Cette observation met en évidence la nécessité de distinguer la simple exactitude ponctuelle d'un modèle de sa capacité à produire des résultats reproductibles et robustes dans différentes situations d'apprentissage.

En apprentissage automatique, la fiabilité se définit comme la capacité d'un modèle à produire des prédictions correctes et stables, au-delà de sa seule exactitude ponctuelle [174]. Or, dans un cadre éducatif, les décisions automatisées fondées sur l'intelligence artificielle, recommander une ressource, déclencher une remédiation ou estimer un risque, reposent bien souvent sur des modélisations partielles de l'état de l'apprenant [160, 151].

Il en résulte que, lorsque le système ne dispose pas d'un volume suffisant de données, qu'il n'accède pas aux données les plus pertinentes, ou qu'il est confronté à des comportements atypiques faiblement représentés dans les données d'entraînement, la fiabilité des prédictions peut être compromise sans que cela ne soit explicitement signalée [150]. Ce manque de fiabilité reste le plus souvent invisible : aucune alerte n'est générée et aucun ajustement n'est proposé. Le modèle peut échouer laissant subsister un risque d'erreurs non détectées et de décisions inadaptées. Cette situation soulève un enjeu central : **faut-il améliorer un modèles avec toujours plus de données, ou bien concentrer l'effort sur les cas pour lesquels le modèle échoue à modéliser correctement ?**

Cette thèse s'articule autour de cette question, en posant l'hypothèse suivante :

> ***H.*** *Un accès conditionnel à des données additionnelles pertinentes permet d'augmenter la fiabilité des modèles des apprenants et d'améliorer significativement la personnalisation.*

À partir de cette hypothèse, trois questions de recherche structurent le travail :

RQ1. Comment identifier les apprenants dont la modélisation est non fiable ?

Dans les systèmes prédictifs, et plus largement dans le champ de l'apprentissage automatique, l'incertitude désigne la capacité du modèle à évaluer la fiabilité de ses propres prédictions [146]. Dans un contexte éducatif, cela se traduit par des situations où le système ne peut pas estimer avec certitude le niveau de l'apprenant, son risque d'échec, ou l'utilité d'une ressource donnée.

Cette incertitude, lorsqu'elle n'est pas explicitement mesurée, peut conduire à des décisions inexactes, inéquitables ou à l'absence totale de personnalisation [80, 146].

Nous distinguons deux types d'incertitude :

— **L'incertitude liée aux données *(epistemic uncertainty)* :** elle correspond à l'incertitude due à un manque, un bruit ou une incohérence dans les données disponibles. Elle traduit une méconnaissance partielle du système étudié et peut théoriquement être réduite si l'on collecte davantage de données de meilleure qualité [82]. Dans le domaine de l'éducation, elle se manifeste par exemple lorsque les traces laissées par un apprenant sont trop rares (qui ne se connecte pas assez, ou ne fait pas assez d'actions) ou atypiques (qui fait des actions qui s'éloignent des actions habituelles), rendant difficile la construction d'un profil fiable.

— **L'incertitude liée au modèle *(aleatoric uncertainty)* :** elle correspond à l'incertitude inhérente au phénomène lui-même ou à la capacité limitée du modèle à le représenter. Elle persiste même lorsque les données sont abondantes et de bonne qualité, et reflète une variabilité intrinsèque que le modèle ne peut pas éliminer [82]. Dans le domaine de l'éducation, elle apparaît lorsque le modèle peine à interpréter correctement certains profils d'apprenants, même bien représentés dans les données, en raison de limites structurelles ou méthodologiques.

L'objectif de cette première question est de développer une capacité de détection automatique de l'incertitude, qu'elles viennent des données ou du modèle. Il ne s'agit pas simplement d'évaluer la performance moyenne d'un modèle, mais d'identifier, les cas où la prédiction risque d'être peu fiable. Cette détection de l'incertitude constitue une condition essentielle : elle détermine si et quand une action d'amélioration ciblée du modèle doit être engagée.

RQ2. Quelles données faut-il mobiliser pour améliorer une modélisation peu fiable ?

Lorsque la modélisation d'un apprenant est identifiée comme peu fiable, il devient nécessaire de déterminer comment améliorer sa fiabilité. Dans notre hypothèse, et pour le cas d'une incertitude liée aux données, cette amélioration ne peut reposer sur une collecte massive, indifférenciée ou systématique de données. L'enjeu ici est de renverser la logique classique des systèmes prédictifs, qui partent du principe que toutes les données sont disponibles dès le départ [130]. Au contraire, elle suppose une sélection ciblée, contextualisée et justifiée des caractéristiques *(features)* les plus susceptibles de réduire l'incertitude du modèle.

Le recours à la sélection de caractéristiques s'impose ici comme une approche plus appropriée que d'autres alternatives couramment utilisées. Contrairement à l'ingénierie de caractéristiques (*feature engineering*), qui requiert la création manuelle ou experte de nouvelles variables, une démarche souvent coûteuse, peu généralisable et dépendante du domaine [113], la sélection vise à identifier parmi les variables déjà disponibles celles qui sont réellement utiles pour réduire l'incertitude [173]. Dans les EIAH, cette orientation est particulièrement pertinente puisque les données n'arrivent pas sous forme de corpus statique mais en flux continus, issus des traces générées par les apprenants sur les plateformes éducatives.

La sélection de caractéristiques permet alors de cibler, au fil du temps, les informations les

plus pertinentes pour améliorer la fiabilité du modèle sans alourdir inutilement le système. Cette approche mobilise deux dimensions fondamentales de la collecte de données pour les EIAH : la frugalité et l'adaptativité.

Une approche frugale : prédire avec moins

Dans ce contexte, la **frugalité** ne se réduit pas à une simple économie de ressources techniques ou financières. Elle correspond à une démarche stratégique qui consiste à mobiliser uniquement les caractéristiques apportant une réelle valeur ajoutée pour la prédiction, en écartant celles dont la contribution est faible ou redondante [61, 17].

Une dimension adaptative : sélectionner en fonction du contexte

Cette approche de frugalité devrait être couplée à une capacité d'adaptation dynamique. Les caractéristiques pertinentes, qui apportent une réelle valeur pour le modèle, ne sont pas toujours les mêmes pour tous les apprenants, par ailleurs, une information pertinente à un instant donné (par exemple, les clics sur une première ressource) peut devenir inutile ou trompeuse à un autre moment (si l'apprenant change de comportement), ce phénomène illustre ce que l'on désigne sous le terme de dérive des caractéristiques (*feature drift*) [164]. Inversement, certaines données non disponibles au début (en raison d'une inactivité ou d'un délai) peuvent devenir déterminantes plus tard [140, 15].

Il s'agit donc de concevoir un système capable d'adapter en continu sa stratégie de sélection des caractéristiques, en fonction de l'évolution du comportement apprenant et de la progression de la modélisation. Ce système doit pouvoir se poser les bonnes questions au bon moment :

— Quelle caractéristique est réellement informative pour améliorer la modélisation ?
— Une caractéristique jugée pertinente à un moment donné l'est-elle toujours lorsque les conditions évoluent ?

Construire un tel mécanisme de sélection frugale et adaptative constitue un levier essentiel pour améliorer la qualité du modèle, tout en maîtrisant la quantité de données mobilisées.

RQ3. Comment intégrer ces données dans le modèle sans le rendre instable ?

Une fois qu'un modèle d'un apprenant a été identifié comme peu fiable, et qu'un besoin spécifique de données a été défini, se pose la question de son intégration dans le modèle existant. L'objectif n'est pas de réentraîner l'ensemble du modèle sur toutes les données, ce qui serait coûteux et potentiellement instable, mais de mettre en œuvre une stratégie de mise à jour, maîtrisée et progressive.

Dans cette perspective, nous mobilisons l'approche dite d'**apprentissage incrémental**[3] (*Incremental Learning*), qui désigne une méthode qui permet à un modèle de s'adapter à de nouvelles données au fil du temps, sans oublier ce qu'il a déjà appris, et sans nécessiter un ré-apprentissage complet [154]. Une difficulté centrale réside toutefois dans l'*oubli catastrophique*, phénomène par lequel l'intégration de nouvelles données entraîne la dégradation, voire l'effacement, des connaissances acquises précédemment [88]. La mise à jour incrémentale doit donc veiller à maintenir un équilibre entre plasticité (capacité à intégrer l'information nouvelle) et stabilité (préservation des acquis antérieurs).

3. Appelé également continual learning ou parfois online learning.

Il s'agit ainsi de renforcer la capacité d'un modèle à évoluer, en intégrant les nouvelles caractéristiques, tout en garantissant la stabilité des performances du système [147].

Cette approche implique plusieurs mécanismes complémentaires :
- **Le déclenchement de la mise à jour** : il convient de définir des critères précis pour décider à quel moment intégrer de nouvelles données. Ces critères peuvent reposer sur le niveau d'incertitude, la nouveauté des données disponibles, ou encore leur potentiel informatif.
- **L'intégration contrôlée de l'information** : la mise à jour ne doit pas introduire d'effet de bord. Le modèle doit être en mesure d'absorber les nouvelles données sans dégrader les prédictions précédemment fiables.
- **La validation de l'impact** : toute mise à jour doit être accompagnée d'un mécanisme d'évaluation. Il s'agit de vérifier que l'intégration des nouvelles données a effectivement réduit l'incertitude ou amélioré la performance prédictive, que ce soit sur le plan local (pour le cas ciblé) ou global (à l'échelle du système).

L'apprentissage incrémental avec un déclenchement conditionnel devient ainsi une brique essentielle du cadre que nous proposons : il permet d'améliorer la modélisation pour les apprenants pour qui c'est nécessaire, sans perte de performances de prédiction, ni surcharger inutilement le processus d'apprentissage.

Dans la continuité de ces trois questions de recherche, la section suivante présente les objectifs poursuivis dans cette thèse ainsi que les principales contributions méthodologiques proposées.

4 Objectifs et contributions

Dans cette thèse, nous poursuivons un objectif central : concevoir une approche ciblée et conditionnelle pour améliorer la modélisation des apprenants dans des systèmes prédictifs, dans un contexte où les données disponibles sont souvent de qualité et quantité variable. Nous proposons ainsi un cadre méthodologique dans lequel la mise à jour du modèle ne se déclenche que lorsque cela est jugé nécessaire, en mobilisant uniquement les données jugées pertinentes, et en garantissant la stabilité du modèle de prédiction.

Les contributions de cette thèse s'organisent autour de trois axes principaux, qui correspondent aux trois questions de recherche détaillées précédemment. Chaque axe donne lieu à une proposition méthodologique.

4.1 Intégrer fiabilité, frugalité et incertitude dans la modélisation des apprenants

Cette contribution vise à répondre à plusieurs limites persistantes dans la littérature sur la prédiction en éducation. Tout d'abord, les travaux existants proposent rarement des mécanismes explicites pour évaluer la fiabilité des prédictions au niveau individuel, alors même que la question de la robustesse et de la confiance des modèles est de plus en plus centrale en apprentissage automatique [19, 174, 74].

Ensuite, la frugalité constitue aujourd'hui un enjeu majeur pour les environnements d'apprentissage virtuels (VLE, Virtual Learning Environments). L'abondance de données issues des

traces d'apprenants ne garantit pas leur pertinence, et leur exploitation sans discernement peut conduire à des traitements coûteux et peu efficaces. Adopter une perspective d'IA frugale revient à privilégier une utilisation sélective et raisonnée des informations, en fonction de leur utilité réelle pour la modélisation [44, 21].

Enfin, les modèles éducatifs existants traitent encore de façon insuffisante la question de l'incertitude : les indicateurs internes (variance, entropie) sont étudiés dans la littérature scientifique [94, 4, 70], mais rarement mobilisés comme déclencheurs pour guider l'actualisation sélective des modèles des apprenants.

Face à ces limites, nous proposons deux contributions :

— Un modèle frugal en matière de consommation de données, qui privilégie une utilisation sélective et conditionnelle afin de préserver la fiabilité prédictive sans recourir systématiquement à des données additionnelles. Ce modèle vise également à identifier pour quels apprenants l'ajout de données additionnelles devient nécessaire pour maintenir la qualité des prédictions, prolongeant ainsi les travaux récents sur la frugalité en apprentissage supervisé.
— Un score d'incertitude combiné, intégrant à la fois les signaux internes du modèle (variance, désaccord) et ceux liés aux données (entropie des distributions prédictives), tout en tenant compte de leur évolution temporelle. Ce score permet d'identifier les apprenants les plus à risque d'erreur et de déclencher une mise à jour ciblée au moment opportun.

Ce premier apport permet de poser un critère explicite pour décider quand un modèle nécessite une mise à jour ciblée. Il constitue le déclencheur du reste du processus, ouvrant la voie à des modèles éducatifs à la fois plus robustes, plus économes en données et plus transparents.

4.2 Un algorithme de sélection dynamique de caractéristiques pour données déséquilibrées et temporelles

Cette contribution consiste à proposer un algorithme de sélection dynamique de caractéristiques. Il s'agit de déterminer, dans un contexte donné, quelles données seraient réellement utiles pour lever l'incertitude identifiée et améliorer le modèle. Elle s'inscrit dans une logique de frugalité des données sélectionnées, où la pertinence prévaut sur la quantité : seules les informations jugées informatives pour la situation et le profil considérés sont mobilisées, évitant ainsi un recours excessif à des données non contributives. Les travaux existants en sélection de caractéristiques se heurtent à plusieurs limites lorsqu'ils sont appliqués aux environnements éducatifs numériques. D'une part, la majorité des approches adoptent une logique statique, où les caractéristiques sont choisies une fois pour toutes en amont, sans prise en compte de l'évolution des apprenants ni de la dynamique des données dans le temps [41, 123]. Cela réduit fortement leur pertinence dans les contextes séquentiels comme les VLEs, où de nouvelles traces apparaissent chaque semaine et où certaines perdent progressivement leur valeur prédictive.

D'autre part, les algorithmes de sélection dynamique existants développés dans d'autres domaines (santé, publicité, bio-informatique) s'appuient souvent sur des méthodes coûteuses, comme les approches à base de Markov blanket [161], d'algorithmes évolutionnaires [56] ou de renforcement[164], qui restent peu adaptés à des environnements massifs ou temps réel. Enfin, les méthodes à base d'information mutuelle [124, 169] sont plus efficaces mais présentent deux faiblesses majeures : elles ignorent en grande partie le problème du déséquilibre de classes (fa-

vorisant la classe majoritaire au détriment des classes peu représentées) et elles reposent sur des seuils fixes, qui ne s'adaptent pas aux changements de distribution dans les flux temporels [176, 178].

Une autre difficulté tient à la gestion de la dérive des caractéristiques (*features drift)*, encore rarement prise en compte dans les méthodes de sélection dynamique. Or, dans les environnements éducatifs, la pertinence des caractéristiques évolue fortement au fil du temps : certaines, initialement informatives, perdent progressivement leur valeur prédictive, tandis que d'autres acquièrent de l'importance [172].

S'ajoute à cela le déséquilibre de classes, souvent sous-estimé dans les approches existantes. La sélection de caractéristiques tend naturellement à privilégier la classe majoritaire, ce qui accentue la marginalisation des classes minoritaires. Ce biais est particulièrement problématique dans les EIAH, où les tâches de prédiction d'échecs ou d'abandons concernent précisément des catégories d'apprenants moins représentées, mais cruciales à identifier [97].

Pour répondre à ces limites, nous proposons un algorithme de sélection dynamique de caractéristiques, spécifiquement conçu pour les contraintes suivantes :

— les données évoluent dans le temps (les traces se renouvellent de manière continue)
— toutes les caractéristiques disponibles ne sont pas pertinentes, ce qui impose une sélection dynamique et non redondante
— la performance du modèle dépend de la capacité à intégrer ou à écarter, au bon moment, les caractéristiques réellement informatives.

Notre algorithme sélectionne dynamiquement les caractéristiques pertinentes en combinant :
— une **mesure de pertinence**, adaptée aux données déséquilibrées,
— un **seuil adaptatif**, tenant compte de la distribution temporelle des données,
— un **mécanisme de stabilité de pertinence**, garantissant la continuité et la cohérence des caractéristiques sélectionnées dans le temps.

Cet algorithme permet de maintenir une sélection frugale, contextuelle et réactive, essentielle pour alimenter la mise à jour du modèle sans surcharge inutile.

4.3 Un mécanisme de déclenchement dynamique d'apprentissage incrémental

Enfin, la dernière contribution porte sur l'intégration des données nouvellement sélectionnées. Les approches traditionnelles reposent sur des modèles statiques qui nécessitent un réentraînement complet dès l'arrivée de nouvelles données, ce qui entraîne des coûts computationnels élevés, une latence peu compatible avec les environnements éducatifs en temps réel, et parfois une dégradation de la stabilité [53, 154]. Cette contribution s'inscrit dans une logique de frugalité computationnelle, en favorisant une mise à jour incrémentale et ciblée du modèle, capable d'intégrer de nouvelles informations de manière progressive sans réinitialisation complète. Si l'apprentissage incrémental a connu un développement important en intelligence artificielle [147, 96], il reste encore peu exploité dans le champ éducatif, où seules quelques expérimentations récentes ont montré son potentiel [93, 92]. De plus, les mécanismes de déclenchement existants reposent souvent sur des critères arbitraires ou périodiques, sans s'appuyer sur un signal d'incertitude indiquant qu'une mise à jour est réellement nécessaire [77, 78].

Pour dépasser ces limites, nous proposons un mécanisme de déclenchement dynamique d'apprentissage incrémental, articulé en trois étapes :

— la **détection d'une prédiction incertaine**, validée par le score défini dans le premier axe.
— la **sélection d'un ensemble minimal et informatif de caractéristiques**, déterminé par l'algorithme du second axe.
— le **déclenchement de la mise à jour**, à l'aide d'un apprentissage incrémental, garantissant que l'intégration des nouvelles données améliore la modélisation sans altérer la stabilité du système.

Ce troisième axe vise une amélioration ciblée et progressive, tout en renforçant la capacité du modèle à s'adapter en continu aux évolutions des apprenants et de leurs données au fil du temps.

5 Organisation de la thèse

Cette thèse est organisée en plusieurs chapitres, qui suivent une progression allant de l'introduction des enjeux à la présentation des contributions méthodologiques et expérimentales.

Le chapitre 2 : État de l'art revient sur les environnements d'apprentissage numériques et les approches de modélisation des apprenants. Il approfondit également quatre notions clés au cœur de ce travail : la fiabilité et l'incertitude dans les modèles prédictifs, la frugalité dans l'usage des données, la sélection dynamique des caractéristiques et l'apprentissage incrémental.

Le chapitre 3 : Jeux de données de la littérature en éducation présente un panorama des jeux de données éducatives les plus utilisés dans la recherche en fouille de données éducatives et en analyse des apprentissages. Il décrit ensuite les ensembles de données mobilisés dans cette thèse, en détaillant leurs caractéristiques, leurs variables et leurs spécificités en matière de temporalité et de comportements d'apprentissage.

Le chapitre 4 : Métrique temporelle de l'incertitude et modélisation frugale pour des prédictions fiables propose un cadre d'analyse visant à identifier les modélisations d'apprenants incertaines. Il introduit des indicateurs pour évaluer la fiabilité, et explore les conditions dans lesquelles l'intégration de données supplémentaires devient nécessaire, dans une logique frugale.

Le chapitre 5 : Sélection dynamique des caractéristiques pertinentes pour la prédiction présente l'algorithme *CI-DFS*, conçu pour identifier et actualiser, dans un contexte temporel, les caractéristiques les plus utiles à la prédiction.

Le chapitre 6 : Apprentissage incrémental et déclenchement adaptatif décrit une stratégie de mise à jour progressive et conditionnelle des modèles, permettant d'intégrer de nouvelles données sans réentraîner le modèle complet et en limitant les effets d'oubli catastrophique.

Enfin, la **Conclusion générale** synthétise les apports de la thèse, discute ses limites et ouvre sur de nouvelles perspectives de recherche.

Chapitre 2

État de l'art

Introduction

Les technologies numériques ont profondément transformé les pratiques d'enseignement et de formation, en particulier à travers la généralisation des environnements virtuels d'apprentissage (VLE, *Virtual Learning Environments*) [76]. Ces plateformes ne se limitent plus à diffuser des contenus : elles organisent les activités pédagogiques, encadrent les parcours, et capturent de manière automatisée une multitude de traces numériques [156]. Ces traces incluent des données d'interaction (clics, accès aux ressources, messages sur les forums), des résultats d'évaluation, et parfois des éléments contextuels ou sociodémographiques [91, 46, 130].

Les traces issues des environnements d'apprentissage numériques sont mobilisées dans de nombreux usages : elles permettent d'analyser les comportements d'étude, de modéliser les parcours, ou encore d'évaluer les niveaux d'engagement des apprenants [175, 160]. Parmi ces applications, la prédiction de la réussite ou du décrochage académique occupe une place centrale, soutenue par des travaux sur l'analytique de l'apprentissage et l'exploration de séquences temporelles [49, 112, 69]. Dans ce cadre, la capacité de prédire de manière précoce les risques de décrochage s'impose comme un enjeu majeur, afin de permettre des interventions rapides, ciblées et personnalisées, capables de soutenir la réussite académique et d'accompagner les trajectoires individuelles [27, 22].

Pourtant, malgré des résultats prometteurs en conditions expérimentales, les modèles prédictifs appliqués à des situations réelles demeurent confrontés à plusieurs défis. Les traces disponibles sont hétérogènes, les comportements des apprenants évoluent dans le temps, et les données peuvent être bruitées ou incomplètes. Ces facteurs réduisent la robustesse des modèles, fragilisent la fiabilité des prédictions et compliquent leur mobilisation dans des contextes éducatifs concrets [12, 29, 41]. Ces constats nourrissent directement la problématique de cette thèse, en faisant émerger plusieurs défis majeurs qui viennent structurer la suite de notre analyse.

Un premier défi concerne la fiabilité et l'incertitude des modèles prédictifs. Dans le domaine éducatif, la plupart des systèmes ne gèrent pas l'incertitude et ne la quantifient pas : ils peuvent ainsi produire des résultats erronés sans indiquer leur degré de fiabilité, faute de mécanismes explicites de mesure ou de contrôle. Identifier et traiter cette incertitude, qu'elle provienne de la qualité des données ou du processus d'inférence, demeure essentiel pour concevoir des modèles capables de signaler leurs zones de fragilité et de soutenir des décisions pédagogiques éclairées

[75, 141, 146].

Un autre défi réside dans la frugalité, entendue à la fois comme une gestion raisonnée des données et comme une optimisation des traitements algorithmiques. Les approches existantes tendent à mobiliser l'ensemble des caractéristiques disponibles, indépendamment de leur pertinence, de leur coût d'acquisition ou de leur sensibilité. Or, cette logique soulève des enjeux éthiques liés à la protection des données, mais aussi des contraintes pratiques de temps de calcul et d'infrastructure, sans oublier le risque de surapprentissage lié à une surcharge informationnelle [61, 148].

La question de la sélection dynamique des caractéristiques illustre un troisième défi majeur. Dans la plupart des systèmes actuels, les caractéristiques sont choisies en amont et demeurent fixes, alors que leur utilité peut évoluer au fil du parcours ou varier selon les profils d'apprenants. Dans un contexte marqué par l'arrivée continue de nouvelles données, la capacité à adapter les caractéristiques retenues, en fonction du moment et des besoins spécifiques, devient indispensable [178, 9].

Enfin, un dernier défi tient à l'intégration continue des données dans les modèles. L'apprentissage en ligne est par nature un processus évolutif, mais de nombreux modèles restent conçus pour des ensembles figés. Le manque d'approches incrémentales empêche de tirer parti du flux progressif de données et limite la capacité des systèmes à maintenir leur pertinence face aux transformations des comportements d'apprentissage [42, 96].

Ces défis dessinent le cadre conceptuel dans lequel s'inscrit cette thèse. Ils justifient l'attention portée à la détection des cas incertains, à la sélection dynamique et contextualisée des données, ainsi qu'à l'intégration incrémentale de ces données, afin de renforcer la robustesse et la fiabilité des prédictions dans des environnements éducatifs complexes et en constante évolution.

Notre état de l'art s'organise en deux temps, présentés dans la figure 2.1, Dans une première partie, nous nous intéressons aux environnements virtuels d'apprentissage et aux travaux menés autour des traces éducatives, de la modélisation des apprenants et des approches prédictives dans les VLEs. Ce premier volet nous permet de faire resortir les limites spécifiques aux contextes éducatifs. Dans un second temps, nous élargissons la focale pour explorer quatre axes transversaux aux défis identifiés : la détection de l'incertitude et du manque de fiabilité, la frugalité des données, la sélection dynamique d'informations et les approches d'apprentissage incrémental.

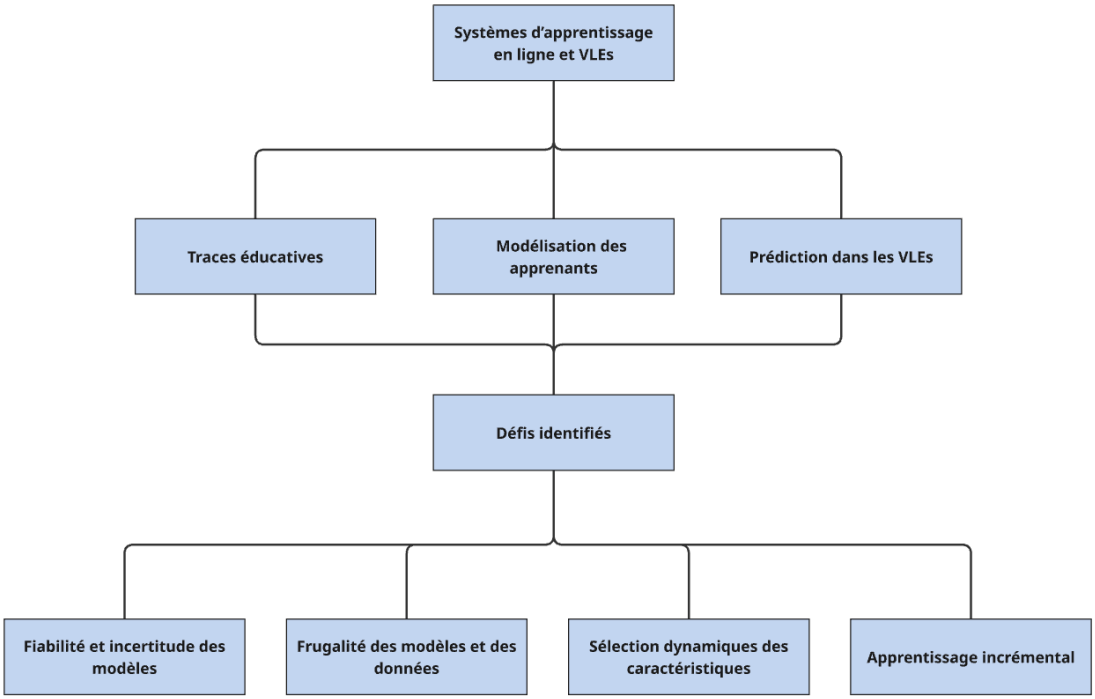

FIGURE 2.1 – Organisation de l'état de l'art

2.1 Systèmes d'apprentissage en ligne

Dans cette première section, nous présentons les systèmes d'apprentissage en ligne, en nous intéressant à leur fonctionnement, aux données qu'ils génèrent, et à la manière dont ils permettent de modéliser et de prédire les parcours des apprenants. Cette analyse nous permettra d'identifier les limites actuelles des systèmes et de poser les bases des problématiques abordées dans la suite de ce travail.

2.1.1 Définition

Les systèmes d'apprentissage en ligne désignent des plateformes numériques conçues pour organiser, encadrer et soutenir les activités pédagogiques dans des contextes d'enseignement à distance, hybride ou enrichi par la technologie [71, 37]. Ils constituent aujourd'hui une composante structurante des systèmes éducatifs numériques, permettant de centraliser les contenus, les interactions, les évaluations et les données d'apprentissage.

Un VLE se caractérise par la combinaison de plusieurs fonctions essentielles : la gestion de contenus pédagogiques (documents, vidéos, quiz), les outils de communication entre enseignants et apprenants (forums, messagerie, classes virtuelles), les dispositifs d'évaluation automatisée (QCM, devoirs notés), ainsi que des outils de suivi individualisé (tableaux de bord de progression ou de performance) [34, 98]. Il existe plusieurs formes de VLEs, qui diffèrent selon leur niveau de

sophistication, leur finalité et leur degré d'adaptation aux besoins de l'apprenant. Le tableau 2.1 ci-dessous présente une typologie synthétique des principales catégories de VLEs et leurs fonctions typiques.

Table 2.1 – Typologie des environnements virtuels d'apprentissage

Type de VLE	**Fonctions typiques**
LMS (Learning Management System)	Organisation de cours, gestion des utilisateurs, diffusion de contenus, évaluations, suivi des notes
LCMS (Learning Content Management System)	Création, gestion et partage de contenus pédagogiques réutilisables, souvent intégrés dans des LMS
Plateformes de MOOC	Apprentissage autonome à grande échelle, accès libre aux cours, certifications, forums entre pairs
LXP (Learning Experience Platform)	Personnalisation des parcours, recommandations de contenus, apprentissage informel et hybride
Systèmes adaptatifs intelligents	Modélisation dynamique des apprenants, adaptation en temps réel des contenus ou des exercices
VLE immersifs (Metavers, VR)	Immersion 3D, interaction spatialisée, scénarios simulés, apprentissage expérientiel

Tous ces environnements partagent une caractéristique commune : la capacité à enregistrer automatiquement les actions des utilisateurs. Un clic, une soumission, un message échangé peut devenir une trace numérique exploitable. Cette propriété sera explorée plus en détail dans la section suivante.

2.1.2 Traces numériques et potentiel analytique

La montée en puissance des VLEs s'accompagne d'un changement de paradigme fondamental : l'activité des apprenants ne se limite plus à la simple consultation de contenus, elle devient aussi une source continue de données. Une action effectuée dans un VLE, qu'il s'agisse de la navigation, de la participation à un forum, de la soumission d'un devoir ou de la consultation d'une ressource, est enregistrée de manière automatique, horodatée, et structurée sous forme de fichiers de logs [128, 116]. Ces traces numériques constituent aujourd'hui un matériau central pour comprendre, modéliser et accompagner les parcours d'apprentissage.

Cette capacité à capturer les interactions a transformé les VLEs en véritables plateformes d'observation. Elle permet non seulement de documenter le comportement des apprenants, mais aussi d'alimenter des modèles prédictifs, des visualisations, ou des dispositifs d'adaptation personnalisée [160, 76]. Les travaux en fouille de données éducatives *(Educational Data Mining - EDM)* et l'analyse d'apprentissage *(Learning Analytics - LA)* s'appuient en grande partie sur l'exploitation de ces données, qu'ils structurent et catégorisent selon leur nature, leur temporalité et leur utilité analytique [128].

Typologie des traces collectées dans les VLEs :

— *Traces de navigation (clickstream)* : elles renseignent sur les séquences de pages visitées, la fréquence de connexion, le temps passé sur chaque activité ou la profondeur d'exploration. Ces données sont souvent mobilisées comme métriques d'engagement ou de fatigue, et se

prêtent bien à des modélisations temporelles [134].

— *Traces de consultation de ressources* : elles incluent les logs de visionnage de vidéos (démarrage, interruptions, relectures), les documents ouverts ou téléchargés, ou encore l'accès aux modules de cours. Elles permettent d'identifier des motifs de révision, des habitudes de travail ou des préférences de formats [116].

— *Traces d'évaluation* : elles regroupent les résultats obtenus aux quiz, les délais entre tentative et soumission, le nombre d'essais, ou encore les types de rétroaction reçue. Elles sont largement utilisées comme caractéristiques explicatives ou cibles dans les modèles de prédiction de la réussite académique [128, 160].

— *Traces de forum et d'interactions sociales* : elles incluent les messages publiés, les réponses apportées, les mentions ou les interactions avec les pairs. Ces éléments révèlent des dynamiques sociales, la participation à une communauté d'apprentissage, voire un sentiment d'appartenance [34].

— *Traces liées à l'autorégulation* : plus rares, elles se manifestent par des actions telles que le marquage de progression, l'ajout de notes personnelles, la fixation d'objectifs ou l'utilisation de rappels. Elles sont précieuses pour analyser les dimensions métacognitives du comportement apprenant [72].

Ces différentes catégories de traces sont généralement indexées par un double contexte : un *contexte temporel*, qui permet de reconstituer des trajectoires (séquences d'activités, récurrence, périodicité), et un *contexte pédagogique*, qui associe chaque trace à une activité, un module, un niveau ou une semaine donnée. Cette double ancrage rend possible des analyses longitudinales fines, tant au niveau individuel que collectif.

En somme, les traces numériques capturées par les VLEs offrent un potentiel analytique considérable. Elles permettent de documenter les comportements d'apprentissage au plus près de l'activité réelle, dans une logique de personnalisation, de prédiction ou de soutien pédagogique. Toutefois, leur exploitation soulève aussi des défis méthodologiques, techniques et éthiques, que nous aborderons dans les sections suivantes.

2.1.3 Modélisation des apprenants à partir de traces

La modélisation de l'apprenant (*learner modeling*) désigne le processus qui consiste à construire, mettre à jour et exploiter une représentation formelle de l'état, du comportement ou des connaissances d'un apprenant à partir des traces générées dans un environnement numérique d'apprentissage [83]. Cette représentation peut être explicite (sous la forme d'un profil, d'un tableau de bord, d'un graphe) ou implicite (encodée dans un vecteur latent ou une distribution de probabilité), mais elle vise toujours à refléter la dynamique individuelle de l'apprentissage en vue de personnaliser les interactions, guider l'accompagnement ou nourrir les dispositifs d'analyse à grande échelle [57, 45].

Dans les VLEs, cette modélisation repose sur l'exploitation de données variées : caractéristiques de l'apprenant, activités réalisées, performances évaluées, interactions sociales, voire signaux d'autorégulation ou d'engagement émotionnel. Ces dimensions, souvent hétérogènes et partiellement observables, sont agrégées pour former un modèle dynamique capable de suivre

l'évolution du parcours d'un apprenant dans le temps [3, 101].

La figure 2.2 illustre les principaux composants d'un modèle apprenant, structuré autour de cinq dimensions complémentaires : les caractéristiques individuelles, les données personnelles, les savoirs acquis, les interactions et l'état d'apprentissage.

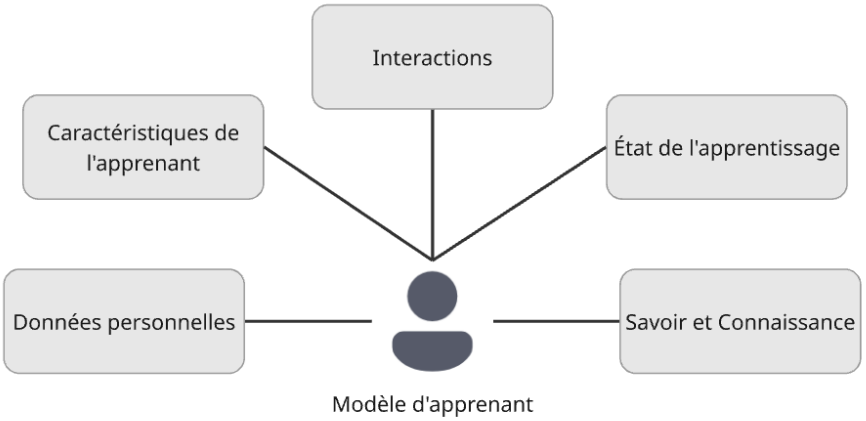

FIGURE 2.2 – Composants d'un modèle d'apprenant

Méthodes de modélisation : un paysage en évolution

Les méthodes utilisées pour construire ce modèle varient selon la nature des données, l'objectif visé et les contraintes du système. On distingue généralement plusieurs grandes approches, résumées dans le tableau 2.2.

TABLE 2.2 – Typologie des approches de modélisation des apprenants

Approche	**Objectif principal**	**Références**
Caractéristiques apprenants	Représenter les traits cognitifs, motivationnels, sociodémographiques	[3, 45]
Modélisation de la connaissance (Knowledge Tracing)	Estimer la maîtrise de concepts à partir des réponses	[16, 25]
Modèles comportementaux	Identifier des motifs d'activité dans les traces d'usage	[54, 131]
Modèles multi-dimensionnels	Combiner comportements, émotions, engagement, performance	[26, 101]
Modèles adaptatifs personnalisés	Adapter dynamiquement contenus et feedbacks à l'état de l'apprenant	[144, 157]

Ce tableau illustre la variété des approches de modélisation des apprenants. Certaines se concentrent sur les caractéristiques individuelles, comme les traits cognitifs, motivationnels ou sociodémographiques [3, 45]. D'autres, issues du courant du *Knowledge Tracing*, visent à estimer la maîtrise progressive de concepts à partir des réponses des étudiants [16, 25]. Ces deux familles, centrées respectivement sur le profil et sur la connaissance, constituent les bases historiques de

la modélisation.

Progressivement, de nouvelles approches ont émergé pour capter davantage la dimension dynamique des apprentissages. Les modèles comportementaux cherchent à identifier des motifs d'activité dans les traces numériques [54, 131], tandis que les modèles multi-dimensionnels tentent de combiner plusieurs aspects comme l'engagement, les émotions et la performance [26, 101]. Ces approches sont prometteuses, mais elles se heurtent à la complexité d'intégrer des données hétérogènes et souvent bruitées.

Enfin, les modèles adaptatifs personnalisés visent à aller plus loin, en ajustant dynamiquement les contenus et les feedbacks en fonction de l'état de l'apprenant [144, 157]. Ils incarnent l'ambition d'une individualisation poussée, mais restent encore limités par leur dépendance à des données complètes et de qualité. Ainsi, chaque type de modèle apporte un éclairage particulier, mais aucun ne résout à lui seul les défis liés à la fiabilité, à la frugalité et à l'évolution continue des traces éducatives.

Applications et usages de la modélisation des apprenants

La modélisation des apprenants est aujourd'hui intégrée dans de nombreux systèmes d'apprentissage en ligne, avec des objectifs qui vont bien au-delà du simple suivi. Ils sont mobilisés pour :

— *Personnaliser les parcours* : en ajustant les contenus, les ressources ou les niveaux de difficulté selon le profil de l'apprenant, ces modèles permettent d'augmenter la pertinence des recommandations pédagogiques [144, 87].

— *Alimenter des tableaux de bord* : dans une logique de transparence et de rétroaction, les modèles sont parfois rendus visibles aux enseignants ou aux apprenants via des visualisations interactives (Open Learner Models) [87].

— *Soutenir la détection précoce* : en analysant les écarts entre comportements attendus et observés, ils peuvent permettre d'anticiper un risque de décrochage ou de désengagement, et de déclencher des interventions ciblées [31, 47].

— *Piloter les systèmes adaptatifs* : les plateformes intelligentes exploitent ces modèles pour ajuster dynamiquement le scénario pédagogique, en fonction de la progression, du rythme ou de la disponibilité cognitive de l'apprenant [101, 157].

Ces usages positionnent la modélisation de l'apprenant comme une brique centrale de l'architecture des environnements numériques d'apprentissage. Elle constitue aussi un point d'entrée essentiel pour les systèmes de prédiction, qui s'appuient sur les dynamiques captées par ces modèles pour anticiper les trajectoires futures des apprenants.

2.1.4 Prédiction dans les environnements virtuels d'apprentissage

Avec l'évolution des plateformes éducatives vers des *environnements virtuels d'apprentissage intelligents*, la collecte systématique de traces numériques ouvre la voie à une variété d'analyses automatisées. Ces systèmes ne se contentent plus de diffuser des contenus ou de structurer des parcours : ils embarquent des modules d'analyse de données, de recommandation, de visualisa-

tion, et de prédiction d'échec ou de décrochage. L'objectif est de passer d'un pilotage réactif à un accompagnement proactif, en anticipant les situations critiques ou les opportunités pédagogiques [12, 11, 87].

Dans cette section, nous nous concentrons sur la dimension prédictive des VLEs, qui consiste à estimer, à partir des traces passées et présentes, des événements ou états futurs concernant les apprenants. Cette capacité prédictive constitue un levier stratégique pour la personnalisation des parcours, la détection précoce du décrochage ou encore l'optimisation de la progression individuelle [125, 26].

Définition et objectifs de la prédiction

La prédiction dans les VLEs désigne l'ensemble des méthodes algorithmiques visant à anticiper des résultats ou états éducatifs à partir des données issues de l'activité des apprenants. Les cibles prédictives les plus fréquentes incluent : la réussite académique, le risque de décrochage, le niveau d'engagement, ou encore la persévérance dans un cours [42, 22, 13]. Ces prédictions peuvent être mobilisées :

- par les enseignants ou tuteurs, pour déclencher des interventions ciblées ;
- par les plateformes, pour adapter dynamiquement les ressources ou l'agencement pédagogique ;
- par les décideurs, pour piloter l'allocation de ressources ou identifier des profils récurrents à risque [27, 115].

Types de prédiction

Les prédictions peuvent être regroupées selon la nature de la cible :

- *Classification binaire* : succès / échec, engagement / désengagement, abandon / persévérance [47, 134].
- *Classification multiclasse* : niveau d'engagement (faible, moyen, élevé), types de comportements, classification des apprenants risques [30, 159].
- *Régression* : score attendu, taux de complétion, temps restant estimé [26].

Plusieurs études se concentrent sur la classification binaire, en raison de son adéquation avec les cas d'usage critiques comme la détection du décrochage [125, 8, 13].

Algorithmes pour la prédiction dans les VLEs

Les prédictions s'appuient majoritairement sur des algorithmes d'apprentissage supervisé. Ces modèles cherchent à établir une correspondance entre des traces observées (ex. nombre de connexions, soumissions de devoirs, participation aux forums) et une étiquette cible (ex. réussite, échec, engagement). Le choix de l'algorithme dépend fortement du contexte (taille des données, hétérogénéité, besoin d'explicabilité) et des contraintes pratiques (temps de calcul, robustesse, facilité d'interprétation).

Les approches les plus classiques incluent notamment les **arbres de décision**, qui reposent sur une décomposition récursive des données selon les attributs les plus discriminants pour aboutir à des règles explicables. Leur capacité à générer des modèles facilement interprétables en fait un choix fréquent pour les systèmes d'alerte précoce dans les VLEs [39, 104].

D'autres approches, comme les **k-plus proches voisins (k-NN)**, fondent la prédiction sur la similarité entre apprenants : la classe d'un nouvel individu est déterminée en fonction de celles observées chez les instances les plus proches dans l'espace des caractéristiques. Cette méthode non paramétrique, simple et intuitive, s'avère efficace sur des ensembles homogènes de petite taille, mais ses performances se dégradent rapidement en présence de données de forte dimensionnalité [8, 121].

Le **classificateur Naïve Bayes**, quant à lui, repose sur le théorème de Bayes et suppose l'indépendance conditionnelle entre les variables explicatives. Malgré cette hypothèse simplificatrice, il reste performant dans des contextes de classification binaire, grâce à sa rapidité d'exécution et à sa robustesse sur de petits volumes de données [66, 138].

Le tableau 2.3 présente une synthèse critique des algorithmes les plus couramment utilisés dans la littérature, en précisant leur contexte d'usage, leurs avantages, leurs limites et les références associées.

TABLE 2.3 – Comparatif critique des algorithmes supervisés pour la prédiction dans les VLEs

Algorithme	**Contexte d'usage**	**Avantages**	**Limites**	**Références**
Régression logistique	Petits jeux de données, besoin d'interprétabilité	Simple, rapide, interprétable, robuste sur petits échantillons	Suppose linéarité, mauvaise gestion des interactions complexes	[5, 8]
Arbre de décision	Systèmes d'alerte explicables, faible complexité	Facile à comprendre, génère des règles explicables	Sensible au bruit et au surapprentissage	[39, 104]
Forêt aléatoire	MOOCs, systèmes à données variées	Stable, efficace sur données hétérogènes, réduit surapprentissage	Moins interprétable, coûteux en calcul	[125, 13]
Boosting (XGBoost, LightGBM)	MOOCs, données bruitées ou déséquilibrées	Très bonnes performances, gère bien les déséquilibres	Peu explicable, dépendant des paramètres, difficile à ajuster	[22, 23]
SVM	Cas spécifiques à haute dimensionnalité	Bon pouvoir de séparation, flexible avec noyaux	Complexe à paramétrer, peu scalable	[142, 41]
k-plus proches voisins (k-NN)	Petits ensembles homogènes, sans apprentissage	Sans entraînement, intuitif, rapide sur petits volumes	Faible performance en dimension élevée, dépend des distances	[8, 121]
Naïve Bayes	Classification de texte, forums étudiants	Efficace sur texte, faible coût de calcul	Suppose indépendance des variables, sensible au déséquilibre	[66, 138]
Réseaux de neurones	Apprentissage séquentiel, comportements complexes	Capture des dépendances complexes, adapté aux séquences	Peu interprétable, nécessite beaucoup de données, long entraînement	[152, 145, 60]

2.1.5 Structuration temporelle des données

Dans les environnements virtuels d'apprentissage, les données ne sont pas disponibles en une seule fois, mais sont générées progressivement au fil du temps, à mesure que l'apprenant interagit avec la plateforme. Cette dimension temporelle est essentielle, car l'apprentissage est un processus évolutif, fait de phases d'exploration, de régularité, de rupture ou de désengagement [129, 33, 109].

Pour analyser ces trajectoires, les traces numériques sont souvent regroupées par unités temporelles fixes, comme les semaines de cours. Chaque période regroupe alors les actions réalisées :

connexions, ressources consultées, devoirs soumis, scores obtenus, messages publiés. Cette structuration permet d'observer les variations d'activité, d'identifier des tendances (hausse ou chute d'engagement) et de construire des métriques dynamiques utilisables pour la prédiction [128, 175].

Plusieurs choix méthodologiques influencent la qualité des analyses temporelles :

— **La granularité temporelle** : une unité trop courte génère du bruit, tandis qu'une unité trop longue risque de masquer des signaux faibles [129].

— **L'alignement pédagogique** : pour comparer les parcours, il est recommandé d'aligner les données sur la structure du cours (e.g. semaines pédagogiques) plutôt que sur un calendrier absolu [159].

— **La gestion des absences de données** : le silence d'un apprenant peut traduire un désengagement, mais aussi une simple pause ou un usage non instrumenté ; il convient donc de distinguer les absences significatives des lacunes techniques [109].

Cette structuration permet également de formuler des prédictions anticipées : par exemple, dès la troisième semaine, il devient possible d'estimer le risque d'échec final à partir des comportements observés jusque-là. Les approches traditionnelles peuvent modéliser cette évolution de manière explicite, en créant des variables temporelles cumulées ou différenciées par semaine. D'autres modèles, comme les réseaux neuronaux séquentiels ou les architectures à mémoire, exploitent directement la séquence des données pour capter les dépendances temporelles [109].

Ainsi, prendre en compte la dimension temporelle ne se limite pas à un simple découpage des données. Il s'agit d'un véritable enjeu de modélisation, qui conditionne la pertinence des prédictions, la précocité des alertes, et la possibilité d'intervenir au bon moment.

2.1.6 Discussion

L'état de l'art met en évidence la maturité croissante des systèmes d'apprentissage en ligne et le potentiel analytique des traces qu'ils génèrent [76, 156]. Les travaux existants décrivent de manière convergente les familles de modélisation des apprenants (profils, modélisation de l'évolution des connaissances (*knowledge tracing)*, modèles comportementaux et multi-dimensionnels, systèmes adaptatifs) et les usages prédictifs dominants (réussite, décrochage, engagement) [128, 160, 69]. L'intérêt pour la prédiction s'affirme, avec des résultats prometteurs en contexte expérimental [27, 22].

Dans le même temps, plusieurs défis persistent. D'abord, la qualité réelle des données demeure hétérogène : bruit, valeurs manquantes, signaux contextuels partiels, et évolutions rapides des comportements [12]. Une part significative des études reste évaluée *hors ligne* en mode *batch*, sur toutes les données, avec l'hypothèse que ces données-là sont statiques et toujours disponible à l'avance, sans protocole temporel strict ni vérification de la stabilité des performances dans la durée [29, 41]. Les sections consacrées à la structuration temporelle montrent pourtant que les trajectoires temporels et les phénomènes de dérive exigent des modèles sensibles au temps et à la séquencialité [129, 109, 175].

Au niveau des modèles, la question de la fiabilité reste insuffisamment traitée. Peu de travaux mesurent explicitement l'incertitude des prédictions (calibration, entropie, détection des cas incertains), alors même que des erreurs peuvent conduire à des décisions pédagogiques inadaptées

[75, 141, 146]. Par ailleurs, le compromis entre performance et explicabilité n'est pas stabilisé : des algorithmes efficaces (comme l'apprentissage profond) demeurent difficiles à interpréter et à auditer dans des contextes réels [22].

La littérature souligne aussi un manque de frugalité dans l'usage des données et des calculs. L'agrégation «maximaliste» de caractéristiques [4] n'est pas toujours corrélée à un gain de performance et accroît les coûts éthiques (minimisation des données), opérationnels (temps de calcul, infrastructure) et méthodologiques (surapprentissage, sur-spécialisation) [61, 148]. De ce point de vue, les travaux plaident pour des stratégies de collecte et d'entraînement plus sobres, pilotées par l'utilité réelle pour la décision.

Un autre point important concerne la sélection dynamique des caractéristiques. Nombreuses études fixent leurs caractéristiques en amont de l'expérimentation, alors que leur pertinence varie selon le temps, les contextes d'activité et les profils [178, 9]. Dans des VLEs où les données arrivent en flux, l'absence d'un mécanisme d'actualisation des caractéristiques utiles limite la précision et la stabilité des modèles au fil du temps.

Enfin, la mise à jour incrémentale des modèles reste peu mobilisée, alors que l'apprentissage humain est un processus continu. Les synthèses récentes montrent l'intérêt d'approches en flux (streaming, continual/online learning) pour absorber la dérive, préserver les acquis et maintenir la pertinence sans réentraîner intégralement les systèmes [42, 96, 147]. Cette exigence est d'autant plus forte que les VLEs évoluent rapidement (contenus, calendriers, publics), rendant obsolètes des modèles figés.

Sur le plan de la reproductibilité, l'usage récurrent d'un petit nombre de jeux de données (souvent publics) facilite la comparaison mais pose la question de la généralisation à d'autres contextes institutionnels et pédagogiques [91, 130, 116].

Au total, la littérature souligne le potentiel des VLEs pour la personnalisation et la détection précoce, mais elle laisse ouvertes des questions cruciales : (i) comment *détecter et traiter l'incertitude* des modèles pour fiabiliser les décisions [146] ; (ii) comment *réduire l'empreinte de données et de calcul* sans dégrader la qualité prédictive [61] ; (iii) comment *sélectionner dynamiquement* les caractéristiques vraiment utiles au moment opportun [178] ; et (iv) comment *intégrer en continu* les nouvelles données pour maintenir des modèles pertinents dans le temps [42, 147]. Ces défis structurent la suite de cette thèse et orientent nos contributions vers des cadres méthodologiques capables d'articuler fiabilité, frugalité, adaptativité et incrémentalité dans des environnements éducatifs réels.

2.1.7 Synthèse de la section

L'analyse des travaux récents sur les environnements virtuels d'apprentissage a permis de dégager plusieurs constats fondamentaux. Si les traces numériques capturées par les plateformes offrent un potentiel considérable pour modéliser et prédire les parcours des apprenants, leur exploitation soulève des défis majeurs. Les modèles actuels peinent à composer avec la l'**incomplétude des profils**, la **capacité à exprimer l'incertitude**, l'**usage non raisonné des ressources**, et l'**actualisation dynamique**.

4. L'utilisation de toutes les données disponibles.

Ces limites structurelles réduisent la robustesse des prédictions, augmentent les risques d'erreurs injustes, et restreignent la capacité des systèmes à s'adapter à la diversité réelle des trajectoires d'apprentissage.

Vers un changement de paradigme

Au-delà du choix d'un algorithme performant, il devient essentiel de repenser le cadre de conception des modèles prédictifs en éducation. Cela implique notamment :

- d'intégrer des mécanismes d'*adaptation aux flux de données*, pour suivre l'évolution temporelle des comportements ;
- de prendre en compte l'*'incertitude*, afin de moduler la confiance accordée à chaque prédiction ;
- d'adopter une logique de *frugalité raisonnée*, en sélectionnant les données réellement utiles à la décision ;
- d'anticiper les erreurs possibles sur des profils minoritaires ou atypiques, pour garantir une *équité algorithmique*.

Ces exigences dessinent les contours d'une approche plus contextuelle, adaptative et raisonnée, sensible à la qualité des données et à la diversité des apprenants.

Résumé de la première partie de l'état de l'art

La première partie de cet état de l'art a permis de poser les bases contextuelles et critiques de la prédiction dans les environnements virtuels d'apprentissage. Elle a mis en évidence les failles méthodologiques récurrentes des modèles actuels, notamment en termes de fragmentation, d'incertitude, de frugalité et de temporalité. Ces limites, souvent interdépendantes, fragilisent la fiabilité et l'équité des systèmes prédictifs lorsqu'ils sont appliqués à des contextes éducatifs réels.

Résumé de la deuxième partie

La deuxième partie de notre revue explorera précisément ces quatre axes comme autant de leviers pour repenser la conception des systèmes prédictifs. Chaque section sera consacrée à l'un de ces défis(incertitude, frugalité, sélection dynamique des données, et apprentissage incrémental), en s'appuyant sur les travaux les plus récents, les méthodes émergentes, et les conditions de leur application au contexte éducatif.

2.2 Fiabilité et incertitude des modèles d'apprentissage automatique

Nous commençons cette partie par l'étude conjointe de la fiabilité et de l'incertitude, deux dimensions encore trop souvent négligées dans les modèles éducatifs. La fiabilité renvoie à la capacité d'un modèle à produire des prédictions stables et cohérentes dans le temps [74], tandis que l'incertitude traduit le degré de confiance accordé à chacune de ces prédictions [146]. Mieux comprendre et quantifier ces deux aspects permet non seulement de renforcer la confiance globale dans les systèmes prédictifs, mais aussi d'identifier plus finement les cas à risque et d'adapter en conséquence les interventions pédagogiques [174].

2.2.1 Définition de la fiabilité

Dans le contexte de prédiction, la fiabilité d'un modèle prédictif désigne sa capacité à maintenir, sur la durée et face à la variabilité des données, un niveau de performance stable et cohérent. Elle se distingue des métriques classiques de classification (précision, rappel, aire sous la courbe ROC) [12, 5], qui mesurent uniquement l'exactitude ponctuelle d'un modèle sur un jeu de données statique.

On peut définir la fiabilité comme une *propriété temporelle* du modèle : soit une séquence de prédictions $\hat{y}i, t$ pour un étudiant i au cours du temps t. Un modèle est dit fiable si la fréquence des prédictions correctes $\hat{y}_{i,t} = y_{i,t}$ se maintient avec une faible variance au fil du temps, malgré l'arrivée de nouvelles données ou de conditions changeantes. Cette robustesse temporelle implique non seulement la stabilité statistique des performances globales, mais aussi la cohérence des décisions au niveau individuel [74, 174].

Dans le domaine de l'éducation, cette exigence est essentielle car l'objectif n'est pas seulement de prédire correctement un événement isolé (par exemple, l'échec à une évaluation), mais de produire des signaux continus et exploitables pour accompagner les trajectoires d'apprentissage. Des travaux récents soulignent ainsi que la valeur d'un modèle dépend de sa capacité à conjuguer *précocité*, *stabilité* et *consistance* des prédictions [29, 125, 104].

La fiabilité peut donc être comprise comme une dimension complémentaire aux métriques traditionnelles de performance : elle traduit la résistance d'un modèle aux perturbations, sa constance au fil du temps et sa capacité à fournir des prédictions pédagogiquement exploitables dans des environnements réels.

2.2.2 Évaluer la fiabilité : apports de la littérature

La question de la fiabilité des modèles prédictifs a donné lieu à plusieurs approches méthodologiques dans la littérature. Celles-ci visent à dépasser l'évaluation statique de la performance pour intégrer la dimension temporelle et la stabilité des prédictions. Dans des contextes éducatifs, cette perspective est particulièrement importante, car l'objectif n'est pas seulement de détecter les étudiants à risque à un instant donné, mais de le faire suffisamment tôt et de manière suffisamment cohérente pour que l'information soit exploitable dans la durée.

Parmi les propositions existantes, l'*Earliness Stability Score (ESS)*, définit dans l'équation 2.1, occupe une place centrale. Développée pour évaluer la qualité des systèmes de prédiction en éducation, cette métrique combine deux dimensions complémentaires : la précocité et la stabilité

des prédictions [29].

— La **précocité** (*Earliness)* mesure la capacité d'un modèle à formuler une prédiction correcte le plus tôt possible dans le déroulement du cours, ce qui est crucial pour maximiser l'impact d'une intervention pédagogique. Elle correspond au premier instant où le modèle prédit correctement l'issue pour un étudiant donné.
— La **stabilité** (*Stability* évalue la cohérence des prédictions au fil du temps et reflète la capacité du modèle à maintenir une décision correcte sur plusieurs observations successives. Elle est définie comme la longueur moyenne des séquences consécutives de prédictions correctes.

Ces deux dimensions sont combinées dans un score unique, l'ESS est calculé comme suit :

$$ESS = \frac{2 \times (Earliness \times Stability)}{Earliness + Stability} \tag{2.1}$$

L'ESS permet ainsi d'équilibrer le besoin d'agir tôt avec celui de garantir une continuité dans les prédictions.

D'autres métriques, issues de la littérature en apprentissage automatique dans des domaines autres que l'éducation, abordent la fiabilité sous des angles complémentaires. Certains travaux insistent sur la robustesse des modèles face à des perturbations ou à des données bruitées, en cherchant à quantifier leur résilience [19]. D'autres proposent des cadres plus holistiques, qui visent à évaluer la fiabilité comme un ensemble de propriétés englobant la stabilité des résultats, la cohérence au niveau individuel et la transparence des mécanismes de décision [48, 74]. Ces perspectives soulignent que la fiabilité ne peut être réduite à une simple mesure numérique, mais qu'elle doit être envisagée comme un faisceau de critères permettant de juger de la solidité et de la crédibilité des modèles.

Dans le champ éducatif, ces approches rappellent que la valeur d'un système prédictif ne réside pas uniquement dans sa capacité à fournir une prédiction correcte isolée, mais dans sa faculté à offrir des signaux fiables et exploitables dans le temps. Une prédiction instable, même correcte, peut fragiliser la confiance des enseignants et limiter l'usage pratique du modèle. À l'inverse, un système dont les prédictions sont précoces et stables offre une base plus robuste pour guider les décisions.

2.2.3 Définition de l'incertitude

L'incertitude prédictive désigne la capacité d'un modèle à exprimer, au moment de sa décision, le degré de confiance qu'il accorde à sa propre prédiction [82]. Contrairement à une probabilité de classe produite par un modèle de classification, qui peut être numériquement élevée même lorsque le modèle est mal calibré ou peu informé, l'incertitude vise à quantifier la fiabilité réelle de la sortie, en tenant compte des zones d'ignorance, des biais d'entraînement, ou encore de la structure des données disponibles [82, 4]. Deux formes fondamentales d'incertitude sont communément distinguées dans la littérature en apprentissage automatique [68] :

— **L'incertitude épistémique** découle d'un manque d'information ou d'ignorance sur certaines zones de l'espace de représentation. Elle est particulièrement élevée lorsque le modèle est confronté à des données rares, nouvelles ou éloignées de celles vues pendant l'entraînement. Elle est théoriquement réductible : on peut la diminuer en acquérant davantage de données représentatives ou en raffinant l'architecture du modèle [94].

— **L'incertitude aléatoire** reflète la variabilité intrinsèque du phénomène observé. Même avec des données complètes et un modèle parfait, il subsiste une part d'imprévisibilité, liée à des facteurs non observables ou à une variabilité naturelle des comportements. Cette incertitude est dite irréductible [82].

Dans le contexte éducatif, cette distinction revêt une importance particulière. L'incertitude épistémique peut refléter l'atypicité d'un apprenant (parcours singulier, faible activité, profil mal représenté), tandis que l'incertitude aléatoire peut correspondre à une instabilité comportementale, à des transitions pédagogiques floues, ou à des stratégies changeantes. Ignorer ces incertitudes revient à supposer que toute prédiction est également fiable, ce qui est risqué, notamment lorsqu'une intervention pédagogique en dépend.

La quantification de l'incertitude dans les systèmes d'apprentissage automatique a fait l'objet de nombreuses recherches ces dernières années, notamment dans des domaines critiques comme la santé, l'automobile, ou la sécurité. En éducation, ces méthodes commencent à être adaptées, mais leur intégration reste encore partielle. Nous présentons ici les principales approches probabilistes, ensemblistes, ainsi que les métriques d'incertitude.

2.2.4 Approches probabilistes et bayésiennes

Les méthodes dites *Bayesian Deep Learning* cherchent à estimer une distribution sur les poids du modèle plutôt qu'une estimation ponctuelle. L'idée est d'approximer la vraisemblance de chaque sortie en tenant compte de l'incertitude sur les paramètres. Les méthodes notables incluent :

— **Monte Carlo Dropout** [68] : le *dropout* est une technique de régularisation consistant à désactiver aléatoirement une proportion de neurones pendant l'entraînement afin d'éviter le surapprentissage et de favoriser la robustesse du modèle. La méthode de Monte Carlo Dropout reprend ce principe en maintenant le *dropout* actif à l'inférence, ce qui introduit de la variabilité dans les poids du réseau. En répétant plusieurs passes du modèle, on obtient une distribution de sorties dont la moyenne et la variance permettent d'estimer l'incertitude épistémique.

La prédiction finale est donnée par la moyenne des T sorties obtenues :

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_{\mathbf{W}_t}(x) \tag{2.2}$$

où $f_{\mathbf{W}_t}(x)$ désigne la sortie du modèle pour un jeu de poids $\mathbf{W}_t$ échantillonné par *dropout*.

L'incertitude est estimée par la variance de ces sorties :

$$\mathrm{Var}[\hat{y}] = \frac{1}{T} \sum_{t=1}^{T} f_{\mathbf{W}_t}(x)^2 - \left(\frac{1}{T} \sum_{t=1}^{T} f_{\mathbf{W}_t}(x) \right)^2 \tag{2.3}$$

Ainsi, la moyenne reflète la prédiction attendue, tandis que la variance indique le degré de confiance du modèle : une variance élevée traduit une incertitude plus forte dans la

prédiction, souvent liée à un manque d'exemples similaires dans les données d'entraînement.

— **Bayesian Neural Networks (BNNs)** [32] : dans un réseau de neurones bayésien, chaque poids n'est pas représenté par une valeur fixe, mais par une distribution de probabilité. Plutôt que d'apprendre un seul ensemble de paramètres, le modèle apprend à estimer la probabilité de chaque configuration de poids, ce qui permet de quantifier directement l'incertitude des prédictions.

La distribution prédictive s'écrit alors :

$$p(y \mid x, \mathcal{D}) = \int p(y \mid x, \mathbf{W})\, p(\mathbf{W} \mid \mathcal{D})\, d\mathbf{W} \tag{2.4}$$

où $p(\mathbf{W} \mid \mathcal{D})$ représente la distribution postérieure des poids après observation des données $\mathcal{D}$. Comme cette intégrale est impossible à calculer directement, on l'approxime en échantillonnant plusieurs jeux de poids $\mathbf{W}_t$ selon une distribution approchée $q_\theta(\mathbf{W})$:

$$p(y \mid x, \mathcal{D}) \approx \frac{1}{T} \sum_{t=1}^{T} p(y \mid x, \mathbf{W}_t) \tag{2.5}$$

Les BNNs offrent ainsi un cadre probabiliste rigoureux pour représenter l'incertitude, mais leur entraînement reste plus coûteux et complexe que celui des réseaux classiques.

Ces méthodes sont particulièrement adaptées à des contextes avec peu de données ou des profils atypiques, mais leur complexité computationnelle limite encore leur adoption dans les systèmes éducatifs réels.

2.2.5 Ensembles de modèles (Deep Ensembles)

Les *Deep Ensembles* reposent sur l'idée d'entraîner plusieurs modèles indépendants sur la même tâche, chacun à partir d'une initialisation différente [94, 70]. Chaque modèle f_{θ_t} apprend sa propre représentation des données, et la combinaison de leurs prédictions permet d'obtenir à la fois une estimation plus robuste et une mesure d'incertitude.

La prédiction finale correspond à la moyenne des sorties des T modèles entraînés :

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_{\theta_t}(x) \tag{2.6}$$

et l'incertitude associée est estimée par la variance inter-modèles :

$$\mathrm{Var}[\hat{y}] = \frac{1}{T} \sum_{t=1}^{T} f_{\theta_t}(x)^2 - \left(\frac{1}{T} \sum_{t=1}^{T} f_{\theta_t}(x) \right)^2 \tag{2.7}$$

Cette approche simple à implémenter se distingue par sa robustesse empirique : elle combine les bénéfices du bagging [5] et de la diversité des modèles sans recourir à des hypothèses bayésiennes complexes. Cependant, son principal inconvénient réside dans le coût computationnel lié à l'entraînement multiple, bien que ce surcoût soit souvent compensé par une meilleure calibration et une meilleure fiabilité des prédictions.

5. Une méthode d'ensemble qui entraîne plusieurs modèles sur des échantillons aléatoires du même jeu de données, puis combine leurs prédictions pour obtenir un résultat plus stable et précis.

2.2.6 Métriques d'incertitude

Il existe plusieurs métriques quantitatives pour estimer l'incertitude à partir des sorties d'un modèle :

— **Entropie prédictive** [135] : L'entropie mesure la dispersion d'une distribution de probabilité. Dans un cadre de classification, elle reflète le degré d'hésitation du modèle entre les classes. Elle est maximale lorsque toutes les classes sont équiprobables et minimale lorsqu'une seule domine :

$$H(y|x) = -\sum_{c=1}^{C} p(y = c|x) \log p(y = c|x) \tag{2.8}$$

Une entropie élevée signale une incertitude importante, même si la probabilité prédite la plus élevée semble convaincante. Elle est particulièrement utile pour identifier les cas ambigus ou hors distribution.

— **Variance inter-prédictions (désaccord)** [94] : Dans les ensembles de modèles ou les approches bayésiennes approximatives (ex. Monte Carlo Dropout), on effectue plusieurs prédictions sur une même instance et on calcule la variance entre les sorties :

$$\mathrm{Var}[p(y|x)] = \frac{1}{T} \sum_{t=1}^{T} \left(p_t(y|x) - \bar{p}(y|x)\right)^2 \tag{2.9}$$

Une variance élevée entre prédictions signale un désaccord entre modèles et reflète une forte incertitude épistémique.

Ces métriques sont souvent utilisées en combinaison avec des mécanismes de *rejection*, d'alerte, ou de déclenchement conditionnel d'une action complémentaire (comme la sollicitation d'un humain).

2.2.7 Applications en éducation

Plusieurs travaux récents explorent l'application de ces techniques dans les environnements d'apprentissage. Thuy et al. proposent par exemple une approche combinant entropie prédictive et visualisation afin de rendre les modèles plus explicables dans le cadre de la prédiction de performance [141]. Dans une perspective complémentaire, Johora et al. identifient les cas récurrents de mauvaise classification en s'appuyant sur une estimation de l'incertitude appliquée aux prédictions de réussite [84]. Hooshyar et al. montrent quant à eux comment l'incertitude peut être exploitée pour déclencher l'intervention d'un tuteur humain dans une architecture hybride associant intelligence artificielle et expertise humaine [80]. Enfin, Agyemang et al. soulignent que, malgré leur potentiel pour la personnalisation et la détection de cas-limites, les mécanismes d'incertitude demeurent encore peu intégrés dans les systèmes éducatifs réellement déployés [7].

Ces études convergent vers un constat commun : dans les contextes éducatifs où les décisions ont un impact direct sur les parcours des apprenants (alerte, soutien, réorientation), l'intégration de l'incertitude constitue un levier essentiel pour renforcer à la fois la fiabilité et l'éthique des modèles prédictifs.

2.2.8 Discussion

Si la littérature a introduit certaines mesures temporelles de fiabilité, comme l'*Earliness Stability Score (ESS)* [29], les travaux existants en éducation se limitent encore à des métriques d'incertitude essentiellement ponctuelles, qui fournissent une estimation instantanée de la confiance du modèle [135, 94]. Or, dans un contexte où l'apprentissage est un processus continu, l'incertitude devrait également être appréhendée comme une dynamique évolutive, susceptible de varier selon les semaines de cours, la nature des activités ou le volume de données disponibles.

L'absence de métriques temporelles d'incertitude entraîne deux limites principales. Premièrement, elle ne permet pas d'évaluer si la confiance d'un modèle se maintient de manière cohérente ou fluctue au fil du temps. Deuxièmement, elle restreint la conception de systèmes capables de déclencher des interventions conditionnelles en fonction non seulement d'un score de risque, mais aussi de l'évolution de l'incertitude. Ainsi, un modèle peut présenter des performances globalement satisfaisantes tout en produisant des prédictions instables ou contradictoires qui réduisent sa valeur pratique pour les enseignants et les apprenants.

Comparée à d'autres domaines sensibles comme la santé ou les transports autonomes, où des métriques temporelles d'incertitude commencent à émerger [4, 82], la recherche en éducation reste encore en retrait sur cette question. Développer des métriques intégrant à la fois *précocité*, *stabilité* et *incertitude* apparaît dès lors comme une étape clé pour renforcer la fiabilité et l'utilité des modèles prédictifs dans des environnements éducatifs réels.

Résumé

La fiabilité et l'incertitude apparaissent comme deux dimensions complémentaires mais encore peu exploitées dans les modèles prédictifs en éducation. La fiabilité concerne la stabilité des prédictions dans le temps, tandis que l'incertitude exprime le degré de confiance associé à chaque décision. Or, la plupart des travaux restent centrés sur des mesures ponctuelles, sans suivi longitudinal. Développer des métriques temporelles d'incertitude constitue ainsi une étape clé pour renforcer la transparence et l'utilité pratique des systèmes prédictifs.

2.3 Frugalité des modèles et des données

Nous abordons à présent une autre dimension structurante des systèmes prédictifs en éducation : la frugalité. Celle-ci renvoie à la capacité de concevoir des modèles économes en données, en calcul et en ressources, tout en maintenant un niveau de performance adapté aux objectifs pédagogiques.

2.3.1 Définition

La frugalité en intelligence artificielle, ou *Frugal AI*, désigne un ensemble de pratiques visant à concevoir des modèles prédictifs performants tout en minimisant l'usage des ressources disponibles. Cette approche vise un équilibre entre efficacité algorithmique, efficience énergétique, accessibilité matérielle et sobriété environnementale [17, 21]. Elle s'oppose à la logique dominante du *bigger is better* et s'inscrit dans une dynamique éthique, soutenable et souvent plus réaliste dans les environnements à ressources contraintes.

En apprentissage supervisé, un modèle frugal ne cherche pas la performance maximale en toutes circonstances, mais une performance *suffisante* obtenue avec un minimum de données, de calculs, de mémoire et d'énergie [95]. Cette perspective rejoint des préoccupations croissantes en matière d'impact environnemental du numérique, de généralisation hors contexte, et de réduction des biais liés à la centralisation excessive des données.

2.3.2 Axes de frugalité

La littérature identifie trois axes principaux structurant la frugalité en IA [21, 165] :

— **Frugalité des données (input frugality)** : elle vise à limiter la quantité d'exemples et de caractéristiques nécessaires à l'apprentissage. Cela passe par des méthodes de sélection de caractéristiques, de réduction dimensionnelle, d'acquisition active ou de raisonnement adaptatif sur les sources.

— **Frugalité du processus d'apprentissage** : elle concerne l'optimisation des ressources requises pour entraîner un modèle (CPU, GPU, RAM, temps). Elle inclut les techniques d'apprentissage incrémental ou d'apprentissage par transfert optimisé.

— **Frugalité du modèle** : elle vise à concevoir des architectures compactes (réseaux peu profonds, quantification, compression, distillation), capables d'inférer rapidement avec une faible empreinte mémoire et énergétique.

La frugalité est aujourd'hui reconnue comme une orientation stratégique à part entière du développement de l'intelligence artificielle, portée par des collectifs académiques, industriels et institutionnels [17, 21].

2.3.3 Réduction des données nécessaires

L'un des leviers de la frugalité consiste à apprendre efficacement à partir d'un volume limité de données. Cette approche, au cœur de la frugalité algorithmique, vise à maximiser l'information extraite de chaque exemple, en s'appuyant sur des stratégies de sélection, d'échantillonnage ou de transfert de connaissances.

— **Sélection de caractéristiques pertinentes** : Les travaux récents ont montré que la réduction du nombre de caractéristiques en entrée constitue une voie majeure d'économie

de données, ces approchaes n'ont pas été initialement conçues pour assurer une frugalité mais viennent d'un besoin de sélection des caractéristiques les plus pertinentes. Zhou et al. [178] proposent, par exemple, une méthode de sélection dynamique de caractéristiques fondée sur l'information mutuelle et la décision en ligne. Cette approche permet d'évaluer en continu la contribution de chaque caractéristique à la tâche de prédiction, en ajustant le sous-ensemble de caractéristiques utilisées selon leur pertinence et leur redondance. Elle s'inscrit dans une logique d'auto-optimisation du flux de données.

— **Apprentissage actif** : Dans un contexte où l'annotation des données peut être coûteuse, l'apprentissage actif consiste à laisser le modèle sélectionner les exemples les plus informatifs à étiqueter. Deschamps et al. [55] montrent que cette approche permet de réduire jusqu'à 70% le besoin d'étiquetage dans des tâches de classification, tout en conservant une précision comparable à un apprentissage exhaustif. En combinant exploration et exploitation, le modèle apprend ainsi à prioriser les données les plus utiles à sa propre amélioration.

— **Apprentissage à partir de peu d'exemples** : Le *few-shot learning* et ses variantes, comme les réseaux prototypiques, visent à généraliser à partir d'un très petit nombre d'exemples par classe. Lin et al. [102] illustrent cette approche dans un contexte de reconstruction visuelle, en intégrant des connaissances préalables sous forme de priors neuronaux. Cette logique d'apprentissage par analogie permet d'exploiter la similarité structurelle entre tâches pour réduire la dépendance à de larges ensembles de données annotées.

2.3.4 Réduction des ressources d'entraînement

La frugalité ne se limite pas aux données : elle concerne aussi la réduction du coût computationnel associé à l'apprentissage. L'objectif est d'optimiser l'usage des ressources matérielles (processeur, mémoire, énergie) tout en maintenant la performance du modèle.

— *Apprentissage incrémental* : Comme le montrent Cholet et al. [44], l'apprentissage incrémental permet de mettre à jour le modèle au fil du temps sans réentraînement global. Cette approche est particulièrement utile pour les environnements évolutifs, où les données arrivent en flux. En ne traitant que les échantillons nouveaux ou pertinents, le modèle réduit considérablement la charge de calcul et le stockage nécessaires, tout en maintenant la continuité des connaissances acquises.

— *Apprentissage par transfert optimisé* : Zhuang et al. [179] soulignent qu'il est souvent inutile de réentraîner l'intégralité d'un réseau pré-entraîné. Le *fine-tuning* sélectif, consistant à n'ajuster que les couches finales ou spécifiques à la tâche cible, permet de réduire jusqu'à 80% du coût de calcul par rapport à un entraînement complet. Cette approche est au cœur des stratégies frugales d'adaptation de modèles existants à de nouveaux contextes.

— *Compression d'optimiseur et mini-batching adaptatif* : Les travaux de Zmushko et al. [180] introduisent des variantes frugales des optimisateurs classiques (comme Adam ou SGD), capables d'adapter dynamiquement la taille des lots ou de compresser les gradients transmis. Ces techniques réduisent la redondance dans les mises à jour des paramètres et permettent d'entraîner des modèles complexes sur des machines à faible capacité de calcul.

2.3.5 Modèles légers et déployables

Enfin, la frugalité se traduit par la conception de modèles à la fois compacts, efficaces et adaptés à des environnements contraints en ressources (appareils mobiles, capteurs éducatifs ou plateformes embarquées).

— *Distillation de connaissances (Knowledge distillation)* : Cette approche consiste à transférer les connaissances d'un modèle complexe (*teacher*) vers un modèle plus léger (*student*). Moslemi et al. [111] montrent que la distillation permet de conserver plus de 90% de la performance initiale tout en divisant la taille du modèle par dix. Le modèle léger apprend à reproduire les probabilités de sortie du modèle enseignant, ce qui favorise une compression efficace sans perte majeure d'information.

— *TinyML et Edge AI* : Dans la continuité des approches frugales, TinyML vise à déployer des modèles d'apprentissage directement sur des microcontrôleurs ou des dispositifs à très faible consommation énergétique. Abadade et al. [2] présentent des applications dans des environnements déconnectés, où les modèles doivent fonctionner sans accès au cloud. Singh et al. [136] montrent que l'Edge AI, en rapprochant le calcul de la source de données, permet à la fois de réduire la latence, les coûts de transfert et l'empreinte énergétique, tout en renforçant la confidentialité des données.

2.3.6 Applications en éducation

Les travaux consacrés à la frugalité en éducation restent pour l'instant marginaux et concernent davantage la gestion des ressources numériques ou l'optimisation de dispositifs techniques que les modèles prédictifs eux-mêmes. À notre connaissance, aucune proposition spécifique n'a été formulée pour intégrer la frugalité comme principe de conception dans les systèmes de prédiction éducative. Les approches existantes se limitent le plus souvent à réduire la complexité des algorithmes [103, 117] ou à sélectionner un sous-ensemble de données pour des raisons pratiques [168, 20], sans élaborer un cadre méthodologique dédié. Ce constat souligne l'originalité et la nécessité d'explorer la frugalité comme un axe de recherche à part entière pour concevoir des modèles prédictifs à la fois efficaces, sobres et adaptés aux contraintes des environnements d'apprentissage réels.

2.3.7 Discussion

La frugalité constitue un axe de recherche émergent qui répond à des enjeux techniques, éthiques et pédagogiques. Dans le champ de l'éducation, elle permettrait de concevoir des modèles prédictifs capables de fonctionner dans des environnements contraints, tout en limitant l'usage de données sensibles et en réduisant l'empreinte computationnelle. Pourtant, les travaux actuels restent fragmentaires. La plupart privilégient des ajustements techniques, comme l'emploi de modèles plus simples ou interprétables [103, 117], ou la réduction du volume de données mobilisées par des méthodes de sélection [168], sans proposer un cadre méthodologique global.

Cette situation révèle deux limites majeures : d'une part, l'absence de métriques pour évaluer la frugalité dans le contexte éducatif ; d'autre part, la difficulté à concilier sobriété des modèles et maintien d'un niveau de performance suffisant pour guider les décisions pédagogiques. En ce sens, la frugalité ne devrait pas être perçue comme un simple compromis de performance, mais comme une véritable philosophie de conception visant à produire des systèmes prédictifs à la fois efficaces, accessibles et soutenables.

Résumé

La frugalité ne se réduit pas à une contrainte technique : elle constitue une approche de conception adaptée aux contextes éducatifs réels. Elle favorise l'accessibilité, et la robustesse des modèles, tout en garantissant leur durabilité. Mais elle implique aussi des compromis, invitant à repenser les objectifs et à concevoir des systèmes adaptatifs capables de s'ajuster aux contraintes de données et de ressources.

2.4 Sélection dynamiques des caractéristiques

Nous explorons à présent la question de la sélection dynamique des caractéristiques, un mécanisme clé pour construire des modèles plus adaptatifs, plus économes et mieux ajustés à la variabilité des données dans les environnements d'apprentissage numérique.

2.4.1 Définition

La sélection dynamique des caractéristiques *(Dynamic Feature Selection, DFS*[6]*)* désigne un ensemble de techniques qui permettent à un modèle d'apprentissage de choisir de manière adaptative les caractéristiques les plus utiles à un instant donné, pour une instance donnée [173, 15, 176]. Contrairement aux approches classiques où un sous-ensemble fixe de caractéristiques est déterminé une fois pour toutes avant l'apprentissage (feature selection statique), la DFS procède par interrogation progressive : les caractéristiques sont sélectionnées séquentiellement en fonction de leur contribution attendue à la réduction de l'incertitude ou de l'erreur prédictive.

Elle s'inscrit dans une logique de décision conditionnelle, proche des processus de test séquentiel, tenant compte à la fois des données déjà observées et du coût potentiel d'acquisition de nouvelles informations [52, 169]. Ce paradigme est particulièrement pertinent dans des environnements contraints en ressources ou dans des systèmes temps réel, comme le streaming éducatif, la surveillance médicale ou les diagnostics embarqués.

2.4.2 Classification des caractéristiques

La littérature distingue généralement les caractéristiques selon leur pertinence prédictive et leur redondance, ce qui permet de mieux orienter les stratégies de sélection. Cette typologie, inspirée des travaux de Yu & Liu [170], repose sur une analyse fine des relations entre caractéristiques et cible (présentée dans la Figure 2.3), et se structure en quatre catégories principales :

1. **Caractéristiques non pertinentes** : elles ne présentent aucune corrélation directe ou indirecte avec la variable cible. Leur inclusion introduit du bruit et dégrade les performances générales du modèle.
2. **Caractéristiques faiblement pertinentes et redondantes** : bien qu'elles contiennent un minimum d'information utile, leur contribution est largement couverte par d'autres caractéristiques. Elles augmentent inutilement la complexité du modèle.
3. **Caractéristiques faiblement pertinentes mais non redondantes** : elles apportent une information originale, souvent marginale, qui peut améliorer la performance dans certaines conditions, notamment pour des cas atypiques.
4. **Caractéristiques fortement pertinentes** : elles sont directement corrélées avec la variable cible, avec un pouvoir prédictif élevé. Ce sont des caractéristiques clés dans toute tâche de classification ou de régression supervisée.

La combinaison des catégories (3) et (4) est généralement considérée comme constituant un sous-ensemble optimal, car elle assure à la fois la pertinence de l'information et l'absence de redondance superflue. Cette distinction joue un rôle essentiel dans les algorithmes de sélection basés sur l'information mutuelle, les techniques de *Markov blankets* ou les modèles de coût adaptatif [114, 100].

6. Nous utiliserons cet acronyme dans le reste du manuscrit.

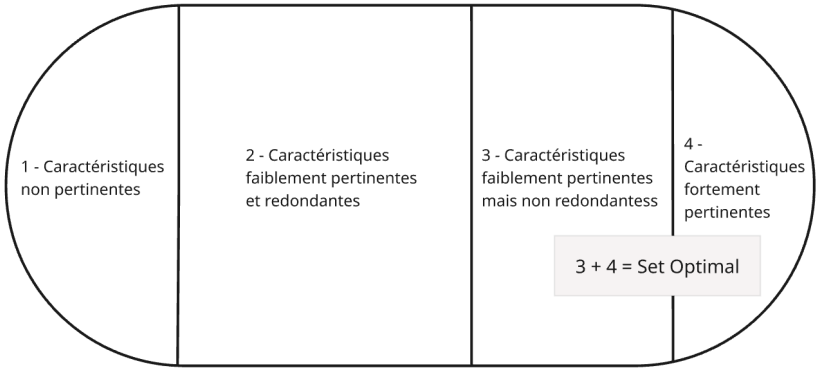

FIGURE 2.3 – Typologie des caractéristiques selon leur pertinence et redondance [170]

2.4.3 Panorama des approches existantes

Les méthodes de sélection dynamique des caractéristiques se déclinent en plusieurs familles théoriques, chacune proposant des mécanismes spécifiques pour identifier, à la volée, les caractéristiques les plus utiles à la prédiction. Ces approches se distinguent par leur niveau de complexité, leur efficacité computationnelle, leur capacité d'interprétation, et leur impact sur la performance finale. Nous présentons ici les cinq grandes classes de techniques les plus représentatives.

Le Tableau 2.4 synthétise les principales caractéristiques de ces approches selon cinq dimensions clés : complexité, temps d'entraînement, explicabilité, et gain de précision attendu. Cette lecture comparative permet d'éclairer les compromis méthodologiques à considérer lors de l'implémentation d'un système de DFS en contexte éducatif.

TABLE 2.4 – Comparaison des principales approches de sélection dynamique des caractéristiques

Famille de méthode	**Complexité algorithmique**	**Temps d'entraînement**	**Niveau d'explicabilité**	**Gain en précision**
Théorie du Markov Blanket	$O(n^3)$ à $O(n^4)$	Modéré (filtrage statistique)	Élevé (structure causale explicite)	Moyen à élevé
Algorithmes évolutionnaires	$O(g \times n \times m)$ à $O(g \times n^2)$	Élevé (générations multiples)	Moyen (exploration heuristique)	Élevé (optimisation globale)
Théorie de l'information	$O(n^2)$ à $O(n^3)$	Modéré	Élevé (mesures interprétables)	Élevé (bonne balance pertinence/redondance)
Ensembles approximatifs (Rough Sets)	$O(n^2)$	Moyenne à élevée	Élevé (logique de règles claire)	Moyen
Apprentissage par renforcement	$O(T \times S \times A)$ ou $O(T \times n^2)$	Élevé (processus séquentiel complexe)	Faible à moyen	Élevé (stratégie optimisée conditionnellement)

Remarque : n = nombre de caractéristiques, m = taille de l'échantillon, g = nombre de générations (dans les algorithmes évolutionnaires), T = nombre d'itérations d'apprentissage, S = nombre d'états, et A = nombre d'actions considérées dans le processus d'apprentissage par renforcement.

Théorie de la couverture de Markov

Les méthodes basées sur le *Markov Blanket* cherchent à identifier un sous-ensemble minimal de caractéristiques conditionnellement suffisantes pour prédire la variable cible, en exploitant les dépendances probabilistes entre caractéristiques. Ce cadre théorique, inspiré des réseaux bayésiens, permet d'isoler les caractéristiques causales les plus pertinentes tout en éliminant les redondances [120, 162]. Elles offrent une forte explicabilité grâce à une interprétation causale des relations entre les caractéristiques et la classe [122]. Mais, elles supposent une estimation fiable des dépendances conditionnelles, ce qui devient complexe sur de grandes dimensions ou en présence de bruit [155, 140].

Algorithmes évolutionnaires

Les approches évolutionnaires (comme les algorithmes génétiques) explorent l'espace des sous-ensembles de caractéristiques par sélection naturelle, en optimisant une fonction objectif liée à la performance prédictive [114]. Elles sont capables d'identifier des solutions globalement optimales, même dans des espaces complexes et non linéaires. Elles excellent en optimisation globale et sont robustes aux corrélations complexes [56], mais leur coût computationnel est élevé (multiples générations, évaluations redondantes), et leur logique de sélection est moins interprétable [140].

Méthodes fondées sur la théorie de l'information

Très répandues dans la DFS, ces approches s'appuient sur l'information mutuelle pour estimer la pertinence et la redondance des caractéristiques. Les algorithmes tels que SAOLA, OSFSMI ou OFS-Density utilisent des seuils pour conserver les caractéristiques les plus informatives et éliminer les redondantes [169, 52, 176]. Les métriques utilisées sont mathématiquement robustes, interprétables, et efficaces dans les environnements de flux, mais leur sensibilité aux déséquilibres de classes et aux caractéristiques corrélées mal identifiées peut réduire leur efficacité dans certains contextes éducatifs.

Théorie des ensembles approximatifs *(Rough Set Theory)*

La théorie des ensembles approximatifs propose une approche logique et transparente pour sélectionner les caractéristiques les plus discriminantes. Elles s'appuient sur des règles d'approximation pour distinguer les éléments appartenant avec certitude ou probabilité à une classe donnée [18, 166, 177]. Un avantage est qu'elles génèrent des règles interprétables, même sans information probabiliste ou métrique. Malheureusement, les performances peuvent se dégrader sur de très grands ensembles ou en présence de bruit, et son usage reste marginal dans les systèmes en ligne.

Apprentissage par renforcement

Plus récemment, des travaux ont formulé la sélection dynamique de caractéristiques comme un problème de décision séquentielle. Le modèle apprend à sélectionner une variable à chaque étape pour maximiser une récompense, typiquement liée à l'amélioration de la précision ou à la réduction du coût. Cette approche est particulièrement adaptée aux environnements interactifs [164]. Elle offre une forte capacité d'adaptation et des performances optimisées, mais la formation de l'agent est complexe, nécessite beaucoup de données et souffre encore d'un manque d'explicabilité.

2.4.4 Limites identifiées

Malgré leur potentiel adaptatif, les méthodes de sélection dynamique des caractéristiques présentent plusieurs limitations structurelles lorsqu'elles sont transposées au contexte de la prédiction éducative.

— **Insensibilité aux déséquilibres inter-classes** : la plupart des algorithmes de sélection dynamique, notamment ceux fondés sur l'information mutuelle (MI), calculent des scores globaux sans prendre en compte l'impact différencié d'une variable sur chaque classe. Cette approche pose problème dans les contextes fortement déséquilibrés, comme le décrochage scolaire, où les apprenants à risque (classe minoritaire) sont sous-représentés. Même les tentatives de prise en compte de l'équité, telles que les approches dites de *Fair DFS* de Duan et al. [58], restent souvent centrées sur la protection de caractéristiques sensibles (genre, origine, etc.) et n'adressent pas pleinement le déséquilibre structurel dans les données d'apprentissage.

— **Rigidité des seuils de sélection** : de nombreuses approches reposent sur des seuils statiques pour décider de la pertinence ou de la redondance des caractéristiques. Cette rigidité entrave l'adaptation du modèle à des évolutions contextuelles : une variable pertinente en début de parcours peut perdre sa valeur prédictive au fil du temps, sans être automatiquement désactivée [169, 124].

— **Absence de traitement du *feature drift*** : très peu de méthodes intègrent un mécanisme explicite de détection ou de repondération des caractéristiques dont la contribution évolue négativement avec le temps. Ce phénomène, appelé *drift* des caractéristiques, correspond à une modification du lien statistique entre une caractéristique et la cible, due à un changement de comportement, de stratégie d'apprentissage ou de contexte. Son absence de prise en compte peut entraîner une accumulation de caractéristiques devenues obsolètes, alourdissant inutilement le modèle et dégradant sa précision [85].

— **Complexité algorithmique et coût de déploiement** : les approches plus expressives, comme celles fondées sur l'apprentissage par renforcement, ou sur les algorithmes évolutionnaires, offrent des capacités d'adaptation accrues mais au prix d'une complexité computationnelle élevée, d'un temps d'entraînement prolongé, et d'un besoin de paramétrage fin. Ces exigences rendent leur intégration difficile dans des plateformes éducatives en ligne soumises à des contraintes de ressources ou de réactivité [164, 56].

— **Opacité des méthodes** : plusieurs stratégies de sélection dynamique, notamment celles basées sur des réseaux de neuronnes profonds, souffrent d'un manque d'explicabilité. Cette "effet boîte noire" complique l'analyse des décisions prises par le modèle et limite leur acceptabilité dans des contextes pédagogiques où la transparence et la justification des interventions sont essentielles [15, 126].

Ces limites soulignent la nécessité de concevoir des approches plus sensibles à la dynamique temporelle des caractéristiques, au déséquilibre entre les classes, à l'évolution contextuelle, et aux contraintes d'implémentation réelles. Elles justifient également l'exploration de modèles hybrides, capables de combiner explicabilité, adaptativité et efficacité algorithmique.

2.4.5 Applications de la sélection dynamique en éducation

Les approches de sélection de caractéristiques appliquées aux données éducatives se situent aujourd'hui à la frontière entre deux logiques : d'une part des méthodes encore largement *statiques*, et d'autre part des propositions plus récentes qui introduisent une véritable dimension *dynamique*.

Dans la logique statique, plusieurs travaux se concentrent sur la construction et l'optimisation d'un ensemble de caractéristiques retenues une fois pour toutes. Qiu et al. [123] proposent par exemple un cadre combinant génération, sélection et exploitation de caractéristiques pour prédire l'abandon dans les MOOCs. Youssef et al. [168] recourent à une sélection efficace des attributs associée à la compétition entre algorithmes, tandis qu'Ardchir et al. [20] présentent une méthode optimisée pour améliorer la prédiction du décrochage. Ces travaux soulignent l'importance de réduire la redondance et de limiter la complexité des modèles, mais ils restent conçus dans une perspective figée, adaptée à un jeu de données particulier.

À l'inverse, d'autres recherches s'orientent vers des approches plus dynamiques ou flexibles. Al-Zawqari et al. [10] montrent qu'il est possible d'abandonner le *feature engineering* manuel au profit d'une sélection fondée sur l'interprétabilité des modèles, ce qui rend les prédictions plus transférables entre cours et environnements. Yu et al. [171] introduisent une approche semi-supervisée capable de tirer parti de données partiellement étiquetées, facilitant une adaptation progressive aux conditions réelles. Enfin, Malik et al. [108] vont plus loin avec un modèle de sélection dynamique basé sur l'évolution d'ensembles de caractéristiques, intégrant un ajustement adaptatif des seuils de pertinence selon l'évolution des données.

Ainsi, si les méthodes statiques ont permis des avancées significatives dans la prédiction de la réussite et du décrochage, les approches dynamiques ouvrent la voie à des systèmes plus flexibles, capables de s'ajuster à la nature évolutive des traces d'apprentissage. Elles restent toutefois encore peu nombreuses et appellent à une formalisation plus systématique dans le contexte éducatif.

2.4.6 Discussion

Si la sélection dynamique des caractéristiques offre un potentiel important, son application au domaine éducatif présente encore plusieurs limites spécifiques. Dans la littérature générale sur la DFS, de nombreux travaux insistent sur la difficulté à gérer le déséquilibre entre classes, la rigidité des seuils de sélection et l'absence de mécanismes explicites pour traiter le *drift* des caractéristiques au fil du temps [169, 52, 173]. Ces limites sont particulièrement problématiques dans l'éducation, où les apprenants à risque de décrochage représentent une minorité cruciale à identifier, et où la pertinence des traces évolue fortement selon la progression du cours.

Du côté des applications éducatives, plusieurs approches restent ancrées dans une logique statique. Qiu et al. [123], Youssef et al. [168] et Ardchir et al. [20] montrent l'intérêt de réduire la redondance des caractéristiques et de simplifier les modèles, mais leurs méthodes ne prennent pas en compte la nature séquentielle et évolutive des données issues des environnements virtuels d'apprentissage. Même les propositions plus récentes, comme celle d'Al-Zawqari et al. [10], qui supprime le *feature engineering* manuel au profit d'une sélection fondée sur l'interprétabilité, ou celle de Yu et al. [171], qui introduit une approche semi-supervisée, ne permettent pas de répondre pleinement aux contraintes temporelles et à l'adaptation en continu. Enfin, l'approche de Malik et al. [108], qui introduit une sélection dynamique par évolution d'ensembles de caractéristiques, reste limitée par sa complexité computationnelle et par le manque de validations

dans des contextes éducatifs diversifiés.

Ainsi, malgré des avancées réelles, les travaux actuels ne répondent pas encore aux défis propres à l'éducation : gestion des déséquilibres structurels (apprenants à risque), adaptation aux données en flux et explicabilité suffisante pour être intégrés dans des dispositifs pédagogiques. La sélection dynamique en éducation demeure donc un champ à explorer, nécessitant de dépasser les limites des approches existantes pour proposer des solutions adaptées aux environnements d'apprentissage numériques.

Résumé

La sélection dynamique des caractéristiques apparaît comme un levier essentiel pour construire des modèles éducatifs plus adaptatifs et économes. Alors que la majorité des travaux existants restent centrés sur des approches statiques [123, 168, 20], quelques contributions récentes introduisent davantage de flexibilité [10, 171, 108]. Toutefois, ces méthodes ne répondent pas encore pleinement aux défis spécifiques de l'éducation : gestion des données en flux, déséquilibres entre classes et nécessité d'une explicabilité pédagogique. La sélection dynamique demeure ainsi un champ en émergence, dont la formalisation reste à approfondir pour les environnements d'apprentissage réels.

2.5 Apprentissage incrémental

Nous explorons à présent la question de l'apprentissage incrémental, un mécanisme clé pour développer des modèles capables de s'adapter en continu aux nouvelles données produites dans les environnements numériques. Contrairement aux approches traditionnelles, qui nécessitent de réentraîner le modèle sur l'ensemble du jeu de données, l'apprentissage incrémental permet une mise à jour progressive, plus économe en ressources et mieux ajustée à l'évolution des comportements des apprenants.

2.5.1 Définition

L'apprentissage incrémental, également appelé *incremental learning*, *continual learning* ou parfois *online learning*, désigne un paradigme dans lequel un modèle est entraîné de manière progressive, en intégrant de nouvelles données à mesure qu'elles deviennent disponibles, sans nécessiter une remise à zéro complète ou un réentraînement total du modèle sur l'ensemble des données passées [53, 154]. Cette approche permet de concevoir des systèmes adaptatifs, capables d'apprendre continuellement dans des environnements dynamiques, sous contraintes de ressources, ou dans des contextes où la totalité des données ne peut être stockée ou exploitée simultanément. Elle se distingue des approches dites *batch* par son caractère itératif, temporellement sensible et potentiellement illimité.

Plusieurs sous-familles ou variantes d'apprentissage incrémental sont identifiées dans la littérature. Van de Ven et al. [147] proposent une typologie en trois catégories :

— **Task-Incremental Learning** : les tâches sont distinctes dans le temps (par exemple, des prédictions sur différents domaines) et le modèle sait à quelle tâche correspond chaque donnée. Ce cadre convient bien à l'adaptation séquentielle à des contextes spécifiques, mais reste éloigné de la réalité des systèmes continus sans frontières nettes entre tâches.

— **Class-Incremental Learning** : les classes cibles ne sont pas connues dès le départ, elles apparaissent progressivement, et le modèle doit être capable de les reconnaître et les distinguer sans supervision explicite. Ce scénario est courant dans la vision artificielle ou la détection d'anomalies.

— **Domain-Incremental Learning** : la tâche reste constante, mais les distributions de données évoluent dans le temps (nouvelles modalités, changements de contexte, etc.). C'est notamment le cas dans les environnements éducatifs, où le comportement des apprenants change avec le temps.

Dans notre cas, l'intérêt se porte spécifiquement sur le **domain-incremental learning**, parfois qualifié de *data-incremental learning*, où les classes sont fixées, la tâche de prédiction est définie (par exemple : prédire l'abandon), et seule la base de données s'enrichit progressivement avec le temps. Ce scénario est particulièrement pertinent dans les systèmes d'apprentissage en ligne, où les interactions se succèdent au fil des semaines ou des modules, et où les modèles doivent s'adapter sans être réentraînés à chaque nouvelle trace d'activité.

2.5.2 Limites et défis actuels

Malgré ses atouts, l'apprentissage incrémental soulève plusieurs défis majeurs :

— **L'oubli catastrophique** : c'est le phénomène le plus documenté. Le modèle tend à oublier les connaissances acquises précédemment lorsqu'il est exposé à de nouvelles données, surtout si celles-ci ne représentent qu'une sous-partie du domaine global [53, 96]. Des stratégies telles que le *replay* (représenter les anciennes données via un buffer) ou la consolidation des poids (*elastic weight consolidation*) ont été proposées, mais leur efficacité dépend du contexte [106].

— **Le *concept drift*** : bien que l'apprentissage incrémental soit conçu pour gérer des environnements évolutifs, il demeure particulièrement vulnérable aux dérives soudaines dans la distribution des données. Ces drifts peuvent refléter des transformations profondes dans les comportements, les contextes d'interaction ou les configurations d'apprentissage, rendant obsolètes certaines relations précédemment apprises [14, 42]. Des approches récentes, comme DYSON [78], s'attachent à répondre à ce défi en proposant des mécanismes de réorganisation dynamique de l'espace des caractéristiques, permettant ainsi au modèle de s'adapter de manière continue aux changements de structure latente dans les flux de données.

— **Le *feature drift*** : l'évolution du jeu de caractéristiques et des caractéristiques lui-même constitue un défi. L'arrivée ou la disparition de certaines caractéristiques dans le flux nécessite des modèles capables de s'adapter structurellement [42, 44].

— **Le surapprentissage local** : l'incrémentation excessive peut conduire à une sensibilité accrue aux derniers exemples reçus, au détriment de la généralisation. Cela survient notamment si les données récentes ne sont pas représentatives de la tendance globale [158].

— **Le moment du déclenchement des mises à jour** : dans la plupart des cadres, les mises à jour sont appliquées à chaque nouvelle donnée. Or, cela peut s'avérer inutile ou coûteux si la donnée n'est pas informative. Ce problème soulève la question du *retraining triggering*, encore peu explorée dans la littérature, et que nous identifierons comme une piste méthodologique clé [67, 107].

2.5.3 Applications en contexte éducatif

Dans le domaine éducatif, l'apprentissage incrémental reste relativement peu exploré, malgré son adéquation structurelle avec la temporalité des parcours apprenants. Labba & al. (2022, 2024) [93, 92] soulignent l'importance de choisir le bon moment pour réentraîner un modèle prédictif dans un système d'alerte, en fonction des données nouvellement disponibles. Ils proposent des heuristiques de déclenchement basées sur des mesures de variation du profil d'un étudiant. Ces travaux posent les bases d'une approche plus parcimonieuse et sensible aux contextes dans lesquels les traces évoluent progressivement. De même, Casalino et al. [38] ou Allogmany et al. [14] mettent en avant l'intérêt de mécanismes d'apprentissage en ligne pour suivre l'évolution des comportements d'apprenants, en intégrant des méthodes de clustering adaptatif ou de détection de drift.

2.5.4 Discussion

L'apprentissage incrémental offre une réponse naturelle à la temporalité des données éducatives, mais son application concrète reste encore limitée. La littérature générale a largement documenté des problèmes comme l'oubli catastrophique [53, 96], la vulnérabilité aux dérives conceptuelles [42, 78], ou encore la difficulté à gérer l'apparition et la disparition de caractéristiques [44]. Ces limites sont particulièrement sensibles en éducation, où les données sont hétérogènes, bruitées et soumises à de fortes variations contextuelles selon les cours ou les cohortes d'apprenants.

Dans les applications éducatives, quelques travaux pionniers explorent des approches incrémentales [38, 14, 92, 93], mais ils se concentrent surtout sur le suivi temporel des performances ou sur la détection de variations dans les profils. Ces propositions soulignent l'importance de choisir le bon moment pour mettre à jour un modèle, mais elles restent encore limitées. En particulier, la question du *retraining triggering*, bien identifiée dans la littérature générale [67, 107], n'a pas encore trouvé de déclinaison méthodologique robuste dans le domaine éducatif.

Enfin, une limite centrale concerne l'absence de couplage avec la sélection dynamique des caractéristiques. Alors que l'apprentissage incrémental vise à adapter le modèle aux nouvelles données, il reste souvent aveugle à la pertinence réelle des traces intégrées. Or, sans mécanisme pour filtrer ou hiérarchiser les données informatives, les mises à jour risquent de renforcer le bruit ou de provoquer un surapprentissage local [158].

Ainsi, si l'apprentissage incrémental apparaît comme un cadre prometteur pour des systèmes prédictifs plus souples et économes, son potentiel reste encore sous-exploité en éducation. Son efficacité dépendra de la capacité à articuler trois dimensions encore peu reliées : la gestion du moment des mises à jour, l'intégration de mécanismes de sélection dynamique des données, et la préservation de la stabilité du modèle face aux évolutions rapides des comportements d'apprentissage.

Résumé

L'apprentissage incrémental constitue une approche prometteuse pour adapter les modèles prédictifs aux flux continus de données en éducation. Il permet de réduire les coûts de réentraînement et d'accompagner l'évolution des comportements apprenants. Toutefois, les défis de l'oubli catastrophique, du drift conceptuel, du surapprentissage local et du déclenchement pertinent des mises à jour restent peu explorés dans ce domaine. De plus, l'absence de couplage avec la sélection dynamique des caractéristiques limite aujourd'hui son efficacité. Son intégration future repose donc sur une articulation fine entre stabilité, pertinence des données et adaptabilité en temps réel.

Conclusion de l'état de l'art

L'ensemble de cette revue a permis de mettre en lumière les dynamiques actuelles de recherche à l'intersection des environnements virtuels d'apprentissage et des systèmes prédictifs alimentés par des traces numériques. La première partie s'est attachée à décrire les caractéristiques des VLEs, les formes de données qu'ils génèrent, et les approches de modélisation de l'apprenant qu'ils permettent de déployer. Elle a montré combien ces environnements offrent un potentiel analytique riche, tout en révélant des limites structurelles importantes : données dispersées, profils incomplets, absence d'actualisation temporelle des modèles.

Dans un second temps, l'analyse s'est orientée vers quatre leviers méthodologiques transversaux susceptibles de répondre à ces défis : la fiabilité, l'incertitude, la frugalité, la sélection dynamique des caractéristiques et l'apprentissage incrémental. Ces approches, bien qu'issues de directions différentes, partagent une même volonté d'adapter les systèmes aux contraintes réelles : flux continus de données, ressources limitées, diversité des trajectoires, et nécessité d'agir avec prudence face aux cas incertains. Chacune a été analysée dans sa définition, ses déclinaisons, ses apports et ses limites, en lien étroit avec les spécificités du contexte éducatif.

Ce panorama met en évidence un manque d'articulation entre ces différents leviers. Peu de travaux explorent les synergies possibles entre l'estimation de l'incertitude, la sélection conditionnelle de données, et la mise à jour progressive des modèles. Or, dans des systèmes éducatifs où chaque prédiction peut avoir des conséquences concrètes, il devient essentiel de concevoir des modèles capables non seulement de faire des choix informés, mais aussi de reconnaître ce qu'ils ne savent pas, de solliciter l'information juste, et d'apprendre au fil de l'eau, sans recommencer depuis zéro.

C'est dans cet espace que se situe notre contribution et à partir des constats dégagés dans cet état de l'art, les chapitres suivants chercheront à définir un cadre méthodologique combinant incertitude, frugalité, sélection dynamique et apprentissage incrémental, dans une logique d'adaptation continue et raisonnée aux données d'apprentissage.

Chapitre 3

Jeux de données de la littérature en éducation

L'étude des comportements d'apprentissage et la prédiction de la réussite des étudiants s'appuient sur une grande variété de jeux de données, issus aussi bien de plateformes publiques que de dispositifs institutionnels ou expérimentaux. Ces jeux de données sont largement exploités dans la recherche en **fouille de données éducatives** (*Educational Data Mining*, EDM) et en **analyse des apprentissages** (*Learning Analytics*, LA). Ils proviennent le plus souvent d'environnements virtuels d'apprentissage (VLEs) ou de plateformes de cours en ligne, et combinent plusieurs types d'informations : traces d'interaction des apprenants avec les ressources, résultats d'évaluation, données contextuelles ou démographiques.

Ils constituent un socle essentiel pour le développement, la validation et la comparaison des modèles prédictifs appliqués à l'éducation. Cette section présente d'une part les principaux jeux de données de référence utilisés dans la littérature scientifique, et d'autre part ceux retenus pour les expérimentations réalisées dans le cadre de cette thèse.

3.1 Jeux de données de référence en EDM et LA

Les travaux en EDM et LA se sont appuyés sur plusieurs bases de données devenues des références dans le domaine, chacune associée à des usages et à des objectifs de recherche spécifiques. Ces ensembles de données ont permis de tester la robustesse des modèles prédictifs, d'étudier le décrochage étudiant, ou encore de mieux comprendre les trajectoires d'apprentissage.

— **KDD Cup 2010 (ASSISTments)[137]** : L'un des premiers jeux de données de référence en EDM provient de la compétition internationale KDD Cup 2010. Il est issu de la plateforme *ASSISTments*, utilisée dans les écoles américaines. Ce jeu contient plusieurs millions de lignes d'interactions entre élèves et exercices de mathématiques, avec pour chaque tentative le résultat (réussi ou non), le temps de réponse et le type de compétence mobilisée. Il a servi de base au développement d'algorithmes de *knowledge tracing*, comme le *Bayesian Knowledge Tracing (BKT)* ou le *Deep Knowledge Tracing (DKT)*, permettant de modéliser l'évolution des connaissances des apprenants au fil du temps.

— **EdNet [43]** : Proposé par Choi et al. (2020), est aujourd'hui l'un des plus vastes ensembles

de données éducatives au monde, avec plus de 100 millions d'interactions collectées auprès de 700 000 apprenants sur une plateforme coréenne de soutien scolaire. Il inclut des informations fines sur les comportements d'étude : navigation entre ressources, réponses aux exercices, révisions, et pauses dans les sessions d'apprentissage. EdNet a contribué à la mise au point de modèles séquentiels avancés, comme les réseaux à mémoire longue (LSTM) ou les Transformers éducatifs, capables de capturer les dépendances temporelles dans les traces d'interaction.

— **KDD Cup 2015 (MoocCN - XuetangX) [64]** : La compétition KDD Cup 2015 a introduit le premier jeu de données massif de type MOOC, issu de la plateforme chinoise *XuetangX*. Il contient les interactions de milliers d'apprenants dans plusieurs cours en ligne, sur une période de six semaines. Les données incluent le nombre de vidéos vues, les soumissions de quiz, les participations aux forums et les résultats finaux. Ce jeu de données est devenu un standard pour l'étude du décrochage précoce et de la prédiction de réussite dans les MOOC, en raison de son caractère temporel, ouvert et structuré.

— **OULAD (Open University Learning Analytics Dataset) [91]** : Le jeu de données OULAD, publié par l'Open University du Royaume-Uni , est l'un des plus utilisés dans la littérature. Il regroupe les données de plusieurs modules d'enseignement à distance, suivis par plus de 30 000 étudiants sur plusieurs années. OULAD combine des informations démographiques, des notes d'évaluation et des traces d'activité sur la plateforme d'apprentissage virtuelle. Sa structure longitudinale, couvrant jusqu'à 38 semaines, en fait une base de référence pour l'étude du décrochage, de l'engagement et de la réussite dans les environnements d'apprentissage ouverts.

— **OpenEdX et Coursera Datasets [63, 79]** : Les plateformes internationales *Coursera* et *edX* ont également permis la publication partielle de jeux de données issus de MOOC, souvent dans le cadre de concours ou de collaborations de recherche. Ces ensembles sont utilisés pour analyser la progression des apprenants, la complétion des cours et les patterns d'interaction, bien qu'ils soient souvent limités en taille ou soumis à des contraintes d'accès et d'anonymisation.

— **Autres ensembles spécifiques.** D'autres jeux de données, tels que *CSEDM* (Computer Science Education Data Mining) [119], *Algebra2005* [89], ou encore les ensembles de données de Khan Academy [1], sont utilisés pour des contextes disciplinaires précis (programmation, mathématiques, sciences). Ces bases permettent d'explorer la granularité des interactions entre étudiants et systèmes éducatifs, et de concevoir des modèles spécialisés selon le type d'activité ou de compétence.

Dans l'ensemble, ces bases de données constituent un socle essentiel pour la recherche en EDM et LA. Elles permettent de reproduire, comparer et valider des approches de modélisation sur des contextes variés, allant des environnements institutionnels (comme OULAD) aux plateformes massives et ouvertes (comme XuetangX ou EdNet).

3.2 Jeux de données choisies

Les expérimentations menées dans le cadre de cette thèse reposent exclusivement sur des **datasets publics**, choisis pour leur qualité, leur exhaustivité et leur adoption dans la communauté scientifique. Il n'a pas été possible de s'appuyer sur les données internes de l'entreprise partenaire : le double rachat de la structure pendant la période de recherche a provoqué une instabilité organisationnelle (changements d'équipes, redéfinition des priorités, évolutions des outils de stockage des données) qui n'a pas permis de sécuriser les accès ni de garantir la continuité des collectes. Dans ces conditions, l'utilisation de données propriétaires aurait limité la portée des analyses et compromis la reproductibilité des résultats.

Afin de pallier cette contrainte, deux **jeux de données ouverts** ont été retenus pour leurs caractéristiques complémentaires :

— **OULAD (Open University Learning Analytics Dataset)**

— **MoocCN (XuetangX)**

Ces deux jeux de données couvrent des **contextes pédagogiques contrastés** :

— OULAD représente un suivi **institutionnel** et **avec un suivi temporel étendu**, où les apprenants sont inscrits dans des modules formalisés sur des durées allant jusqu'à 38 semaines.

— MoocCN illustre un suivi **massif** et **intensif**, concentré sur des sessions courtes de 6 semaines, typiques des MOOCs.

Leur disponibilité publique et leur **documentation détaillée** permettent de reproduire les analyses et de comparer les résultats avec un large éventail de travaux existants [91, 64]. Ces datasets constituent donc une base solide pour **évaluer et comparer les approches développées dans cette thèse**, tout en assurant un haut niveau de transparence méthodologique.

3.2.1 Problématique étudiée : la prédiction du décrochage

L'axe principal de modélisation retenu dans cette thèse est celui de la **prédiction du décrochage**, un sujet fondateur en *Educational Data Mining* (EDM) et en *Learning Analytics* (LA). Cette problématique, abondamment explorée depuis une décennie, vise à anticiper les abandons d'apprenants avant la fin d'un cours ou d'un programme, afin de permettre des interventions pédagogiques ciblées [91, 64, 41]. Dans les environnements d'apprentissage à distance, le décrochage constitue un phénomène massif et récurrent : les études récentes montrent qu'entre 30 et 50 % des apprenants quittent un module avant son terme, en particulier dans les contextes ouverts et massifs comme les MOOCs [125].

Au-delà de son importance pédagogique, cette problématique présente un intérêt scientifique majeur. Le décrochage n'est pas un événement ponctuel, mais un processus progressif, marqué par des changements subtils dans les comportements d'apprentissage. Ces changements peuvent se manifester par une baisse de l'engagement sur la plateforme, une participation décroissante aux évaluations ou une irrégularité dans la consultation des ressources. La modélisation prédictive de ce phénomène repose ainsi sur l'analyse de signaux temporels et comportementaux qui

traduisent l'évolution du rapport de l'étudiant à son apprentissage [49, 156].

Un autre atout essentiel de cette problématique réside dans la dimension interprétative qu'elle offre. Les indicateurs mobilisés (e.g. nombre de clics, régularité d'accès, fréquence de soumission, progression hebdomadaire) renvoient à des comportements observables et pédagogiquement compréhensibles. Cette transparence permet un dialogue direct entre la modélisation algorithmique et l'analyse éducative, favorisant des interprétations exploitables pour les enseignants et les concepteurs de dispositifs [54, 143].

Ce choix se distingue également par ce qu'il écarte délibérément : contrairement à d'autres axes de recherche comme la recommandation de ressources ou la génération de parcours personnalisés, la prédiction du décrochage ne dépend pas d'un contexte pédagogique particulier ni d'un contenu spécifique. Les systèmes de recommandation et de personnalisation reposent en effet sur des structures curriculaires propres à chaque institution, sur des typologies de ressources variables et sur des objectifs d'apprentissage contextualisés, ce qui limite leur transférabilité et leur généricité. À l'inverse, le décrochage peut être étudié dans tout environnement d'apprentissage, qu'il soit académique ou informel, car il repose sur des signaux comportementaux universels et comparables entre plateformes. Cette généralité confère à la prédiction du décrochage une valeur pour évaluer la fiabilité et la frugalité des modèles prédictifs dans des contextes hétérogènes.

En somme, le choix de la prédiction du décrochage s'impose comme un compromis équilibré entre **pertinence éducative**, **généricité scientifique** et **applicabilité technique**. Il offre un cadre d'expérimentation robuste pour examiner comment un modèle peut évoluer dans le temps, gérer la rareté de l'information et conserver un niveau de fiabilité élevé, même dans des contextes d'apprentissage ouverts et évolutifs.

3.3 OULAD (Open University Learning Analytics Dataset)

3.3.1 Présentation générale et contexte institutionnel

Le dataset **OULAD (Open University Learning Analytics Dataset)** a été publié par *The Open University* (Royaume-Uni) pour soutenir la recherche en *Learning Analytics* et en *prédiction du décrochage* [91]. Cette institution, pionnière de l'enseignement supérieur à distance en Europe, propose depuis plusieurs décennies des modules de formation en ligne et accueille chaque année des dizaines de milliers d'apprenants. Le profil des étudiants est particulièrement hétérogène :

— Beaucoup sont des adultes en reprise d'études ou en reconversion professionnelle ;
— L'âge moyen est supérieur à 30 ans et une part significative n'a jamais suivi d'études universitaires auparavant ;
— Les étudiants jonglent souvent avec d'autres responsabilités (emploi, famille), ce qui explique un **taux de décrochage élevé** (plus de 35%).

OULAD couvre les données collectées entre 2013 et 2014 pour **32 593 étudiants**, inscrits dans **22 présentations** (sessions) de **7 modules** (cours). C'est un jeu de données riche et complet car il inclut :

— Des **données démographiques** détaillées ;

— Les **interactions sur le Virtual Learning Environment (VLE)** : clics sur les ressources pédagogiques, forums, quiz, etc. ;
— Les **évaluations** (formatives et sommatives) avec les scores et les pondérations ;
— Des **métadonnées sur les modules et les ressources**.

3.3.2 OULAD dans la littérature

Le dataset OULAD a donné lieu à une large variété d'approches méthodologiques dans la littérature. Plusieurs travaux inclus dans la bibliographie de cette thèse s'appuient explicitement sur ses données pour développer et tester de nouveaux modèles ou méthodes d'analyse. Parmi eux :

— **Towards An Online Incremental Approach to Predict Students Performance** [93] propose une approche d'apprentissage incrémental pour la prédiction du décrochage étudiant, s'appuyant sur les données longitudinales hebdomadaires de OULAD. Les auteurs y exploitent des modèles adaptatifs permettant une mise à jour continue sans réentraînement complet.

— **Fair Streaming Feature Selection** [58] introduit une méthode de sélection de caractéristiques prenant en compte la classe minoritaire (décrocheurs), et évalue son efficacité sur OULAD dans un cadre de flux de données.

— **Using Behavioral Primitives to Model Students' Digital Behavior** [54] s'appuie sur les traces VLE pour dériver des primitives comportementales (patterns de navigation, fréquence, régularité) utilisées dans des modèles prédictifs interprétables.

— **A New Way to Characterize Learning Datasets** [143] propose un cadre d'analyse des jeux de données éducatifs, en identifiant des métriques de complexité et de diversité sur les données d'OULAD. Ce travail positionne le dataset comme un cas d'étude riche et représentatif des plateformes d'apprentissage.

Ces contributions soulignent la pertinence de OULAD pour développer, comparer et évaluer des approches centrées sur la prédiction, l'équité et la compréhension des comportements numériques en environnement d'apprentissage.

3.3.3 Organisation et schéma relationnel

OULAD est structuré en **11 fichiers CSV**, organisés en tables relationnelles interconnectées (voir Figure 3.1).

Les fichiers composant OULAD sont listés dans le Tableau 3.1, avec leur description détaillée.

3.3.4 Cours sélectionné

Pour cette thèse, le cours **CCC (2014J)** a été retenue comme corpus principal d'analyse. Plusieurs éléments motivent ce choix. Tout d'abord, il s'agit de l'une des plus grandes cohortes du dataset OULAD, avec **2 498 étudiants inscrits**, ce qui garantit une masse critique suffisante pour entraîner et évaluer des modèles prédictifs. Ensuite, la durée du module s'étend sur **38 semaines (soit 269 jours)**, offrant ainsi un **suivi temporel particulièrement riche** qui permet d'observer l'évolution des comportements d'apprentissage sur le long terme. Par ailleurs,

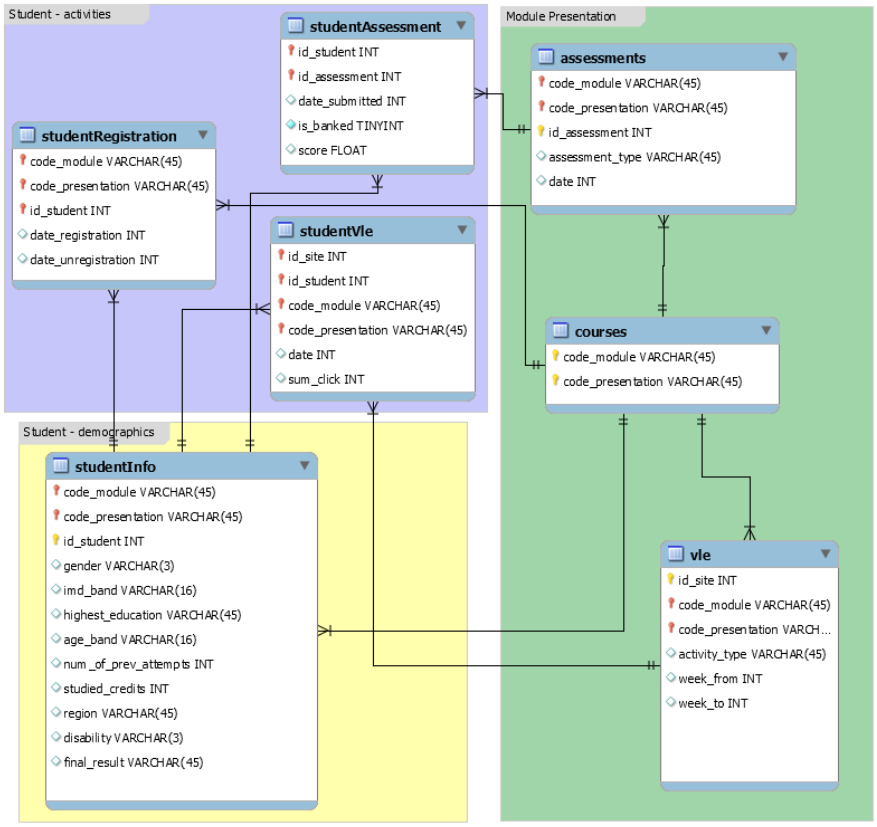

FIGURE 3.1 – Schéma relationnel d'OULAD

Fichier	**Description**
`studentInfo.csv`	Profil des étudiants et statut final (`final_result`) : `gender`, `age_band`, `highest_education`, `region`, `disability`, `studied_credits`, `num_of_prev_attempts`.
`studentRegistration.csv`	Dates d'inscription (`date_registration`) et de désinscription (`date_unregistration`) par module.
`studentVle.csv`	Logs VLE (clics horodatés) avec la variable `sum_click`.
`studentAssessment.csv`	Résultats des évaluations obtenus par chaque étudiant (scores).
`assessments.csv`	Informations sur les évaluations (type, date, pondération).
`courses.csv`	Métadonnées sur les modules et présentations.
`vle.csv`	Liste des ressources pédagogiques (`activity_type`) et période de disponibilité (`week_from`, `week_to`).

TABLE 3.1 – Résumé des fichiers composant OULAD et leur contenu

CCC (2014J) se distingue par la **diversité des ressources pédagogiques mobilisées** (vidéos, forums, quiz, supports de cours), ainsi que par la **structure variée des évaluations**, intégrant à la fois des devoirs intermédiaires et un examen final. Enfin, ce cours présente une particularité intéressante sur le plan analytique : Il s'agit en outre du module avec la plus forte proportion d'apprenants ayant également suivi d'autres cours sur la même période. Cette configuration rend possible l'analyse de trajectoires éducatives inter-modules, ce qui constitue un atout pour des travaux centrés sur la continuité de l'apprentissage.

3.3.5 Statistiques descriptives du module CCC (2014J)

Répartition des étudiants par statut final

L'une des premières métriques à considérer dans l'analyse du module CCC (2014J) est la répartition des apprenants selon leur statut final à l'issue du cours. Le dataset OULAD encode cette information à travers la variable `final_result`, qui distingue plusieurs catégories : `Pass`, `Fail`, `Withdrawn` et `NotAttempted`. Pour les besoins de cette étude, ces catégories ont été agrégées selon une logique binaire permettant d'opposer deux groupes analytiques : les **non-décrocheurs**, correspondant aux étudiants ayant terminé le cours avec succès ou non (`Pass ou Fail`), et les **décrocheurs**, s'étant désinscrits (`Withdrawn`), ou n'ayant pas participé aux évaluations finales (`NotAttempted`).

Cette binarisation permet d'étudier de manière plus lisible les phénomènes de persévérance ou de décrochage au sein du module. Le tableau 3.2 présente la distribution quantitative des deux groupes sur l'ensemble des 2 498 étudiants inscrits à CCC (2014J), tandis que la figure 3.2 en fournit une visualisation synthétique.

Statut final	**Nombre d'étudiants**	**Pourcentage**
Non-Décrocheurs (Pass)	1 540	61,7%
Décrocheurs (Fail, Withdrawn, Not Attempted)	958	38,3%
Total	**2 498**	**100%**

TABLE 3.2 – Répartition des statuts finaux dans CCC (2014J)

On observe que près de **38,3% des étudiants** n'ont pas mené le module jusqu'à son terme, ce qui confirme le caractère structurel du décrochage dans ce type d'enseignement à distance. Ce taux est cohérent avec les statistiques générales de l'Open University, où l'hétérogénéité des profils, la flexibilité des parcours et les contraintes personnelles des apprenants peuvent engendrer un engagement discontinu. À l'inverse, les **61,7% restants** ont complété l'ensemble du module et validé les évaluations, constituant ainsi une population de référence pour les modèles de persévérance. Cette répartition équilibrée permet de disposer d'un cadre pertinent pour entraîner et évaluer des modèles binaires de prédiction du décrochage.

Engagement et interactions

L'analyse des interactions numériques constitue un axe fondamental dans l'étude des comportements d'apprentissage à distance. Dans le dataset OULAD, chaque clic réalisé par un étudiant sur une ressource pédagogique du VLE est enregistré et horodaté. Cela permet de reconstituer finement le rythme d'engagement hebdomadaire au sein du module CCC (2014J), sur une durée de 38 semaines.

La figure 3.3 illustre la distribution des interactions au fil des semaines pour l'ensemble des étudiants. On y observe une dynamique rythmée par les échéances académiques, avec des

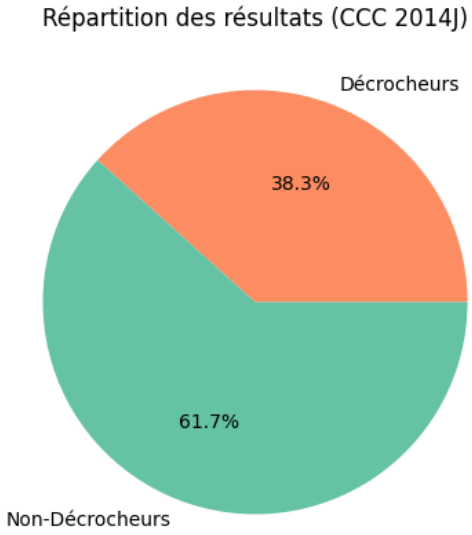

FIGURE 3.2 – Visualisation de la répartition des statuts finaux dans CCC (2014J)

pics d'activité notables autour des semaines 3, 10, 21 et 31, correspondant aux dates clés des évaluations intermédiaires et finale. Ce phénomène est typique des cours en ligne, où les étudiants tendent à intensifier leur engagement à l'approche des échéances notées.

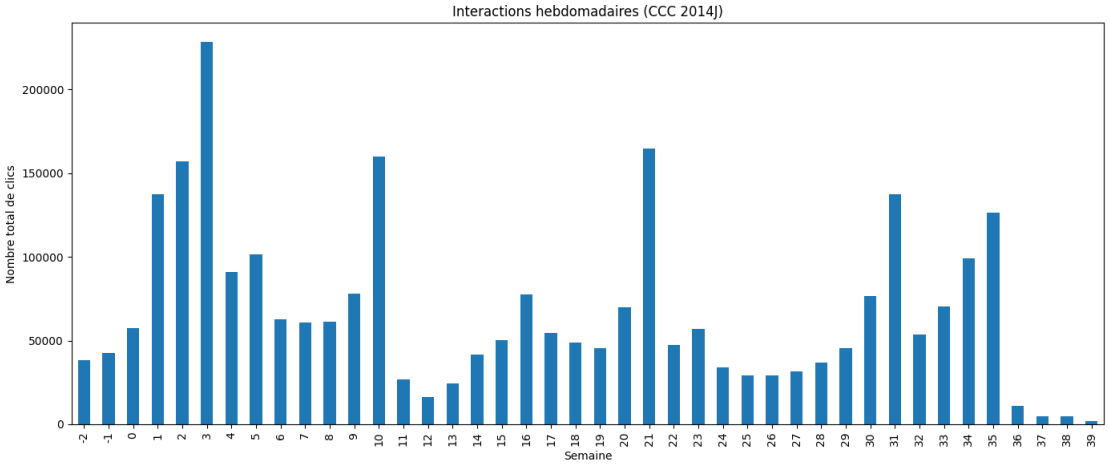

FIGURE 3.3 – Interactions hebdomadaires dans le module CCC (2014J)

Au-delà de ces variations globales, la figure 3.4 affine l'analyse en comparant le nombre total de clics par étudiant selon leur statut final. Elle révèle une différence marquée entre les deux groupes : les **non-décrocheurs** se caractérisent par un engagement régulier et soutenu, avec un volume médian de clics nettement supérieur à celui des **décrocheurs**, dont l'activité reste souvent faible voire sporadique. La dispersion est également plus importante parmi les non-décrocheurs,

certains étudiants très engagés ayant effectué plusieurs milliers d'interactions.

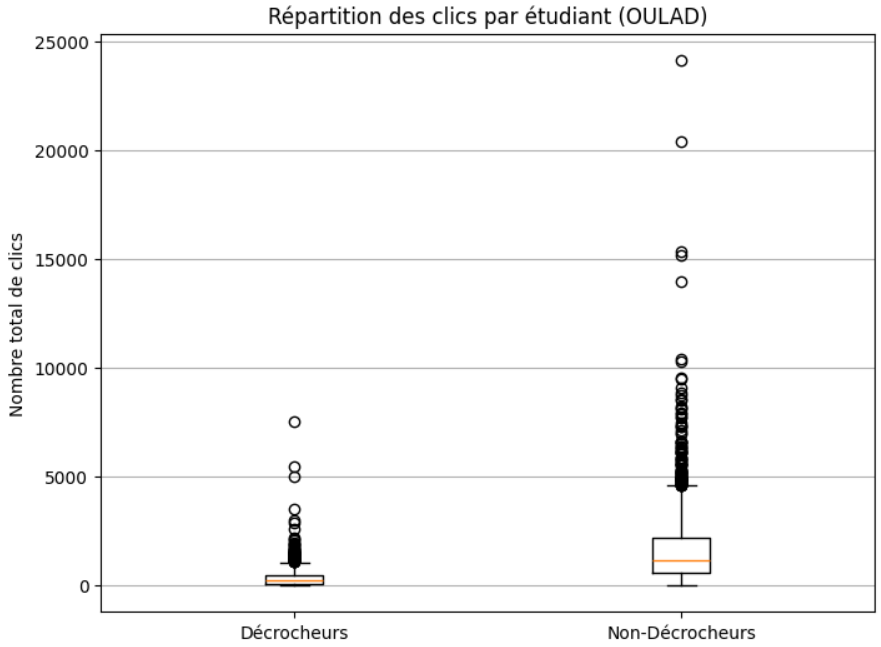

FIGURE 3.4 – Répartition du nombre total de clics par étudiant selon le statut (Décrocheurs vs Non-Décrocheurs)

Ces résultats confirment l'intérêt des traces d'interaction comme métriques prédictifs du décrochage : un volume faible ou décroissant de clics dans les premières semaines peut signaler un désengagement précoce. À l'inverse, une activité régulière est généralement corrélée à la persévérance et à la réussite dans le module.

Participation et performance aux évaluations

L'engagement des apprenants dans les activités évaluatives constitue une métrique clé de persévérance et de réussite. Le module CCC (2014J) propose plusieurs évaluations réparties tout au long du cours, combinant devoirs intermédiaires et examen final. L'analyse de la participation à ces évaluations permet de mettre en évidence des dynamiques différenciées entre les décrocheurs et les non-décrocheurs.

La figure 3.5 présente le nombre d'étudiants ayant soumis au moins une évaluation chaque semaine, selon leur statut final. On y observe une tendance nette : les **décrocheurs participent majoritairement aux premières évaluations**, mais leur engagement chute drastiquement à partir de la semaine 5. Cette désaffection progressive témoigne d'un désengagement académique précoce, parfois amorcé dès les premières semaines du module. À l'inverse, les **non-décrocheurs maintiennent une participation soutenue** tout au long des 38 semaines, avec une couverture quasi complète des devoirs proposés.

La figure 3.6 complète cette analyse en comparant les **scores moyens hebdomadaires** obtenus par les deux groupes d'étudiants. Il apparaît que les non-décrocheurs non seulement

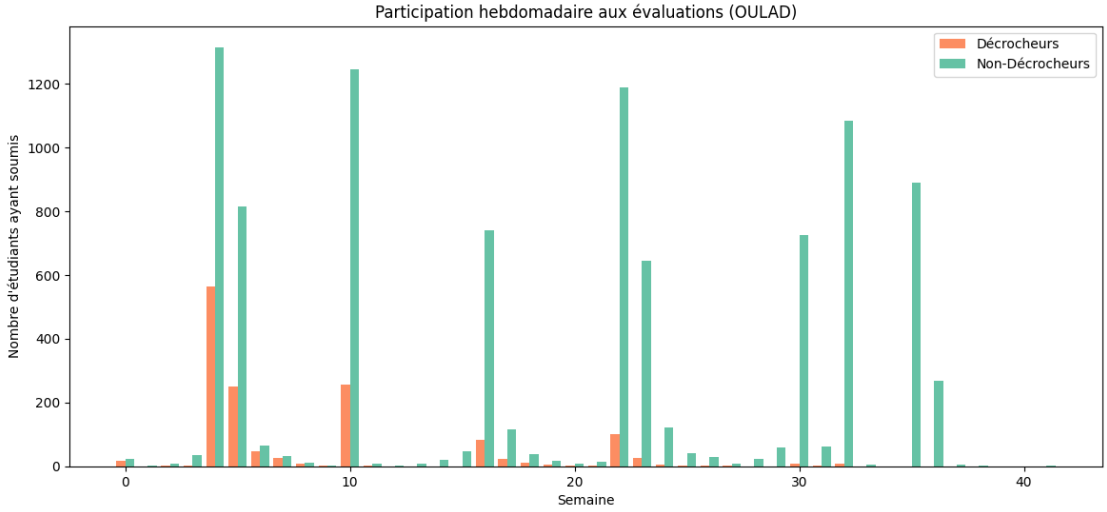

FIGURE 3.5 – Participation hebdomadaire aux évaluations (Décrocheurs vs Non-Décrocheurs)

participent davantage, mais obtiennent également de **meilleures performances** sur l'ensemble des évaluations. L'écart de score entre les deux groupes se creuse particulièrement à partir de la mi-parcours, traduisant un différentiel d'investissement, de préparation et de continuité pédagogique.

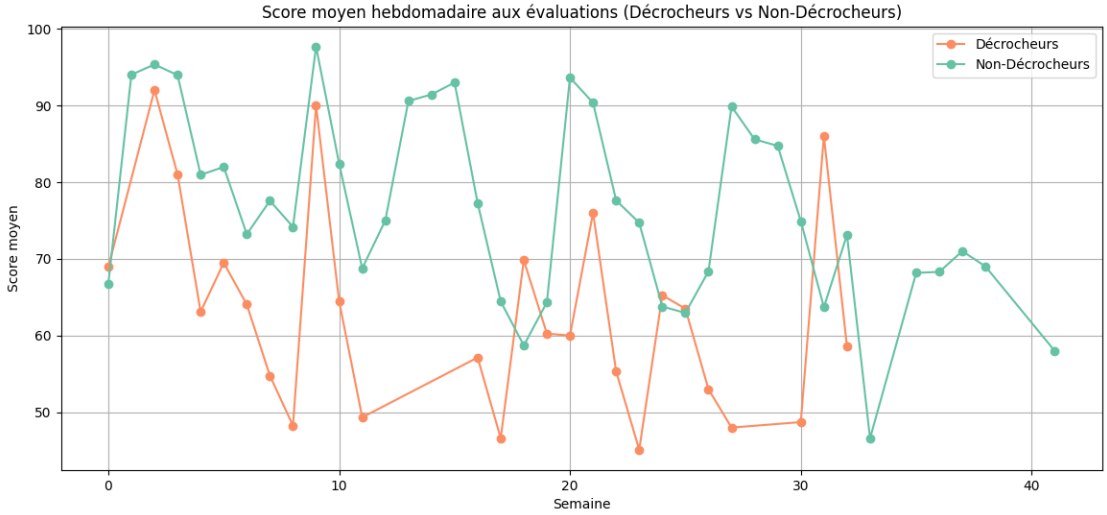

FIGURE 3.6 – Score moyen hebdomadaire aux évaluations (Décrocheurs vs Non-Décrocheurs)

Ces observations soulignent la complémentarité entre la dimension comportementale (participation) et la dimension académique (performance) pour caractériser le risque de décrochage. Un modèle de prédiction robuste devra intégrer ces deux facettes pour capter les signaux faibles d'un désengagement en cours de formation.

Analyse démographique

Au-delà des traces comportementales et académiques, les variables démographiques fournies par le dataset OULAD permettent d'approfondir la compréhension des profils étudiants. Deux dimensions particulièrement informatives ont été analysées dans le cadre de cette thèse : les dates d'inscription aux modules et le niveau d'éducation déclaré au moment de l'entrée en formation.

La figure 3.7 montre la distribution des dates d'inscription des étudiants au module CCC (2014J), exprimées en nombre de jours avant le début effectif du cours. La majorité des apprenants s'inscrivent entre **50 et 100 jours avant le démarrage**, ce qui témoigne d'un engagement anticipé et potentiellement planifié. Une fraction non négligeable s'inscrit néanmoins à moins de 30 jours du lancement, ce qui peut indiquer une démarche plus opportuniste ou moins préparée, hypothèse qu'il est possible d'explorer en croisant ces données avec les dynamiques d'engagement observées par la suite.

Par ailleurs, le niveau d'éducation antérieur constitue un facteur déterminant dans la capacité d'adaptation aux exigences de l'apprentissage en ligne. Comme le montre la figure 3.8, près de **la moitié des étudiants** déclarent un niveau équivalent à l'*A Level* (équivalent du baccalauréat au Royaume-Uni), tandis que les autres se répartissent entre des niveaux inférieurs (éducation secondaire) et supérieurs (études universitaires). Cette diversité éducative reflète le positionnement particulier de l'Open University, qui attire à la fois des apprenants en reprise d'études, des autodidactes et des profils plus académiques. Elle justifie l'importance d'adopter des modèles prédictifs capables de prendre en compte cette hétérogénéité de départ.

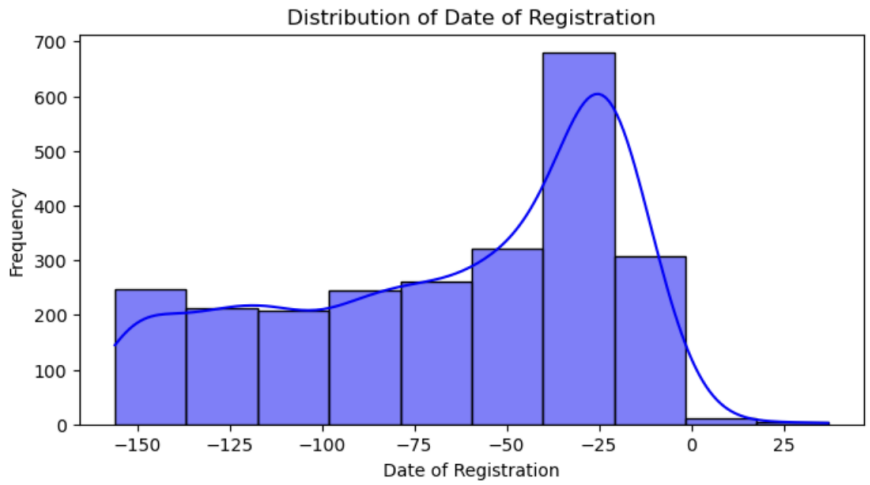

FIGURE 3.7 – Distribution des dates d'inscription des étudiants (CCC 2014J)

Analyse des types d'activités

Les données issues du VLE permettent non seulement de quantifier le volume d'interactions hebdomadaires, mais aussi d'identifier les types d'activités pédagogiques les plus mobilisées au fil du temps. Chaque clic est associé à un type de ressource défini dans le fichier `vle.csv`, ce qui

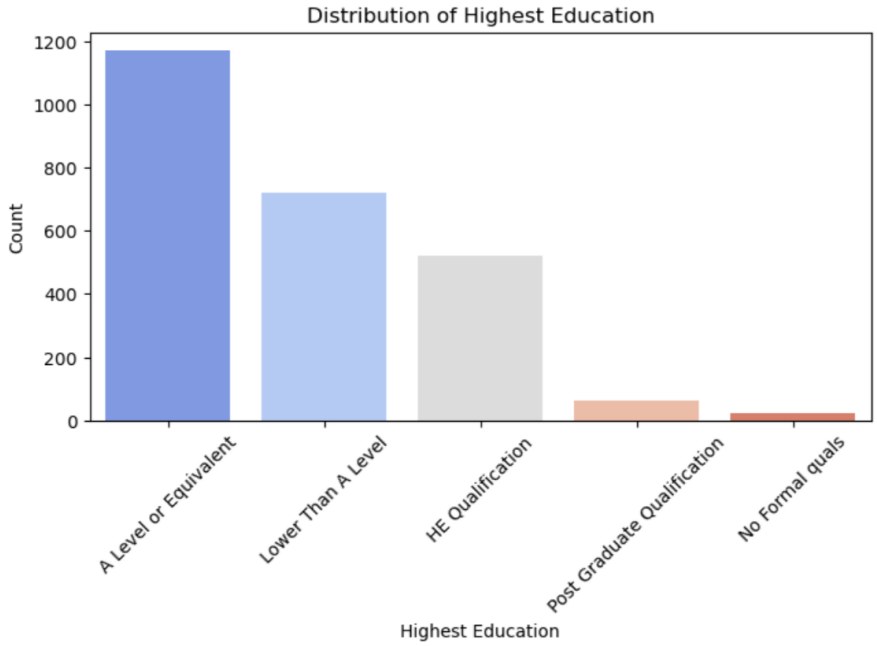

FIGURE 3.8 – Répartition du plus haut niveau d'éducation déclaré (CCC 2014J)

autorise une analyse fine de la temporalité pédagogique du module.

La figure 3.9 représente une **heatmap des types d'activités** consultées semaine par semaine par l'ensemble des étudiants inscrits au module CCC (2014J). Chaque ligne correspond à une catégorie d'activité (telle que `quiz`, `resource`, `homepage`, `forum`, etc.), et l'intensité de la couleur traduit le volume relatif d'interactions.

Cette visualisation fait apparaître une hiérarchie claire dans les usages pédagogiques : les **quiz dominent très largement les interactions**, concentrant l'essentiel des clics des étudiants, notamment en amont des semaines d'échéances. Ils sont suivis par les **pages de cours** (`homepage`) et les ressources pédagogiques téléchargeables (`resource`). D'autres activités, telles que la participation aux forums ou la consultation de glossaires, apparaissent plus marginales.

La temporalité des interactions est également instructive : certaines ressources ne sont consultées qu'en début de module (introductives), tandis que d'autres sont revisitées à intervalles réguliers (supports récurrents). Cette analyse renforce l'idée selon laquelle la nature et le calendrier des ressources influencent directement le comportement d'engagement des apprenants, un aspect que les modèles prédictifs doivent être en mesure d'intégrer.

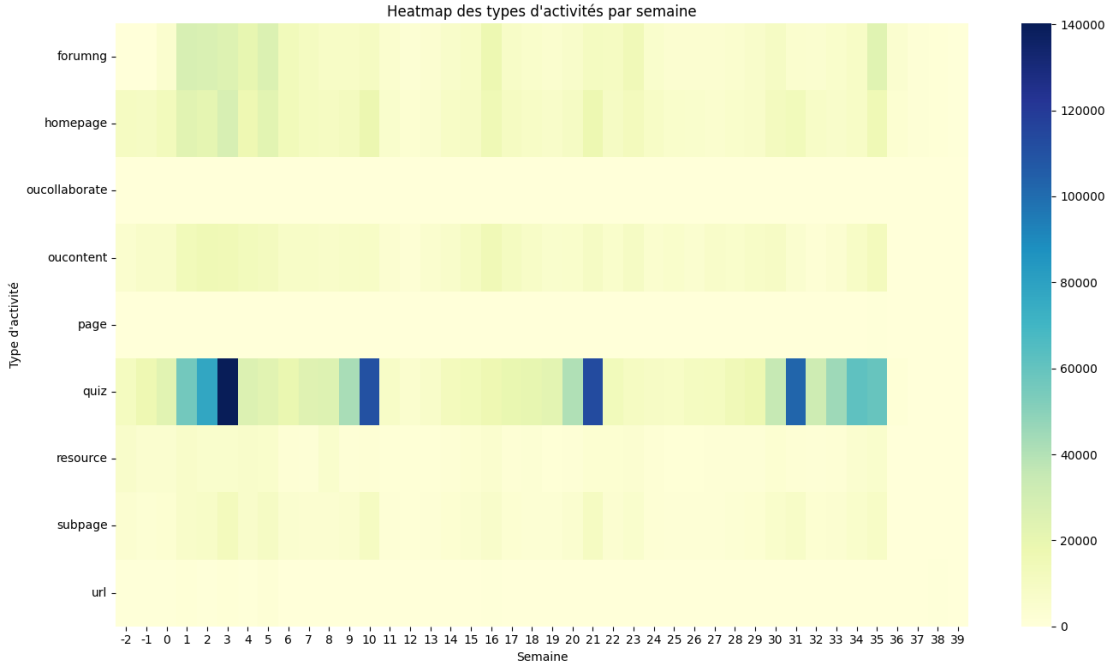

FIGURE 3.9 – Heatmap des types d'activités consultées par semaine dans le module CCC (2014J)

3.4 MoocCN (XuetangX)

3.4.1 Présentation générale et origine

Le dataset **MoocCN** est dérivé de la plateforme chinoise **XuetangX**, développée par l'université Tsinghua en collaboration avec le ministère chinois de l'Éducation. Lancée en 2013, XuetangX est aujourd'hui l'une des plus grandes plateformes de cours en ligne ouverts et massifs au niveau mondial, avec des millions d'utilisateurs inscrits. Le dataset a été rendu public à l'occasion de la **KDD Cup 2015**, une compétition scientifique internationale visant à stimuler l'innovation dans l'analyse de données éducatives à grande échelle [64].

MoocCN représente une des initiatives les plus ambitieuses de mise à disposition de données d'apprentissage issues de MOOCs. Il a depuis été largement adopté dans les travaux de recherche sur le décrochage, la prédiction de la persistance, la modélisation du comportement d'engagement, ou encore les approches fédérées d'analyse distribuée. Le dataset se distingue par les caractéristiques suivantes :

— **Un volume massif de données** : des millions de lignes de logs horodatés, enregistrant de manière exhaustive les interactions des étudiants avec la plateforme pendant plusieurs semaines.

— **Une granularité fine des actions** : chaque clic est associé à une action typée (lecture, pause, clic sur un quiz, soumission d'examen, participation au forum, etc.), ce qui permet une modélisation détaillée du comportement des apprenants.

— **Un suivi temporel structuré** : les données sont organisées semaine par semaine, permettant une analyse longitudinale des parcours étudiants avec un bon niveau de synchronisation temporelle entre les individus.

— **Des annotations explicites du décrochage** : chaque ligne de log est associée à un label binaire indiquant si l'étudiant a terminé le cours ou s'il a décroché (`truth` = 0 ou 1), ce qui facilite l'usage pour les tâches de classification supervisée.

Contrairement à des jeux de données éducatifs plus institutionnels (comme OULAD), MoocCN offre une perspective typiquement *MOOC* : massification, diversité des profils, taux élevé de décrochage, et engagement centré autour des vidéos et des évaluations formatives. Ces caractéristiques en font un complément idéal pour évaluer des approches robustes et généralisables dans des contextes d'apprentissage ouverts et à grande échelle.

3.4.2 MoocCN dans la littérature

Le dataset MoocCN (issu de XuetangX) est devenu une ressource de référence dans la recherche en *Learning Analytics*, en particulier pour l'analyse du décrochage, des comportements d'engagement et des stratégies pédagogiques adaptatives dans les MOOCs. Plusieurs travaux récents ont exploité ce corpus pour tester et affiner des approches méthodologiques variées. Nous présentons ici quatre études représentatives de cette diversité :

— **Mubarak et al. (2021)** proposent une approche d'apprentissage profond pour prédire la performance des étudiants dans les MOOCs. Leur modèle repose sur une architecture temporelle à base de *sliding windows* et de réseaux neuronaux profonds, appliquée à MoocCN pour capter les régularités comportementales menant au décrochage [112].

— **Gong et al. (2022)** introduisent *HinCRec-RL*, un système de recommandation s'appuyant sur l'apprentissage par renforcement dans des graphes hétérogènes. Leurs expérimentations sur MoocCN montrent que l'intégration des relations complexes entre étudiants, contenus et concepts améliore significativement la personnalisation des parcours [73].

— **Roh (2024)** conçoit un modèle de *Graph Neural Network* baptisé SIG-Net, capable d'extraire des sous-graphes d'interaction pour prédire le décrochage. Le dataset XuetangX est utilisé pour valider la capacité du modèle à détecter les structures relationnelles associées à un désengagement progressif [127].

— **Fachola et al. (2023)** mobilisent MoocCN pour expérimenter des approches fédérées visant à préserver la confidentialité des apprenants. Leur article intitulé *Federated Learning for Data Analytics in Education* montre que l'entraînement distribué sur des segments de données hétérogènes permet d'obtenir des performances comparables aux modèles globaux tout en renforçant la protection des données [62].

Ces contributions illustrent la richesse du dataset MoocCN pour explorer des problématiques variées allant de la prédiction à la personnalisation, en passant par l'interprétation du comportement collectif ou l'apprentissage distribué. Elles confirment son rôle central dans l'évaluation empirique des méthodes d'IA en contexte MOOC.

3.4.3 Structure et format des données

Le dataset MoocCN se présente sous la forme de fichiers CSV correspondant à des logs d'interaction enregistrés automatiquement sur la plateforme XuetangX. Chaque ligne représente une action effectuée par un étudiant à un instant donné. La structure des données est uniforme et bien documentée, ce qui facilite son exploitation dans des tâches de modélisation comportementale ou de classification supervisée.

Chaque enregistrement contient les attributs suivants :

Attribut	**Description**
`user_id`	Identifiant unique de l'étudiant, utilisé pour agréger les actions au niveau individuel.
`enroll_id`	Identifiant du couple étudiant–cours, permettant de distinguer plusieurs inscriptions éventuelles sur différents modules.
`course_id`	Code du cours concerné par l'action (ici, un seul cours a été sélectionné pour l'analyse).
`time`	Horodatage de l'action, au format date–heure, permettant d'ordonner les événements dans le temps et de construire des fenêtres temporelles glissantes.
`action`	Nature de l'interaction effectuée par l'étudiant (lecture de vidéo, clic sur un lien, soumission de quiz, consultation de document, etc.).
`object`	Identifiant ou nom de la ressource associée à l'action (exemple : nom d'une vidéo ou ID d'un exercice).
`truth`	Variable binaire indiquant si l'étudiant a terminé le cours (`0` = non-décrocheur) ou s'il a décroché (`1` = décrocheur).

TABLE 3.3 – Structure des attributs contenus dans les fichiers de logs de MoocCN

Les valeurs prises par la variable `action` sont particulièrement riches et couvrent l'ensemble des interactions pédagogiques disponibles sur la plateforme, notamment :

— les actions liées à la vidéo : `play_video`, `pause_video`, `seek_video`, `stop_video`
— les actions d'apprentissage : `click_courseware`, `problem_check`, `problem_check_correct`
— les interactions sociales : `discussion`, `forum` ;
— d'autres actions spécifiques liées à la navigation ou à la progression

Ce format de données, à la fois compact et expressif, permet d'effectuer des analyses séquentielles, temporelles ou basées sur la fréquence des événements. Il est particulièrement adapté pour entraîner des modèles de prédiction du décrochage ou pour détecter des motifs d'engagement au fil des semaines.

3.4.4 Cours sélectionné

Le cours retenu pour les expérimentations est identifié par le code `course-v1:TsinghuaX+10610224X+2016_T1`. Ce module a été sélectionné pour plusieurs raisons qui en font un cas d'étude pertinent et représentatif des dynamiques pédagogiques observées dans les MOOCs :

— Il s'étale sur une durée de **6 semaines**, ce qui correspond à la durée standard de nombreux MOOCs et permet une analyse temporelle suffisamment riche tout en restant contenue.

— Il regroupe un volume conséquent d'apprenants avec un total de **3 700 étudiants**.

— Il présente une diversité importante d'interactions pédagogiques, incluant notamment le visionnage de vidéos, la soumission de quiz, la consultation de contenus de cours, ainsi que des activités sociales comme les forums de discussion.

Les fichiers fournis sont divisés en deux sous-ensembles distincts :

— un jeu d'entraînement (*train*) couvrant les 6 premières semaines d'activité avec les labels de vérité (`truth`) associés

— un jeu de test (*test*) structuré de manière similaire, permettant d'évaluer les modèles prédictifs dans des conditions proches de celles de déploiement

Cette structuration bimodale permet d'entraîner, de valider et de comparer les approches proposées dans des conditions réalistes, tout en respectant une séparation stricte entre les phases d'apprentissage et de généralisation.

3.4.5 Statistiques descriptives du cours sélectionné

Répartition des décrocheurs et non-décrocheurs

La Figure 3.10 illustre la répartition des apprenants selon leur statut final, en distinguant les décrocheurs (étiquetés `1`) des non-décrocheurs (étiquetés `0`), pour les deux sous-ensembles *train* et *test*. Ces visualisations permettent d'apprécier l'équilibre des classes au sein des données, un aspect fondamental pour le développement et l'évaluation de modèles prédictifs robustes.

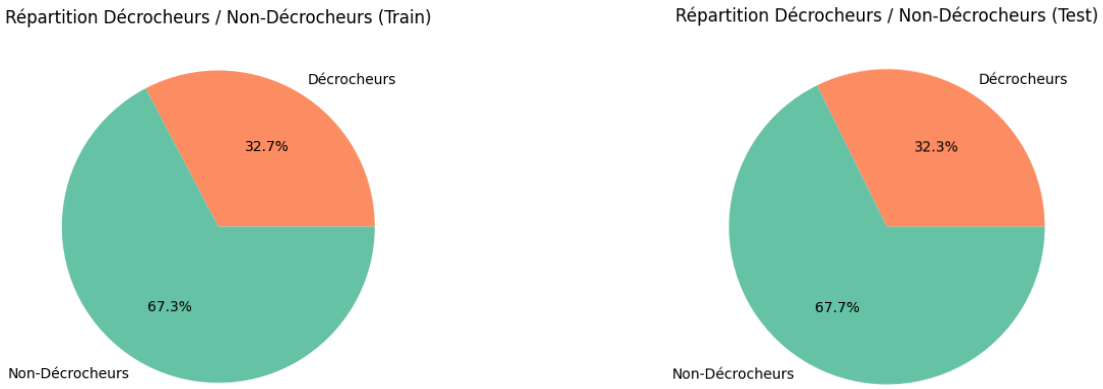

FIGURE 3.10 – Répartition des classes (Décrocheurs vs Non-Décrocheurs) pour le jeu d'entraînement (gauche) et le jeu de test (droite).

Le Tableau 3.4 fournit les valeurs numériques exactes correspondant à cette répartition. On y observe que le jeu d'entraînement contient **854 décrocheurs** (soit environ 32,7%) et **1 759 non-décrocheurs**, tandis que le jeu de test comprend **351 décrocheurs** et **736 non-décrocheurs**,

ce qui représente un taux de décrocheurs très similaire (32,3%).

Statut	**Train**	**Test**	**Total**
Non-Décrocheurs	1 759	736	2 495
Décrocheurs	854	351	1 205
Total	**2 613**	**1 087**	**3 700**

TABLE 3.4 – Répartition des décrocheurs et non-décrocheurs dans les jeux d'entraînement et de test

Ces proportions sont cohérentes avec les statistiques généralement observées dans les environnements de type MOOC, où les taux de décrochage sont fréquemment supérieurs à 30%. Cette répartition justifie l'intérêt porté au problème de la prédiction du décrochage, en tant qu'enjeu central pour améliorer l'engagement et la persévérance des apprenants dans les dispositifs massifs en ligne.

Volume d'interactions hebdomadaires selon la classe.

La Figure 3.11 présente le nombre total d'interactions enregistrées chaque semaine, en distinguant les décrocheurs et les non-décrocheurs, pour les jeux d'entraînement et de test. Ce graphique met en évidence l'évolution hebdomadaire de l'engagement global des deux groupes d'apprenants tout au long des six semaines du cours.

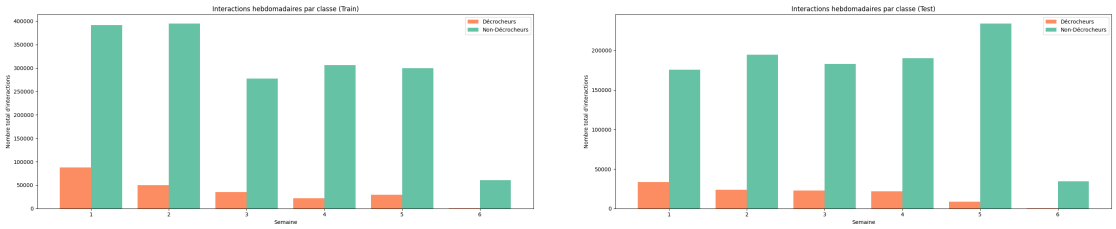

FIGURE 3.11 – Distribution du nombre total de clics par étudiant pour le jeu d'entraînement (gauche) et le jeu de test (droite).

Ces visualisations mettent en évidence une forte hétérogénéité des comportements d'engagement, à la fois dans le volume et dans la dispersion des interactions. On constate que :

— la majorité des **décrocheurs** se situent dans la zone inférieure du graphique, avec un nombre total de clics généralement inférieur à 500. Cela reflète une activité très limitée sur la plateforme, souvent concentrée sur les premiers jours du cours

— les **non-décrocheurs**, en revanche, affichent des niveaux d'engagement beaucoup plus élevés et une plus grande variabilité. Certains étudiants très actifs dépassent les **40 000 clics** dans le jeu d'entraînement et même **60 000 clics** dans le jeu de test. Ces cas extrêmes suggèrent un investissement soutenu dans l'ensemble des ressources pédagogiques.

L'analyse de cette distribution renforce l'hypothèse que le volume total d'interactions peut constituer une métrique précoce pertinent pour différencier les trajectoires d'apprentissage.

Clics moyens par étudiant selon la classe

La Figure 3.12 présente l'évolution hebdomadaire du nombre moyen de clics par étudiant, distinguée selon le statut final des apprenants (*décrocheur* vs *non-décrocheur*) pour les jeux d'entraînement et de test. Cette mesure permet de mieux appréhender la dynamique de l'engagement au fil des semaines, en tenant compte du nombre d'apprenants encore actifs à chaque étape.

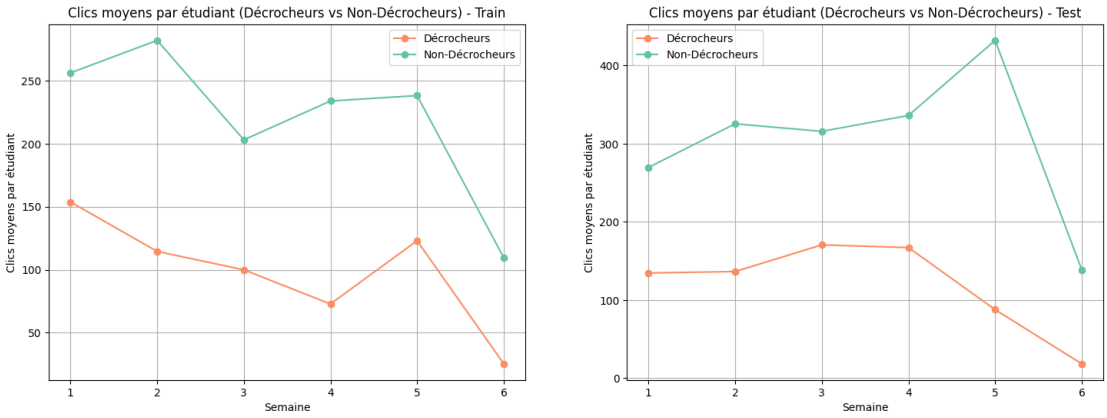

FIGURE 3.12 – Clics moyens par étudiant et par semaine pour les décrocheurs et les non-décrocheurs (jeu d'entraînement à gauche, jeu de test à droite).

L'analyse révèle un écart constant et significatif entre les deux groupes dès la première semaine :

— Les **non-décrocheurs** maintiennent un niveau d'interaction relativement élevé tout au long du cours, avec une activité particulièrement intense en semaine 2 et en semaine 5. Ces pics correspondent vraisemblablement aux moments où sont publiées des ressources clés ou des évaluations majeures. Leur constance dans le temps témoigne d'un engagement stable, soutenu par une exploration régulière des contenus proposés.

— Les **décrocheurs**, en revanche, manifestent un désengagement rapide : après un pic initial en semaine 1, leur activité chute brutalement dès la deuxième semaine, suggérant une rupture précoce dans le parcours d'apprentissage. Cette tendance est similaire dans les deux jeux de données, ce qui renforce la robustesse de cette observation.

Cette figure illustre clairement l'intérêt de suivre les trajectoires d'interaction sur une base hebdomadaire pour identifier les signaux faibles de décrochage. Elle appuie l'idée que les premières semaines sont déterminantes, et qu'un modèle prédictif efficace doit être sensible à la baisse rapide d'activité dès les débuts du cours.

Répartition des clics par étudiant.

La Figure 3.13 illustre la distribution du nombre total de clics par étudiant, pour les jeux d'entraînement et de test. Ces visualisations en boîte permettent de comparer l'intensité des interactions entre décrocheurs et non-décrocheurs. Chaque point correspond à un apprenant, et son positionnement reflète son niveau d'activité cumulé sur les six semaines du cours.

On observe que les non-décrocheurs présentent des volumes de clics nettement plus élevés que les décrocheurs, avec une variabilité importante au sein de chaque groupe. Les décrocheurs se concentrent majoritairement autour de faibles valeurs de clics, tandis que les non-décrocheurs incluent des profils très engagés atteignant plusieurs dizaines de milliers d'interactions.

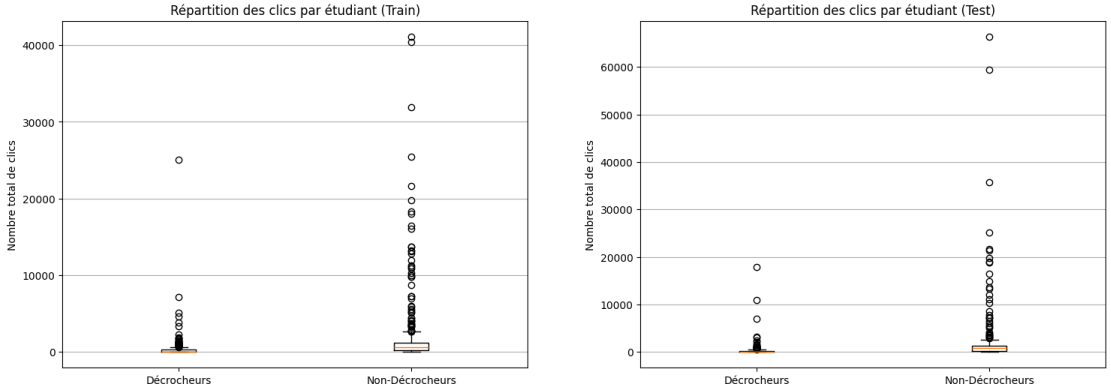

FIGURE 3.13 – Répartition du nombre total de clics par étudiant selon la classe (train à gauche, test à droite).

Participation hebdomadaire aux évaluations.

La Figure 3.14 illustre l'évolution hebdomadaire du nombre d'étudiants ayant soumis au moins une évaluation, en distinguant les décrocheurs des non-décrocheurs, pour les jeux d'entraînement et de test. Cette représentation permet de comparer les dynamiques de participation entre les deux groupes au fil des six semaines du cours.

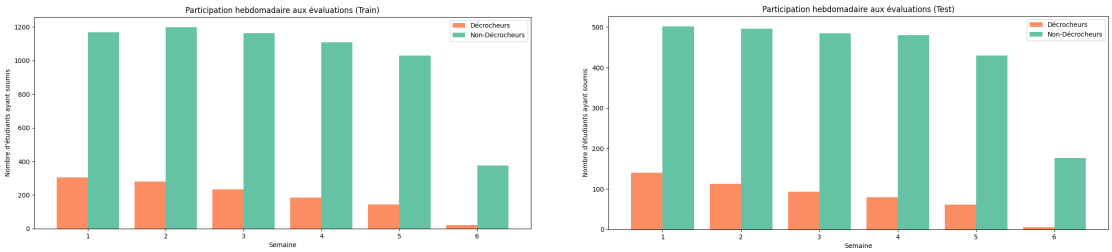

FIGURE 3.14 – Évolution du nombre d'étudiants ayant soumis une évaluation par semaine (train à gauche, test à droite).

On constate une nette divergence dans les comportements d'engagement évaluatif. Les non-décrocheurs maintiennent une participation soutenue et relativement constante tout au long du cours, ce qui témoigne d'une implication régulière dans les activités formatives. En revanche, les décrocheurs montrent un schéma très différent : après une première semaine relativement active, leur taux de soumission chute rapidement dès la deuxième semaine. Cette désaffection progressive traduit un désengagement précoce vis-à-vis des évaluations, souvent considéré dans la littérature comme un des premiers signaux de décrochage imminent. Cette tendance se vérifie dans les deux jeux de données, ce qui renforce la robustesse de cette observation comportementale.

Répartition hebdomadaire des types d'activités

La Figure 3.15 illustre, sous forme de heatmaps, la répartition des différents types d'activités réalisées chaque semaine dans le jeu d'entraînement (à gauche) et le jeu de test (à droite). Ces visualisations permettent d'appréhender la dynamique globale de l'engagement sur la plateforme.

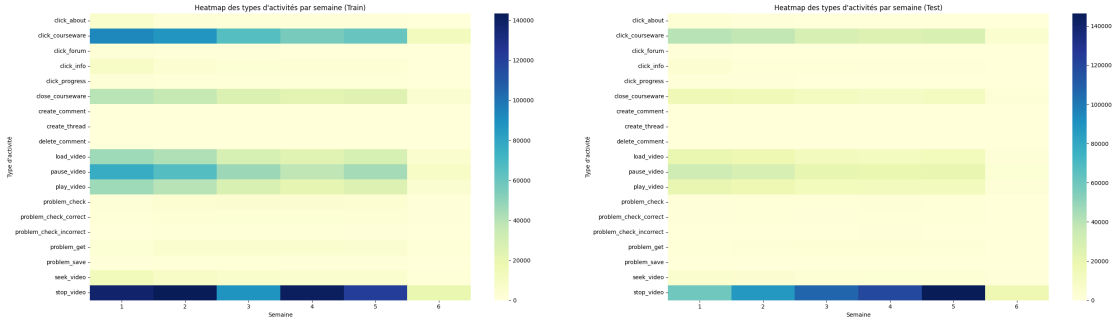

FIGURE 3.15 – Répartition hebdomadaire des types d'activités dans les jeux d'entraînement (gauche) et de test (droite)

Les activités les plus représentées sont celles liées au visionnage des vidéos, en particulier `stop_video`, qui domine nettement toutes les autres catégories, suivie de `pause_video`, `play_video`, `seek_video` et `load_video`. Cette forte densité d'interactions audiovisuelles témoigne du rôle central des vidéos dans la progression pédagogique du cours.

Les clics sur les contenus pédagogiques (`click_courseware`, `close_courseware`) sont également bien représentés et répartis de manière relativement stable au fil des semaines. Ces actions correspondent à la navigation dans les supports de cours et à la fermeture des ressources consultées.

À l'inverse, les interactions liées à la participation sociale ou collaborative, telles que `create_comment`, `create_thread` ou `click_forum`, sont très peu fréquentes. Cela laisse penser que les apprenants exploitent peu les fonctionnalités de discussion ou d'échange mises à disposition.

Enfin, les activités associées aux évaluations (`problem_check`, `problem_check_correct`, `problem_save`, etc.) apparaissent avec une intensité plus modérée, concentrées principalement entre les semaines 2 et 5. Elles traduisent la participation aux exercices formatifs et sommatifs, bien qu'en quantité bien inférieure à celle observée pour les vidéos.

Cette analyse met en lumière une forte prépondérance des comportements centrés sur la consommation de contenu, au détriment des activités interactives ou collaboratives, et une structure temporelle marquée par des pics autour des ressources audiovisuelles.

3.5 Synthèse

Dans ce chapitre nous avons présenté de manière approfondie les deux jeux de données publics mobilisés dans le cadre de cette thèse : **OULAD** et **MoocCN**. Issus de contextes pédagogiques fondamentalement différents, ils constituent un socle empirique solide pour tester, comparer et affiner les approches méthodologiques proposées.

— **OULAD**, issu de l'*Open University* britannique, se distingue par la richesse de ses variables (données démographiques, historiques d'interactions sur le VLE, évaluations notées), ainsi que par la longueur des parcours suivis (jusqu'à 38 semaines). Il reflète un environnement institutionnel formel, où l'engagement des apprenants s'inscrit dans une logique structurée, accompagnée et prolongée dans le temps.

— **MoocCN**, dérivé de la plateforme chinoise *XuetangX*, offre un aperçu contrasté typique des contextes MOOC : cours de courte durée (6 semaines), masse d'apprenants hétérogènes, interactions majoritairement centrées sur les vidéos, les quiz en ligne et une forte volatilité de l'engagement.

L'analyse descriptive conduite sur ces deux datasets a permis de dégager plusieurs tendances récurrentes et instructives :

— une différence nette d'activité, de persistance et de régularité entre les apprenants ayant terminé le cours et ceux ayant décroché

— un désengagement observable dès les premières semaines, confirmant leur rôle critique dans la prédiction précoce

— un poids important des traces d'interaction (clics) et des évaluations (soumissions et résultats) comme signaux prédictifs.

Ces observations, largement alignées avec les constats établis dans la littérature, confirment l'importance d'une modélisation fine et contextualisée des trajectoires d'apprentissage. Elles offrent un cadre de référence concret pour guider la construction, l'entraînement et l'évaluation des modèles présentés dans les chapitres suivants.

Chapitre 4

Métrique temporelle d'incertitude et modélisation frugale pour des prédictions fiables

Ce chapitre aborde trois dimensions de la modélisation prédictive en éducation : la **fiabilité**, l'**incertitude** et la **frugalité**. Il propose d'examiner comment ces aspects peuvent contribuer à mieux comprendre la robustesse et les limites des modèles utilisés dans des contextes d'apprentissage en ligne. La section 4.2 présente un **cadre de mesure de l'incertitude et de la fiabilité** , combinant variance, entropie et dynamique temporelle au travers de notre contribution, la métrique *TWPUS*. La section 4.3 explore la **frugalité des données**, en s'intéressant à la manière dont l'utilisation de données additionnelles peut être conditionnée par la confiance accordée aux prédictions au travers de notre contribution le modèle **FEP**, et introduit une approche heuristique de la **fiabilité**, fondée sur la cohérence entre prédictions intermédiaires et résultats finaux. Enfin, les sections 4.4 et 4.5 présentent les expérimentations menées sur les deux jeux de données publics (OULAD et MOOC-CN), et leurs résultats, ainsi qu'une discussion (section 4.6) et une conclusion (section 4.7) visant à situer les apports et les limites de l'approche proposée.

4.1 Introduction

Les modèles prédictifs connaissent un essor considérable et sont de plus en plus mobilisés pour anticiper des événements critiques à partir de données historiques et de flux d'interactions variés [19, 48]. Leur évaluation repose généralement sur des métriques globales tels que la précision, le rappel ou encore l'aire sous la courbe ROC [24]. Ces métriques constituent des repères essentiels pour comparer des approches, mais elles demeurent insuffisantes pour apprécier la robustesse réelle des systèmes lorsqu'ils sont confrontés à la variabilité des données, à leur évolution temporelle et à leur hétérogénéité structurelle [28, 105].

Comme discuté dans l'état de l'art (chapitre 2), trois dimensions apparaissent centrales pour dépasser ces limites : la **fiabilité**, l'**incertitude** et la **frugalité**. La fiabilité renvoie à la capacité d'un modèle à maintenir des prédictions cohérentes et stables au-delà des conditions expérimentales idéalisées [74, 40]. L'incertitude exprime le degré de confiance que le modèle accorde à ses propres décisions, fournissant une lecture plus nuancée des résultats et des erreurs possibles [146, 141]. Quant à la frugalité, désigne une approche de modélisation qui vise à optimiser l'usage

des données et des ressources, en recherchant le meilleur compromis entre la quantité d'information mobilisée et la performance obtenue. Elle s'inscrit dans une démarche de conception efficace, où seules les données réellement informatives sont exploitées [44, 165].

Ce chapitre repose sur deux contributions complémentaires : une métrique temporelle d'incertitude (TWPUS), visant à caracteriser l'évolution de la confiance du modèle, un modèle frugal (FEP), qui détermine, sur la base de cette incertitude, si l'intégration de données additionnelles est nécessaire. Ces deux contributions s'articulent autour d'un objectif commun : adapter le modèle aux situations où les données disponibles sont insuffisantes, incohérentes ou non informatives. L'analyse s'appuie sur une série d'expérimentations menées à partir de jeux de données publics, afin d'examiner ces dimensions de manière intégrée et d'en discuter les apports, les limites et les perspectives.

4.2 Incertitude des prédictions

L'évaluation de l'incertitude d'un modèle prédictif vise à estimer le degré de confiance associé à ses décisions. Contrairement aux métriques de performance globales, qui mesurent la qualité moyenne des prédictions sur un ensemble de données, l'incertitude est définie à un niveau individuel et dynamique [146]. Elle permet d'indiquer, pour chaque élément (dans notre cas un apprenant) et à chaque instant, si la prédiction fournie par le modèle est produite avec une forte confiance ou, au contraire, dans une situation de doute élevé.

Toutefois, l'incertitude à elle seule ne constitue pas un levier de confiance suffisant : un modèle peut afficher une confiance élevée tout en étant peu fiable (sûr de ses erreurs), ou au contraire se montrer hésitant alors que ses prédictions demeurent correctes et stables [94, 70]. La fiabilité et l'incertitude doivent donc être considérées comme deux dimensions complémentaires pour caractériser la valeur opérationnelle d'un système prédictif.

Plusieurs métriques existent dans la littérature pour quantifier cette incertitude. Dans ce travail, nous retenons deux métriques complémentaires largement utilisées : la variance des sorties issues d'ensembles de modèles et l'entropie des distributions de probabilité prédictives, définis et détaillées dans la section 2.2.6.

4.2.1 Interprétation en contexte éducatif

Ces deux métriques apportent des perspectives complémentaires. La variance met en évidence le degré de consensus entre plusieurs modèles, ce qui renseigne sur la stabilité de la décision. L'entropie, quant à elle, traduit le degré d'hésitation d'un modèle donné face à un exemple particulier. Dans un cadre éducatif, ces mesures permettent d'identifier les situations où le modèle risque de se tromper ou de fournir un signal peu exploitable. Leur intégration dans une analyse temporelle ouvre la voie à la détection de trajectoires d'incertitude et à la construction de profils interprétables, qui seront présentés dans la section suivante.

4.2.2 Calibration des seuils de variance et d'entropie

La variance et l'entropie constituent deux mesures de l'incertitude. Toutefois, pour qu'elles puissent être exploitées, il est nécessaire de définir des seuils permettant de distinguer les prédictions considérées comme fiables de celles jugées incertaines. La calibration joue ici un rôle

central : elle consiste à relier les valeurs d'incertitude calculées par le modèle aux erreurs effectivement observées. Autrement dit, une valeur faible de variance ou d'entropie ne doit pas seulement refléter un désaccord réduit entre modèles ou une forte concentration de probabilité sur une classe ; elle doit correspondre à une probabilité réelle plus élevée que la prédiction soit correcte.

Concrètement, chaque prédiction est associée à une variable d'erreur binaire, définie de la manière suivante :

$$error_{i,t} = \begin{cases} 1 & \text{si } \hat{y}_{i,t} \neq y_i \\ 0 & \text{sinon} \end{cases} \tag{4.1}$$

où $\hat{y}_{i,t}$ est la prédiction du modèle pour l'exemple i à l'instant t et y_i la classe réelle.

La calibration consiste alors à analyser la correspondance entre la distribution des valeurs d'incertitude et la fréquence observée des erreurs. Lorsque ce lien est bien établi, une prédiction annoncée comme incertaine est effectivement associée à une probabilité d'erreur plus élevée, tandis qu'une prédiction confiante est justifiée par une probabilité d'erreur réduite [86, 167]. À l'inverse, un modèle mal calibré peut donner une confiance excessive à des prédictions erronées ou, au contraire, signaler inutilement des situations comme incertaines. La calibration conditionne donc l'usage fiable de la variance et de l'entropie comme baromètres de confiance, en permettant de transformer des mesures abstraites en métriques opérationnelles pour la prise de décision.

Les mesures d'incertitude (variance et entropie) sont alors considérées comme des scores, et une analyse ROC est conduite afin d'évaluer leur capacité à discriminer les bonnes et les mauvaises prédictions. Le seuil optimal τ est déterminé en maximisant l'indice de Youden [167] définit comme suit :

$$J = \text{Sensibilité} + \text{Spécificité} - 1 \tag{4.2}$$

où

$$\text{Sensibilité} = \frac{VP}{VP + FN} \quad \text{et} \quad \text{Spécificité} = \frac{VN}{VN + FP} \tag{4.3}$$

avec VP les vrais positifs, VN les vrais négatifs, FP les faux positifs et FN les faux négatifs.

Le recours à l'indice de Youden se justifie par plusieurs raisons. Tout d'abord, il fournit un critère simple et interprétable qui combine en une seule valeur les deux dimensions essentielles d'un classificateur : sa capacité à détecter correctement les cas positifs (sensibilité) et à éviter les fausses alertes (spécificité). Maximiser J revient donc à rechercher un équilibre optimal entre ces deux aspects, ce qui est particulièrement pertinent lorsque l'on souhaite transformer une mesure d'incertitude en signal décisionnel. Contrairement au choix d'un seuil arbitraire, cette approche garantit que le point retenu est celui qui maximise la capacité discriminante globale de la métrique.

Enfin, l'indice de Youden présente l'avantage d'être indépendant de la prévalence des classes, ce qui en fait un critère robuste pour comparer et calibrer des mesures d'incertitude dans des contextes où la proportion d'erreurs et de succès varie. L'utilisation de J permet ainsi de définir des seuils τ^{Var} et τ^{Ent} qui traduisent, de manière cohérente et objectivement calibrée, les niveaux d'incertitude effectivement associés à un risque accru d'erreur.

Pour déterminer les seuils de décision à partir des mesures d'incertitude, on considère l'incertitude comme un score permettant de séparer les prédictions correctes des erreurs. En faisant varier ce seuil, on observe sa capacité à distinguer les deux cas, puis on retient la valeur qui maximise l'indice de Youden. Ce seuil optimal assure le meilleur compromis entre détection des erreurs et limitation des fausses alertes. Cette calibration est effectuée sur un jeu de validation afin de garantir la robustesse du seuil choisi.

En appliquant exactement cette procédure à la distribution des variances, on obtient le seuil τ^{Var}, et en l'appliquant à la distribution des entropies, on obtient le seuil τ^{Ent}. Ces seuils traduisent des niveaux à partir desquels une valeur d'incertitude est effectivement associée à un risque accru d'erreur, et permettent de transformer les mesures d'incertitude en règles décisionnelles opérationnelles.

On définit ensuite une métrique binaire U_i combinant les deux signaux :

$$U_i = \begin{cases} 1 & \text{si } Var_i > \tau^{Var} \text{ ou } Ent_i > \tau^{Ent} \\ 0 & \text{sinon} \end{cases} \tag{4.4}$$

Une prédiction est ainsi marquée comme incertaine si elle franchit l'un des deux seuils calibrés. Le choix d'une métrique binaire répond à une exigence de simplicité et de lisibilité : il permet de transformer des mesures continues en une décision claire, directement exploitable dans les étapes suivantes. L'utilisation de seuils calibrés permet de relier l'incertitude aux erreurs réellement observées, plutôt que de reposer sur un choix arbitraire. La métrique ainsi obtenue reste une simplification, mais il fournit un repère clair pour l'agrégation temporelle présentée dans la section suivante.

4.2.3 Une métrique temporelle : le TWPUS

Si les mesures ponctuelles d'incertitude, telles que la variance ou l'entropie, permettent d'évaluer le degré de confiance associé à une prédiction donnée, elles ne prennent pas en compte l'évolution de cette incertitude dans le temps. Dans un cadre temporel, la dynamique de l'incertitude joue un rôle central : un modèle peut se montrer hésitant dans les premières phases, lorsque peu de données sont disponibles, puis se stabiliser au fil du temps ; à l'inverse, certains cas initialement prévisibles peuvent devenir de plus en plus incertains à mesure que de nouvelles données s'accumulent.

Pour capturer ces trajectoires, nous introduisons une métrique temporelle : le *Time-Weighted Persistent Uncertainty Score* (TWPUS). L'objectif de cette métrique est de mesurer la persistance de l'incertitude au cours du temps, tout en accordant un poids croissant aux instants tardifs, car une incertitude persistante en fin de parcours est plus problématique qu'une incertitude transitoire en début de parcours.

On définit, le TWPUS pour un étudiant i sur un interval T comme suit :

$$TWPUS_i = \frac{\sum_{t=1}^{T} w_t \cdot U_{i,t}}{\sum_{t=1}^{T} w_t} \tag{4.5}$$

où :

— $U_{i,t} \in \{0, 1\}$ est le signal d'incertitude pour l'étudiant i au temps t (1 si l'incertitude dépasse un seuil défini, 0 sinon),

— w_t est un poids temporel défini comme une fonction croissante de t, $w_t = \frac{t}{T}$ pour une pondération linéaire.
— T est la durée totale du parcours.

Le score est normalisé entre 0 et 1. Une valeur proche de 0 traduit un parcours majoritairement confiant, tandis qu'une valeur proche de 1 reflète une incertitude persistante, surtout dans les temps derniers.

Le $TWPUS$ fournit ainsi une information complémentaire aux mesures ponctuelles. Il permet d'identifier non seulement les moments d'incertitude, mais aussi leur persistance et leur gravité relative au fil du temps.

Dans une perspective éducative, la dynamique temporelle est essentielle : un modèle peut hésiter en début de cours, lorsque peu de données sont disponibles, puis se stabiliser au fil des semaines, à l'inverse, certains profils peuvent paraître prévisibles au départ mais devenir de plus en plus incertains à mesure que les données s'accumulent. Le $TWPUS$ met en évidence les étudiants pour lesquels l'incertitude reste élevée même en fin de parcours, signalant un risque particulier de mauvaise prédiction ou de trajectoire atypique. À l'inverse, il distingue les incertitudes transitoires, moins problématiques, qui peuvent être résolues naturellement au fur et à mesure que les données s'accumulent.

4.2.4 Profils d'évolution de l'incertitude

L'analyse de l'incertitude à l'échelle temporelle ne se limite pas à la mesure ponctuelle ou agrégée, mais permet également d'identifier des trajectoires typiques parmi les étudiants. En effet, les signaux d'incertitude du modèle évoluent différemment selon les cas : pour certains une incertitude persistante, pour d'autres le niveau d'incertitude se réduit au fil du temps, tandis qu'une troisième catégorie développe une incertitude croissante à mesure que le cours progresse.

À partir de la métrique TWPUS et des séquences temporelles d'incertitude, nous proposons trois profils :

— **Persistants incertains :** Ce profil correspond à des cas pour lesquels le niveau d'incertitude du modèle reste constamment élevé tout au long du cours. Les prédictions demeurent instables et marquées par un doute permanent. Ce type de trajectoire est particulièrement préoccupant, car il met en évidence une fragilité structurelle du modèle vis-à-vis de ces individus.

— **Résolutifs :** Dans ce profil, le modèle exprime un haut niveau d'incertitude au début du parcours, lorsque peu de données sont disponibles, mais cette incertitude diminue progressivement avec l'accumulation d'informations. Il traduit la capacité du modèle à s'ajuster et à renforcer la confiance de ses prédictions au fil du temps.

— **Émergents incertains :** Ici, le modèle apparaît confiant au départ mais devient progressivement plus incertain à mesure que le temps avance. Ce phénomène peut refléter l'apparition de comportements d'apprentissage atypiques ou de données nouvelles qui perturbent la cohérence du modèle. Ces situations doivent être suivies avec attention, car elles signalent une perte de stabilité des prédictions dans la durée.

La typologie des profils d'incertitude apporte un outil d'interprétation utile. Elle permet de ne pas considérer l'incertitude uniquement comme un signal négatif, mais comme une information dynamique sur l'évolution des apprenants et sur la capacité du modèle à s'adapter. Dans une perspective pédagogique, ces profils peuvent orienter la mise en place d'interventions différenciées : par exemple, une intervention rapide pour les cas persistants, un suivi de consolidation pour les profils résolutifs, ou une analyse approfondie des comportements pour les profils émergents.

Ainsi, l'étude des trajectoires d'incertitude, combinée à des mesures ponctuelles et à la métrique temporelle TWPUS, peut aider à fournir une lecture cohérente et interprétable de la confiance des modèles prédictifs dans un cadre éducatif.

Algorithme d'estimation de l'incertitude

Le processus expérimental est formalisé dans l'algorithme 1, qui combine variance, entropie, calibration et attribution des profils temporels. L'algorithme 1 prend en entrée une série de jeux de données temporels, notés $\{X_1, \ldots, X_T\}$, où chaque X_t correspond aux données cumulées disponibles à l'instant t. Chaque observation est associée à une classe réelle $y_i \in \{0, 1\}$, permettant d'évaluer l'exactitude des prédictions. L'algorithme s'appuie par ailleurs sur un ensemble de M modèles indépendants (un *ensemble* de Random Forest dans notre cas), afin de quantifier l'incertitude à travers la variance et l'entropie des sorties.

Algorithm 1 Estimation temporelle de l'incertitude et identification des profils

Input : Jeux de données $\{X_1, \ldots, X_T\}$; classes réelles $\{y_i\}$; nombre de modèles M
Output : Score d'incertitude persistante $TWPUS_i$ pour chaque apprenant i

for $t \leftarrow 1$ **à** T **do**
  Entraîner un ensemble M_t de M modèles indépendants sur X_t
  for chaque apprenant i **do**
    Calculer la variance $Var_{i,t}$ (Équation 2.9) et l'entropie $H_{i,t}$ (Équation 2.8)
    Générer la prédiction $\hat{y}_{i,t}$ et vérifier si elle est correcte
  end for
  Déterminer les seuils hebdomadaires τ_t^{Var} et τ_t^{Ent} (Equation 4.2)
  for chaque apprenant i **do**
    Attribuer le drapeau d'incertitude $U_{i,t}$ (Équation 4.4)
  end for
end for
for chaque apprenant i **do**
  Calculer $TWPUS_i$ (Equation 4.5)
  Attribuer à l'apprenant i un profil
end for

En sortie, l'algorithme retourne, pour chaque apprenant i, un score d'incertitude persistante $TWPUS_i$, qui synthétise la manière dont les signaux d'incertitude évoluent au cours du temps, en pondérant davantage les instants tardifs où une incertitude élevée est particulièrement problématique.

Le fonctionnement général se déroule en deux étapes principales. Dans un premier temps, pour chaque instant t, un ensemble de modèles M_t est entraîné sur les données disponibles X_t.

Pour chaque apprenant, la variance $Var_{i,t}$ et l'entropie $H_{i,t}$ sont calculées à partir des sorties du modèle, puis comparées à des seuils temporels τ_t^{Var} et τ_t^{Ent} obtenus via une calibration hebdomadaire. Cette étape permet d'assigner un indicateur binaire d'incertitude $U_{i,t}$, signalant si la prédiction doit être considérée comme incertaine à l'instant t.

Dans un second temps, les indicateurs temporels $\{U_{i,1}, \ldots, U_{i,T}\}$ sont agrégés selon la métrique pondérée $TWPUS$, définie à l'équation 4.5, qui mesure la persistance de l'incertitude à l'échelle du parcours complet de l'apprenant. Ce score final constitue la sortie principale de l'algorithme et sert à caractériser les profils temporels d'incertitude présentés dans la section 4.2.4.

Métriques d'évaluation

Afin d'évaluer le cadre d'incertitude temporelle, plusieurs métriques sont suivies au cours des expérimentations :

1. **Évolution des incertitudes** Suivi hebdomadaire de la variance prédictive et de l'entropie prédictive pour chaque étudiant (Équation 2.9 & 2.8).
2. **Taux d'erreur certain/incertain** Comparaison des taux d'erreur entre les prédictions marquées comme certaines ($U_{i,t} = 0$) et celles marquées comme incertaines ($U_{i,t} = 1$) (Équation 4.1 & 4.4).
3. **Calibration des probabilités** Utilisation de courbes de calibration pour vérifier l'alignement entre les probabilités prédites de décrochage et les fréquences réellement observées (Équation 4.2).
4. **Score temporel pondéré TWPUS** Mesure de la persistance de l'incertitude dans le temps, en pondérant davantage les semaines avancées du cours (Équation 4.5).
5. **Profils et transitions d'incertitude** Catégorisation des trajectoires d'incertitude en trois profils (*persistants incertains*, *résolutifs*, *émergents incertains*) et analyse des transitions entre ces profils au fil des semaines à l'aide de diagrammes de Sankey (Sous-section 4.2.4).

4.3 Frugalité des données

La frugalité des données suppose de ne pas mobiliser systématiquement toutes les informations disponibles, mais de déclencher l'usage de données additionnelles uniquement lorsque cela est nécessaire. L'idée centrale est que l'accumulation de données n'est pas toujours synonyme d'amélioration des performances, et qu'il est possible d'obtenir des résultats comparables, voire supérieurs, en limitant l'accès à l'information. Cette logique s'inscrit directement dans l'hypothèse de la thèse, selon laquelle un accès sélectif et ciblé à des données complémentaires permet d'améliorer la qualité des prédictions pour les cas incertains, tout en évitant une consommation inutile de ressources.

4.3.1 Le modèle FEP

Le modèle **FEP (Frugal Early Prediction)** qui repose sur le principe de frugalité guidée par une estimation de la fiabilité de la prédiction. Ce modèle a pour objectif de limiter l'utilisation de données supplémentaires en ne les mobilisant que lorsque cela est réellement nécessaire.

L'idée est d'adapter la consommation de données au niveau de fiabilité estimé des prédictions.

Le fonctionnement de FEP repose sur deux étapes successives, illustrées dans la Figure 4.1.

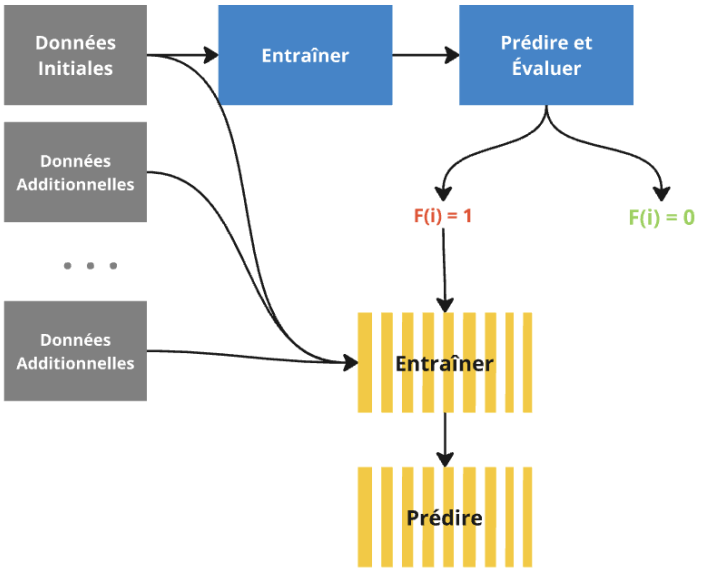

FIGURE 4.1 – Workflow du modèle FEP

La particularité de FEP réside dans le mécanisme de déclenchement $F(i)$, qui permet de décider de manière automatique et individualisée si l'on reste sur une prédiction simple ou si l'on active des données supplémentaires. Ainsi, FEP ne traite pas tous les cas de la même façon : les prédictions jugées fiables restent frugales en données, tandis que les prédictions peu fiables bénéficient d'un apport d'information destiné à renforcer la décision.

— **Phase 1 – Prédiction primaire :** une première prédiction est produite en utilisant uniquement les données initiales. Cette prédiction est accompagnée d'une estimation de fiabilité, à partir de la fonction de déclenchement $F(i)$. Cette étape permet de savoir si le modèle est confiant ou non dans sa réponse.

— **Phase 2 – Activation conditionnelle :** si la prédiction est jugée fiable ($F(i) = 0$), aucune donnée supplémentaire n'est utilisée. En revanche, si la prédiction est jugée incertaine ($F(i) = 1$), FEP active l'intégration de données additionnelles pour recalculer la prédiction avec un ensemble d'informations plus riche.

De cette manière, FEP met en pratique le principe de frugalité guidée par la fiabilité : il cherche à réduire la consommation inutile de données, tout en maintenant la qualité des prédictions grâce à une activation conditionnelle et ciblée.

4.3.2 Frugalité guidée par la fiabilité des prédictions intermédiaires

Une heuristique adaptée aux contextes éducatifs

Si des métriques comme l'*Earliness Stability Score (ESS)* (section 2.2.2) offrent une mesure intégrée de la précocité et de la stabilité, elles présentent néanmoins certaines limites. En particulier, elles se concentrent sur les trajectoires de prédictions correctes dans le temps, sans évaluer directement la cohérence entre les prédictions intermédiaires et l'issue finale. Or, dans un contexte éducatif, la valeur d'un modèle se mesure aussi à sa capacité à aligner des prédictions locales, produites au cours du processus d'apprentissage, avec le résultat final attendu, tel que l'échec ou le décrochage.

Pour répondre à cette limite, nous proposons une heuristique qui consiste à utiliser les prédictions intermédiaires comme baromètre de la fiabilité d'un modèle. L'idée est que si la prédiction d'une performance intermédiaire (par exemple la réussite à une évaluation de mi-parcours) est correcte, alors la fiabilité de la prédiction finale (par exemple le décrochage ou la réussite au cours) peut être renforcée. À l'inverse, une incohérence persistante entre prédictions intermédiaires et finale peut signaler une fragilité du modèle, indépendamment de ses performances globales.

Notre proposition peut être formalisée de la manière suivante. Pour un étudiant i, soit $P_{int}(i)$ la prédiction associée à l'évaluation ou l'activité intermédiaire (e.g. quiz, questionnaire). La fiabilité heuristique $Q(i)$ se définit alors comme :

$$Q(i) = \begin{cases} 1 & \text{si } P_{int}(i) = \text{Vrai}_{int}(i) \\ 0 & \text{sinon} \end{cases} \tag{4.6}$$

où $\text{Vrai}_{int}(i)$ désigne le résultat réel à l'évaluation intermédiaire. Cette mesure agit comme une métrique de cohérence locale. Elle est ensuite mobilisée comme baromètre pour juger de la qualité des prédictions finales.

Cette approche ne remplace pas les métriques existantes mais les complète, en proposant un angle d'analyse différent. Elle met en évidence le rôle structurant des prédictions intermédiaires comme points de référence pour juger de la fiabilité d'un modèle. En effet, une prédiction finale n'est pas seulement le résultat d'un calcul ponctuel : elle s'inscrit dans une trajectoire composée d'estimations successives qui traduisent la manière dont le modèle ajuste sa compréhension des données. Observer la cohérence de ces étapes intermédiaires permet de repérer si le modèle progresse de façon stable vers une conclusion ou, au contraire, s'il oscille de manière chaotique.

Déclenchement de la frugalité

Cette heuristique est intéressante car elle introduit une dimension dynamique et explicative. Elle ne se limite pas à évaluer la performance a posteriori, mais renseigne sur la consistance interne du processus de prédiction. Elle permet ainsi d'identifier les situations où une prédiction correcte pourrait reposer sur une suite instable d'estimations, et donc être moins fiable qu'il n'y paraît. Elle contribue également à rapprocher l'analyse de la logique humaine d'évaluation : tout comme un raisonnement gagne en crédibilité lorsqu'il s'appuie sur des étapes cohérentes, une prédiction devient plus robuste lorsqu'elle se construit sur une série d'estimations intermédiaires convergentes.

La frugalité guidée par la fiabilité, que nous défendons dans cette thèse, repose sur l'idée que l'usage de données supplémentaires ne doit pas être systématique, mais conditionné par une métrique interne reflétant la robustesse des prédictions. Dans ce cadre, l'heuristique introduite précédemment, fondée sur la cohérence entre prédictions intermédiaires et résultats observés, constitue un élément central pour cette métrique. Lorsque les prédictions suivent une trajectoire correcte et convergente, la prédiction finale peut être jugée suffisamment valide pour se limiter aux données initiales. À l'inverse, lorsque des incohérences ou des erreurs apparaissent au fil de la trajectoire, cela traduit une fragilité et justifie l'activation de données additionnelles.

Avec $Q(i)$ la fiabilité heuristique associée à l'individu i, la fonction de déclenchement de la frugalité $F(i)$ peut être définie comme suit :

$$F(i) = \begin{cases} 1 & \text{si } Q(i) = 0 \\ 0 & \text{si } Q(i) = 1 \end{cases} \tag{4.7}$$

où $F(i) = 1$ signifie que des données additionnelles doivent être mobilisées pour l'individu i, tandis que $F(i) = 0$ indique que les données initiales suffisent.

4.3.3 Frugalité guidée par l'incertitude

L'incertitude constitue un second levier possible pour déclencher la frugalité. Plutôt que de considérer uniquement la cohérence entre prédictions intermédiaires et résultats observés, il est possible de s'appuyer directement sur les métriques d'incertitude produites par le modèle, tels que la variance des sorties d'ensembles de modèles, l'entropie des distributions prédictives (sous section 2.2.6) ou la métrique temporelle TWPUS (sous section 4.2.4).

Lorsqu'un étudiant est associé à des prédictions de confiance, caractérisées un TWPUS faible, le modèle dispose d'un niveau de certitude suffisant et il n'est pas nécessaire de mobiliser des données supplémentaires. En revanche, si l'incertitude dépasse certains seuils, cela indique que le modèle hésite ou reste fragile. Dans ce cas, l'accès à des données additionnelles peut être déclenché afin d'améliorer la certitude de la prédiction.

La fonction de déclenchement de frugalité F peut également être définie par l'incertitude au lieu des prédictions intérmédiaires comme suit :

$$F(i) = \mathbf{1}\Big(Var_i^t > \tau_v \ \vee \ H_i^t > \tau_h \ \vee \ TWPUS_i > \tau_T\Big) \tag{4.8}$$

où :
— Var_i^t est la variance des probabilités prédictives de l'étudiant i à l'instant t t,
— H_i^t est l'entropie associée à la prédiction de i à l'instant t,
— $TWPUS_i$ est le score temporel d'incertitude de l'étudiant i,
— τ_v, τ_h et τ_T sont des seuils définis pour la variance, l'entropie et le TWPUS.

Cette approche rend la frugalité dynamique : elle ne repose pas sur une décision statique mais sur une évaluation continue de la confiance du modèle. Elle présente plusieurs avantages : elle cible précisément les cas incertains, elle réduit la consommation de données inutiles pour les cas confiants, et elle introduit une transparence supplémentaire en reliant l'activation de nouvelles données à des signaux mesurables d'incertitude.

4.4 Expérimentation

L'expérimentation est conçue pour évaluer les trois dimensions introduites dans ce chapitre : la **fiabilité**, la **frugalité** et l'**incertitude**. Ces dimensions sont intégrées dans un protocole commun qui simule le déroulement temporel d'un cours, en produisant chaque semaine des prédictions mises à jour à partir des données disponibles jusqu'à cette date.

La première expérimentation a pour objectif d'évaluer le modèle FEP en considérant simultanément deux dimensions : la fiabilité des prédictions et la consommation de données. Pour ce faire, FEP est comparé à des modèles de référence, de manière à analyser l'effet d'un recours sélectif à des données additionnelles par rapport à une utilisation systématique.

La seconde expérimentation, que nous détaillons dans la suite, se concentre sur l'incertitude. Son objectif est d'évaluer non seulement l'exactitude des prédictions, mais aussi la confiance que l'on peut leur accorder, à travers des métriques dédiées et une calibration hebdomadaire destinée à relier les signaux d'incertitude à la probabilité réelle d'erreur.

4.4.1 Expérimentation sur la frugalité et la fiabilité (modèle FEP)

Cadre expérimental

L'expérimentation s'appuie sur les jeux de données *OULAD* et *MoocCN*, déjà introduits dans le chapitre 3. Pour chaque étudiant, les données du cours suivi constituent les données initiales, tandis que les données additionnelles sont définies par le dernier cours antérieur complété.

Trois modèles sont comparés et détaillés dans la Figure 4.2 :
— le modèle **SS** (Single Source), n'utilisant que les données initiales (4.2.a),
— le modèle **MS** (Multi Source), intégrant systématiquement toutes les données disponibles (4.2.b),
— le modèle **FEP**, qui conditionne l'ajout de données supplémentaires à une estimation de la fiabilité de la prédiction issue du modèle SS (4.2.c).

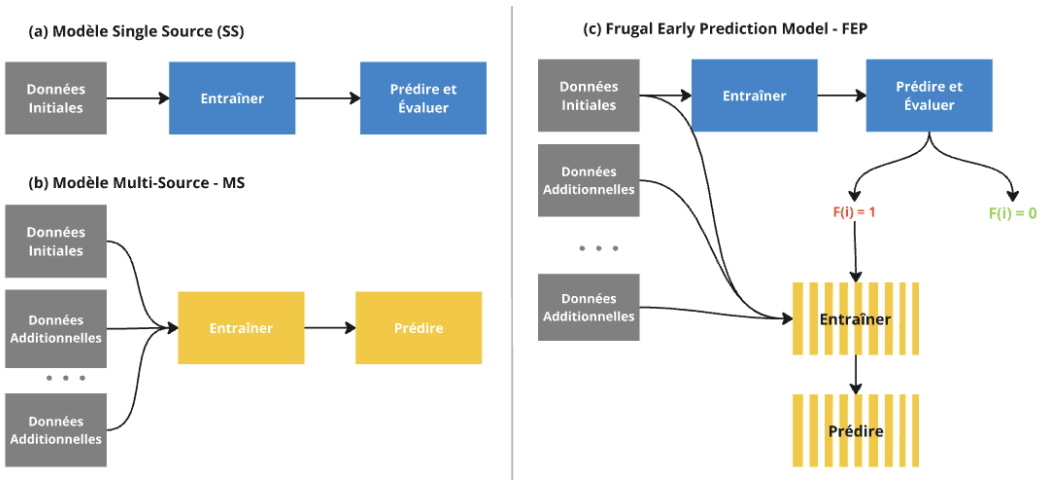

FIGURE 4.2 – Workflow des modèles SS-MS et FEP

L'ensemble des modèles est implémenté avec l'algorithme **XGBoost**, choisi après comparaison avec SVM, arbres de décision et forêts aléatoires, car il s'est avéré être le plus performant et robuste. Les expériences consistent à prédire, à différents instants du cours, la réussite ou l'échec à l'évaluation suivante.

Le fonctionnement du modèle FEP (figure 4.2.c) repose sur une logique en deux étapes. Dans un premier temps, la prédiction est effectuée à partir des données initiales (**SS**). Si la confiance estimée dans cette prédiction, calculée à l'aide de l'**heuristique définie précédemment** (sous section 4.3.2.0), est jugée suffisante, aucune donnée supplémentaire n'est sollicitée.
Dans le cas contraire, une seconde phase est déclenchée : des données additionnelles issues du dernier cours suivi sont intégrées, et la prédiction est réévaluée à l'aide du modèle multi-sources (**MS**). Une fois collectées, ces données restent disponibles pour les prédictions futures, de sorte que le coût d'accès n'est supporté qu'une seule fois par étudiant.

Ce fonctionnement est détaillé dans la Figure 4.3, qui met en évidence différents scénarios d'étudiants selon que des données supplémentaires sont activées ou non, et la manière dont elles influencent la fiabilité de la prédiction.

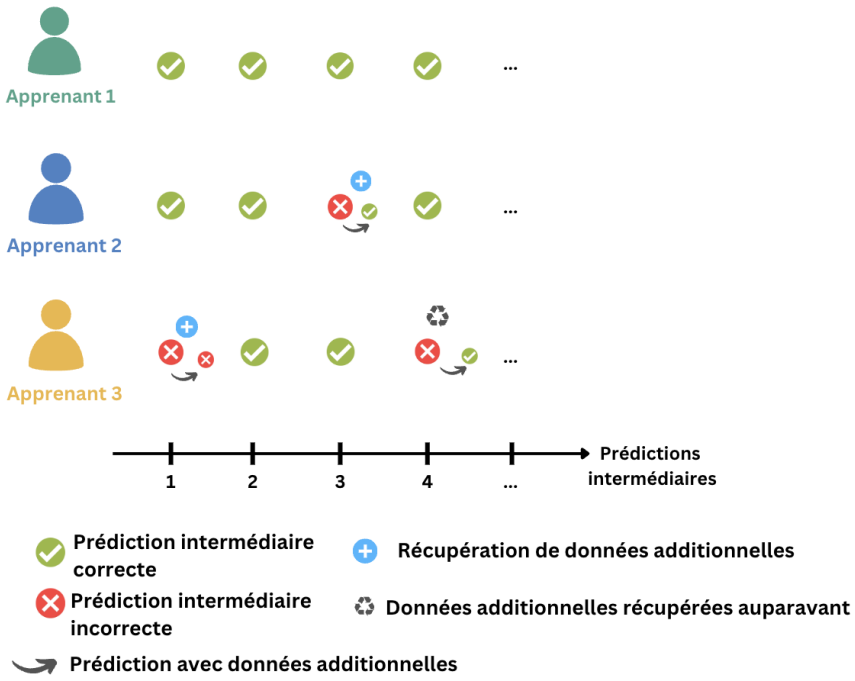

FIGURE 4.3 – Illustration du principe d'évaluation intermédiaire comme métrique de fiabilité.

Pour **l'apprenant 1**, les prédictions intermédiaires sont systématiquement correctes. La trajectoire est cohérente et stable, et aucune incohérence locale n'est détectée. Dans ce cas, le modèle considère que la prédiction finale peut être établie avec un niveau de confiance suffisant, sans recours à des données additionnelles. L'apprenant suit ainsi un parcours entièrement frugal.

Pour **l'apprenant 2**, une première série de prédictions correctes est suivie d'une erreur inter-

médiaire. Cette erreur déclenche la récupération de données additionnelles (symbolisée par l'icône +), puis une nouvelle prédiction est effectuée à partir de ces informations complémentaires. Les données additionnelles récupérées sont ensuite stockées et réutilisées pour les prédictions ultérieures (icône de recyclage), évitant ainsi des sollicitations répétées et préservant la frugalité du modèle.

Pour **l'apprenant 3**, la trajectoire est plus instable, avec plusieurs prédictions incorrectes successives. Une première erreur conduit à l'activation des données additionnelles, mais celles-ci ne suffisent pas immédiatement à stabiliser les prédictions. Dès qu'une nouvelle incohérence apparaît, le modèle réutilise les données déjà collectées pour recalculer la prédiction, sans avoir à solliciter de nouvelles sources. Ce comportement traduit un profil plus difficile à prédire, où la frugalité est maintenue grâce à la non-répétition des appels aux données externes.

Dans son ensemble, ce schéma met en évidence le rôle central de la cohérence temporelle des prédictions intermédiaires dans le déclenchement ciblé de la frugalité. Il illustre comment l'approche proposée distingue les cas stables, pour lesquels les données initiales suffisent, des cas fragiles ou incertains qui nécessitent l'intégration de données supplémentaires pour garantir la robustesse des prédictions finales.

Métriques d'évaluation

L'évaluation de FEP suit une approche multi-dimensionnelle :

— **Exactitude (Accuracy)** : proportion de prédictions correctes parmi l'ensemble des prédictions.
— **Earliness Stability Score (ESS)** : mesure conjointe de l'anticipation (earliness) et de la stabilité (stability) des prédictions correctes dans le temps.
— **Consommation de données (Data Consumption)** : quantité totale de données utilisées par étudiant, permettant de mesurer le gain de frugalité en comparaison avec les modèles SS et MS.

Cette combinaison de métriques permet d'évaluer simultanément la performance prédictive, la stabilité temporelle et l'efficacité en termes de consommation de données, afin de valider le caractère frugal et fiable du modèle FEP.

4.4.2 Expérimentation sur l'incertitude et la fiabilité

Après une première expérimentation consacrée à la frugalité et la fiabilité des modèles prédictifs à travers le FEP (Frugal Early Prediction), nous proposons une seconde expérimentation centrée cette fois sur l'incertitude. L'objectif est d'évaluer non seulement l'exactitude des prédictions, mais également la fiabilité avec laquelle elles sont produites, en s'appuyant sur la robustesse et la calibration des modèles.

Configuration expérimentale

L'incertitude est estimée à partir d'ensembles de modèles entraînés chaque semaine. Concrètement :

— Un ensemble de $M = 5$ *Random Forest* est entraîné indépendamment à chaque semaine t, en utilisant toutes les données cumulées disponibles jusqu'à la semaine t.

— Les probabilités issues des modèles sont utilisées pour calculer deux mesures complémentaires : la **variance prédictive** et l'**entropie prédictive**.

— Des seuils hebdomadaires spécifiques (τ_v, τ_h) sont calibrés par analyse ROC avec maximisation de la statistique de Youden, afin d'aligner les signaux d'incertitude avec les erreurs réellement observées.

— Les signaux binaires $U_{i,t} \in \{0, 1\}$ sont agrégés dans le score temporel TWPUS, qui valorise davantage l'incertitude persistante en fin de parcours.

— Enfin, les trajectoires d'incertitude sont regroupées en trois profils interprétables à partir du $TWPUS$: *persistants incertains*, *résolutifs*, et *émergents incertains*.

Les expérimentations ont été menées sur deux jeux de données publics : *MoocCN - XuetangX* (6 semaines, 3 700 étudiants, taux de décrochage ∼33%) et *OULAD* (CCC-2014J, 38 semaines, 2 498 étudiants, taux de décrochage ∼30%). Chaque semaine, un nouveau modèle est entraîné sur les données cumulées, et les incertitudes sont évaluées sur un ensemble de test tenu à part.

4.5 Résultats

Cette section présente les résultats expérimentaux obtenus sur les trois dimensions introduites dans ce chapitre : la fiabilité, la frugalité et l'incertitude. Pour chacune d'elles, nous analysons les performances des modèles considérés, en mobilisant les métriques adaptées et en comparant les tendances observées sur les deux jeux de données étudiés. Ces résultats permettent d'évaluer de manière conjointe la robustesse prédictive, l'efficacité dans l'usage des données et la capacité à quantifier la confiance associée aux prédictions.

4.5.1 Fiabilité des prédictions

La première dimension analysée concerne la fiabilité des modèles. La Figure 4.4 illustre les résultats obtenus sur OULAD, où trois configurations sont comparées : le modèle à source unique (SS Model), le modèle frugal guidé par l'heuristique intermédiaire (FEP Model), et le modèle multi-sources (MS Model).

L'analyse des résultats sur le jeu de données OULAD confirme la supériorité du modèle frugal (FEP) par rapport aux approches à source unique (SS) et multi-sources (MS). Dès les premières évaluations, le modèle FEP se distingue par une exactitude moyenne d'environ 0,95 au jour 7, contre 0,86 pour le modèle MS et 0,84 pour le modèle SS. Cet écart précoce, de l'ordre de 9 à 11 points de pourcentage, indique que la stratégie d'activation conditionnelle des données additionnelles permet d'améliorer la qualité prédictive dès les premières phases du cours, où les informations disponibles sont encore limitées.

Cette tendance se maintient au fil du temps : autour du jour 67, le modèle FEP atteint un pic d'exactitude proche de 0,97, surpassant le modèle MS (0,93) et le modèle SS (0,91). Même dans les périodes plus avancées (jours 158 à 214), lorsque tous les modèles disposent d'un volume de données suffisant, le FEP conserve une performance légèrement supérieure (écart moyen de 5 à 8 points) et plus stable. Ces résultats montrent que l'approche frugale ne dégrade pas la précision dans les phases tardives, tout en optimisant son utilisation des sources d'information durant les

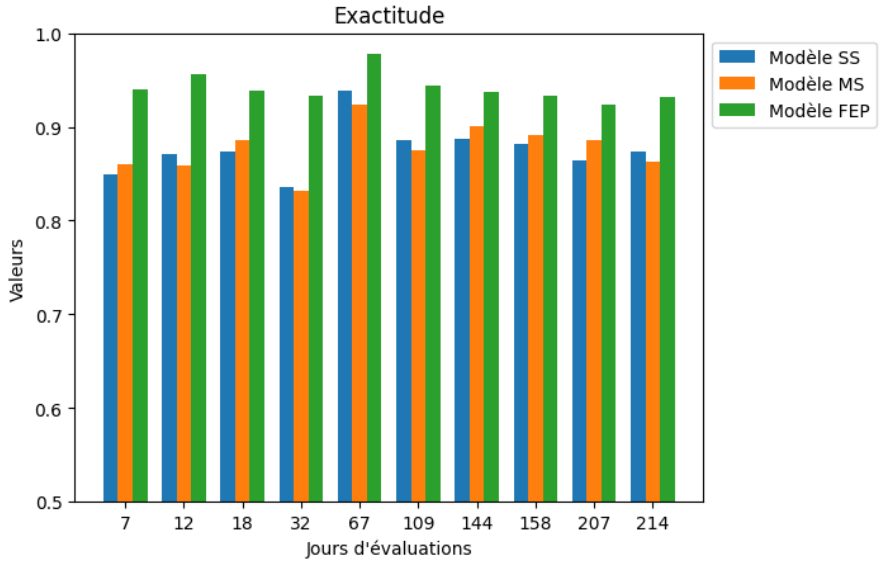

FIGURE 4.4 – Exactitude des modèles au fil des semaines (OULAD).

étapes intermédiaires.

Cette observation est confirmée par l'évolution de l'*Earliness Stability Score (ESS)* (Figure 4.5).

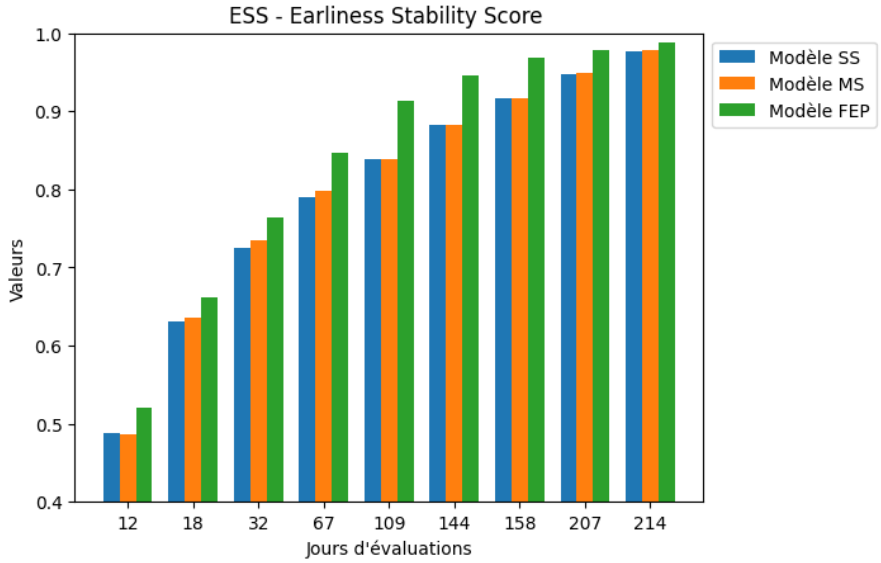

FIGURE 4.5 – Évolution de l'ESS des modèles au fil des semaines (OULAD).

Le modèle FEP obtient un ESS initial d'environ 0,52 au jour 12, soit 4 points de plus que

le modèle SS, traduisant une meilleure précocité et stabilité des premières prédictions. Au fil des semaines, le score ESS du FEP augmente plus rapidement : il atteint 0,9 dès le jour 109, contre 0,84 pour le modèle MS et 0,81 pour le modèle SS, puis dépasse 0,9 à la fin du cours. Ce comportement souligne la capacité du modèle à maintenir une cohérence temporelle élevée tout en produisant des prédictions fiables plus tôt que les approches de référence.

En somme, ces résultats montrent que l'activation sélective des données additionnelles, guidée par la fiabilité heuristique, permet au modèle FEP d'obtenir des performances à la fois plus précises, plus stables et plus précoces, sans recourir à une utilisation systématique et coûteuse de toutes les sources disponibles.

Des résultats comparables sont observés sur le jeu de données MOOC-CN (Figures 4.6 et 4.7).

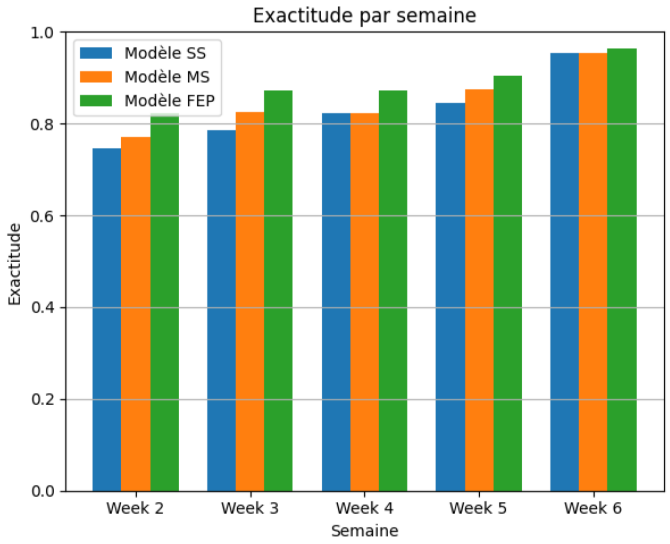

FIGURE 4.6 – Exactitude des modèles au fil des semaines (MOOC-CN).

Sur ce jeu plus court, le modèle FEP conserve un avantage clair sur les modèles à source unique (SS) et multi-sources (MS). Dès la deuxième semaine, il atteint une exactitude moyenne d'environ 0,82, contre 0,75 pour le modèle SS et 0,79 pour le modèle MS. Cette supériorité se maintient jusqu'à la fin du cours, où les trois modèles convergent autour de 0,95, confirmant que FEP permet d'atteindre des performances élevées plus rapidement.

L'analyse de l'*Earliness Stability Score (ESS)* (figure 4.7) révèle une tendance similaire : le FEP affiche un ESS initial d'environ 0,89, supérieur de 3 à 4 points à celui du modèle MS et de 5 points à celui du modèle SS. Cette différence traduit une meilleure cohérence temporelle des prédictions et une capacité accrue à stabiliser les décisions dès les premières semaines.

Ces résultats montrent que l'activation conditionnelle des données additionnelles, guidée par l'heuristique de fiabilité, constitue un mécanisme efficace pour renforcer la robustesse et la stabilité du modèle, même dans des contextes d'apprentissage plus courts ou moins riches en données.

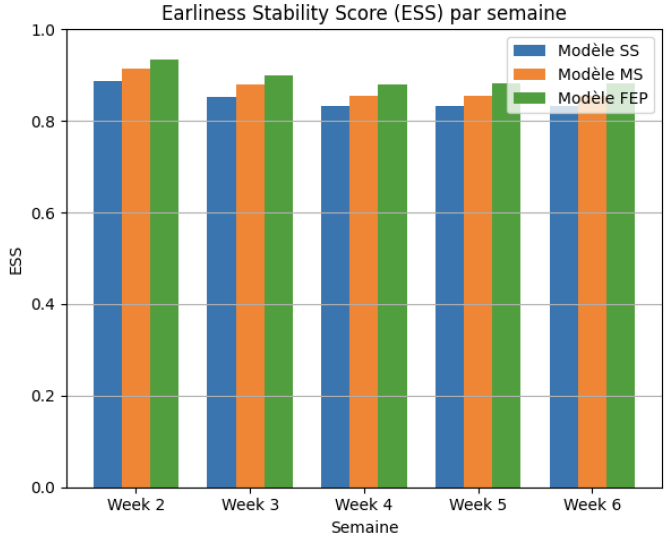

FIGURE 4.7 – Évolution de l'ESS des modèles au fil des semaines (MOOC-CN).

Dans l'ensemble, l'analyse de l'exactitude et de l'ESS sur deux corpus distincts confirme que :

— la fiabilité ne peut être réduite à une performance ponctuelle, mais doit être envisagée comme une trajectoire temporelle

— les modèles guidés par des métriques de fiabilité (ici l'heuristique intermédiaire) atteignent une meilleure cohérence temporelle

— dans un cadre éducatif, cette robustesse accrue permet de déclencher plus tôt des interventions pédagogiques tout en réduisant le risque de signaux instables.

4.5.2 Frugalité

La deuxième dimension évaluée concerne la frugalité, entendue comme la capacité à limiter la consommation de données tout en maintenant un haut niveau de performance et de fiabilité. Dans nos expérimentations, cette frugalité est visée à travers le modèle *Frugal Early Prediction (FEP)*, qui n'active des données additionnelles que lorsque la fiabilité heuristique détecte une incohérence.

Sur le jeu de données *OULAD*, le modèle FEP active des données additionnelles pour seulement 323 étudiants sur 694 (soit 46%). La consommation totale est ainsi réduite à $1.46n$ données contre $2n$, où n est le nombre d'étudiants, pour le modèle multi-sources, ce qui correspond à une économie d'environ 27%. Malgré cette réduction significative, FEP atteint une précision et un ESS supérieurs au modèle à source unique et le modèle multi-sources. Des tendances similaires apparaissent sur le jeu de données *MOOC-CN*. Selon les semaines, FEP n'active des données additionnelles que pour 15 à 30% des étudiants, avec une diminution progressive jusqu'à moins de 5% en fin de parcours. Cette activation sélective permet de limiter l'intégration des données additionnelles tout en conservant des performances proches du modèle multi-sources, et supé-

rieures à celles du modèle simple.

Afin de synthétiser ces observations, le tableau 4.1 présente une comparaison des trois modèles selon leur part d'activation des données, leur consommation relative et le gain obtenu en termes de réduction de la quantité d'informations mobilisées. Il illustre la capacité du modèle FEP à maintenir une performance élevée tout en diminuant significativement la charge de données utilisée.

TABLE 4.1 – Évaluation de la frugalité sur les jeux de données *OULAD* et *MOOC-CN*

Jeu de données	**Modèle**	**% d'activation des données**	**Consommation relative**	**Gain estimé**
OULAD	SS	0%	$1.0n$	Référence minimale
	MS	100%	$2.0n$	—
	FEP	46%	$1.46n$	**−27%**
MOOC-CN	SS	0%	$1.0n$	Référence minimale
	MS	100%	$2.0n$	—
	FEP	15–30% (début) → 5% (fin)	$\approx 1.25n$	**−35 à −40%**

Ces résultats confirment que la frugalité n'est pas seulement une contrainte, mais peut être intégrée comme un mécanisme adaptatif, guidé par des signaux internes de fiabilité. Elle permet de concentrer l'effort de collecte et de traitement sur les cas réellement fragiles, offrant ainsi un compromis efficace entre robustesse prédictive et économie de ressources.

4.5.3 Incertitude et fiabilité

La troisième dimension évaluée est l'incertitude, qui reflète le degré de confiance des modèles dans leurs prédictions. Contrairement à la fiabilité, qui décrit la robustesse temporelle des résultats, l'incertitude vise à identifier les cas où le modèle hésite ou exprime un doute. Elle constitue ainsi une métrique complémentaire, permettant de comprendre pourquoi certaines prédictions échouent et d'anticiper les situations fragiles.

Résultats OULAD

L'évolution de l'entropie (Figure 4.8) met clairement en évidence des trajectoires différenciées selon les profils d'étudiants définis grâce au $TWPUS$.

Les **persistants incertains** conservent un niveau d'entropie élevé tout au long du cours, oscillant autour de $0,90$ à la première semaine et se maintenant au-dessus de $0,70$ en fin de période. Cette stabilité relative, associée à une lente diminution d'environ vingt centièmes sur la durée totale, indique une incertitude structurelle : ces apprenants restent difficiles à modéliser même après plusieurs semaines d'observation.

Les **résolutifs**, à l'inverse, présentent une décroissance nette et régulière de leur entropie, passant de valeurs proches de $0,85$ à la semaine 1 à environ $0,25$ à la semaine 38. La baisse est particulièrement marquée pendant les quinze premières semaines, période au cours de laquelle le modèle semble affiner rapidement sa confiance à l'égard de ces étudiants.

Les **émergents incertains** se situent dans une position intermédiaire : leur entropie diminue progressivement de $0,90$ à environ $0,40$, mais reste systématiquement supérieure à celle

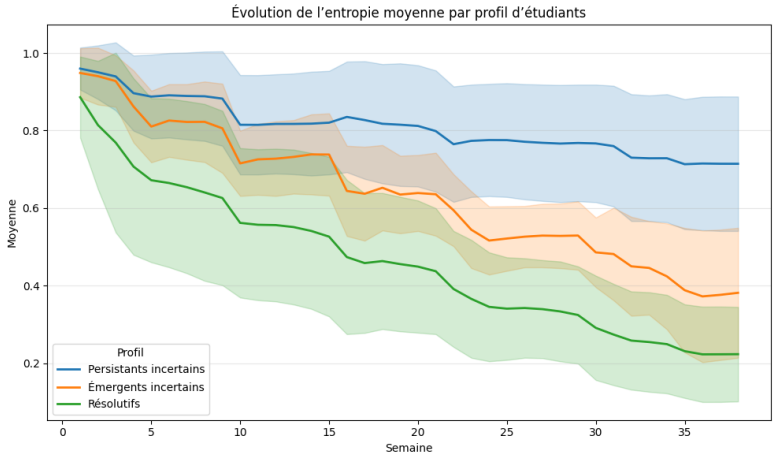

FIGURE 4.8 – Évolution de l'entropie moyenne par profil d'étudiants (OULAD).

des résolutifs. Ces trajectoires confirment que les trois profils ne se distinguent pas uniquement par leur niveau d'incertitude initial, mais aussi par leur dynamique de stabilisation dans le temps.

L'analyse de la variance (Figure 4.9) complète cette observation en mettant en évidence la dispersion interne à chaque profil.

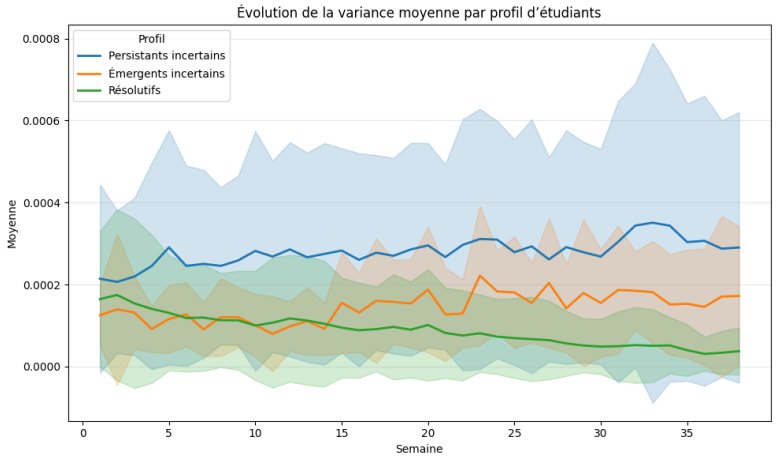

FIGURE 4.9 – Évolution de la variance moyenne par profil d'étudiants (OULAD).

Les **persistants incertains** affichent la variance la plus élevée, comprise entre $0,0002$ et $0,0006$ tout au long du cours, avec plusieurs pics dépassant $0,0007$ vers les semaines 20 et 30. Cette amplitude traduit des fluctuations importantes dans la confiance du modèle. Les **résolutifs** présentent une variance plus faible, généralement inférieure à $0,0002$, et tendant vers des valeurs proches de zéro en fin de cours. Les **émergents incertains** se situent à un niveau intermédiaire, avec une variabilité modérée autour de $0,0003$, traduisant une transition progressive entre les deux extrêmes. Ces résultats confirment que les profils ne diffèrent pas seulement par leur niveau

moyen d'incertitude, mais également par la stabilité temporelle de leurs prédictions.

La courbe de calibration (Figure 4.10) montre une relation globalement linéaire entre le niveau d'entropie et la probabilité d'erreur.

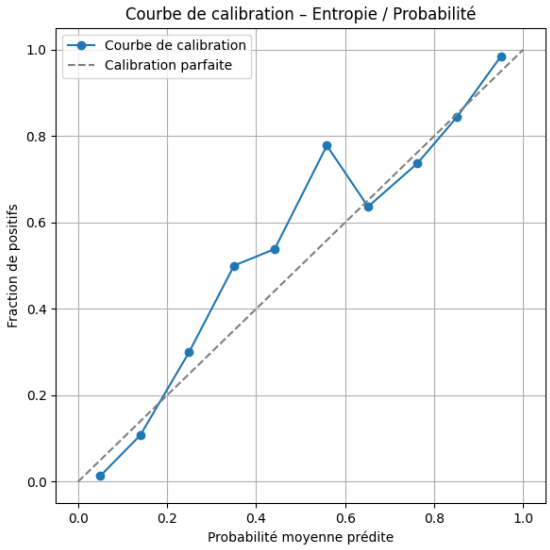

FIGURE 4.10 – Courbe de calibration entre entropie et probabilité observée (OULAD).

Pour des probabilités prédites autour de $0,2$, la fraction d'erreurs observées reste inférieure à $0,25$, tandis qu'au-delà de $0,8$, elle dépasse $0,85$. Le modèle tend donc à être bien calibré : une entropie élevée correspond effectivement à une probabilité d'erreur plus forte. Une légère sur-prédiction apparaît autour de $0,5$–$0,6$, où les erreurs sont un peu plus fréquentes que prévu, mais elle reste contenue et n'affecte pas la tendance générale. Cette cohérence entre incertitude estimée et erreur réelle valide la pertinence de l'entropie comme indicateur de fiabilité.

L'analyse du taux d'erreur par indicateur d'incertitude (Figure 4.11) confirme ces observations. Les prédictions marquées comme **incertaines** présentent un taux d'erreur compris entre $0,30$ et $0,42$ selon les semaines, avec un pic autour de la semaine 20. À l'inverse, les prédictions jugées **certaines** montrent des erreurs nettement inférieures, oscillant entre $0,05$ et $0,15$, puis tombant sous $0,05$ en fin de période. L'écart moyen entre les deux catégories dépasse $0,20$ points sur l'ensemble du cours, démontrant la capacité du modèle à distinguer efficacement les prédictions fiables des plus risquées.

Résultats MOOC-CN

Sur le jeu de données MOOC-CN (Figure 4.12), l'évolution de l'entropie met en évidence des différences de trajectoire similaires à celles observées sur OULAD, mais sur une période plus courte.

Les **persistants incertains** conservent une incertitude élevée et stable, avec une moyenne d'environ $0,56$ à la première semaine et une légère augmentation jusqu'à $0,61$ à la sixième semaine. Cette faible variation positive traduit une confiance du modèle qui reste limitée pour ce

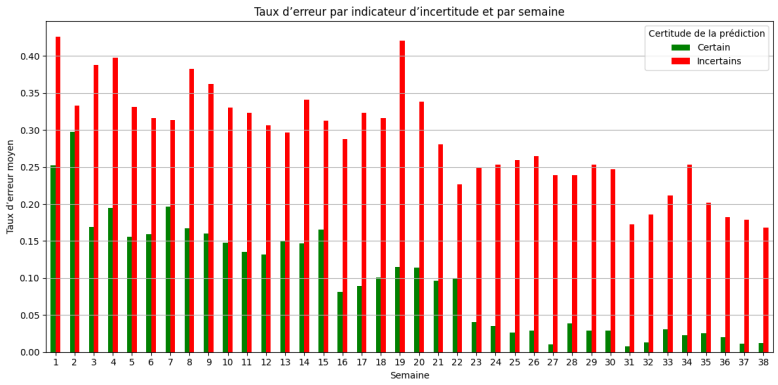

FIGURE 4.11 – Taux d'erreur selon l'indicateur d'incertitude au fil des semaines (OULAD).

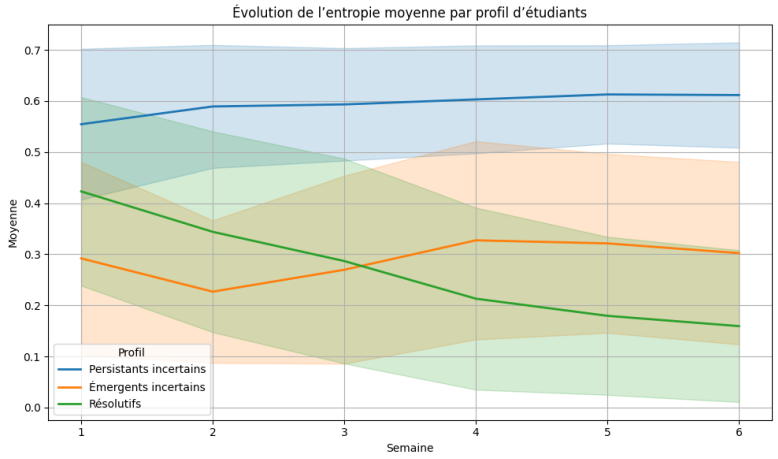

FIGURE 4.12 – Évolution de l'entropie moyenne par profil d'étudiants (MOOC-CN).

profil d'apprenants.

Les **résolutifs**, à l'inverse, présentent une décroissance continue de leur entropie, passant d'environ $0,42$ à la semaine 1 à $0,15$ à la semaine 6, soit une réduction d'environ vingt-sept centièmes, correspondant à une baisse de près de 65 %. Les **émergents incertains** se situent entre ces deux extrêmes : ils débutent autour de $0,29$, descendent légèrement à $0,23$ à la semaine 2, avant de remonter à $0,33$ à la semaine 4, puis se stabilisent autour de $0,30$. Bien que leur entropie tende à diminuer en fin de période, elle reste systématiquement supérieure à celle des résolutifs, avec un écart moyen d'environ quinze centièmes.

Ces résultats suggèrent que la réduction de l'incertitude est rapide pour les apprenants résolutifs, tandis qu'elle demeure lente et incomplète pour les autres profils.

L'analyse de la variance (Figure 4.13) confirme la hiérarchie observée sur l'entropie. Les **persistants incertains** présentent la variance la plus élevée, comprise entre 0,0005 et 0,0010, avec un pic proche de 0,0011 à la cinquième semaine. Cette amplitude traduit des fluctuations

importantes dans la confiance du modèle, suggérant une grande hétérogénéité des comportements prédictifs.

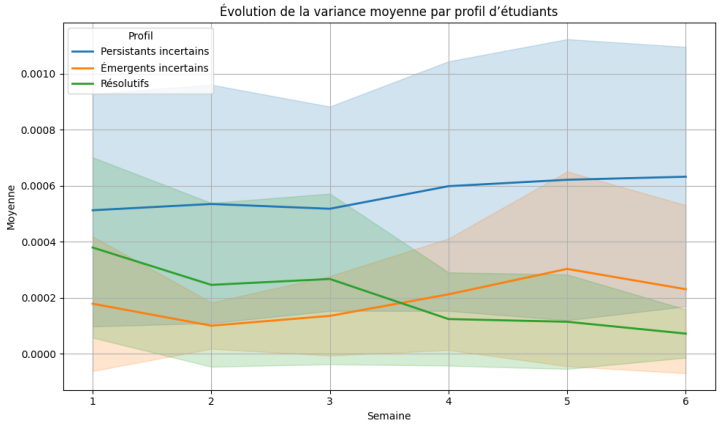

FIGURE 4.13 – Évolution de la variance moyenne par profil d'étudiants (MOOC-CN).

Les **résolutifs** conservent des valeurs plus faibles, autour de 0,0004 à la première semaine, qui diminuent progressivement pour atteindre environ 0,0001 à la sixième semaine. Les **émergents incertains**, enfin, oscillent dans une bande intermédiaire comprise entre 0,00015 et 0,0003, avec un léger pic à la semaine 5. Cette hiérarchie cohérente entre les profils, visible à la fois sur l'entropie et la variance, confirme que la stabilité prédictive suit la même structure que la confiance du modèle.

La courbe de calibration (Figure 4.14) montre une bonne correspondance entre les probabilités prédites et les taux d'erreur observés.

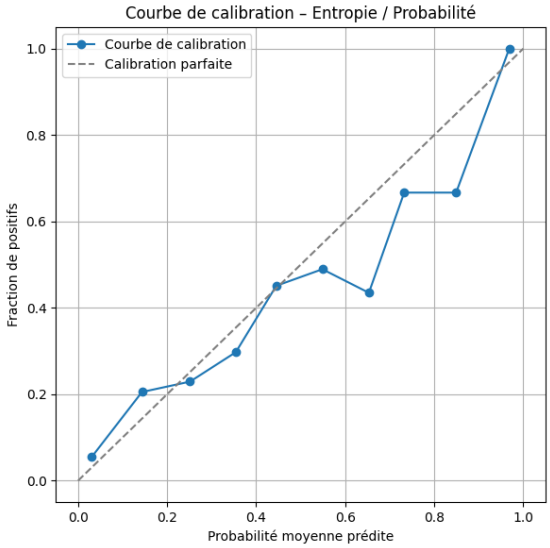

FIGURE 4.14 – Courbe de calibration entre entropie et probabilité observée (MOOC-CN).

Pour les prédictions dont la probabilité moyenne se situe autour de 0,20, la proportion d'erreurs observée est d'environ 0,25, tandis qu'elle atteint 0,80 lorsque la probabilité prédit une incertitude supérieure à 0,80. Le modèle est donc globalement bien calibré, les points s'alignant de manière satisfaisante sur la diagonale de calibration parfaite.

Une légère sur-estimation est néanmoins perceptible pour des probabilités comprises entre 0,50 et 0,60, où le taux d'erreur réel est d'environ dix centièmes plus élevé que prévu, mais cette déviation reste marginale et ne remet pas en cause la validité de la calibration.

La Figure 4.15 distingue nettement les prédictions selon leur niveau de certitude. Les prédictions marquées comme **incertaines** présentent des taux d'erreur compris entre 0,35 et 0,42 selon la semaine, tandis que les **certaines** se situent entre 0,07 et 0,20. L'écart moyen entre les deux catégories dépasse ainsi vingt points de pourcentage, ce qui démontre que l'indicateur d'incertitude identifie efficacement les prédictions les plus fragiles. En fin de période, les erreurs des prédictions certaines chutent sous la barre des 0,08, tandis que celles des incertaines demeurent supérieures à 0,35, confirmant la stabilité du signal de confiance.

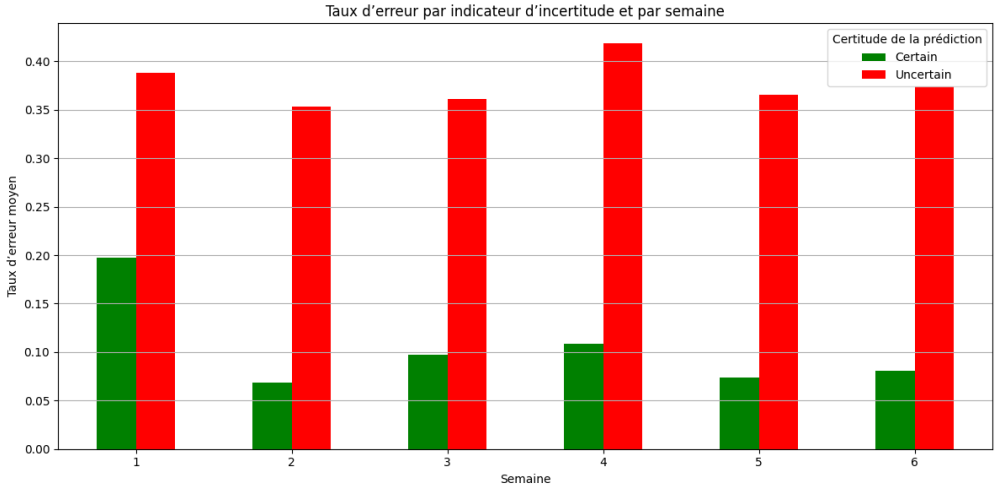

FIGURE 4.15 – Taux d'erreur selon l'indicateur d'incertitude au fil des semaines (MOOC-CN).

Enfin, le diagramme de Sankey (Figure 4.16) illustre la dynamique d'évolution des profils d'incertitude au fil des semaines. La majorité des étudiants conservent leur profil d'une semaine à l'autre, les flux intra-profils représentant entre 70 et 75 % des transitions, ce qui traduit une forte stabilité globale. Cependant, plusieurs évolutions significatives apparaissent : une partie des **persistants incertains**, estimée à environ 15 %, évolue vers le profil **résolutif** à partir de la première semaine, indiquant une amélioration progressive de la confiance du modèle pour ces apprenants. Inversement, une fraction plus restreinte des **résolutifs** (environ 8 à 10 %) bascule vers le profil des **émergents incertains** en fin de période, traduisant une légère instabilité dans la prédiction de certains étudiants.

Ces transitions, majoritairement descendantes, témoignent d'un processus de stabilisation progressive de la confiance du modèle, où certains apprenants initialement incertains finissent par être mieux prédits à mesure que les observations s'accumulent.

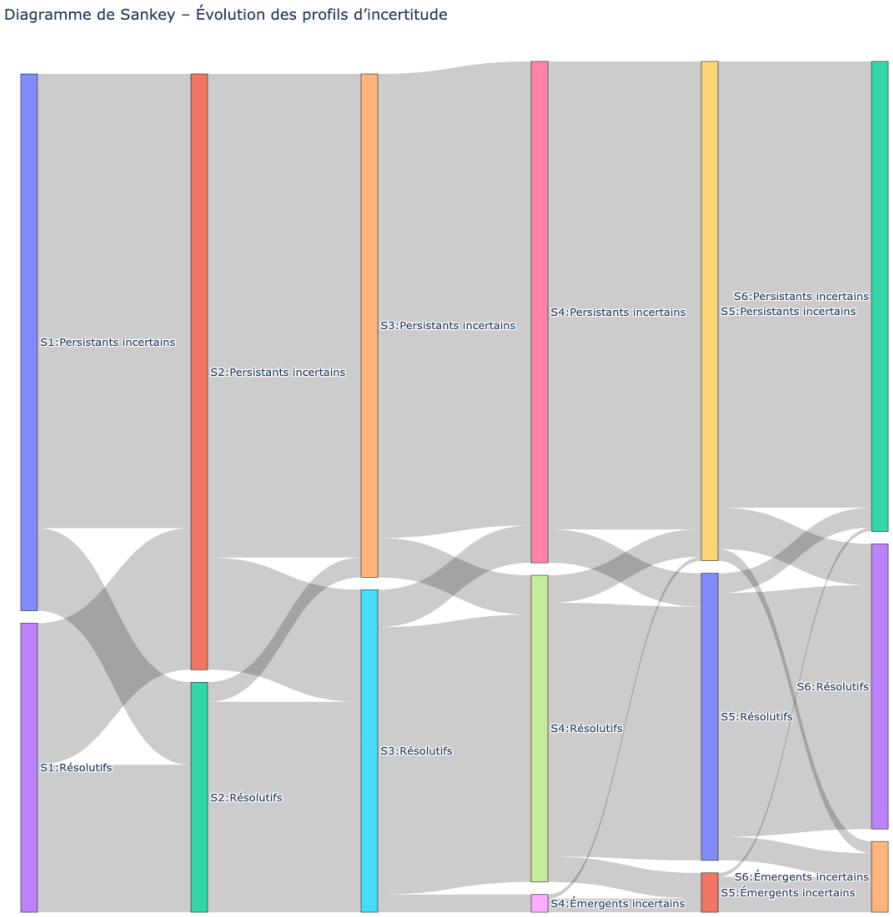

FIGURE 4.16 – Transitions entre profils d'incertitude au fil des semaines (MOOC-CN).

4.6 Discussion

Les résultats obtenus dans ce chapitre permettent d'évaluer de manière conjointe la fiabilité, l'incertitude et la frugalité des modèles prédictifs dans un contexte éducatif. Pris dans leur ensemble, ils confirment la pertinence d'une approche intégrée où ces trois dimensions ne s'opposent pas, mais se complètent pour renforcer la robustesse et la transparence des systèmes prédictifs. Néanmoins, au-delà de ces constats positifs, plusieurs limites méritent d'être discutées, afin de situer la portée réelle de ces travaux.

L'analyse de la fiabilité, d'abord, montre que la performance d'un modèle ne peut être évaluée uniquement à travers des métriques globales telles que la précision ou le rappel. La cohérence temporelle des prédictions apparaît comme un indicateur bien plus révélateur de la stabilité du modèle. L'heuristique introduite dans ce travail, fondée sur la cohérence entre prédictions intermédiaires et résultats finaux, a permis de détecter des divergences locales que des mesures

agrégées ne capturent pas. Cependant, cette approche présente des limites structurelles : elle dépend étroitement de la nature et de la qualité des jalons intermédiaires disponibles. Si ces jalons sont mal calibrés, peu discriminants ou peu corrélés à la performance finale, l'évaluation de la fiabilité peut devenir trompeuse. En d'autres termes, la fiabilité mesurée ici reflète autant la stabilité du modèle que la richesse informative du dispositif d'apprentissage sous-jacent. Dans certains contextes, un modèle pourrait être jugé instable non pas parce qu'il l'est réellement, mais parce qu'il s'appuie sur des données intermédiaires insuffisantes ou incohérentes.

Les résultats obtenus sur la frugalité apportent également des enseignements importants, mais appellent une lecture nuancée. Le modèle FEP montre qu'il est possible de réduire la consommation de données de près d'un tiers tout en maintenant, voire en améliorant, les performances. Ce constat valide le postulat d'une utilisation raisonnée et conditionnelle des données. Cependant, il soulève plusieurs questions. D'une part, le mécanisme d'activation conditionnelle repose sur un signal binaire de fiabilité, dont la sensibilité et la spécificité influencent directement le niveau de frugalité atteint. Une sous-activation peut conduire à ignorer des situations réellement fragiles, tandis qu'une sur-activation peut déclencher inutilement l'intégration de données supplémentaires. Ces deux phénomènes, faux négatifs et faux positifs de déclenchement, ont un impact direct sur la balance entre économie de ressources et stabilité prédictive. D'autre part, la frugalité telle qu'elle est mise en oeuvre ici reste essentiellement quantitative : elle réduit le volume de données mobilisées, sans encore discriminer finement quelles sources d'information apportent la plus forte valeur ajoutée. L'efficacité réelle de la frugalité dépendrait d'une compréhension plus fine de la valeur marginale de chaque source, ce qui ouvre la voie à des travaux ultérieurs sur la sélection dynamique des caractéristiques.

L'analyse de l'incertitude, pour sa part, offre un regard complémentaire sur la robustesse des modèles. Les mesures d'entropie et de variance montrent une cohérence satisfaisante avec les erreurs observées, confirmant leur utilité pour diagnostiquer les limites du modèle. Les profils temporels identifiés (persistants incertains, émergents incertains et résolutifs), se retrouvent dans les deux jeux de données étudiés, suggérant une transférabilité encourageante. Toutefois, ces métriques ne capturent qu'une partie de la complexité du phénomène. La variance rend compte de l'incertitude épistémique, liée à la connaissance limitée du modèle, tandis que l'entropie décrit davantage l'incertitude aléatoire, inhérente aux données elles-mêmes. Or, dans un contexte éducatif, ces deux formes d'incertitude s'entremêlent. Les comportements d'apprentissage peuvent être irréguliers ou ambigus sans pour autant traduire une faiblesse du modèle. Inversement, un modèle mal calibré peut afficher une confiance élevée sur des cas erronés, produisant une « fausse certitude ». Les résultats obtenus, bien qu'encourageants, doivent donc être interprétés avec prudence : la bonne calibration observée dans nos expérimentations ne garantit pas une fiabilité universelle. Elle dépend de la structure des données, du type de modèle employé et de la stabilité du processus d'entraînement.

Au-delà de ces constats spécifiques, la confrontation des trois dimensions révèle une tension entre performance, explicabilité et efficience. La fiabilité favorise la cohérence temporelle mais requiert des suivis longs et détaillés ; la frugalité réduit la redondance informationnelle mais peut accentuer les erreurs locales ; l'incertitude améliore la transparence mais introduit un surcoût de calcul et une complexité supplémentaire d'interprétation. Le principal apport du cadre proposé réside dans sa capacité à équilibrer ces dimensions plutôt qu'à les optimiser isolément. Les modèles n'ont pas pour ambition de maximiser la précision absolue, mais de fournir une estimation plus *honnête* de leur confiance et de leur robustesse. En ce sens, la valeur ajoutée de ce chapitre est moins dans les gains de performance que dans l'émergence d'un paradigme de *mo-*

délisation adaptative, où le modèle devient capable de qualifier la fiabilité de ses propres décisions.

Cependant, plusieurs limites doivent être reconnues pour garantir une lecture rigoureuse des résultats. Sur le plan méthodologique, la dépendance à des seuils de calibration globaux (calculés via l'indice de Youden) peut limiter l'adaptabilité du modèle à d'autres contextes. Une calibration dynamique, sensible à la semaine ou au profil d'apprenant, offrirait une meilleure granularité. Par ailleurs, la granularité hebdomadaire des analyses lisse probablement des fluctuations plus fines : des changements rapides de comportement, ou des dérives de caractéristiques (*feature drift*), peuvent survenir entre deux mesures sans être détectés. Sur le plan de la validité externe, les deux corpus utilisés (*OULAD* et *MOOCCN*) diffèrent à plusieurs égards. *OULAD*, ancré dans un contexte universitaire, couvre 38 semaines d'apprentissage structuré avec des traces régulières et complètes, tandis que *MOOC-CN*, issu de XuetangX, s'étend sur six semaines seulement, avec des interactions plus hétérogènes et un cadre d'autoformation. Cette complémentarité renforce la cohérence interne des résultats, mais leur transférabilité à d'autres contextes tels que la formation continue, l'apprentissage hybride ou présentiel, reste à confirmer, ces environnements étant souvent moins instrumentés et plus variables dans leurs traces d'activité. Enfin, la question de l'équité n'a pas été explorée dans ce chapitre : la fiabilité et l'incertitude peuvent varier selon les groupes d'apprenants, leurs parcours ou leur exposition aux activités, et ces différences pourraient introduire des biais dans le déclenchement de la frugalité ou dans l'interprétation des signaux de confiance.

Ces réserves n'invalident pas les résultats, mais invitent à une lecture prudente et à un élargissement des perspectives expérimentales. Elles soulignent la nécessité d'une évaluation orientée *utilité*, où les gains de fiabilité ou de frugalité sont mis en relation avec les coûts computationnels, les bénéfices pédagogiques et les effets potentiels sur la décision éducative. Elles suggèrent également d'explorer des approches temporelles continues (modèles bayésiens séquentiels, apprentissage incrémental) permettant d'actualiser les seuils de confiance en temps réel et d'adapter la frugalité aux conditions courantes d'apprentissage. Enfin, le couplage entre incertitude et équité apparaît comme un axe de recherche prometteur : un modèle conscient de ses zones d'incertitude pourrait, à terme, devenir un instrument de justice algorithmique, en signalant lui-même les cas où ses décisions risquent d'être inéquitables.

4.7 Conclusion du chapitre

Ce chapitre a proposé un cadre complet pour analyser conjointement la fiabilité, la frugalité et l'incertitude des modèles prédictifs dans un contexte éducatif. Il a montré que ces trois dimensions, souvent étudiées séparément, peuvent être intégrées de manière cohérente au sein d'un même processus d'évaluation. Les résultats empiriques, obtenus sur les jeux de données OULAD et MOOC-CN, confirment que la fiabilité doit être comprise comme une trajectoire temporelle et non comme une mesure ponctuelle, que l'incertitude peut servir d'indicateur opérationnel de confiance, et que la frugalité conditionnelle offre une voie efficace pour réduire la consommation de données sans compromettre la performance.

Les analyses critiques ont également mis en évidence plusieurs limites : la dépendance aux seuils globaux de calibration, la sensibilité aux jalons intermédiaires et à la qualité des données, la granularité hebdomadaire qui peut masquer certaines dérives, et la restriction du cadre expérimental à des contextes massifs en ligne. Ces limites ouvrent néanmoins des perspectives claires pour la suite : développer des calibrations adaptatives, explorer la généralisation à d'autres types de contextes et intégrer des approches incrémentales capables de gérer l'évolution des données dans le temps.

En définitive, ce chapitre introduit une étape importante dans le cadre de la thèse. Il établit les bases d'une *modélisation adaptative* où le modèle apprend non seulement à prédire, mais aussi à reconnaître les conditions de sa propre fiabilité. Ce paradigme servira de fondement aux deux chapitres suivants, consacrés respectivement à la sélection dynamique des caractéristiques et à l'apprentissage incrémental, qui prolongent la logique ici développée : apprendre à choisir, à s'ajuster et à se corriger dans un environnement d'apprentissage évolutif.

Chapitre 5

Sélection dynamique des caractéristiques pertinentes pour la prédiction

Ce chapitre s'inscrit dans le cadre plus large de la prédiction précoce du décrochage dans les environnements virtuels d'apprentissage (VLE). Il propose une contribution centrée sur la sélection dynamique des caractéristiques, particulièrement adaptée aux contextes de données déséquilibrées, évolutives et massives.

L'algorithme introduit, nommé **CI-DFS** (Class-Imbalanced Dynamic Feature Selection), repose sur trois piliers méthodologiques : une mesure de pertinence pondérée selon la distribution des classes, un seuil adaptatif calculé à partir de la distribution des valeurs des caractéristiques, et un mécanisme de gestion temporelle de la dérive via une fenêtre glissante. Contrairement aux méthodes classiques de sélection statique ou à seuil fixe, CI-DFS permet d'identifier à chaque instant t les caractéristiques les plus informatives, en tenant compte de l'évolution des comportements d'apprentissage au fil du temps.

Ce travail prolonge les recherches menées dans cette thèse sur l'amélioration de la qualité prédictive dans les systèmes éducatifs adaptatifs. Il vient compléter les contributions précédentes par une approche frugale de réduction de dimension dynamique, pensée pour des modèles entraînés de manière incrémentale.

Le chapitre est structuré comme suit : après une introduction des enjeux liés à la sélection dynamique, la section méthodologie détaille le fonctionnement de CI-DFS (section 5.2), les métriques d'évaluation mobilisées ainsi que les algorithmes comparateurs (section 5.3). Deux sections sont ensuite consacrées à l'application de l'algorithme sur les jeux de données OULAD (section 5.4) et MoocCN (section 5.5). Suivent une section dédiée aux résultats expérimentaux (section 5.6), comprenant une analyse d'ablation, puis une discussion critique des résultats (section 5.7) avant la conclusion (section 5.8).

5.1 Introduction

Prédire le décrochage des étudiants dans les VLEs constitue un enjeu central pour les institutions éducatives. En effet, l'identification précoce des apprenants à risque de décrochage, permet de déclencher des actions de soutien ciblées, réduisant ainsi les inégalités d'apprentissage et favorisant la réussite académique [47, 81]. La plupart des modèles prédictifs existants exploitent des traces comportementales extraites des plateformes éducatives : temps passé sur les ressources, clics, activités soumises, résultats intermédiaires [69, 156]. Ces modèles reposent généralement sur un ensemble statique de caractéristiques, défini de manière globale avant la phase d'entraînement. Or, cette approche présente deux limites majeures dans les contextes réels d'apprentissage en ligne : la nature temporelle et l'hétérogénéité des données d'interaction.

Tout d'abord, les données éducatives évoluent au fil du temps. Certaines caractéristiques (comme la participation aux forums) peuvent être pertinentes en début de parcours, tandis que d'autres (comme les résultats d'évaluation) gagnent en importance à mesure que l'année avance [42]. Dans un cadre dynamique, conserver un ensemble fixe de caractéristiques à exploiter, entraîne soit une sous-utilisation des nouvelles informations disponibles, soit une surcharge inutile de caractéristiques redondantes ou obsolètes. Ensuite, les données issues des VLEs sont très souvent déséquilibrées : les cas de réussite sont sur-représentés, rendant difficile la détection des décrocheurs (classe minoritaire), en particulier dans les premières semaines où peu d'information est disponible [168]. Dans d'autres domaines tels que la finance, la publicité ou la santé, la question de la sélection dynamique de caractéristiques (Dynamic Feature Selection, DFS) a été abordée de manière plus systématique. Plusieurs approches se basent sur des mesures d'information mutuelle (MI) pour sélectionner à chaque étape les caractéristiques les plus informatives [124, 169]. SAOLA, par exemple, est une méthode de DFS reconnue pour sa scalabilité, combinant pertinence et redondance en ligne. Toutefois, comme présenté dans l'état de l'art, ces algorithmes présentent trois limites lorsqu'ils sont transposés au domaine éducatif : (1) ils ne traitent pas le déséquilibre des classes, (2) ils reposent sur des seuils statiques de sélection, et (3) ils n'intègrent pas la notion de dérive temporelle (feature drift), qui reflète la perte de pertinence d'une caractéristique avec le temps [178].

Face à ces constats, ce chapitre introduit une approche alternative, CI-DFS (Class-Imbalanced Dynamic Feature Selection), qui prend en compte simultanément les enjeux du déséquilibre, de la variabilité temporelle et de la sélection adaptative. L'objectif est d'adapter le modèle au fil de l'arrivée des données, en ne retenant que les caractéristiques les plus pertinentes et non redondantes à chaque instant t, tout en s'assurant que les signaux liés au décrochage ne soient pas noyés par la dominance des classes majoritaires. CI-DFS se distingue par l'utilisation d'une information mutuelle pondérée par classe, d'un seuil adaptatif basé sur des percentiles dynamiques, et d'un mécanisme de gestion de la dérive des caractéristiques par fenêtre glissante.

La suite du chapitre présentera d'abord la méthodologie détaillée de l'algorithme, puis les résultats expérimentaux sur deux jeux de données représentatifs (OULAD et MoocCN), avant de discuter les apports, les limites et les perspectives de généralisation.

5.2 Méthodologie

5.2.1 Définition du problème

Dans les VLEs, les plateformes génèrent un flux continu de données à partir des interactions des apprenants : ouverture de ressources, soumissions de devoirs, accès aux évaluations, participation aux forums, etc. Ces traces constituent une base précieuse pour l'analyse prédictive, notamment pour anticiper le décrochage des apprenants. Toutefois, ces données présentent trois caractéristiques majeures qui complexifient leur exploitation : elles sont *temporelles*, *déséquilibrées*, et *croissantes en volume.*

Temporalité

Les données d'interaction évoluent jour après jour : de nouvelles activités sont proposées, certaines ressources ne sont disponibles qu'à partir d'un instant donné, et le comportement des apprenants se transforme au fil du temps (e.g., baisse de motivation, accélération des performances). Ainsi, les caractéristiques disponibles au temps t peuvent différer de celles observées aux temps précédents, tant en nature qu'en importance prédictive. Une approche figée risquerait de conserver des caractéristiques obsolètes ou d'ignorer des signaux émergents.

Déséquilibre des classes

Dans la majorité des formations, la classe des apprenants qui terminent la formation reste dominante, tandis que les abandons représentent une proportion minoritaire mais critique. Or, de nombreux algorithmes de sélection de caractéristiques sont insensibles à ce déséquilibre et privilégient les caractéristiques qui améliorent l'exactitude globale, au détriment de la capacité à identifier les apprenants en difficulté. Cela pose un problème fondamental en termes d'équité et d'utilité prédictive.

Croissance du volume de données

À mesure que le cours avance, le nombre de caractéristiques disponibles augmente. Cela entraîne une explosion de l'espace de représentation, une hausse du coût computationnel, un risque de surapprentissage et une baisse de la lisibilité des modèles. Une approche naïve qui intégrerait systématiquement toutes les caractéristiques historiques serait à la fois inefficace et difficilement interprétable.

Le problème que nous adressons dans ce chapitre peut donc se formuler ainsi :

> À chaque instant t, il s'agit de déterminer un sous-ensemble S_t de caractéristiques parmi les nouvelles caractéristiques disponibles F_t et l'ensemble cumulé des caractéristiques précédentes $F_{1:t-1}$, de manière à maximiser la capacité prédictive du modèle, en particulier sur la classe minoritaire (décrochage), tout en garantissant la frugalité du modèle, son adaptabilité temporelle et la maîtrise des ressources mobilisées.

Dans un VLE, les données évoluent continuellement au fil du temps : de nouvelles caractéristiques apparaissent, d'autres perdent leur pertinence, et la distribution des classes peut se déséquilibrer davantage. La tâche de sélection dynamique des caractéristiques vise à maintenir la capacité du modèle à s'adapter à ces évolutions, tout en limitant la consommation de données et de ressources. Contrairement aux approches statiques qui s'appuient sur un ensemble fixe

de variables, cette approche cherche à opérer une mise à jour incrémentale du sous-espace de caractéristiques, en intégrant progressivement les informations nouvelles et en éliminant celles devenues redondantes ou inutiles.

Cette tâche de sélection dynamique s'inscrit donc dans un cadre non stationnaire où la distribution des données et la pertinence des variables évoluent continuellement. Elle doit donc être capable d'ajuster en continu le sous-ensemble S^t, en intégrant les nouvelles caractéristiques informatives et en supprimant celles devenues obsolètes ou redondantes, afin de maintenir un compromis optimal entre performance, stabilité temporelle et efficience des ressources.

5.2.2 Présentation détaillée de l'algorithme CI-DFS

L'algorithme CI-DFS (Class-Imbalanced Dynamic Feature Selection) vise à assurer une sélection progressive, frugale des caractéristiques utiles à la prédiction du décrochage. Il repose sur une adaptation continue aux nouvelles données, à chaque pas de temps t, et combine trois mécanismes complémentaires :

— une **évaluation de pertinence** pour mieux gérer le déséquilibre entre classes

— un **seuil adaptatif** pour éviter les critères de sélection figés

— une **actualisation temporelle des caractéristiques** pour écarter celles devenues peu informatives.

À chaque instant t, CI-DFS prend en entrée un ensemble de nouvelles caractéristiques candidates F_t, les évalue selon leur utilité prédictive pour la tâche en cours (prédiction du décrochage), et met à jour l'ensemble S_t des caractéristiques sélectionnées, tout en maintenant un historique de leurs scores de pertinence.

L'algorithme 2 formalise l'ensemble de ces étapes en décrivant le processus complet de sélection dynamique à chaque pas de temps t.

Détail de l'algorithme :
En entrée, l'algorithme considère : F_t, l'ensemble des nouvelles caractéristiques candidates à l'instant t ; S_{t-1}, l'ensemble des caractéristiques déjà retenues à l'instant $t-1$; R_{t-1}, l'historique des valeurs de pertinence des caractéristiques à $t-1$; C, l'ensemble des classes ; λ, le seuil de redondance ; k, le percentile utilisé pour la sélection ; et w, la taille de la fenêtre temporelle.

En sortie, l'algorithme renvoie l'ensemble des caractéristiques sélectionnées S_t ainsi que les valeurs de pertinence mises à jour R_t.

Ligne 1 : Initialisation de l'ensemble courant de caractéristiques sélectionnées S_t en copiant S_{t-1}, l'ensemble issu de l'itération précédente.

Ligne 2 : De la même manière, l'historique des pertinences R_t est initialisé avec les valeurs de R_{t-1}.

Ligne 3 : Calcul du seuil de pertinence θ, défini comme le $k^{\text{ème}}$ percentile des valeurs présentes dans R_t. Ce seuil est utilisé pour filtrer les nouvelles caractéristiques candidates.

Algorithm 2 CI-DFS

Input : F_t ensemble des nouvelles caractéristiques candidates à un instant donné t ; S_{t-1} caractéristiques retenues depuis un instant donné $t-1$; R_{t-1} historique des valeurs de pertinence des caractéristiques à $t-1$; C ensemble des classes ; λ seuil de redondance ; k percentile ; w taille de la fenêtre.

Output : Caractéristiques sélectionnées S_t ; valeurs de pertinence mises à jour R_t

// Initialisation

1: $S_t \leftarrow S_{t-1}$

2: $R_t \leftarrow R_{t-1}$

// Calcule le k^e percentile de la pertinence des caractéristiques à l'instant $t-1$

3: $\theta \leftarrow Percentile_k(R_t)$

4: **for** each nouvelle caractéristique f_i in F_t **do**

// Vérifier la pertinence des caractéristiques - Fonction 1

5: $(relevant, relvalue) \leftarrow$ check_relevance(f_i, C, R_t, θ)

6: **if** *relevant* **then**

// Vérifier la redondance des caractéristiques avec celles déjà identifiées - Fonction 2

7: *redundant* $\leftarrow$ check_redundancy(f_i, S_t, λ)

8: **if not** *redundant* **then**

9: $S_t \leftarrow S_t \cup \{f_i\}$

10: $R_t[f_i] \leftarrow new(relvalue)$

11: **end if**

12: **end if**

13: **end for**

// Gestion de la dérive - Fonction 3

14: $\theta \leftarrow Percentile_{(100-k)}(R_t)$

15: $(S_t, R_t) \leftarrow$ manage_drift(S_t, C, R_t, w, θ)

// Renvoyer les caractéristiques sélectionnées et leurs pertinences

16: **return** (S_t, R_t)

Lignes 4–5 : Pour chaque nouvelle variable f_i dans F_t, on évalue sa pertinence via la fonction *check_ relevance* (fonction 1). Cette fonction retourne un booléen (`relevant`) et une valeur de pertinence associée (`relvalue`).

Lignes 6–7 : Si la variable est jugée pertinente (i.e. `relevant = true`), on vérifie si elle est redondante par rapport aux caractéristiques déjà présentes dans S_t, en appelant *check_ redundancy* (fonction 2).

Lignes 8–10 : Si la variable n'est pas redondante, elle est ajoutée à l'ensemble S_t, et sa valeur de pertinence est enregistrée dans R_t.

Lignes 11–12 : Fin des blocs conditionnels : seules les caractéristiques pertinentes et non redondantes sont conservées.

Ligne 13 : Fin de la boucle sur les nouvelles caractéristiques. À ce stade, toutes les caractéristiques candidates ont été évaluées.

Ligne 14 : Nouveau calcul d'un seuil de pertinence, ici défini comme le $(100-k)^{\text{ème}}$ percentile. Ce seuil bas est utilisé pour identifier les caractéristiques dont la pertinence a diminué.

Ligne 15 : La fonction *manage_ drift* (fonction 3) examine les caractéristiques de S_t sur une fenêtre temporelle de taille w. Les caractéristiques dont la pertinence moyenne est inférieure à θ sont supprimées.

Ligne 16 : L'ensemble S_t mis à jour et l'historique R_t sont retournés en sortie.

L'approche CI-DFS permet ainsi d'assurer une sélection dynamique, sensible à la qualité des caractéristiques, à leur évolution dans le temps, et à la redondance, tout en limitant la taille du modèle.

5.2.3 Évaluation de la pertinence : *check_ relevance*

La fonction *check_ relevance* permet de déterminer si une nouvelle caractéristique f_i est suffisamment informative pour être ajoutée à l'ensemble sélectionné S_t. Cette évaluation repose sur la notion d'**information mutuelle pondérée** (WMI), introduite dans le cadre de l'algorithme CI-DFS pour traiter le déséquilibre entre classes.

Information mutuelle

L'information mutuelle (MI), définie dans la littérature comme une mesure de dépendance entre deux caractéristiques aléatoires [51, 149], est utilisée ici pour quantifier la capacité prédictive d'une variable f_i vis-à-vis de la cible Y (décrochage ou non). L'équation 5.1 ci-dessous correspond à la définition classique de l'information mutuelle entre une caractéristique f_i et la variable cible Y :

$$MI(f_i; Y) = \sum_{x \in \mathrm{Val}(f_i)} \sum_{y \in \{0,1\}} P(x, y) \cdot \log \left(\frac{P(x, y)}{P(x) \cdot P(y)} \right) \tag{5.1}$$

Comme souligné dans [132], cette métrique est influencée par la distribution marginale des classes. En contexte de données éducatives déséquilibrées, une variable qui prédit bien la classe majoritaire (souvent le non-abandon) aura une MI élevée, même si elle est peu utile pour identifier les abandons.

WMI – Information mutuelle pondérée

Pour pallier cette limitation, CI-DFS utilise une métrique ajustée de la littérature, la **Weighted Mutual Information (WMI)**, qui pondère la contribution de chaque classe en fonction de sa fréquence dans les données [132, 97]. L'équation 5.2 présente la version pondérée utilisée dans CI-DFS pour corriger le biais induit par la distribution déséquilibrée des classes :

$$WMI(f_i; C) = \sum_{j=1}^{|C|} (1 - p(c_j)) \sum_{v \in \mathrm{val}(f_i)} \log \left(\frac{P(v, c_j)}{P(v) \cdot P(c_j)} \right) \tag{5.2}$$

où :
— $C = \{c_0, c_1\}$ est l'ensemble des classes (ici. : non-décrochage et abandon),
— $p(c_j)$ est la probabilité empirique de la classe c_j,
— la pondération $1 - p(c_j)$ permet de survaloriser la classe minoritaire.

Cette formulation, permet de rétablir l'équilibre dans la mesure de pertinence, en donnant une chance équitable aux caractéristiques associées à des cas rares.

Implémentation algorithmique

La fonction 1 décrit le processus de vérification de la pertinence d'une caractéristique à l'aide du score WMI.

Entrées :

- f_i : caractéristique candidate à évaluer. Il s'agit d'une nouvelle caractéristique issue du flux de données à l'instant t, dont on cherche à déterminer la pertinence.
- C : ensemble des classes cibles (par exemple, *décrocheur* / *non-décrocheur*), utilisées pour estimer la relation entre la caractéristique et la variable à prédire.
- R_t : file d'historique contenant les valeurs de pertinence des caractéristiques déjà sélectionnées (S_t). Elle permet d'estimer la distribution actuelle de la pertinence.
- θ : seuil dynamique de pertinence, généralement défini comme le k^e percentile des valeurs présentes dans R_t.

Sorties :

- *relevant* : booléen indiquant si la caractéristique candidate est considérée comme pertinente (`True`) ou non (`False`).
- *relevance_value* : valeur numérique de pertinence de la caractéristique, calculée via la *Weighted Mutual Information (WMI).*

Description du fonctionnement :

La fonction `check_relevance` évalue la pertinence d'une nouvelle caractéristique f_i vis-à-vis de la variable cible en deux étapes principales. Tout d'abord, elle calcule la WMI (équation 5.2) entre f_i et les classes C.

Ensuite, la valeur obtenue est comparée au seuil dynamique θ, estimé à partir de la distribution des pertinences précédemment observées dans R_t. Si la valeur de pertinence *relevance_value* dépasse ce seuil, la caractéristique est jugée informative et marquée comme pertinente (`relevant = True`).

La fonction retourne enfin le couple $(relevant, relevance_value)$, permettant à l'algorithme principal de décider si la caractéristique doit être ajoutée à l'ensemble S_t.

Fonction 1 check_relevance

Input : f_i la caractéristique candidate ; C l'ensemble des classes ; R_t la file d'historique des valeurs de pertinence des caractéristiques S_t ; θ le seuil de pertinence

Output : *relevant* : booléen indiquant la pertinence de la caractéristique ; *relevance_value* : valeur de pertinence

// Calculer la valeur WMI pour f_i (Équation 2)

1: $relevance_value \leftarrow \text{WMI}(f_i, C)$

// Calculer le seuil dynamique en tête des files d'attente in R_t

2: $relevant \leftarrow relevance_value > \theta$

3: **return** $relevant, relevance_value$

5.2.4 Détection de la redondance : *check_ redundancy*

Une fois qu'une caractéristique f_i est jugée pertinente au sens de la WMI, il est crucial de vérifier qu'elle n'apporte pas une information déjà couverte par une ou plusieurs caractéristiques déjà présentes dans l'ensemble sélectionné S_t. En effet, l'introduction de caractéristiques redondantes augmente le risque de surapprentissage, complique l'interprétation du modèle et détériore la frugalité recherchée.

La fonction *check_ redundancy* repose sur une mesure de la corrélation absolue entre la variable candidate f_i et les caractéristiques déjà sélectionnées. Si l'une d'entre elles dépasse un seuil fixé λ, f_i est considérée comme redondante et rejetée.

La corrélation linéaire de Pearson est un indicateur usuel de redondance entre caractéristiques continues. Elle permet de détecter les duplications implicites (ex. : nombre de clics et temps passé), sans alourdir le calcul. Dans notre contexte, les caractéristiques sont binaires, discrètes ou agrégées.

Implémentation algorithmique

La fonction 2 implémente la détection de redondance par corrélation absolue, afin d'écarter les caractéristiques redondantes de l'ensemble S_t.

Entrées :
- f_i : caractéristique candidate à évaluer, nouvellement considérée pour intégration dans l'ensemble des caractéristiques sélectionnées.
- S_t : ensemble des caractéristiques déjà sélectionnées à l'instant t. Ces caractéristiques ont été jugées pertinentes lors des itérations précédentes.
- R_t : file d'historique des valeurs de pertinence associées aux caractéristiques présentes dans S_t, utilisée pour le suivi de leur évolution.
- λ : seuil de redondance définissant la corrélation maximale autorisée entre deux caractéristiques. Une valeur élevée de λ rend le filtrage plus permissif, tandis qu'une valeur faible favorise la compacité du modèle.

Sortie :
- *redundant* : booléen indiquant si la caractéristique candidate est redondante (`True`) ou non (`False`) par rapport à celles déjà sélectionnées.

Description du fonctionnement :

La fonction `check_redundancy` a pour objectif d'éviter la sélection de caractéristiques fortement corrélées entre elles, afin de maintenir un ensemble S_t informatif et non redondant.

Pour chaque caractéristique f_j déjà présente dans S_t, la fonction calcule le coefficient de corrélation $\rho(f_i, f_j)$ entre f_i et f_j. Si la valeur absolue de cette corrélation $|\rho(f_i, f_j)|$ dépasse le seuil fixé λ, la caractéristique f_i est considérée comme redondante et la boucle est interrompue (`break`). Dans le cas contraire, si aucune corrélation supérieure au seuil n'est trouvée, la caractéristique est jugée non redondante.

La fonction retourne finalement le booléen *redundant*, utilisé par l'algorithme principal pour décider si f_i doit être ajoutée à l'ensemble S_t ou écartée.

Fonction 2 check_redundancy

Input : f_i la caractéristique candidate ; S_t les caractéristiques sélectionnées ; R_t la file d'historique des valeurs de pertinence des caractéristiques de S_t ; λ le seuil de redondance
Output : *redundant* : booléen indiquant la redondance de la caractéristique

```
redundant ← false
for each f_j ∈ S_t do
    if |ρ(f_i, f_j)| ≥ λ then
        redundant ← true
        break
    end if
end for
return redundant
```

Choix du seuil λ

Le seuil λ doit être suffisamment élevé pour ne pas éliminer des caractéristiques faiblement corrélées mais complémentaires.

5.2.5 Gestion du drift temporel : *manage_drift*

Une caractéristique initialement informative peut devenir moins pertinente, voire inutile, au fil du temps. Ce phénomène, appelé *feature drift*, est fréquent dans les systèmes éducatifs, où l'importance d'une variable peut dépendre du stade d'avancement du cours (ex. : la première connexion, la participation aux forums en début de semestre, etc.).

Pour y faire face, CI-DFS implémente une gestion du drift via une fenêtre glissante : chaque variable f_j sélectionnée conserve un historique de ses scores de pertinence récents. Si la moyenne glissante de ces scores descend sous un certain seuil adaptatif, la variable est automatiquement retirée de l'ensemble S_t.

Définition

À chaque pas de temps t, CI-DFS met à jour le score de pertinence de chaque $f_j \in S_t$ à l'aide de la métrique WMI. Ce score est ajouté en tête de file $R_t[f_j]$, qui conserve uniquement les w derniers scores. CI-DFS calcule ensuite la moyenne de cette fenêtre. Si celle-ci est inférieure à un seuil θ_{drift}, alors f_j est considérée comme obsolète et supprimée.

Implémentation algorithmique

L'algorithme 3 détaille le processus de suppression des caractéristiques devenues obsolètes au fil du temps.

Entrées :

— S_t : ensemble des caractéristiques actuellement sélectionnées à l'instant t. Chaque caractéristique de cet ensemble possède un historique de pertinence associé.
— C : ensemble des classes cibles utilisées pour calculer la pertinence via la *Weighted Mutual Information (WMI).*
— R_t : file d'historique contenant les valeurs successives de pertinence pour chaque caractéristique de S_t, sur les itérations précédentes.

— w : taille de la fenêtre glissante utilisée pour le suivi temporel de la pertinence. Elle détermine la profondeur historique considérée pour le calcul des moyennes.
— θ : seuil de pertinence dynamique servant à identifier les caractéristiques devenues obsolètes ou moins informatives.

Sortie :

— S_t : ensemble de caractéristiques mis à jour, obtenu après suppression des variables dont la pertinence moyenne est inférieure au seuil θ.
— R_t : file d'historique mise à jour.

Description du fonctionnement :

La fonction `manage_drift` a pour objectif de gérer le drift temporel des caractéristiques, c'est-à-dire la perte progressive de pertinence de certaines variables au fil du temps.

Pour chaque caractéristique f_j de l'ensemble S_t, la fonction commence par recalculer sa pertinence courante avec la variable cible C. Cette nouvelle valeur est ajoutée à l'historique $R_t[f_j]$. Si la longueur de cette file dépasse la taille maximale w, la valeur la plus ancienne est supprimée afin de ne conserver que les w dernières observations.

Ensuite, la pertinence moyenne de f_j est calculée sur la fenêtre courante. Si cette moyenne devient inférieure au seuil θ, la caractéristique est considérée comme ayant subi une dérive (ou *drift*) : elle est alors retirée de l'ensemble S_t, et son historique est supprimé.

Ainsi, la fonction `manage_drift` permet de maintenir au cours du temps un ensemble de caractéristiques à la fois compact et informatif, en éliminant automatiquement celles qui perdent en influence prédictive au fur et à mesure de l'évolution du flux de données.

Fonction 3 manage_drift

Input : S_t caractéristiques sélectionnées ; C classes ; R_t file d'historique des valeurs de pertinence des caractéristiques de S_t ; w taille de la fenêtre ; θ seuil de pertinence
Output : Ensemble de caractéristiques mis à jour S

for each feature $f_j \in S_t$ **do**
  $relevance_value \leftarrow \text{WMI}(f_j, C)$
  $R_t[f_j].push(relevance)$
  if $R_t[f_j].length > w$ **then**
    $R_t[f_j].pop()$
  end if
  // Calculer la pertinence moyenne
  $avg_relevance \leftarrow avg(R_t[f_i])$
  if $avg_relevance < \theta$ **then**
    $S_t \leftarrow S_t \setminus f_j$ // La caractéristique f_j est retirée de l'ensemble des caractéristiques.
    $R_t[f_j] \leftarrow$ NULL
  end if
end for
return S_t, R_t

Rôle du paramètre w

La taille de la fenêtre permet de réguler la sensibilité du mécanisme : une petite fenêtre permet des réactions rapides, mais risque de produire des suppressions trop fréquentes ; une grande fenêtre permet une meilleure stabilité, au risque de conserver trop longtemps des caractéristiques devenues inutiles.

5.2.6 Vue globale et complémentarité des composants

L'algorithme CI-DFS s'inscrit dans un paradigme de sélection adaptative de caractéristiques, avec une actualisation dynamique des données et sensible aux biais structurels des données éducatives. Il repose sur une articulation cohérente de trois composants fondamentaux, qui ensemble, ont pour objectif d'obtenir un système robuste, frugal et performant.

La WMI (Weighted Mutual Information)

est la brique fondatrice de l'approche. Elle permet de compenser le déséquilibre des classes, omniprésent dans les contextes de décrochage. Contrairement à l'information mutuelle classique qui privilégie les prédicteurs des classes majoritaires, la WMI donne une chance équitable aux caractéristiques qui discriminent efficacement les cas de décrochage, même minoritaires. Elle constitue ainsi une base fiable pour une évaluation équitable de la pertinence.

Le seuil adaptatif par percentile k

introduit une capacité d'adaptation dynamique du modèle. Plutôt que de fixer arbitrairement un seuil unique, CI-DFS ajuste ce seuil à la distribution actuelle des scores de pertinence. Cette démarche permet au modèle de se calibrer automatiquement : il devient plus exigeant lorsque les scores sont élevés, plus souple lorsque les données sont plus pauvres. Le percentile k devient ainsi un véritable levier de régulation du compromis entre ouverture à la nouveauté et exigence de qualité.

La gestion temporelle du drift

assure un renouvellement progressif de l'ensemble S_t en éliminant les caractéristiques devenues obsolètes. Grâce à une moyenne glissante sur les w derniers pas de temps, le système ne conserve que les caractéristiques dont la pertinence reste suffisamment stable dans le temps. Ce mécanisme protège le modèle contre l'accumulation excessive, réduit le bruit, et renforce sa capacité à s'ajuster à des dynamiques d'apprentissage évolutives (par exemple : démarrage lent puis intensification, ou décroissance progressive de l'engagement).

Complémentarité

Pris isolément, chacun de ces composants apporte une contribution utile. Mais c'est leur combinaison qui confère à CI-DFS sa puissance :

— La WMI assure la *justesse de la mesure*, en respectant les contraintes du problème
— Le seuil adaptatif k garantit la *souplesse du choix* face à l'hétérogénéité des flux
— Le drift temporel impose une *discipline de maintenance* et de renouvellement permanent

Ensemble, ces trois mécanismes rendent possible une sélection dynamique de haute qualité, orientée vers les apprenants à risque, tout en limitant les redondances, la stagnation et la croissance incontrôlée du modèle.

La section suivante décrira les métriques utilisées pour évaluer l'impact de chacun de ces leviers, tant en termes d'efficacité prédictive que de coût computationnel.

5.3 Métriques d'évaluation

L'évaluation d'un algorithme de sélection de caractéristiques dynamique dans un contexte éducatif doit reposer sur des indicateurs à la fois prédictifs, explicatifs et opérationnels. Dans cette optique, nous adoptons un ensemble de métriques couvrant trois dimensions complémentaires : *la qualité prédictive*, *la frugalité du modèle*, et *la stabilité temporelle*.

Qualité prédictive

Nous évaluons la capacité du modèle à prédire correctement les cas de décrochage, tout en maintenant une performance globale raisonnable. Trois métriques sont utilisées :

— **Exactitude (*Accuracy*)** : proportion de prédictions correctes, toutes classes confondues. Bien qu'intuitive, cette métrique est peu informative dans les contextes de classes déséquilibrées.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

où TP et TN désignent respectivement les vrais positifs (décrochage bien prédit) et vrais négatifs (non-décrochage bien prédit).

— **Sensibilité *(Sensitivity (Recall))*** : capacité à identifier correctement les étudiants qui vont abandonner. C'est une métrique centrale dans notre étude, car l'objectif est d'agir précocement sur les cas à risque.

$$Sensitivity(Recall) = \frac{TP}{TP + FN}$$

Une faible sensibilité signifie que de nombreux abandons passent inaperçus.

— **Spécificité *(Specificity)*** : proportion de non-décrocheurs correctement détectés. Elle complète la sensibilité en indiquant la stabilité du modèle sur la classe majoritaire.

$$Specificity = \frac{TN}{TN + FP}$$

Une faible spécificité signifie que de nombreux apprenants sont identifiés à tort comme risquant l'abandon, alors qu'ils poursuivront effectivement leur parcours.

Interprétation

De notre point de vue, un bon modèle ne cherche pas à maximiser uniquement l'exactitude globale, mais à trouver un équilibre entre sensibilité et spécificité. Dans les VLEs, la priorité est donnée à la sensibilité, car les faux négatifs (décrochage non détecté) sont les plus dommageables pédagogiquement.

Frugalité du modèle

CI-DFS vise explicitement à réduire le volume de données mobilisées. Deux métriques sont donc utilisées pour évaluer cette dimension :

— **Nombre moyen de caractéristiques sélectionnées par semaine** : mesure la taille de l'ensemble S^t au fil du temps. Un modèle plus frugal sélectionne moins de caractéristiques tout en maintenant ses performances.

— **Temps de traitement (par étape)** : durée moyenne nécessaire pour effectuer la sélection à l'instant t, exprimée en secondes. Elle reflète la scalabilité de l'approche.

Résumé

L'évaluation de CI-DFS repose sur un ensemble de métriques complémentaires, visant à mesurer à la fois la performance prédictive, l'équité de prédiction entre classes (sensibilité, spécificité), le nombre de caractéristiques sélectionnées et son efficacité computationnelle. Le tableau suivant synthétise les indicateurs mobilisés dans le protocole expérimental. Le tableau 5.1 résume les principales métriques utilisées dans notre protocole d'évaluation :

Métrique	**Objectif**
Exactitude globale	Évaluer la performance globale toutes classes confondues
Sensibilité	Détecter correctement les abandons
Spécificité	Identifier correctement les non-abandons
Nb. de features	Mesurer la frugalité (compacité du modèle)
Temps de traitement	Estimer le coût computationnel

TABLE 5.1 – Résumé des métriques utilisées pour l'évaluation de CI-DFS

5.3.1 Méthodes comparatives : SAOLA et baseline sans sélection

L'évaluation expérimentale de CI-DFS repose sur une comparaison avec deux approches de référence, représentatives des extrêmes en matière de sélection de caractéristiques :

— **SAOLA** (*Scalable and Accurate Online Learning Algorithm*) [169], un algorithme dynamique reconnu pour sa capacité à gérer les flux de données à haute dimension

— Un **modèle baseline sans sélection**, désigné par *AllFeatures*, qui applique un apprentissage sur la totalité des caractéristiques disponibles à l'instant t sans aucune sélection

SAOLA : sélection dynamique à seuil fixe

SAOLA est un algorithme dynamique de sélection de caractéristiques, conçu pour les environnements en ligne. Chaque nouvelle variable est évaluée dès son apparition selon deux critères : sa *pertinence* vis-à-vis de la cible, et sa *non-redondance* vis-à-vis des caractéristiques déjà sélectionnées.

Définition La pertinence est mesurée via l'information mutuelle classique (MI) entre la variable et la cible Y. Une variable f est ajoutée à l'ensemble sélectionné S si :

— $MI(f, Y)$ dépasse un seuil fixe θ (typiquement 0,01) ;
— Et si $MI(f, f_j) < \lambda$ pour toute variable f_j déjà présente dans S.

Algorithme Le fonctionnement de SAOLA peut être résumé comme suit :

Fonction 4 SAOLA – sélection incrémentale avec seuil fixe

Input : f : nouvelle caractéristique ; S : ensemble courant ; θ : seuil de pertinence ; λ : seuil de redondance
Output : Mise à jour éventuelle de S

```
Calculer MI(f, Y)
if MI(f, Y) > θ then
    for chaque f_j ∈ S do
        if MI(f, f_j) ≥ λ then
            return S                          ▷ Redondance détectée
        end if
    end for
    S ← S ∪ {f}                               ▷ Ajout de la variable
end if
return S
```

Forces et limites SAOLA offre un compromis intéressant entre efficacité computationnelle et capacité d'adaptation en ligne. Toutefois, il présente plusieurs limites dans un contexte éducatif :

— Le seuil de pertinence est fixe et ne s'ajuste pas aux dynamiques des données
— L'information mutuelle utilisée n'est pas pondérée : elle ne tient pas compte du déséquilibre des classes
— La redondance est évaluée à l'aide de MI entre caractéristiques, ce qui peut s'avérer coûteux en pratique
— L'algorithme ne prend pas en compte la dérive temporelle des données.

En somme, SAOLA constitue une base comparative solide pour la sélection de caractéristiques dans des environnements à flux de données, mais son mécanisme repose sur une évaluation ponctuelle de la pertinence et de la redondance des variables. Il ne prend pas en compte l'évolution des relations entre caractéristiques et variable cible par classe au fil du temps, ni les phénomènes de dérive ou d'obsolescence des caractéristique (*feature drift)*. Cette absence de prise en compte de la dimension temporelle et contextuelle limite sa pertinence dans les systèmes éducatifs, où les comportements d'apprentissage et la distribution des données changent progressivement au cours du temps.

Modèle baseline *AllFeatures* : sans sélection

Pour évaluer l'apport spécifique de la sélection dynamique, nous utilisons également un modèle sans aucun mécanisme de filtrage. À chaque instant t, toutes les caractéristiques disponibles dans le flux cumulatif $F_{1:t}$ sont transmises au modèle de classification. Ce modèle vise à établir :

— Un **plafond théorique d'information**, en exploitant la totalité des caractéristiques disponibles à chaque étape

— Une **ligne de base de comparaison**, afin de quantifier les gains introduits par CI-DFS, tant en termes de performance que de frugalité ou de temps de traitement

Limites Ce modèle ne prend pas en compte la nature évolutive et contextuelle des données éducatives, dans lesquelles la pertinence des caractéristiques peut varier au fil du temps. Il suppose implicitement que l'ensemble des variables disponibles conserve une valeur informative constante, sans mécanisme d'actualisation ni de suppression des caractéristiques devenues obsolètes. Une telle hypothèse limite sa capacité à s'adapter aux changements progressifs des comportements d'apprentissage ou à l'apparition de nouvelles sources de données.

De plus, il ne répond pas aux exigences de compacité et de scalabilité nécessaires aux environnements opérationnels : l'utilisation simultanée de toutes les caractéristiques disponibles entraîne une croissance rapide du volume de données à traiter, ce qui augmente la latence et la charge mémoire. Enfin, l'absence de prise en compte des contraintes de traitement en ligne empêche son intégration dans des systèmes de prédiction continue. Ce modèle constitue ainsi une borne haute en information brute, mais il reste une solution peu réaliste dans un cadre où la sélection dynamique et la gestion frugale des caractéristiques sont essentielles pour concilier performance, adaptabilité et efficacité computationnelle.

Les sections suivantes présentent une comparaison expérimentale détaillée entre CI-DFS, SAOLA et *AllFeatures*, à l'aide des métriques définies précédemment.

5.4 Évaluation sur le jeu de données OULAD

Pour évaluer la performance de l'algorithme CI-DFS dans un contexte réaliste, nous l'appliquons au jeu de données OULAD [91].

5.4.1 Présentation du jeu de données

Le jeu de données OULAD regroupe des informations issues de sept modules d'apprentissage à distance proposés par l'Open University. Chaque module correspond à un cours complet, étalé sur plusieurs mois.

Pour notre expérimentation, nous avons choisi le module **CCC-2014J**, car c'est le plus riche en données temporelles. Ce module comprend :

— **2 498 étudiants** inscrits
— une durée de **269 jours** (soit environ 38 semaines)
— un taux de décrochage de **38,3%**.

5.4.2 Construction temporelle des caractéristiques

Notre objectif est de simuler un contexte réaliste où les données sont révélées de manière séquentielle, semaine après semaine. Pour cela, nous avons procédé à une re-construction du jeu de données selon les étapes suivantes :

1. **Découpage temporel** : la durée du module est divisée en 38 semaines.
2. **Agrégation hebdomadaire** : pour chaque étudiant, nous générons un vecteur de caractéristiques cumulées semaine par semaine, en intégrant l'ensemble des interactions antérieures.

3. **Génération de caractéristiques** : chaque interaction avec une ressource (ex. : quiz, forum, vidéo) est transformée en une variable binaire ou comptable (e.g., "accès au quiz 12").
4. **Encodage démographique** : les caractéristiques en catégories comme le niveau d'études sont automatiquement encodées et intégrées à chaque étape du flux.

Le résultat est un jeu de données structuré en flux temporels hebdomadaires, dans lequel les caractéristiques disponibles jusqu'à la semaine t sont utilisées pour entraîner un modèle qui prédit le décrochage en fin de parcours.

Le découpage hebdomadaire répond à plusieurs considérations méthodologiques et pédagogiques. D'une part, il correspond au rythme naturel d'organisation des cours en ligne, où les activités et les évaluations sont souvent planifiées sur une base hebdomadaire, permettant ainsi une interprétation cohérente des comportements d'apprentissage. D'autre part, il constitue un compromis pertinent entre granularité temporelle et stabilité des données : un pas de temps plus court (par jour) introduirait une variabilité excessive et du bruit, tandis qu'un pas plus long (par mois ou par session) diluerait les dynamiques d'engagement et retarderait la détection des signaux de décrochage. Ce choix permet donc de suivre l'évolution progressive des apprenants tout en maintenant une cohérence analytique et une robustesse statistique suffisante.

5.4.3 Préparation du jeu d'entraînement et de test

Conformément aux bonnes pratiques, nous avons divisé le jeu de données en deux sous-ensembles :

— **Entraînement (80%)** : utilisé pour faire la sélection de caractéristiques et entraîner les modèles de prédiction ;
— **Test (20%)** : utilisé uniquement pour l'évaluation finale.

À chaque semaine t, un nouveau jeu d'entraînement est généré en utilisant les caractéristiques disponibles à cette date.

5.4.4 Modèle de prédiction utilisé

Pour la phase de prédiction, nous utilisons un classifieur Random Forest, reconnu pour sa robustesse et sa capacité à gérer des données hétérogènes. Les hyperparamètres sont à leurs valeurs par défaut (100 arbres, profondeur illimitée), car l'objectif est de comparer les performances en fonction des ensembles de caractéristiques sélectionnés, et non d'optimiser le modèle lui-même.

5.4.5 Paramétrage de CI-DFS

Les hyperparamètres de CI-DFS ont été calibrés par validation sur l'ensemble d'entraînement :
— **Seuil de redondance** $\lambda = 0,95$: corrélation maximale tolérée entre deux caractéristiques
— **Seuil de pertinence** $k = 25$: seules les caractéristiques dont la pertinence dépasse le $25^{\text{ème}}$ percentile sont sélectionnées
— **Fenêtre de drift** $w = 10$: la pertinence d'une variable est évaluée sur les 10 dernières semaines.

Ces valeurs ont été choisies pour maximiser la sensibilité sans pénaliser la frugalité du modèle.

5.4.6 Paramétrage des modèles baseline

Pour assurer une comparaison équitable, les méthodes SAOLA et AllFeatures utilisent le même classifieur Random Forest. Leurs paramètres sont définis comme suit :

— **SAOLA** :
  — Seuil de pertinence MI : $\theta = 0,01$ (valeur recommandée dans [169])
  — Seuil de redondance MI : $\lambda = 0,9$
  — Pas de suppression des caractéristiques.

— **AllFeatures** :
  — Toutes les caractéristiques disponibles jusqu'à la semaine t sont utilisées ;
  — Aucune sélection, aucune suppression.

Les résultats obtenus sur ce protocole sont présentés dans les sections suivantes.

5.5 Évaluation sur le jeu de données MoocCN

En complément du jeu de données OULAD, nous évaluons les modèles sur un second ensemble issu d'un contexte différent : un MOOC de courte durée, tiré du jeu de données MoocCN. Ce jeu permet d'évaluer l'algorithme dans des conditions, où les informations sont plus limitées dans le temps et la fenêtre d'observation plus restreinte.

5.5.1 Présentation du jeu de données MoocCN

MoocCN est l'une des principales plateformes de MOOCs en Chine. Le jeu de données utilisé provient d'une compétition de prédiction (KDD Cup 2015), et contient des informations anonymisées sur l'engagement des apprenants dans un MOOC de 6 semaines.

— **2 613 étudiants** dans l'ensemble d'entraînement
— **1 087 étudiants** dans l'ensemble de test public
— Taux de décrochage moyen : **32,5%**

Les traces disponibles incluent :
— les événements d'interaction : clics sur le lecteur vidéo, accès aux documents, interactions avec le forum
— des données statiques comme le niveau d'éducation, l'âge ou la localisation géographique

5.5.2 Construction temporelle des caractéristiques

Le jeu de données étant structuré par semaine, nous reproduisons un protocole similaire à celui d'OULAD :

1. **Découpage temporel** : la progression dans le MOOC est divisée en 6 semaines, correspondant à la durée totale du module.

2. **Agrégation hebdomadaire** : pour chaque apprenant, un vecteur cumulatif de caractéristiques est généré à chaque semaine, en tenant compte de l'ensemble des interactions enregistrées depuis le début du cours.
3. **Génération de caractéristiques** : chaque action sur la plateforme (visionnage de vidéo, tentative de quiz, contribution au forum, etc.) est convertie en variable binaire ou numérique, représentant par exemple la fréquence ou la présence d'un événement.
4. **Encodage démographique** : les attributs stables comme le niveau d'études ou le pays sont encodés dès l'entrée et réinjectés de manière constante à chaque étape du processus.

Ainsi, pour chaque étudiant, un vecteur de caractéristiques est généré à chaque semaine t, puis utilisé pour prédire le décrochage final.

5.5.3 Spécificités du protocole expérimental

Contrairement à OULAD, le split train/test est prédéfini dans le jeu de données. Nous respectons ce découpage.

La tâche de prédiction reste binaire : décrochage (1) vs. non-décrochage (0), évalué à la fin de la 6e semaine. Le modèle doit donc prendre des décisions de plus en plus précises au fil des semaines, malgré un horizon temporel réduit.

5.5.4 Paramétrage de CI-DFS sur MoocCN

Les paramètres de CI-DFS ont été réutilisés tels que définis sur OULAD, pour garantir la comparabilité entre contextes :
— Seuil de pertinence $k = 25$
— Seuil de redondance $\lambda = 0,95$
— Fenêtre glissante de pertinence $w = 10$

Ce choix permet d'évaluer la robustesse du modèle sans nécessiter de recalibrage spécifique à chaque jeu de données.

5.5.5 Paramétrage des comparateurs

Comme précédemment, les comparateurs (SAOLA, AllFeatures) utilisent les mêmes représentations temporelles et le même modèle de prédiction (Random Forest) :

— **SAOLA** : MI classique, seuil fixe de pertinence $\theta = 0,01$ et de redondance $\lambda = 0,9$, accumulation sans suppression

— **AllFeatures** : aucune sélection, toutes les caractéristiques hebdomadaires sont conservées

La courte durée du MOOC concentre les traces d'apprentissage sur un laps de temps limité, ce qui permet d'observer rapidement l'impact des méthodes sur la détection précoce des décrocheurs. Les résultats seront comparés à ceux obtenus sur OULAD pour discuter de la généricité de CI-DFS.

5.6 Résultats

Cette section présente l'évaluation expérimentale de l'algorithme CI-DFS sur les jeux de données éducatifs présentés précédemment. CI-DFS est comparé à deux approches de référence : SAOLA et AllFeatures. Trois axes d'analyse sont développés : les performances globales, l'apport de chaque composant de l'algorithme via une étude d'ablation, et une analyse qualitative des caractéristiques retenues.

5.6.1 Comparaison globale des algorithmes

Étude des performances moyennes

Le tableau 5.2 présente les performances moyennes hebdomadaires sur l'ensemble du parcours pour chaque algorithme et chaque jeu de données. Globalement, le modèle **CI-DFS** se distingue par un meilleur compromis entre performance prédictive et frugalité, confirmant son efficacité à sélectionner les caractéristiques les plus pertinentes tout en limitant la redondance.

Sur le jeu de données **OULAD**, CI-DFS atteint une **exactitude moyenne de 0,65** et une **sensibilité de 0,58**, soit une amélioration de **+3,1% en exactitude** et de **+28,9% en sensibilité** par rapport à SAOLA, et de **+4,8%** et **+38,1%** respectivement par rapport au modèle utilisant l'ensemble des caractéristiques.

Cette progression montre que CI-DFS identifie plus efficacement les étudiants à risque, tout en réduisant nettement le nombre de caractéristiques mobilisées (en moyenne **204,6 contre 386,2 pour SAOLA**, soit une **réduction de 47%**, et **−93%** par rapport au modèle complet). Ces résultats traduisent une meilleure capacité de généralisation et une stabilité accrue dans la sélection des caractéristiques, deux éléments cruciaux dans un contexte de flux de données éducatives où la structure d'information évolue au fil des semaines.

Jeu	Méthode	Exactitude	Sensibilité	Nb. caractéristiques
OULAD	CI-DFS	0,65	0,58	204,6
	SAOLA	0,64	0,45	386,2
	AllFeatures	0,63	0,42	3094,4
MoocCN	CI-DFS	0,69	0,84	116,5
	SAOLA	0,71	0,45	358,2
	AllFeatures	0,68	0,39	635,0

TABLE 5.2 – Performances moyennes sur les deux jeux de données (semaines combinées)

Sur **MoocCN**, le modèle CI-DFS confirme sa robustesse, avec une **exactitude moyenne de 0,69**, comparable à celle de SAOLA (0,71), mais une **sensibilité nettement supérieure (0,84 contre 0,45)**, soit un **gain relatif de 86,7%**. Autrement dit, CI-DFS détecte près de **deux fois plus de cas positifs correctement** que SAOLA, sans perte notable de précision globale. De plus, la taille moyenne du sous-ensemble de caractéristiques chute à **116,5**, contre **358,2 pour SAOLA** et **635 pour AllFeatures**, soit une **réduction de 67% et 82%** respectivement.

Ces résultats convergent vers un même constat : **CI-DFS** permet de préserver, voire d'améliorer la performance prédictive tout en divisant par deux à dix le volume de caractéristiques nécessaires. Les différences observées, notamment sur la sensibilité, sont significatives, confirmant la robustesse des écarts observés. En synthèse, ces résultats indiquent que **CI-DFS** permet

d'obtenir des prédictions plus stables, plus efficaces et plus économes en données.

Performances temporelles – OULAD

Les experimentations suivantes ont pour objectif l'analyse de l'évolution hebdomadaire des performances de CI-DFS sur le jeu de données OULAD, comparée à SAOLA et AllFeatures.

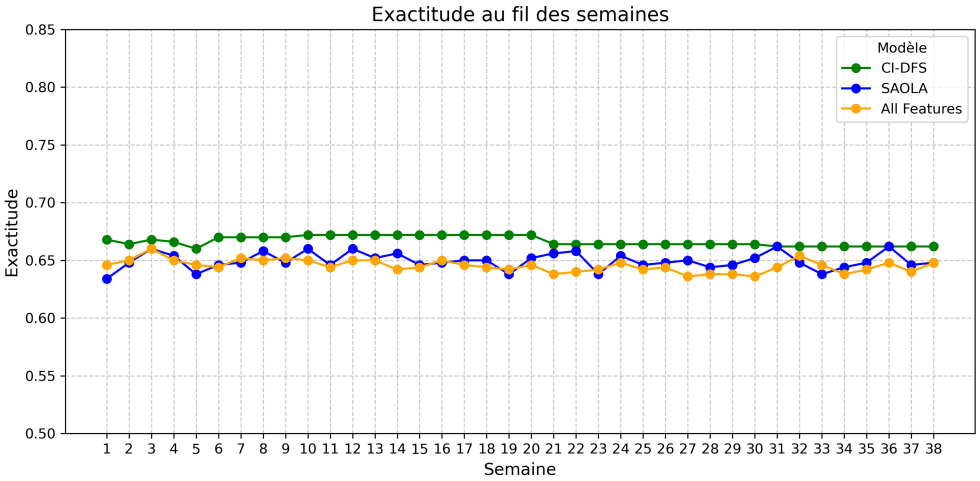

FIGURE 5.1 – Exactitude hebdomadaire sur OULAD.

Comme le montre la figure 5.1, le modèle **CI-DFS** maintient une **exactitude stable autour de 0,65** tout au long des 38 semaines, avec une variabilité hebdomadaire inférieure à **±0,01**. Cette stabilité traduit la **capacité du mécanisme de sélection dynamique de caractéristiques** à conserver un ensemble pertinent de caractéristiques au fil du temps, malgré l'évolution des interactions des apprenants.

Le modèle **SAOLA** présente une exactitude légèrement inférieure (**0,64**), tandis que **AllFeatures** n'atteint que **0,63**. Si l'écart d'exactitude entre les trois approches demeure modeste (de l'ordre de **2 à 3%**), il masque des différences bien plus marquées sur la capacité à détecter les décrocheurs.

Afin d'en rendre compte, l'analyse suivante se concentre spécifiquement sur la **sensibilité**, indicateur plus représentatif de la qualité de la détéction des étudiants à risque.
Comme l'indique la figure 5.2, la sensibilité de **CI-DFS** se maintient au-dessus de **0,55** tout au long des 38 semaines, ce qui traduit une détection constante et fiable des étudiants en décrochage. Cette stabilité met en évidence la capacité du modèle à préserver des caractéristiques pertinentes au fil du temps, même lorsque les données disponibles restent partielles.

En comparaison, la sensibilité de **SAOLA** chute rapidement sous **0,45** dès la semaine 10, tandis que celle du modèle **AllFeatures** demeure globalement inférieure à **0,43**. Ces écarts significatifs confirment que le mécanisme de sélection dynamique pondéré de CI-DFS permet de mieux équilibrer la représentation des classes et d'améliorer la détection des cas de décrochage, là où les approches classiques perdent en efficacité au fil du temps.

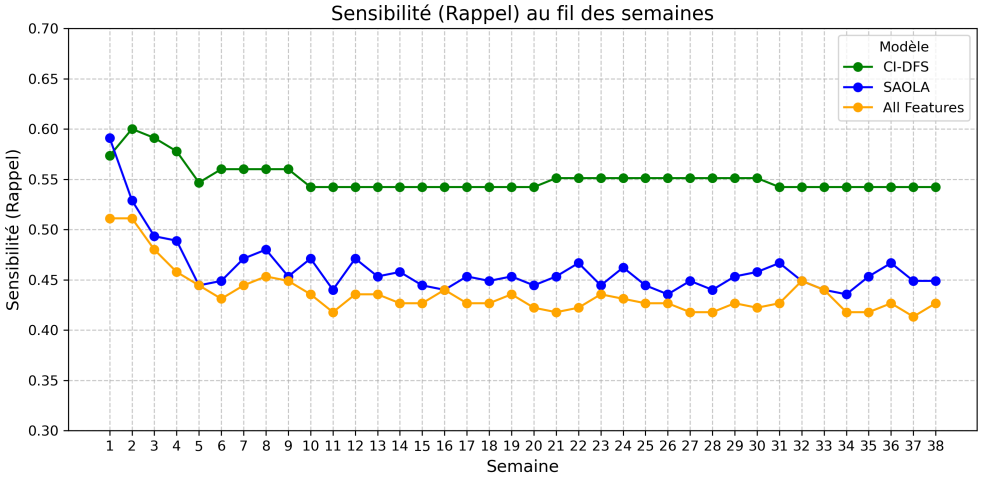

FIGURE 5.2 – Sensibilité hebdomadaire (Rappel).

Sur le jeu de données **OULAD**, cette constance suggère que les caractéristiques initiales, en particulier celles liées aux premières interactions avec la plateforme, conservent une valeur prédictive importante tout au long du parcours.

La figure 5.3 présente l'évolution de la spécificité au fil des semaines. On observe que le modèle **AllFeatures** reste légèrement supérieur, avec des valeurs comprises entre **0,83 et 0,84**, tandis que **SAOLA** se maintient autour de **0,82**.

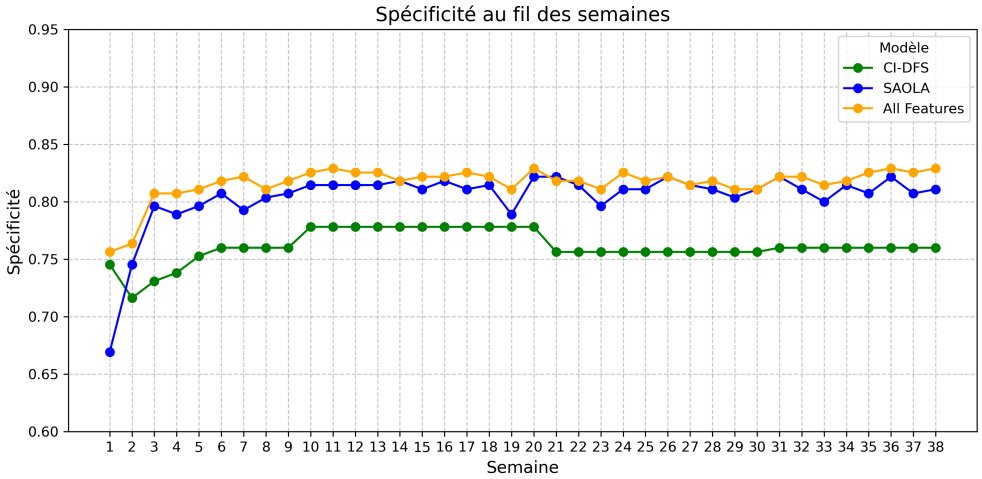

FIGURE 5.3 – Spécificité hebdomadaire.

Le modèle **CI-DFS**, quant à lui, se stabilise autour de **0,76** à partir de la semaine 10. Cette tendance traduit un compromis explicite adopté par CI-DFS : renforcer la détection des appre-

nants en décrochage (haute sensibilité) au prix d'une légère diminution de la capacité à identifier correctement les non-décrocheurs (spécificité).

Ce comportement est cohérent avec la logique de la sélection dynamique de caractéristiques, qui privilégie la préservation de caractéristiques discriminantes pour la classe minoritaire, dans un contexte où l'équilibre entre les deux types d'erreurs (faux positifs et faux négatifs) reste un enjeu central.

La figure 5.4 illustre l'évolution hebdomadaire du nombre de caractéristiques utilisées par les trois approches. On observe que le nombre de caractéristiques sélectionnées par **CI-DFS** reste remarquablement stable, autour de **200 à 250 caractéristiques** sur l'ensemble des 38 semaines. Cette constance traduit la capacité du modèle à maintenir un sous-ensemble compact et pertinent, sans être influencé par l'accumulation progressive de nouvelles caractéristiques au fil du temps.

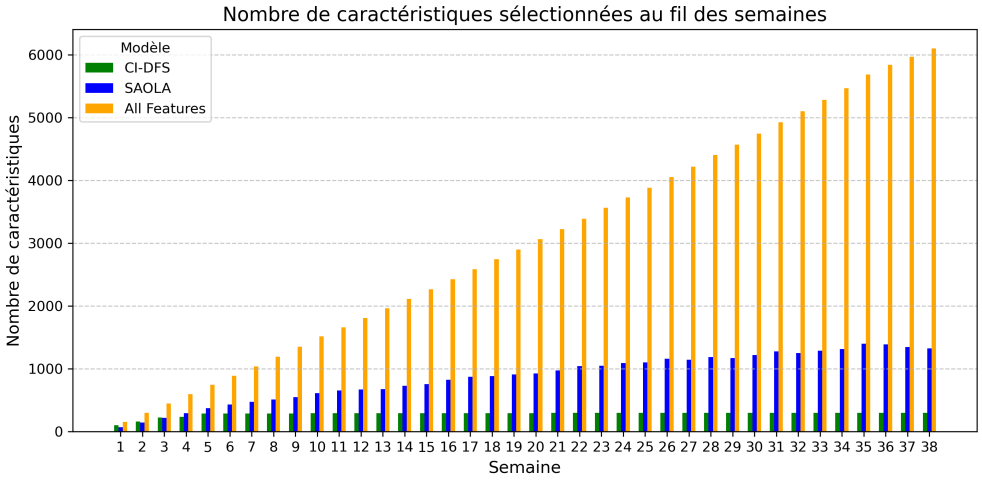

FIGURE 5.4 – Nombre de caractéristiques sélectionnées.

En comparaison, le modèle **SAOLA** présente une croissance régulière : à la première semaine, il mobilise environ **100 caractéristiques**, mais ce nombre augmente presque linéairement pour atteindre plus de **1 100 caractéristiques** en fin de parcours. Cela correspond à une croissance moyenne d'environ **30 caractéristiques supplémentaires par semaine**. Cette dynamique met en évidence l'absence de mécanisme de suppression ou de réévaluation des caractéristiques devenues redondantes, ce qui alourdit progressivement le modèle et limite son efficacité à long terme.

Enfin, l'approche **AllFeatures** illustre les conséquences directes de l'absence de sélection. Le nombre de caractéristiques croît de manière cumulative avec un rythme très soutenu, passant d'environ **100** à la semaine 1 à plus de **6 400** à la semaine 38. Cette évolution correspond à une augmentation moyenne de près de **170 nouvelles caractéristiques par semaine**, ce qui rend cette approche irréaliste dans un contexte de traitement en continu.

Tracer cette évolution semaine par semaine permet donc de comparer les trajectoires des trois approches. Là où AllFeatures illustre un scénario irréaliste et coûteux en ressources, et SAOLA

une stratégie intermédiaire mais non durable, CI-DFS démontre une réelle efficacité en maintenant un volume réduit et constant de caractéristiques. Cette frugalité contrôlée constitue un atout majeur pour les environnements éducatifs où la scalabilité et la réactivité des modèles sont essentielles.

Intéressons-nous maintenant au **temps de sélection des caractéristiques**, présenté à la figure 5.5. Cette métrique permet d'évaluer la capacité d'un algorithme à produire un sous-ensemble pertinent dans des délais raisonnables. On observe que **CI-DFS** maintient un temps de sélection remarquablement faible, de l'ordre de **2 secondes en moyenne**, avec une croissance modérée et régulière au fil des semaines. Cette stabilité montre que l'algorithme gère efficacement l'accumulation progressive des données tout en conservant une exécution rapide.

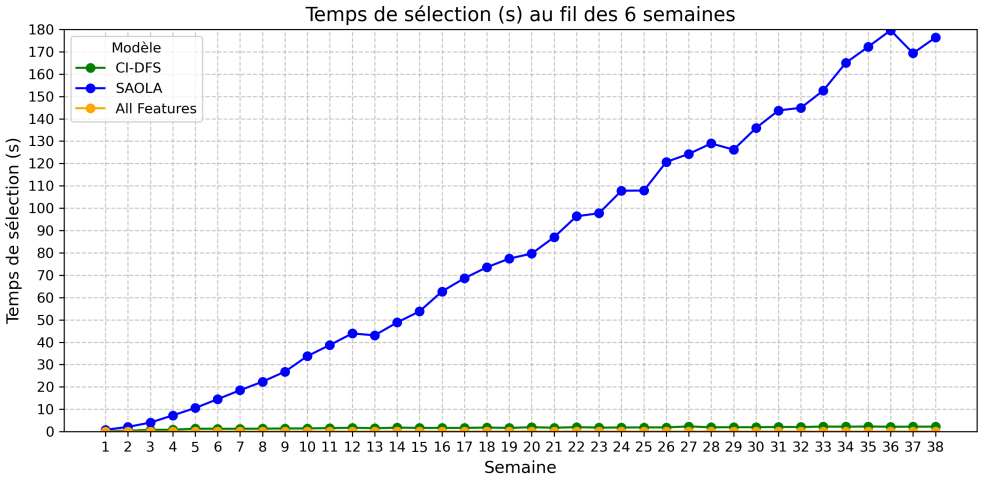

FIGURE 5.5 – Temps de sélection

En comparaison, **SAOLA** présente une évolution beaucoup plus coûteuse : si son temps de calcul reste raisonnable au début du cours, il croît ensuite de manière quasi linéaire, dépassant **100 secondes dès la semaine 20** et atteignant près de **180 secondes** en fin de période. Cette différence met en évidence l'efficacité du mécanisme de gestion dynamique de la pertinence et de la redondance implémenté dans CI-DFS, qui limite la complexité temporelle sans sacrifier la qualité de sélection.

Enfin, l'approche **AllFeatures** apparaît presque invisible sur la figure, son temps de calcul étant négligeable en l'absence d'étape de sélection. Cependant, cette absence de sélection transfère la charge computationnelle à l'étape d'entraînement du modèle, ce qui en réduit l'intérêt dans un cadre où la maîtrise du volume de données est recherchée.

Enfin, la figure 5.6 présente l'évolution du **temps d'entraînement des modèles prédictifs** au fil des semaines. Le modèle **CI-DFS** affiche une **bonne scalabilité**, avec un **temps moyen d'environ 0,15 seconde par semaine**. Cette performance s'explique par le maintien d'un ensemble réduit de caractéristiques, ce qui limite le coût de calcul à mesure que de nouvelles données sont intégrées.

Le modèle **SAOLA** présente des valeurs légèrement plus élevées, comprises entre **0,35 et 0,6**

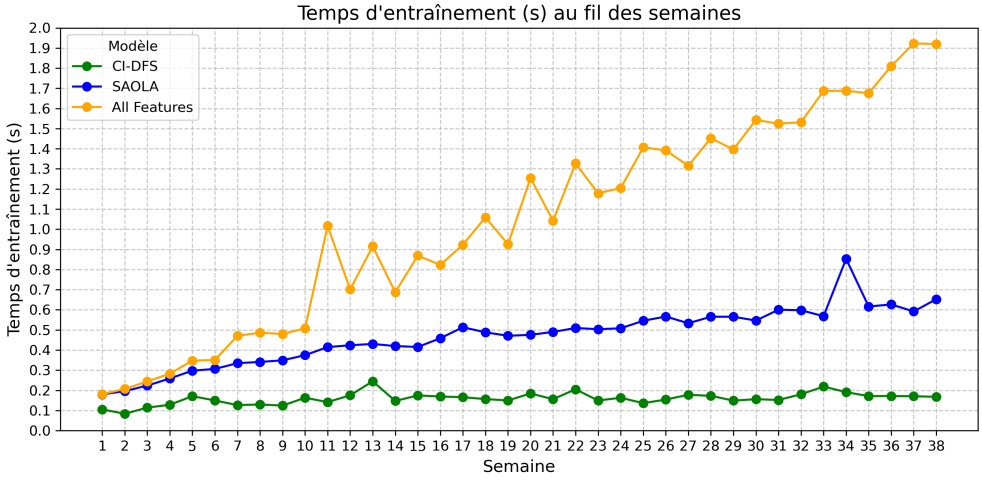

FIGURE 5.6 – Temps d'entraînement du modèle.

seconde, reflétant la croissance progressive du nombre de caractéristiques retenues. Ce comportement traduit une complexité de traitement plus importante, bien que les durées restent modérées.

Enfin, le modèle **AllFeatures** montre une augmentation régulière du temps d'entraînement, atteignant environ **1,8 seconde** en fin de période. Cette progression est directement liée à la croissance du nombre de caractéristiques en entrée, qui alourdit mécaniquement la phase d'entraînement. L'ensemble de ces résultats souligne l'avantage de CI-DFS en termes de **scalabilité temporelle**, grâce à un volume de caractéristiques maîtrisé et à une complexité de calcul contenue.

Performances temporelles – MoocCN

La figure 5.7 illustre l'évolution de la **sensibilité** au fil des six semaines du MOOC-CN. On observe que le modèle **CI-DFS** obtient les meilleures performances tout au long du parcours, avec une **sensibilité moyenne de 0,60**, contre **0,55** pour **SAOLA** et **0,56** pour **AllFeatures**.

La différence la plus marquée apparaît dès la première semaine, où CI-DFS atteint **0,48**, soit une amélioration relative de **+77%** par rapport à SAOLA (**0,27**) et de **+65%** par rapport à AllFeatures (**0,29**). Cette avance initiale témoigne de la capacité du mécanisme de sélection dynamique à extraire rapidement les caractéristiques les plus informatives, même dans un contexte de données limitées.

Contrairement aux deux autres approches, la sensibilité de CI-DFS progresse rapidement entre les semaines 1 et 2 puis se maintient autour de 0,60 jusqu'à la fin du cours, traduisant une détection stable et efficace des apprenants à risque. À l'inverse, SAOLA et AllFeatures affichent une progression plus lente et un niveau de performance inférieur, atteignant respectivement **0,55** et **0,56** en semaine 6, sans parvenir à rattraper l'écart initial.

Ces résultats soulignent la robustesse de CI-DFS face à la variabilité temporelle des données, ainsi que son efficacité accrue dans la détection précoce des décrocheurs. L'approche démontre

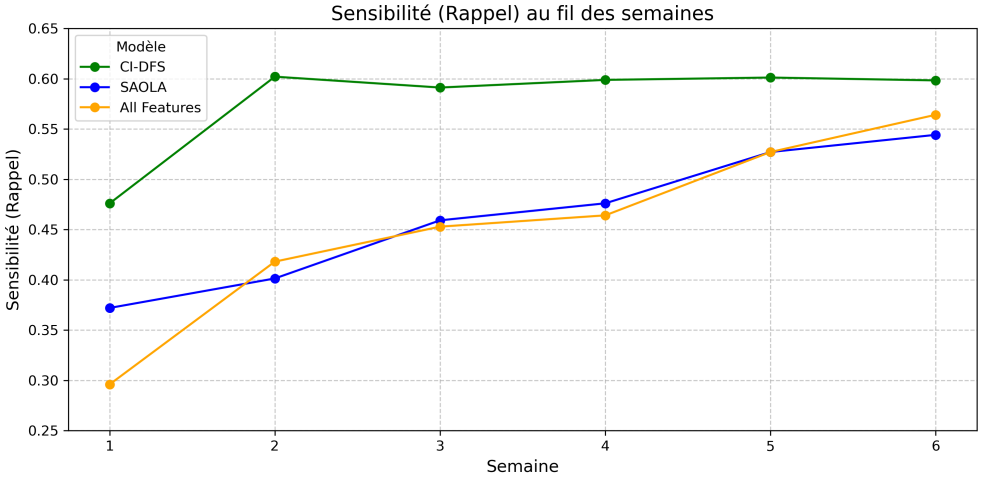

FIGURE 5.7 – Sensibilité sur MoocCN (MoocCN)

qu'une sélection conditionnée et évolutive des caractéristiques permet de mieux capter les signaux faibles liés au désengagement, là où SAOLA et AllFeatures, peinent à s'ajuster à la dynamique du cours.

La figure 5.8 illustre l'évolution de l'exactitude au fil des six semaines du MOOC-CN. Les trois approches présentent des performances proches, oscillant entre **0,71 et 0,79**. Le modèle **CI-DFS** démarre légèrement en dessous de **SAOLA** et **AllFeatures** durant les deux premières semaines, mais comble rapidement cet écart pour atteindre un niveau comparable dès la **troisième semaine**.

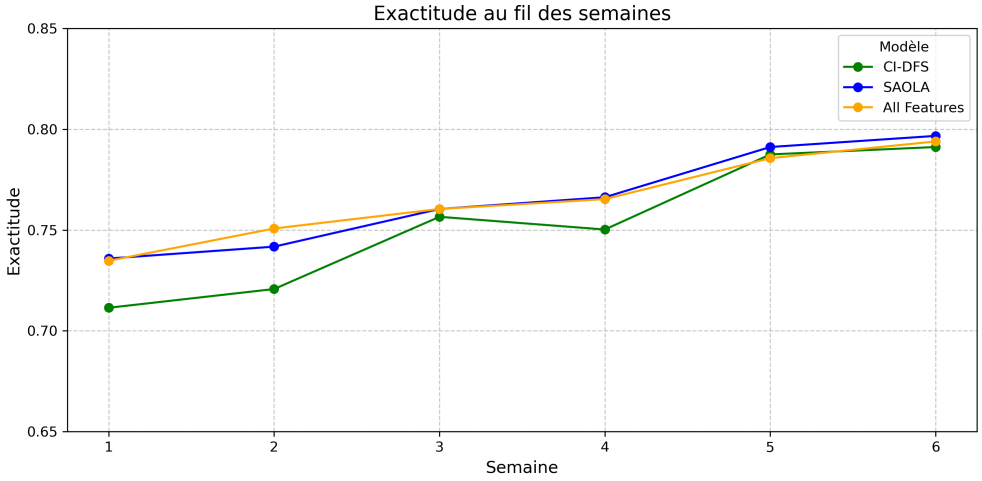

FIGURE 5.8 – Exactitude hebdomadaire sur MoocCN (MoocCN)

À partir de la quatrième semaine, les courbes convergent et se stabilisent autour de **0,79**,

sans différence notable entre les méthodes. Cette convergence indique que, malgré une sélection beaucoup plus restreinte de caractéristiques, **CI-DFS** maintient un niveau de performance équivalent à celui des modèles exploitant un volume de données bien plus important.

Ainsi, l'intérêt principal de CI-DFS ne réside pas dans un gain direct d'exactitude, mais dans sa capacité à préserver des performances élevées avec un coût de traitement considérablement réduit. Ces résultats soulignent que l'apport principal de CI-DFS n'est pas tant une amélioration directe de l'exactitude, mais plutôt sa capacité à atteindre un niveau similaire avec une efficacité bien supérieure en termes de sélection et de traitement des caractéristiques.

La figure 5.9 illustre l'évolution de la spécificité au fil des six semaines du MOOC. Globalement, les trois approches obtiennent des résultats élevés, compris entre 0,77 et 0,93. CI-DFS démarre significativement plus bas (0,80 en semaine 1) que SAOLA (0,89) et AllFeatures (0,94), mais progresse régulièrement pour atteindre environ 0,89 en semaine 6, soit un niveau très proche des deux autres méthodes.

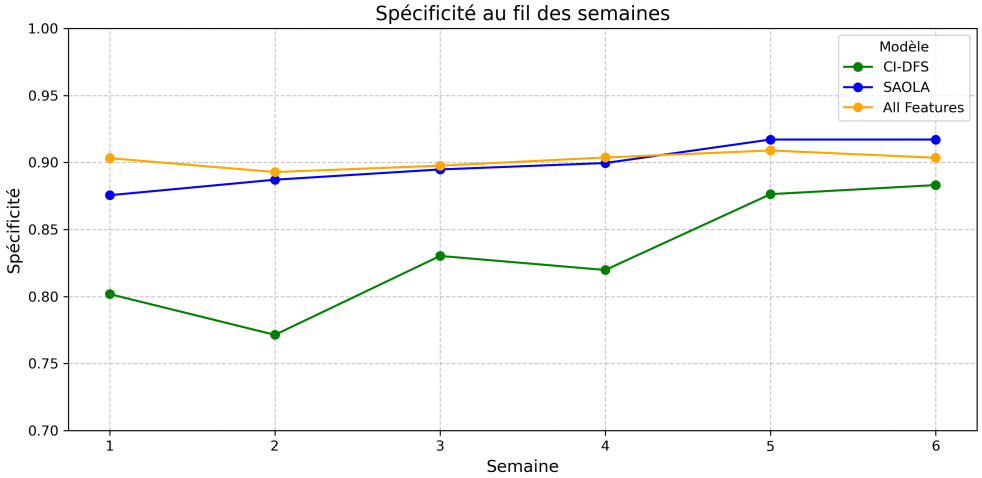

FIGURE 5.9 – Spécificité sur MoocCN (MoocCN)

Cette tendance montre que CI-DFS, bien qu'utilisant un nombre beaucoup plus restreint de caractéristiques, parvient à préserver une bonne capacité à identifier correctement les étudiants qui ne sont pas à risque. En revanche, SAOLA et AllFeatures conservent une avance marginale, mais au prix d'une complexité et d'un coût computationnel nettement plus élevés.

En somme, l'écart initial de spécificité entre CI-DFS et les méthodes classiques se réduit rapidement, confirmant la robustesse de l'approche dans un cadre éducatif où la détection fiable des non-décrocheurs reste essentielle pour limiter les interventions inutiles.

La figure 5.10 montre l'évolution du **nombre de caractéristiques sélectionnées** au cours des six semaines du MOOC-CN. Les écarts entre les approches sont marqués et significatifs. Le modèle **CI-DFS** demeure particulièrement compact, avec **moins de 210 caractéristiques** en semaine 6, contre **plus de 520 pour SAOLA** et **près de 680 pour AllFeatures**. Cela représente une **réduction d'environ 60% par rapport à SAOLA** et de **près de 70% par**

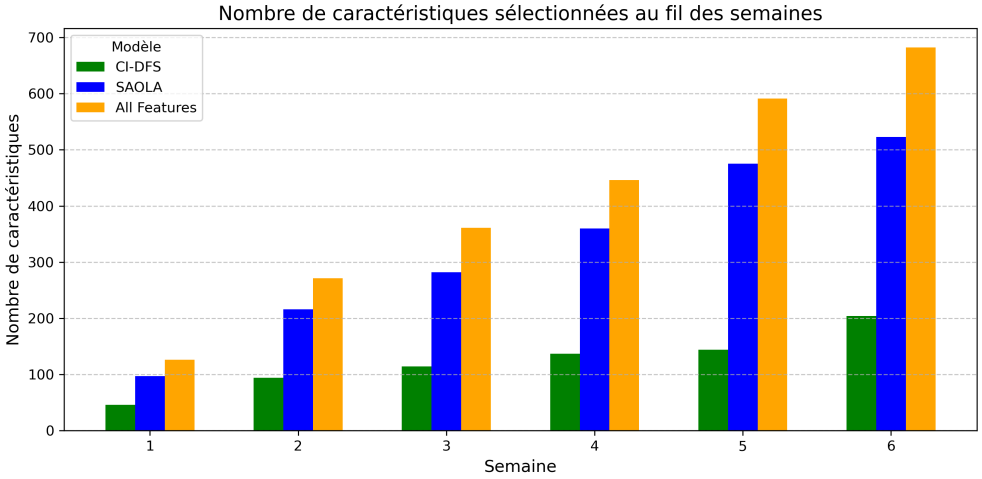

FIGURE 5.10 – Nombre de caractéristiques sélectionnées sur MoocCN

rapport à AllFeatures.

La croissance hebdomadaire reste également bien plus contenue avec CI-DFS, qui ajoute en moyenne **25 à 30 nouvelles caractéristiques par semaine**, soit une progression environ **2,5 fois plus lente que SAOLA** et **3 à 4 fois plus lente qu'AllFeatures**. Ces résultats rejoignent les tendances observées sur OULAD, où CI-DFS parvenait déjà à maintenir un nombre limité de caractéristiques malgré l'allongement progressif de la durée du cours. Dans le cas du MOOC, cette stabilité est d'autant plus marquante que la courte durée de l'expérience accentue l'impact des choix de sélection : CI-DFS démontre sa capacité à rester efficace dès les premières semaines, tandis que les autres approches accumulent rapidement un volume important de variables.

Ainsi, la comparaison entre les deux jeux de données souligne la robustesse de CI-DFS, capable de conserver un ensemble réduit et pertinent de caractéristiques, que ce soit dans un contexte long (OULAD) ou plus court (MoocCN).

La figure 5.11 illustre l'évolution du temps nécessaire à la sélection des caractéristiques au cours des six semaines. CI-DFS conserve un temps de calcul quasi constant, inférieur à une seconde, ce qui traduit une excellente stabilité et une scalabilité adaptée à un déploiement en continu. En revanche, SAOLA voit son temps de sélection croître rapidement, passant de 2 secondes en semaine 1 à près de 49 secondes en semaine 6.

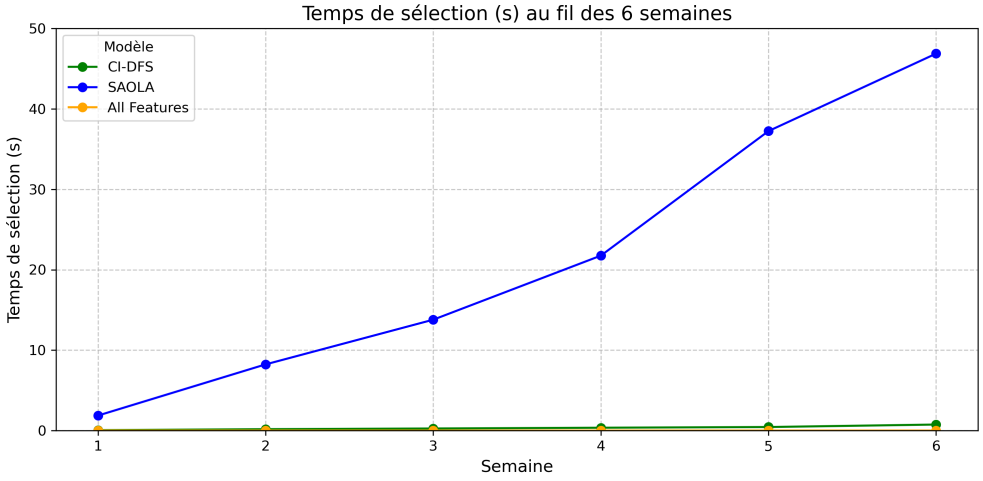

FIGURE 5.11 – Temps de sélection sur MoocCN (MoocCN)

Enfin, La figure 5.12 illustre l'évolution du temps d'entraînement des modèles au fil des six semaines. CI-DFS conserve l'avantage avec une croissance modérée, passant de 0,16 seconde à 0,32 seconde, soit un temps globalement inférieur d'environ 25% à celui des deux autres approches. SAOLA démarre légèrement plus bas qu'AllFeatures, mais converge rapidement, atteignant 0,43 seconde en semaine 6, contre 0,42 seconde pour AllFeatures.

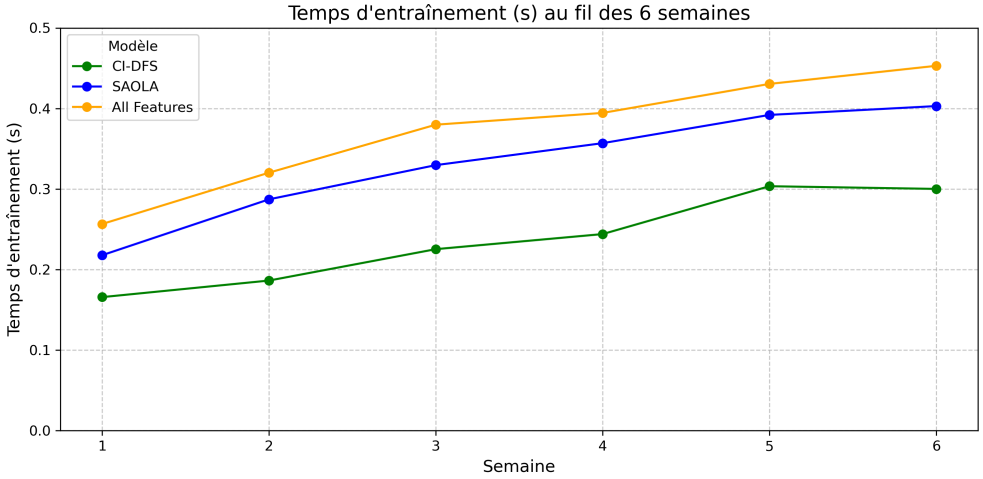

FIGURE 5.12 – Temps d'entraînement sur MoocCN (MoocCN)

Cette tendance confirme que l'efficacité de CI-DFS ne réside pas uniquement dans la réduction du nombre de caractéristiques, mais aussi dans un gain direct en termes de rapidité d'entraînement. Dans un contexte éducatif où les modèles doivent être mis à jour régulièrement, cette économie de temps représente un avantage concret pour le déploiement en continu.

5.6.2 Ablation des composants de CI-DFS

Cette section examine l'effet de la suppression progressive des mécanismes constitutifs de CI-DFS, afin de valider leur utilité respective. Nous testons plusieurs variantes : sans pondération (WMI remplacée par MI), avec seuil fixe, et sans détection de drift. Les performances sont analysées sur les deux jeux de données.

Le tableau 5.3 présente l'impact de chaque mécanisme du CI-DFS sur les performances. Chaque composant (pondération WMI, seuil adaptatif, gestion du drift) contribue significativement à la robustesse du modèle. La suppression de l'un d'entre eux entraîne une dégradation marquée de la sensibilité ou une explosion du nombre de caractéristiques.

Variante	**Sensibilité**		**Nb. caractéristiques**	
	OULAD	MoocCN	OULAD	MoocCN
CI-DFS	0,58	0,84	204	116
WMI → MI	0,46	0,53	194	107
Seuil adaptatif → fixe	0,49	0,61	512	235
Drift → sans drift	0,45	0,48	3300	610

TABLE 5.3 – Impact de chaque composant de CI-DFS sur les performances

Lorsqu'on remplace la **pondération WMI** par une information mutuelle classique (MI), la sensibilité chute de **21% sur OULAD** (de 0,58 à 0,46) et de **37% sur MoocCN** (de 0,84 à 0,53). Cette baisse significative souligne le rôle central de la pondération dans la prise en compte des classes minoritaires et dans la préservation de l'équilibre entre précision et rappel.

Le passage d'un **seuil adaptatif** à un seuil fixe entraîne une dégradation plus modérée de la sensibilité (**-16% sur OULAD** et **-27% sur MoocCN**), mais surtout une inflation marquée du nombre de caractéristiques : **+151% sur OULAD** (de 204 à 512) et **+103% sur MoocCN** (de 116 à 235). Cette évolution confirme que l'adaptation dynamique des seuils est indispensable pour contrôler la croissance de la dimensionnalité sans perte majeure de performance.

Enfin, la suppression du mécanisme de **gestion du drift** a l'impact le plus critique : la sensibilité s'effondre (**-22% sur OULAD** et **-43% sur MoocCN**), tandis que le nombre de caractéristiques explose (**+1517% sur OULAD** et **+426% sur MoocCN**). En l'absence de détection du drift, l'algorithme accumule des caractéristiques obsolètes au fil du temps, ce qui accroît considérablement la complexité du modèle sans gain en performance.

En somme, chaque composant de CI-DFS contribue à l'équilibre entre performance prédictive et contrôle de la dimensionnalité, et leur combinaison est indispensable pour obtenir un modèle à la fois efficace et frugal.

Intéressons-nous maintenant à la sensibilité et à l'impact de l'ablation des composants de CI-DFS au fil du temps.

Sur MoocCN (Figure 5.13), on observe que CI-DFS atteint rapidement une sensibilité proche de 0,60 dès la deuxième semaine, et la maintient de manière stable jusqu'à la fin. Les variantes, qu'il s'agisse de l'utilisation de l'information mutuelle classique, de l'absence de gestion du drift ou du remplacement du seuil adaptatif par un seuil fixe, présentent toutes un retard dans la

montée initiale et plafonnent autour de 0,55 à 0,57. Cela confirme que chacun des mécanismes introduits (pondération de l'information, seuil adaptatif et gestion du drift) contribue à renforcer la détection précoce des décrocheurs.

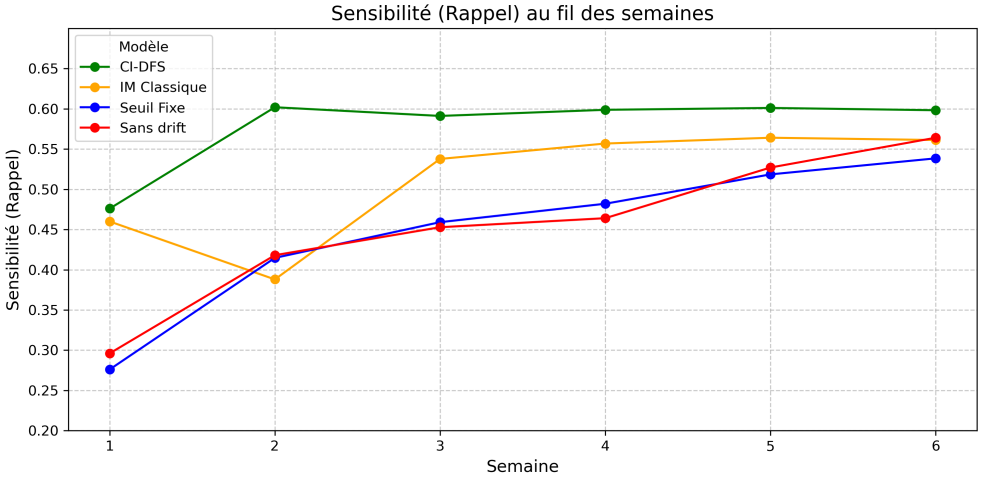

FIGURE 5.13 – Sensibilité sur MoocCN(MoocCN) avec ablation

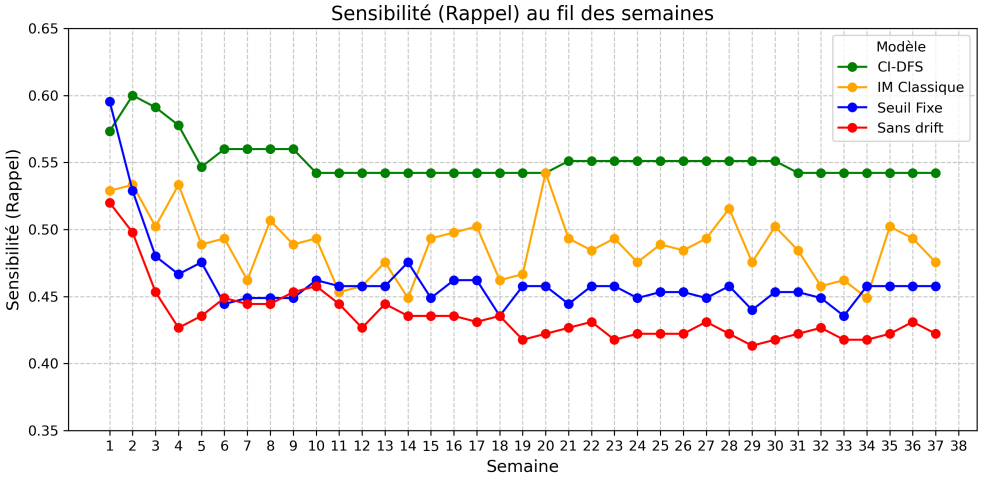

FIGURE 5.14 – Sensibilité sur OULAD avec ablation

Sur OULAD (Figure 5.14), la dynamique est différente : la sensibilité démarre plus haut pour toutes les méthodes, mais CI-DFS reste systématiquement en tête avec une stabilité autour de 0,54–0,55, tandis que les variantes oscillent entre 0,44 et 0,50 avec une variabilité plus marquée. Ici encore, la combinaison complète des mécanismes assure une meilleure robustesse temporelle, en évitant la dégradation observée avec les autres configurations.

Ces résultats suggèrent que l'apport de chaque composant est particulièrement visible sur un

jeu de données court comme MoocCN, où la détection précoce est critique, tandis que sur un jeu plus long comme OULAD, l'avantage réside surtout dans la stabilité et la régularité des prédictions. En résumé, CI-DFS tire sa force de la complémentarité de ses modules, chacun jouant un rôle pour compenser les limites des approches plus simplifiées.

Pour ce qui est du nombre de caractéristiques retenues au fil des semaines, présenté dans les Figures 5.15 (MOOC MoocCN) et 5.16 (OULAD).

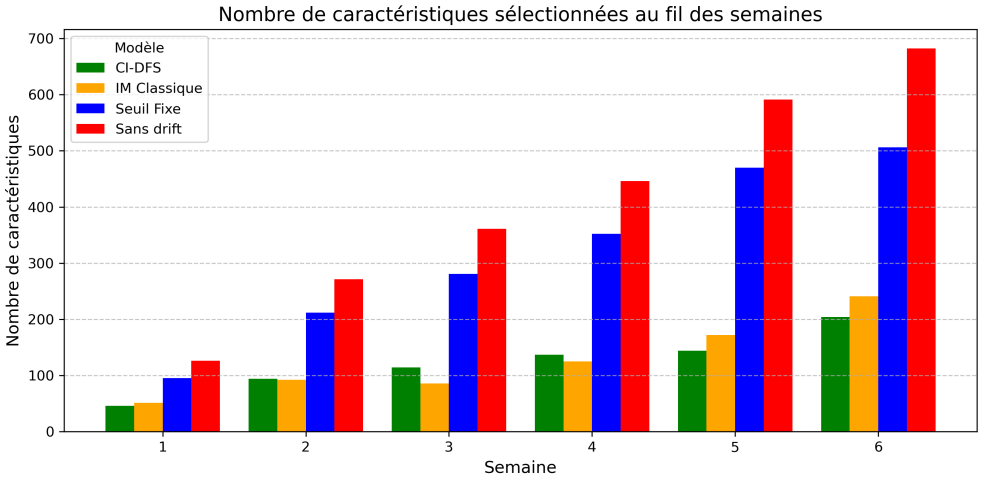

FIGURE 5.15 – Nombre de caractéristiques sur MoocCN(MoocCN) avec ablation

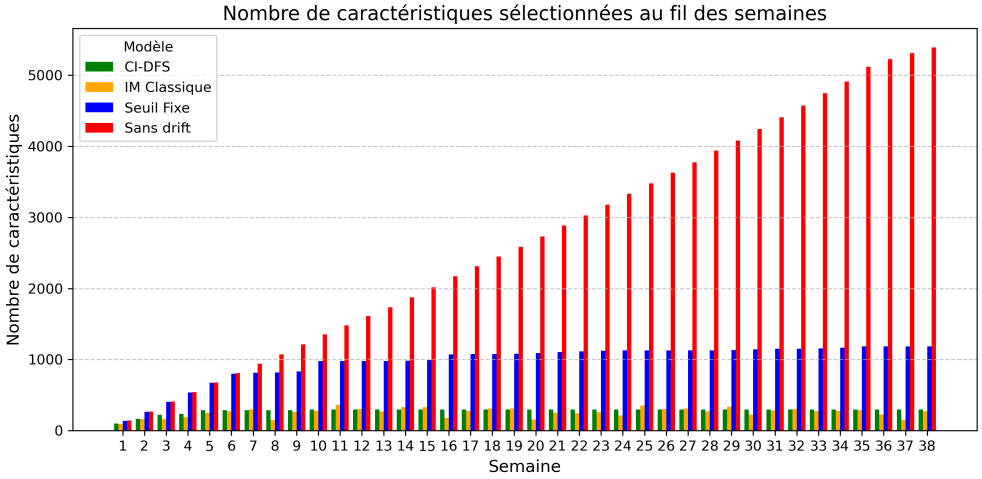

FIGURE 5.16 – Nombre de caractéristiques sur OULAD avec ablation

Sur MoocCN (Figure 5.15), on observe une croissance régulière du nombre de caractéristiques pour toutes les variantes. Cependant, CI-DFS reste largement plus compact (116 caractéristiques en moyenne à la semaine 6), tandis que l'ablation du drift ou de l'adaptation du seuil conduit à

une inflation rapide, dépassant 500 caractéristiques. Cela confirme que la gestion du drift et le seuil adaptatif sont essentiels pour contrôler la taille du sous-ensemble retenu dans des contextes à faible horizon temporel.

Sur OULAD (Figure 5.16), la différence est encore plus marquée : l'absence de gestion du drift entraîne une accumulation massive de variables (plus de 5000 caractéristiques retenues en fin de parcours), contre environ 200 pour CI-DFS. L'ablation du seuil adaptatif produit également un modèle nettement plus lourd (plus de 1000 caractéristiques). Ces résultats mettent en évidence le rôle central de chaque composant de CI-DFS dans la maîtrise de la complexité et la frugalité du modèle sur des flux de données longs et évolutifs.

Enfin, les figures suivantes complètent l'analyse avec l'impact sur le temps de sélection (Figures 5.17 et 5.18).

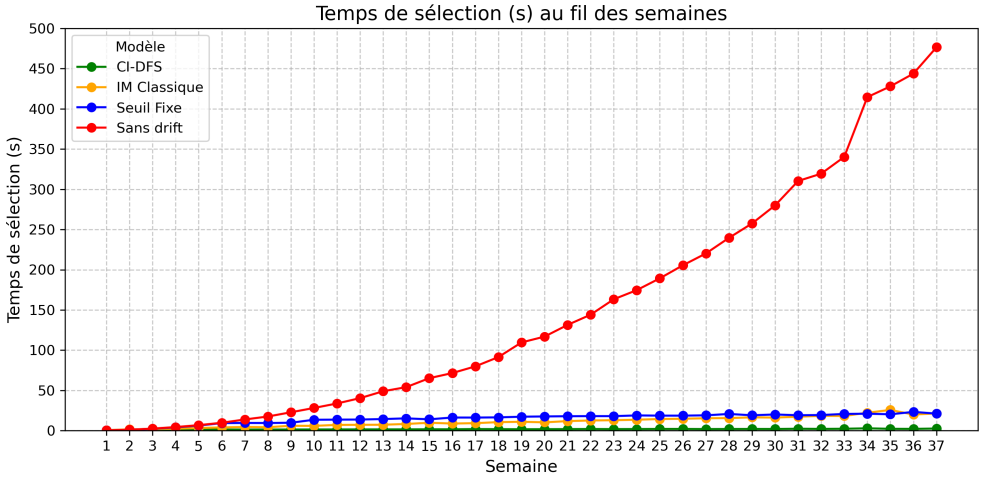

FIGURE 5.17 – Temps de sélection sur OULAD avec ablation.

Sur le jeu de données MoocCN (Figure 5.18), CI-DFS se distingue par une stabilité remarquable, avec un temps inférieur à 2 secondes en moyenne, quand les variantes « No Drift » et « Fixed Threshold » atteignent respectivement près de 9 et 5 secondes à la sixième semaine. Cela confirme l'efficacité de la stratégie combinée (pondération + seuil adaptatif + gestion du drift) pour contenir les coûts computationnels dans des environnements courts et rapides.

Sur OULAD, les écarts se creusent encore davantage (Figure 5.17). Alors que CI-DFS se maintient à un niveau quasi constant ($<$ 2 secondes), la variante « No Drift » explose progressivement pour dépasser 350 secondes en fin de parcours, illustrant le coût d'une accumulation de caractéristiques non régulées. Les versions « Classic MI » et « Fixed Threshold » montrent des croissances plus modérées (10–20 secondes), mais restent nettement moins efficaces que CI-DFS.

5.6.3 Analyse qualitative des caractéristiques

Au-delà des performances quantitatives, nous analysons ici la nature des caractéristiques retenues par CI-DFS selon leur comportement temporel :

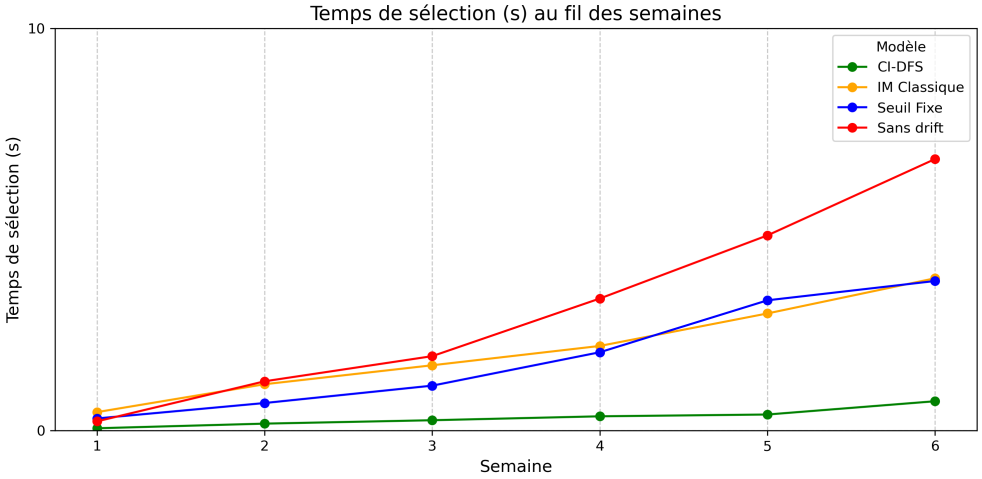

FIGURE 5.18 – Temps de sélection sur MoocCN(MoocCN) avec ablation

— **Caractéristiques stables** : il s'agit de variables structurelles, peu sensibles à l'évolution temporelle et dont l'importance reste constante pour un apprenant donné. Elles sont présentes dès le début du parcours et conservées tout au long de l'apprentissage. C'est par exemple le cas de la *registration date*, qui indique uniquement le moment d'inscription de l'étudiant, ou encore du *education level*, qui reflète un attribut sociodémographique fixé (diplôme, niveau d'études antérieur). Ces caractéristiques, bien que statiques, conservent un rôle prédictif stable dans le temps, car elles capturent des éléments de contexte structurel (profil initial, caractéristiques personnelles) pouvant influencer la réussite ou le risque de décrochage. Leur utilité ne réside donc pas dans la dynamique hebdomadaire des traces, mais dans leur valeur explicative complémentaire, en tant que variables de référence ancrées dans le profil de l'apprenant.

— **Caractéristiques transitoires** : il s'agit de variables dont la pertinence est limitée à une phase spécifique du parcours. Par exemple, la lecture de *PDF* ou l'activité sur les *forums* sont souvent fortement informatives au tout début de la formation, car elles traduisent un engagement initial ou une prise de contact avec les ressources pédagogiques. Toutefois, leur valeur prédictive tend à diminuer au fil du temps : une fois les habitudes d'étude installées, ces comportements ne constituent plus des indicateurs discriminants de persévérance ou de décrochage. Ces caractéristiques présentent donc une importance épisodique, jouant un rôle de signaux précoces, mais devenant progressivement redondantes ou secondaires par rapport à d'autres variables plus structurantes, comme la régularité des soumissions d'évaluations ou l'évolution des résultats obtenus. Il est important de noter que **CI-DFS** adopte un principe de sélection irréversible, lorsqu'une caractéristique est écartée, elle ne peut plus être réintroduite par la suite. Cette décision s'inscrit dans la logique frugale du modèle, qui vise à éviter la réintégration de variables devenues non informatives ou redondantes.

— **Caractéristiques tardives** : ces variables deviennent pertinentes uniquement dans les phases avancées du parcours. C'est le cas, par exemple, des accès tardifs aux *quiz* ou des sessions d'apprentissage intensives concentrées en fin de formation. Elles agissent comme

des signaux de rattrapage, révélant des stratégies compensatoires mises en place par certains apprenants lorsqu'ils cherchent à combler rapidement un retard accumulé ou à maximiser leurs chances de réussite avant une échéance. Leur caractère prédictif est donc lié à une temporalité spécifique : elles n'apparaissent que tardivement dans les traces, mais leur activation constitue souvent un marqueur fort d'un risque imminent d'abandon ou, à l'inverse, d'une mobilisation accrue pour atteindre l'objectif final.

CI-DFS opère donc une sélection adaptative, sensible aux dynamiques temporelles et à la spécificité des signaux faibles, c'est-à-dire des caractéristiques discrètes dont la valeur prédictive n'apparaît qu'à certains moments, ce qui renforce l'interprétabilité du modèle tout en maintenant une structure frugale et efficace.

5.7 Discussion

Les résultats exposés précédemment confirment la pertinence de l'approche CI-DFS pour la sélection dynamique de caractéristiques dans les environnements éducatifs. Cette section propose une analyse synthétique des performances observées, met en lumière le rôle des différents composants, et ouvre des pistes concrètes d'amélioration.

5.7.1 Interprétation globale des résultats

Efficacité prédictive

CI-DFS se distingue par sa capacité à détecter précocement les abandons, avec une sensibilité supérieure à 0,55 sur OULAD et atteignant 0,84 dès la troisième semaine sur MoocCN(MoocCN). Cette performance s'explique par la prise en compte des signaux faibles spécifiques aux apprenants décrocheurs, rendue possible par la pondération WMI et la mise à jour hebdomadaire des caractéristiques sélectionnées. Comparé aux variantes ablatées et aux baselines classiques, qui plafonnent généralement autour de 0,45–0,50 sur OULAD et peinent à dépasser 0,60 sur MoocCN, CI-DFS démontre un avantage significatif en termes de sensibilité précoce et de stabilité dans le temps.

Frugalité des données et stabilité

CI-DFS parvient à maintenir un ensemble de caractéristiques restreint et maîtrisé tout au long du parcours, avec en moyenne 205 caractéristiques retenues sur OULAD et 117 sur MoocCN (MoocCN). Cette frugalité limite le bruit, favorise la lisibilité du modèle, et réduit les temps d'entraînement. Le mécanisme de gestion du drift permet d'éviter l'accumulation de caractéristiques obsolètes, favorisant ainsi une stabilité dans des parcours longs et évolutifs.

Comparaison avec SAOLA

Bien que plus rapide à l'exécution, SAOLA souffre de plusieurs limitations : absence de pondération, seuils statiques, et incapacité à gérer l'évolution temporelle des données. Ces faiblesses se traduisent par une sensibilité inférieure, une croissance du nombre de caractéristiques et un coût computationnel plus élevé à mesure que les semaines s'accumulent.

5.7.2 Apport de chaque composant

L'étude d'ablation permet d'isoler l'effet de chaque mécanisme constitutif pour évaluer son impact :

— La suppression de la pondération WMI entraîne une baisse significative de la sensibilité, pouvant atteindre 12 points selon les jeux de données. Sans cette pondération, les caractéristiques liées aux apprenants décrocheurs, déjà sous-représentés dans les données, perdent en importance relative au profit de caractéristiques plus fréquentes mais moins discriminantes. Le modèle devient alors moins réactif aux comportements atypiques et peine à détecter précocement les cas d'abandon, ce qui réduit l'efficacité des interventions ciblées.

— Le remplacement du seuil adaptatif par un seuil fixe entraîne une perte de flexibilité, une augmentation du nombre de caractéristiques, et une moindre capacité d'adaptation aux changements de comportement. Sur OULAD, le nombre moyen de caractéristiques passe de 205 avec seuil adaptatif à plus de 512 avec un seuil fixe, soit plus du double. Sur MoocCN, on observe un effet similaire, avec 117 caractéristiques en moyenne contre 235. Cette inflation reflète l'incapacité du seuil fixe à désactiver progressivement les caractéristiques devenues moins pertinentes. Sur le plan prédictif, la sensibilité chute également, passant de 0,58 à 0,49 sur OULAD et de 0,84 à 0,61 sur MoocCN, ce qui montre que le modèle peine à s'adapter aux évolutions temporelles des comportements apprenants.

— L'absence de gestion du drift provoque une explosion du nombre de caractéristiques, nuisant à la frugalité du modèle et ralentissant le temps d'entraînement. Sur OULAD, le nombre moyen de caractéristiques passe de 205 pour CI-DFS complet à plus de 3300 sans gestion du drift, soit une multiplication par plus de 15. Sur MoocCN, il grimpe de 117 à plus de 610. Cette accumulation massive reflète l'incapacité à éliminer les caractéristiques devenues obsolètes au fil du temps. Sur le plan computationnel, le temps d'entraînement hebdomadaire augmente légèrement (0,30 seconde avec CI-DFS contre environ 0,48 seconde sans gestion du drift). Cette différence n'est pas critique dans le cadre de la présente expérimentation, mais elle pourrait devenir déterminante sur des jeux de données plus volumineux ou dans des déploiements à grande échelle.

Ces résultats confirment que chacun des mécanismes de CI-DFS est essentiel à la robustesse et à la stabilité globale de l'approche.

5.7.3 Limites et perspectives

Choix fixe des hyperparamètres

Les valeurs des paramètres k, λ et w ont été déterminées empiriquement. Une approche plus adaptative, s'ajustant dynamiquement aux caractéristiques de l'apprenant ou à la distribution des données, pourrait renforcer la flexibilité du modèle.

Cible binaire

Le cadre actuel repose sur une classification binaire (abandon vs. poursuite). Étendre cette approche à des classes multiples, par exemple en distinguant différents types de décrochage ou

niveaux d'engagement, permettrait une modélisation plus fine des comportements d'apprentissage.

Identification des signaux d'alerte

La dynamique des caractéristiques sélectionnées chaque semaine ouvre des perspectives prometteuses pour l'identification précoce des signaux de décrochage. En repérant les caractéristiques activées juste avant une perte d'engagement, CI-DFS pourrait contribuer à construire des indicateurs exploitables pour une intervention pédagogique ciblée.

Apprentissage incrémental du classifieur

CI-DFS a été testé avec un classifieur batch (Random Forest). Son intégration avec des modèles incrémentaux (e.g. Perceptron, SGD, Naïve Bayes incrémental) constituerait une avancée vers une architecture pleinement adaptative, capable de s'ajuster en continu aux flux de données, ce sujet sera étudié dans le prochain chapitre.

5.8 Conclusion du chapitre

Dans ce chapitre nous avons proposé un algorithme de sélection dynamique de caractéristiques, conçu pour répondre aux défis de la prédiction de décrochage dans les VLEs. L'algorithme CI-DFS (Class-Imbalanced Dynamic Feature Selection) repose sur une combinaison de trois mécanismes clés : (i) une pondération des scores de pertinence sensible au déséquilibre des classes (WMI), (ii) un seuil adaptatif basé sur un percentile pour réguler dynamiquement la sélection, et (iii) une stratégie de suppression conditionnée pour gérer l'obsolescence des caractéristiques dans le temps.

L'évaluation expérimentale conduite sur deux contextes distincts, OULAD (38 semaines) et MoocCN (6 semaines), met en évidence la robustesse de CI-DFS. Sur les deux jeux de données, l'algorithme atteint une sensibilité supérieure à celle de toutes les autres méthodes comparées, tout en maintenant un nombre de caractéristiques très réduit et des temps de traitement compatibles avec des déploiements réels. Elle démontre également une meilleure capacité à détecter les abandons précoces sans sacrifier la spécificité.

Les analyses d'ablation confirment le rôle structurant de chaque composant du modèle. La suppression de la pondération WMI ou du seuil adaptatif conduit à une baisse significative des performances, tandis que l'absence de gestion du drift entraîne une explosion du nombre de caractéristiques. CI-DFS s'impose ainsi comme un modèle équilibré, capable de maintenir une frugalité des données sans compromis sur la détection des cas des classes minoritaires.

En complément des résultats quantitatifs, l'analyse qualitative met en évidence la capacité du modèle à sélectionner dynamiquement des caractéristiques stables, transitoires ou tardives selon leur pertinence temporelle. Ce comportement adaptatif, sensible à la trajectoire d'apprentissage de l'apprenant, constitue un atout essentiel pour renforcer l'interprétabilité et la personnalisation des modèles prédictifs.

Cette contribution constitue une première étape vers des systèmes d'analyse plus intelligents, capables d'évoluer avec les données, d'identifier les signaux faibles, et d'optimiser l'usage des ressources computationnelles. Les pistes futures incluent l'extension à des tâches multi-classes, l'intégration dans des pipelines d'apprentissage incrémental de bout en bout, et l'exploitation directe des caractéristiques sélectionnées pour des interventions pédagogiques ciblées sur les apprenants à risque.

Chapitre 6

Apprentissage incrémentale et déclenchement adaptatif

Ce chapitre s'intéresse à l'apprentissage incrémental, une approche conçue pour répondre à la nature évolutive des données. À l'inverse des modèles traditionnels, entraînés de manière statique sur des jeux de données figés, les modèles incrémentaux sont capables d'intégrer progressivement de nouvelles observations, sans nécessiter de réentraînement complet. Cette capacité d'adaptation continue est particulièrement précieuse dans les environnements d'apprentissage virtuels (VLEs), où les données sont générées au rythme des interactions des apprenants.

Le chapitre est structuré de manière progressive. Il s'ouvre sur une introduction aux défis spécifiques de l'apprentissage incrémental dans les VLEs. Nous détaillons ensuite la méthodologie proposée, avant d'en présenter l'application sur deux jeux données présentés au chapitre 3 : OULAD [91] et MoocCN [64]. Les résultats font l'objet d'une analyse fondée sur différentes métriques de performance et d'efficacité, avant d'être discutés de manière critique. Le chapitre se conclut par une synthèse des contributions et l'ouverture de perspectives pour de futurs travaux.

6.1 Introduction

Les VLEs produisent des flux continus d'interactions, structurés dans le temps, au rythme des activités des apprenants [90, 156]. Dans ce contexte dynamique, les modèles d'apprentissage doivent être capables de s'adapter progressivement, sans attendre que l'ensemble des données soit disponible a priori. L'apprentissage incrémental répond à cette exigence en actualisant les modèles au fil de l'eau, à mesure que de nouvelles données deviennent accessibles, tout en conservant les connaissances acquises [77, 53].

De nombreuses stratégies ont été développées pour mettre en œuvre cet apprentissage progressif. Certaines reposent sur des mises à jour systématiques à fréquence fixe, indépendamment de la stabilité des données ou de la pertinence d'une actualisation, ce qui entraîne un gaspillage de ressources computationnelles [93]. D'autres approches s'appuient sur la détection de dérive conceptuelle pour déclencher les mises à jour [28, 133], mais elles se focalisent essentiellement sur les changements dans la distribution des données en entrée, sans considérer les évolutions dans la pertinence des caractéristiques, un aspect pourtant crucial dans les contextes éducatifs, où les contenus, les ressources et les comportements évoluent jour après jour comme discutée dans le

chapitre précédent.

Un autre enjeu majeur réside dans le coût computationnel associé aux réentraînements fréquents. Dans les systèmes éducatifs à grande échelle, l'adaptation du modèle ne peut se faire au prix d'une complexité excessive. La frugalité devient alors un critère central, tant pour des raisons de scalabilité que d'opérabilité [61, 153]. Il s'agit non seulement d'adapter le modèle aux données les plus récentes, mais aussi de le faire de manière ciblée, en identifiant les moments où une mise à jour est réellement nécessaire, sur la base de signaux pertinents et interprétables.

Dans cette perspective, ce chapitre introduit un cadre d'apprentissage incrémental informé, reposant sur un mécanisme de déclenchement adaptatif. Ce dernier combine trois types de signaux : d'une part, les changements observés dans l'espace des caractéristiques actives via une sélection dynamique (ajout, suppression et détéction de dérives) [169], et d'autre part, la détection d'une dégradation de performance mesurée sur un échantillon représentatif issu des données passées. Le modèle n'est mis à jour que si au moins l'un de ces signaux dépasse un seuil prédéfini, ce qui permet de limiter les réentraînements aux seuls moments pertinents. L'approche vise ainsi à concilier robustesse prédictive, réactivité aux situations à risque, et sobriété computationnelle, dans un cadre adapté aux contraintes spécifiques des flux éducatifs.

6.2 Méthodologie

L'objectif de cette méthodologie est de permettre à un modèle prédictif de s'adapter au fil du temps, tout en limitant les mises à jour aux seules situations où elles sont réellement nécessaires. Pour cela, nous proposons un mécanisme de déclenchement fondé sur une double source de signal : la dynamique des caractéristiques utilisées par le modèle, et l'évolution de sa performance sur un échantillon de validation construit en continu. Cette section détaille les différents composants du cadre et leur articulation.

Définition du problème

On considère un flux de données organisé en séries temporelles, noté (X_t, y_t), où t représente l'instant d'observation.

À chaque instant t, on observe un ensemble de n_t apprenants, chacun décrit par un vecteur de caractéristiques de dimension d_t. On définit :

$$X_t = \{x_{i,t}\}_{i=1}^{n_t}, \quad x_{i,t} \in \mathbf{R}^{d_t}$$

où $x_{i,t}$ est le vecteur de caractéristiques de l'apprenant i à l'instant t.

De même, on associe à chaque apprenant une étiquette binaire indiquant son état (décrochage ou non) :

$$y_t = \{y_{i,t}\}_{i=1}^{n_t}, \quad y_{i,t} \in \{0, 1\}$$

Ainsi, X_t peut être vu comme une matrice de dimension $n_t \times d_t$, regroupant l'ensemble des vecteurs de caractéristiques observés à l'instant t, et y_t comme un vecteur de dimension n_t contenant les étiquettes correspondantes.

Contrairement aux approches classiques, l'espace des caractéristiques n'est pas fixe : il peut évoluer dans le temps, certaines variables devenant obsolètes, redondantes ou nouvellement disponibles.

L'objectif est de construire une séquence de modèles prédictifs $(M_0, M_1, \ldots, M_z)$ capables de s'adapter au fil du temps, tout en limitant le nombre de réentraînements. Chaque mise à jour doit être justifiée par un changement significatif dans les données ou dans les performances du modèle.

6.2.1 Vue d'ensemble du cadre proposé

Le cadre proposé repose sur trois modules complémentaires, conçus pour fonctionner en ligne :

— **Sélection dynamique des caractéristiques.** Ce module, détaillé dans le chapitre 5, identifie, à chaque pas de temps, les caractéristiques les plus pertinentes pour la tâche de prédiction. Il prend en compte la pertinence statistique (via l'information mutuelle pondérée), la redondance entre les caractéristiques, et leur stabilité dans le temps.

— **Mécanisme de déclenchement.** Un ensemble de règles est défini pour décider si une mise à jour du modèle est nécessaire. Trois cas sont pris en compte pour déclancher les mises à jour : (i) un changement dans l'ensemble des caractéristiques actives (ajouts ou suppressions), (ii) une dérive dans leurs scores de pertinence, et (iii) une baisse de performance du modèle, mesurée sur un échantillon de validation représentatif construit à partir de données antérieures. Si l'un de ces signaux dépasse un seuil prédéfini, une mise à jour est déclenchée.

— **Protocole de mise à jour incrémentale.** Lorsqu'un déclenchement est validé, le modèle est mis à jour à l'aide des données disponibles au moment t ainsi que d'un sous-ensemble d'exemples des pas de temps précédents, sélectionnés de manière à maintenir un équilibre entre les classes. Cette mémoire tampon (*buffer*) permet d'éviter l'oubli des exemples anciens et de stabiliser les performances au cours du temps.

À chaque pas de temps, les données sont analysées, les caractéristiques sont sélectionnées dynamiquement, et un ensemble de règles détermine si une mise à jour du modèle est nécessaire. L'apprentissage s'effectue alors de façon ciblée, uniquement lorsqu'un signal de changement est détecté.

Ce couplage entre sélection dynamique, déclenchement conditionnel et mise à jour contrôlée permet d'adapter finement le modèle au fil du temps, en évitant les surcoûts liés aux réentraînements systématiques, tout en assurant une capacité de détection robuste des étudiants à risque.

6.2.2 Rôle et gestion de la mémoire tampon

La mémoire tampon (*buffer*) joue un rôle essentiel dans le cadre proposé, en assurant la continuité et la stabilité de l'apprentissage au fil du temps. Elle correspond à un sous-ensemble borné des données historiques, régulièrement mis à jour afin de conserver des exemples représentatifs des comportements passés [77, 53].

Concrètement, à chaque itération, une fraction des observations récentes (X_t, y_t) est intégrée à la mémoire, tandis que les plus anciennes sont remplacées selon une politique de renouvellement (souvent de type *first-in, first-out*). Cette stratégie permet de maintenir un équilibre entre actualité et diversité, tout en maîtrisant la taille du buffer [153].

La mémoire tampon intervient à deux niveaux :

— pour **évaluer la performance courante** du modèle sur des exemples antérieurs, facilitant la détection de baisses de précision susceptibles de déclencher une mise à jour.
— pour **stabiliser l'apprentissage incrémental** lors du réentraînement, en combinant les nouvelles données aux exemples stockés, afin de réduire le risque d'oubli catastrophique et de dérive conceptuelle [44, 93].

6.2.3 Description détaillée de l'algorithme

L'Algorithme 5 présente le fonctionnement du déclenchement incrémental de la sélection dynamique de caractéristiques. Le Tableau 6.1 récapitule l'ensemble des variables utilisées, incluant les paramètres d'entrée, les variables de sortie et les états mis à jour au fil des itérations.

TABLE 6.1 – Variables d'entrée, de sortie intermédiaire de l'algorithme 5

Symbole	**Description**
	Entrées
$\{(X_t, y_t)\}_{t=1}^{T}$	Flux de données (caractéristiques et étiquettes) observé à l'instant t, jusqu'à la fin T.
T	Nombre total d'itérations temporelles (semaines, sessions, etc.).
τ_{add}	Seuil de déclenchement lié à l'ajout de caractéristiques.
τ_{rem}	Seuil de déclenchement lié au retrait de caractéristiques.
τ_{rel}	Seuil de déclenchement lié à la variation de pertinence.
τ_{perf}	Seuil de déclenchement lié à la dégradation de performance sur la mémoire tampon.
M_0	Modèle prédictif initialisé avant la première mise à jour.
S_0	Ensemble initial (éventuellement vide) des caractéristiques actives.
$\mathcal{B}$	Mémoire tampon (buffer) contenant un sous-échantillon d'exemples passés utilisé pour l'évaluation et le rejeu.
	Sorties
M_T	Modèle incrémental final après T itérations (sortie principale).
M_t	État du modèle après la mise à jour (ou réutilisation) au temps t.
S_t	Sous-ensemble de caractéristiques sélectionnées au temps t par CI-DFS.
R_t	Scores de pertinence associés aux caractéristiques sélectionnées au temps t.
Δ_t^{add}, Δ_t^{rem}, Δ_t^{rel}	Variations mesurées (ajout, retrait, pertinence) utilisées pour le déclenchement.
Δ_t^{perf}	Variation de performance de M_{t-1} évaluée sur $\mathcal{B}$.
$\mathcal{B}$ (mis à jour)	Mémoire tampon après ajout d'un sous-ensemble de (X_t, y_t).

Algorithm 5 Apprentissage incrémental avec déclenchement et sélection de caractéristiques

Input : Flux de données $\{(X_t, y_t)\}_{t=1}^{T}$; Seuils $\tau_{\text{add}}, \tau_{\text{rem}}, \tau_{\text{rel}}, \tau_{\text{perf}}$ Modèle initiale M_0, Ensemble de caractéristiques S_t, Mémoire tampon $\mathcal{B}$
Output : Modèle incrémental final M_T

1: Initialiser le modèle M_0, $S_0 \leftarrow \emptyset$, et $\mathcal{B} \leftarrow \emptyset$
2: **for** $t = 1$ **to** T **do**
// Sélection dynamique des caractéristiques à l'instant t
3: $(S_t, \mathrm{R}_t) \leftarrow \text{CI-DFS}(X_t, y_t)$
// Calcul des variations - équations 6.1, et 6.2
4: Calculer $\Delta_t^{\text{add}}, \Delta_t^{\text{rem}}, \Delta_t^{\text{rel}}$
// Evaluer le modèle sur la mémoire tampon - équation 6.3
5: $\Delta_t^{\text{perf}} \leftarrow \text{EvalPerf}(M_{t-1}, \mathcal{B})$
// Vérifier les conditions de déclenchement
6: **if** $(\Delta_t^{\text{add}} > \tau_{\text{add}})$ **or** $(\Delta_t^{\text{rem}} > \tau_{\text{rem}})$ **or** $(\Delta_t^{\text{rel}} > \tau_{\text{rel}})$ **or** $(\Delta_t^{\text{perf}} > \tau_{\text{perf}})$ **then**
// Mettre à jour le modèle incrémental
7: $M_t \leftarrow \text{Train}(M_{t-1}, X_t \cup \mathcal{B}, y_t \cup y_{\mathcal{B}})$
8: Update $\mathcal{B}$ with samples from (X_t, y_t)
9: **else**
// Réutiliser le modèle précédent
10: $M_t \leftarrow M_{t-1}$
11: **end if**
12: **end for**
13: **return** M_T

Description ligne par ligne

- **Lignes 1** : initialisation du modèle M_0, de l'ensemble des caractéristiques S_0 (vide) et de la mémoire tampon $\mathcal{B}$.
- **Lignes 2** : début de la boucle pour chaque pas de temps t.
- **Lignes 3** : sélection dynamique des caractéristiques pertinentes S_t via CI-DFS et calcul de leurs valeurs de pertinence R_t.
- **Ligne 4** : calcul des variations liées aux caractéristiques : ajout, retrait (équation 6.1), évolution de pertinence (équation 6.2)$(\Delta_t^{\text{add}}, \Delta_t^{\text{rem}}, \Delta_t^{\text{rel}})$.
- **Ligne 5** : évaluation des performances du modèle M_{t-1} sur la mémoire tampon $\mathcal{B}$, produisant Δ_t^{perf} (équation 6.3).
- **Lignes 6** : évaluation des conditions de déclenchement. Si une variation dépasse son seuil, une mise à jour incrémentale est réalisée.
- **Lignes 7-8** : le modèle est réentraîné sur les nouvelles données (X_t, y_t) combinées avec le buffer $\mathcal{B}$, et ce dernier est mis à jour.
- **Lignes 9-10** : sinon, aucune mise à jour n'est effectuée et le modèle précédent M_{t-1} est conservé.
- **Lignes 11-12** : fin de condition et de la boucle.
- **Ligne 13** : après T itérations, retour du modèle incrémental final M_T.

6.2.4 Sélection dynamique des caractéristiques

La sélection dynamique des caractéristiques constitue l'un des fondements du cadre proposé. Elle permet de suivre, au fil du temps, l'évolution de la pertinence des variables issues des traces d'apprentissage. Ce module repose sur les principes présentés en détail dans le chapitre 5, adaptés

ici à une intégration en ligne dans un cadre d'apprentissage incrémental.

Ce module permet de maintenir une représentation adaptée, synthétique et évolutive de l'activité des apprenants. Les modifications observées dans l'ensemble S_t ou dans les scores de pertinence R_t fournissent des signaux exploitables pour le déclenchement conditionnel des mises à jour du modèle, abordé dans la section suivante.

6.2.5 Déclenchement de la mise à jour du modèle

Le mécanisme de déclenchement, (ligne 6 dans l'algorithme 5), a pour but d'éviter les mises à jour systématiques du modèle, en n'autorisant une réadaptation que lorsque des signaux significatifs justifient un réentraînement. Trois types de signaux sont analysés à chaque pas de temps t : des changements structurels dans l'espace des variables sélectionnées, une dérive dans leurs scores de pertinence, et une baisse de performance du modèle.

Changement dans l'espace des caractéristiques

On évalue la stabilité de l'espace de représentation en mesurant la proportion de caractéristiques ajoutées et supprimées entre deux pas de temps consécutifs. Soit S_{t-1} l'ensemble des caractéristiques actives à l'instant $t-1$, et S_t celui de l'instant courant. On définit alors :

$$\Delta_t^{\text{ajout}} = \frac{|S_t \setminus S_{t-1}|}{|S_{t-1}|}, \qquad \Delta_t^{\text{suppr}} = \frac{|S_{t-1} \setminus S_t|}{|S_{t-1}|} \tag{6.1}$$

La première définition, Δ_t^{ajout}, indique la proportion de nouvelles caractéristiques intégrées dans la sélection, tandis que la seconde, Δ_t^{suppr}, mesure la proportion de variables abandonnées. Une forte instabilité dans S_t signale une modification structurelle du comportement des apprenants ou des ressources actives.

Dérive des scores de pertinence

Même si l'ensemble S_t reste relativement stable, les scores de pertinence des variables peuvent varier dans le temps. On surveille donc la dérive maximale parmi les caractéristiques communes à S_{t-1} et S_t :

$$\Delta_t^{\text{rel}} = \max_{f \in S_t \cap S_{t-1}} |R_t(f) - R_{t-1}(f)| \tag{6.2}$$

Une telle variation peut révéler une réorganisation dans l'importance relative des traces observées, par exemple en raison d'un changement de phase dans le cours ou de nouveaux comportements d'engagement.

Baisse de performance du modèle

Enfin, on évalue l'exactitude (*accuracy*) du modèle M_{t-1} à l'instant t, sur un ensemble de validation $\mathcal{B}$ composé d'exemples récents. L'exactitude correspond à la proportion d'exemples correctement classés par le modèle par rapport au nombre total d'exemples évalués.

La baisse de performance est alors définie par :

$$\Delta_t^{\text{perf}} = \text{Acc}_{t-1} - \text{Acc}_t \tag{6.3}$$

où Acc_t désigne l'exactitude du modèle à l'instant t. Une diminution supérieure au seuil τ_{perf} indique que le modèle ne s'adapte plus suffisamment au flux de données en cours, justifiant une mise à jour.

Déclenchement global

Le déclenchement d'une mise à jour à l'instant t est alors défini par une règle d'activation binaire T_t :

$$T_t = \begin{cases} 1 & \text{si } \Delta_t^{\text{ajout}} > \tau_{\text{ajout}} \\ & \text{ou } \Delta_t^{\text{suppr}} > \tau_{\text{suppr}} \\ & \text{ou } \Delta_t^{\text{rel}} > \tau_{\text{rel}} \\ & \text{ou } \Delta_t^{\text{perf}} > \tau_{\text{perf}} \\ 0 & \text{sinon} \end{cases} \tag{6.4}$$

Autrement dit, une mise à jour est déclenchée si au moins l'un des quatres condition est valide. Les seuils τ_{ajout}, τ_{suppr}, τ_{rel} et τ_{perf} peuvent être fixés de manière empirique (via validation croisée) ou ajustés dynamiquement à partir de moyennes glissantes.

6.2.6 Mise à jour incrémentale du modèle

Lorsque le déclenchement est activé à l'instant t (i.e. $T_t = 1$), le modèle est mis à jour afin d'intégrer les nouvelles données tout en conservant une mémoire des connaissances passées. L'enjeu est double : adapter le modèle à l'évolution des comportements récents, tout en évitant l'oubli des régularités antérieures, un phénomène bien connu sous le nom d'oubli catastrophique.

Données utilisées : La mise à jour repose sur deux composantes :

- **Les données récentes** (X_t, y_t) : elles représentent les interactions observées au temps t, potentiellement porteuses d'un changement de comportement ou d'un signal nouveau.
- **Une mémoire historique** $\mathcal{B}$: il s'agit d'un sous-ensemble contrôlé des données précédentes. Cette mémoire est maintenue à jour au fil du temps et contient un échantillon représentatif et équilibré, destiné à stabiliser le modèle et à préserver les acquis.

La taille de $\mathcal{B}$ est bornée (par exemple, à 20 % des données passées), et peut être mise à jour elle-même selon une stratégie de renouvellement. L'inclusion explicite de cette mémoire permet de pallier l'effet de dérive et d'assurer une continuité dans l'apprentissage.

Stratégie d'entraînement : Lorsque $T_t = 1$, le modèle M_{t-1} est mis à jour à l'aide des deux sources (X_t, y_t) et $(\mathcal{B}, y_{\mathcal{B}})$, selon la règle suivante :

$$M_t \leftarrow \text{Train}(M_{t-1},\ X_t \cup \mathcal{B},\ y_t \cup y_{\mathcal{B}}) \tag{6.5}$$

Lorsque le déclenchement est activé, le modèle est mis à jour de manière incrémentale, en intégrant les données récemment observées ainsi qu'un sous-ensemble de données antérieures conservées en mémoire. Cette mise à jour vise à adapter le modèle à l'évolution du comportement des apprenants, tout en préservant les connaissances acquises lors des périodes précédentes. Le processus d'apprentissage repose ici sur une dynamique cumulative, sans réinitialisation du modèle.

Réutilisation sans réentraînement

Si le déclenchement $T_t = 0$, aucune mise à jour n'est effectuée. Le modèle précédent est simplement réutilisé :

$$M_t = M_{t-1} \tag{6.6}$$

Le modèle est mis à jour par la règle (6.5) uniquement lorsque le déclenchement T_t est activé. Sinon, il est conservé en l'état. Cette stratégie garantit un apprentissage incrémental, réactif et parcimonieux, adapté aux environnements de flux éducatifs.

Mise à jour conditionnelle du modèle

À chaque pas de temps t, le modèle est réentraîné sur un lot combiné ($X_t \cup \mathcal{B}$) uniquement si le déclenchement T_t est activé. En l'absence de signal fort, le modèle est simplement reconduit.

6.2.7 Métriques et méthodes d'évaluation

L'évaluation du cadre proposé repose sur un ensemble de métriques couvrant deux dimensions : la *qualité prédictive* et la *frugalité du modèle*. Les mesures sont calculées à chaque pas de temps, selon une logique temporelle cohérente avec le flux des données observées.

Qualité prédictive

Nous évaluons la capacité du modèle à produire des prédictions fiables, en particulier sur les apprenants les plus à risque. Deux métriques sont mobilisées :

— L'exactitude globale (*Accuracy*

— La sensibility (*Sensitivity*)

Le suivi conjoint de ces deux indicateurs permet d'analyser les compromis entre couverture des cas critiques et performance globale, en tenant compte du déséquilibre fréquent des classes.

Frugalité du modèle

L'approche proposée vise à limiter les réentraînements du modèle, en les déclenchant uniquement lorsque des signaux de changement sont détectés. Le coût d'adaptation est mesuré par :

— **Nombre cumulé de mises à jour** :

$$\mathrm{MAJ}_T = \sum_{t=1}^{T} [T_t = 1] \tag{6.7}$$

où $T_t = 1$ indique qu'une mise à jour a été effectuée à l'instant t. Cette métrique permet de quantifier le nombre d'itérations d'apprentissage nécessaires, et constitue un indicateur direct de l'économie computationnelle réalisée.

Un bon compromis consiste à maintenir des performances proches de l'optimum tout en minimisant le nombre de mises à jour.

Méthodes comparatives

L'évaluation expérimentale repose sur la comparaison de trois configurations d'apprentissage incrémental, reflétant des stratégies contrastées en matière de déclenchement des mises à jour et de sélection des caractéristiques. Ces méthodes servent de points de référence pour évaluer la

pertinence du cadre proposé.

— **Réentraînement incrémental régulier (sans sélection)** : Avec $t = semaine$, le modèle est mis à jour à chaque semaine à partir de toutes les caractéristiques disponibles, sans déclenchement conditionnel ni filtrage. Cette stratégie constitue une baseline simple.

— **Réentraînement hebdomadaire avec sélection dynamique (CI-DFS)** : à chaque semaine, une sélection dynamique des caractéristiques est effectuée, mais le modèle est tout de même mis à jour systématiquement. Cette variante permet d'évaluer l'impact isolé de la sélection dynamique sur l'apprentissage incrémental.

— **Cadre incrémental avec déclenchement adaptatif (proposition)** : le modèle n'est mis à jour que si un ou plusieurs signaux de variation dans les données dépassent un seuil défini, conformément au mécanisme présenté précédemment. Il intègre à la fois une sélection dynamique des caractéristiques et un déclenchement conditionnel.

Ce protocole comparatif permet d'analyser l'effet combiné de la frugalité structurelle (via la réduction du nombre de variables) et de la frugalité computationnelle (via la réduction du nombre de mises à jour), en observant l'impact sur la performance prédictive et la sensibilité des apprenants à risque.

Les métriques sont comparées à travers trois configurations expérimentales contrastées, afin de mesurer les apports spécifiques du déclenchement adaptatif et de la sélection dynamique des caractéristiques.

6.3 Évaluation expérimentale

L'évaluation du cadre incrémental est menée sur deux jeux de données complémentaires : le **jeu OULAD** [91] et le **jeu MoocCN (XuetangX)**, déjà décrits dans le chapitre 3. Ces corpus, issus de contextes d'apprentissage en ligne distincts, permettent de tester la robustesse du modèle face à des temporalités et des dynamiques comportementales variées.

6.3.1 Jeu de données OULAD

L'expérimentation s'appuie sur le module `CCC-2014J`, qui regroupe 2,498 étudiants observés sur 38 semaines. À chaque semaine t, seules les données disponibles jusqu'à cet instant sont exploitées pour entraîner ou mettre à jour le modèle. Le jeu est structuré en flux hebdomadaires (X_t, y_t), composés de vecteurs d'activité binaires et d'une étiquette finale indiquant le décrochage.

Les caractéristiques catégorielles sont encodées en binaire (*one-hot*), les valeurs manquantes sont imputées ou neutralisées selon leur nature, et les interactions (accès aux ressources, forums, évaluations intermédiaires) sont agrégées par semaine afin de produire une représentation temporelle adaptée à l'apprentissage incrémental.

6.3.2 Jeu de données MoocCN (XuetangX)

Le second corpus provient de la plateforme chinoise XuetangX (KDD Cup 2015) et couvre un cours en ligne d'une durée de six semaines. Bien que plus court, ce jeu présente une dynamique pédagogique soutenue, avec des variations hebdomadaires marquées dans les activités proposées,

constituant ainsi un environnement pertinent pour évaluer la réactivité et la stabilité du modèle.

Les données, traitées selon la même procédure que pour OULAD, incluent des métadonnées sur les apprenants (âge, niveau d'études) et des traces d'interaction (consultation de vidéos, participation à des quiz, activité sur les forums). Elles sont encodées sous forme binaire et organisées en flux hebdomadaires pour permettre une évaluation cohérente du comportement du modèle dans un cadre temporel.

6.3.3 Modèle prédictif et paramétrage

Le modèle utilisé pour la prédiction hebdomadaire du décrochage est une **régression logistique en ligne** [163], choisie pour sa mise à jour incrémentale efficace et sa bonne interprétabilité, construite à l'aide de la bibliothèque `river` [110]. Le classifieur est entraîné par descente de gradient stochastique (SGD), avec un prétraitement par standardisation des variables.

```
Model = StandardScaler() | LogisticRegression(optimizer=SGD())
```

Table 6.2 – Configuration du modèle prédictif

Composant	**Description**
Type de modèle	Régression logistique en ligne
Bibliothèque	`river`
Optimiseur	SGD (descente de gradient stochastique)
Prétraitement	Standardisation des caractéristiques
Mise à jour	Incrémentale, semaine par semaine

Paramétrage du déclenchement

Dans la méthode à déclenchement adaptatif, quatre seuils sont utilisés pour activer la mise à jour du modèle. Ils diffèrent selon le jeu de données, afin de refléter la variabilité propre à chaque contexte.

Table 6.3 – Seuils de déclenchement utilisés pour les mises à jour incrémentales

Seuil	**OULAD**	**MoocCN**
τ_{ajout}	0,45	0,50
τ_{suppr}	0,25	0,25
τ_{rel}	0,38	0,42
τ_{perf}	0,23	0,30

Ces seuils ont été définis empiriquement pour équilibrer la performance prédictive et la frugalité computationnelle.

Paramètres de la sélection CI-DFS

La sélection dynamique des caractéristiques repose sur la méthode CI-DFS, avec des paramètres fixés de manière cohérente entre les deux jeux de données afin de garantir la comparabilité expérimentale.

TABLE 6.4 – Paramétrage de la méthode CI-DFS

Paramètre	**Valeur**
Fenêtre d'évaluation (w)	10
Seuil de redondance (r)	0,95
Seuil adaptatif de pertinence	25e centile historique
Sous-échantillonnage	20 % des données cumulées par semaine

6.4 Résultats

Les performances du cadre incrémental proposé sont évaluées sur les deux jeux de données présentés précédemment, à travers une comparaison avec deux approches de référence : le réentraînement hebdomadaire sans sélection, et le réentraînement hebdomadaire avec sélection dynamique CI-DFS. L'objectif est d'analyser la capacité du mécanisme de déclenchement à maintenir une bonne qualité prédictive tout en réduisant la charge computationnelle liée aux mises à jour.

6.4.1 Résultats sur le jeu de données MoocCN

Les figures 6.1 et 6.2 présentent l'évolution hebdomadaire de l'exactitude (accuracy) et de la sensibilité (sensitivity) pour les trois méthodes comparées, sur les six semaines du cours MoocCN.

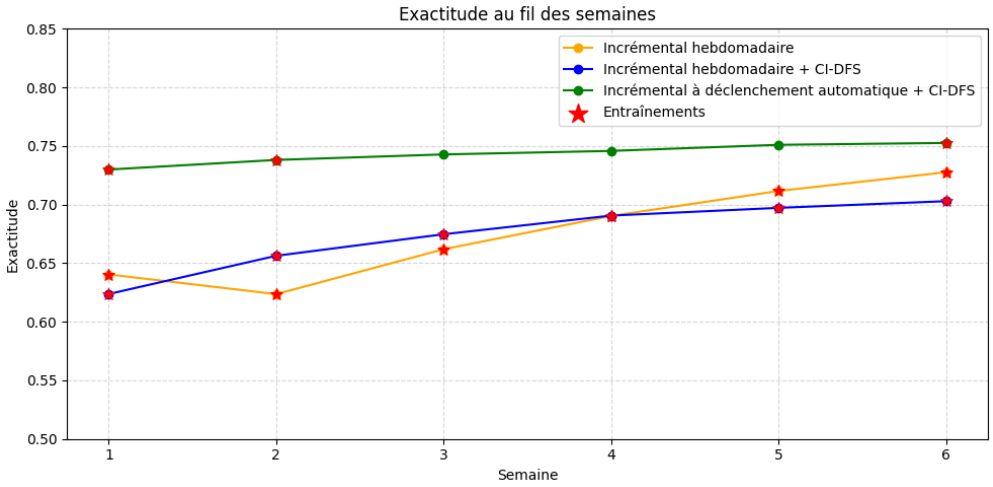

FIGURE 6.1 – Évolution de l'exactitude hebdomadaire (MoocCN)

Le modèle *Weekly Incremental*, qui effectue un apprentissage incrémental à chaque semaine sans filtrage ni sélection, présente une progression régulière de l'exactitude, passant de 63 % à la semaine 1 à 73 % à la semaine 6 (figure 6.1). Cependant, elle montre une chute marquée de la sensibilité dès la semaine 3 (figure 6.2) : la sensibilité chute brutalement à 56,4 % avant d'atteindre à 50,8 % en fin de parcours. Cette évolution suggère une spécialisation progressive du modèle sur les exemples majoritaires, au détriment de la détection des apprenants à risque.

L'ajout de la sélection dynamique de caractéristiques via CI-DFS (*Weekly Incremental + DFS*) produit des résultats contrastés. L'exactitude est légèrement inférieure (70,3 % à la semaine

6), mais la sensibilité est fortement améliorée dès les premières semaines, atteignant 88,7 % à la semaine 3, et se maintenant autour de 86 % jusqu'à la fin (figure 6.2). Cette stabilité suggère que la réduction régulière de l'espace des variables permet de préserver les signaux utiles à la détection du décrochage, sans les noyer dans un bruit croissant.

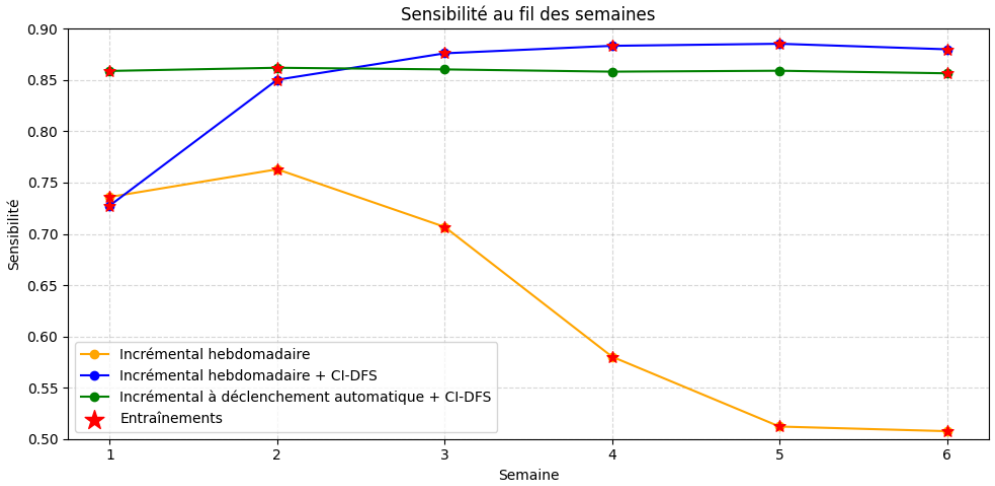

FIGURE 6.2 – Évolution de la sensibilité hebdomadaire (MoocCN)

Enfin, la méthode *Trigger Incremental + DFS*, qui combine sélection dynamique et déclenchement conditionnel, atteint le meilleur compromis. Elle conserve une exactitude élevée et stable, culminant à 75,2 % à la semaine 6 (figure 6.1), tout en maintenant une sensibilité moyenne de 85,6 %. Surtout, elle n'effectue que trois mises à jour ciblées (semaines 1, 2 et 6), concentrées autour des changements structurels les plus marqués dans les données. Ces résultats montrent que le déclenchement permet d'éviter les sur-ajustements aux variations non pertinentes tout en restant réactif aux signaux critiques.

Ce comportement illustre la capacité du cadre proposé à s'adapter de manière parcimonieuse, tout en conservant une forte sensibilité aux apprenants à risque, objectif central dans le contexte de prédiction du décrochage.

6.4.2 Résultats sur le jeu de données OULAD

Les figures 6.3 et 6.4 illustrent l'évolution hebdomadaire de l'exactitude (accuracy) et de la sensibilité (sensitivity) sur les 38 semaines du module sélectionné du jeu de données OULAD, pour les trois méthodes comparées.

La modèle *Weekly Incremental* montre une progression constante de l'exactitude au fil des semaines, atteignant 80,7 % en fin de module (figure 6.3). Toutefois, cette performance globale masque une dégradation marquée de la sensibilité : la sensibilité chute brutalement dès la semaine 3, passant sous les 50 %, pour atteindre seulement 33,6 % à la semaine 38 (figure 6.4). Le modèle devient progressivement aveugle aux apprenants décrocheurs, probablement en raison d'un surajustement aux comportements dominants dans la population.

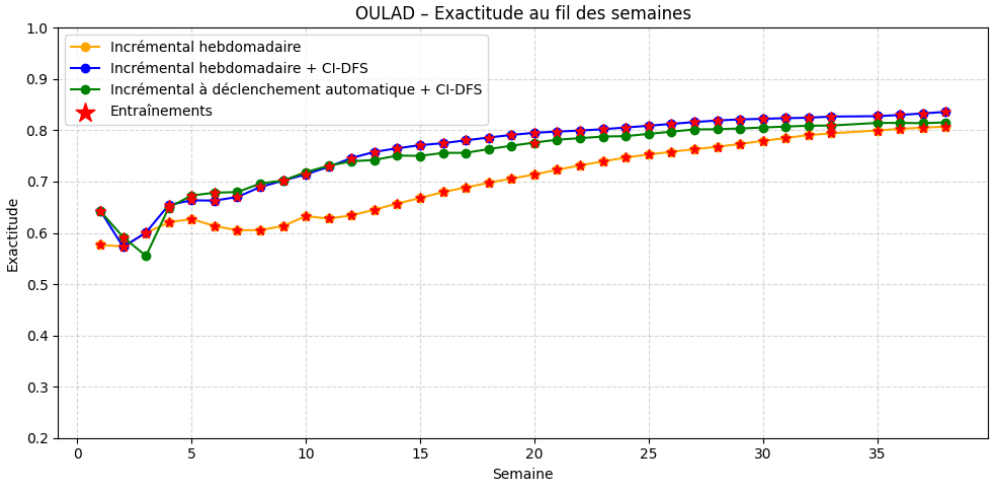

FIGURE 6.3 – Évolution de l'exactitude hebdomadaire (OULAD)

L'ajout d'une sélection dynamique des caractéristiques via CI-DFS (*Weekly Incremental + DFS*) permet de corriger cette dérive. L'exactitude progresse plus rapidement, atteignant 83,6 % à la fin, tandis que la sensibilité reste élevée tout au long du parcours, dépassant 95 % à partir de la semaine 15. Ce résultat montre que la sélection dynamique, en supprimant les caractéristiques inutiles ou bruitées, favorise une meilleure généralisation sur la classe minoritaire. Toutefois, cette approche impose un coût d'apprentissage constant, avec 38 mises à jour successives.

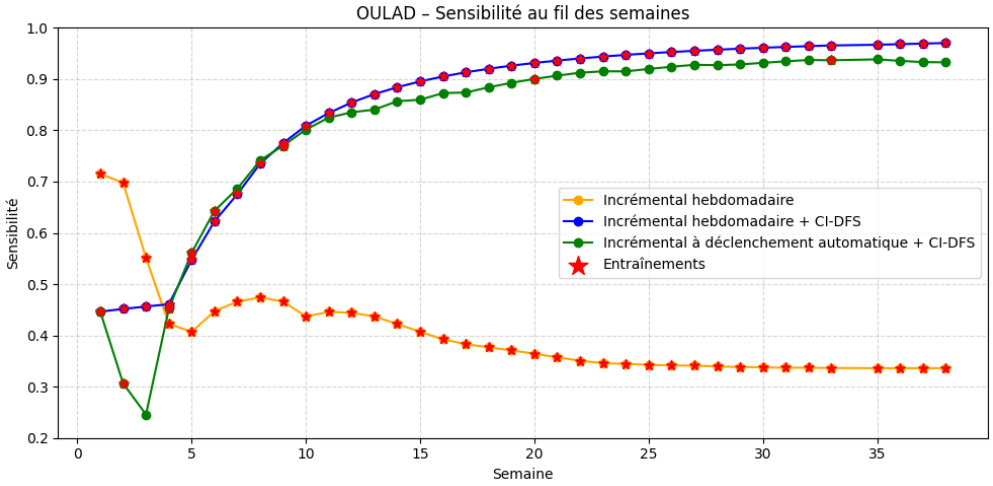

FIGURE 6.4 – Évolution de la sensibilité hebdomadaire (OULAD)

Le cadre *Trigger Incremental + DFS* offre un compromis nettement plus favorable. Il permet de maintenir une exactitude compétitive (81,5 %) tout en atteignant une sensibilité de 93,3 % à la semaine 38. Ce résultat est obtenu avec seulement 8 mises à jour du modèle (semaines 1, 2,

4, 5, 6, 9, 20 et 33), déclenchées de manière ciblée. On observe que ces points de mise à jour correspondent à des ruptures dans les courbes de performance des autres méthodes, suggérant que le déclenchement agit comme un mécanisme de régulation adaptative.

En somme, le cadre proposé permet de préserver la qualité du modèle sur le long terme, tout en réduisant significativement la charge computationnelle. Cette frugalité adaptative constitue un atout décisif dans un contexte éducatif à grande échelle, où les ressources de calcul sont souvent limitées et les données générées en continu.

6.4.3 Résumé comparatif

Le tableau 6.5 présente une synthèse des performances finales obtenues par les trois méthodes sur les deux jeux de données. Il rapporte l'exactitude (accuracy), la sensibilité (sensitivity) sur la classe des décrocheurs, ainsi que le nombre cumulé de mises à jour effectuées par le modèle tout au long de la période d'observation.

TABLE 6.5 – Résultats comparatifs finaux sur MoocCN / OULAD

Méthode	**Accuracy (%)**	**Sensibilité (%)**	**MAJ**
Weekly Incremental	72,7 / 80,7	50,8 / 33,6	6 / 38
Weekly Incremental + DFS	70,3 / **83,6**	**88,0** / **97,0**	6 / 38
Trigger Incremental + DFS	**75,2** / 81,5	85,6 / 93,3	**3** / **8**

Sur les deux corpus, la méthode *Weekly Incremental* obtient une exactitude modérée (72,7 % sur MoocCN et 80,7 % sur OULAD), mais échoue à détecter correctement les apprenants à risque, avec une sensibilité particulièrement faible sur OULAD (33,6 %). Cette stratégie systématique ne parvient donc pas à concilier apprentissage progressif et sensibilité aux cas minoritaires.

L'ajout d'une sélection dynamique des caractéristiques via CI-DFS améliore nettement la sensibilité (88,0 % sur MoocCN, 97,0 % sur OULAD), au prix d'un réentraînement hebdomadaire constant. Cette approche démontre la pertinence de la sélection pour limiter la dérive des modèles sur les classes minoritaires, mais reste coûteuse computationnellement.

La méthode *Trigger Incremental + DFS* combine les avantages des deux précédentes. Elle permet d'atteindre la meilleure exactitude sur MoocCN (75,2 %) et une sensibilité élevée sur les deux jeux (85,6 % / 93,3 %), tout en divisant par deux (MoocCN) ou par près de cinq (OULAD) le nombre de mises à jour nécessaires. Le déclenchement agit ainsi comme un filtre intelligent, ne sollicitant l'adaptation du modèle que lorsque cela est justifié par la dynamique des données.

Ces résultats confirment que le cadre proposé permet un apprentissage plus frugal (en termes de nombre de ré-entrainement) et mieux ciblé, sans compromettre la capacité du système à identifier les apprenants à risque. Il s'adapte aux contraintes des environnements éducatifs réels, en limitant les ressources mobilisées tout en préservant la qualité de service.

6.5 Discussion

Les résultats expérimentaux obtenus sur deux jeux de données contrastés, MoocCN (6 semaines, dynamique instable) et OULAD (38 semaines, progression plus linéaire), permettent

d'éclairer les apports et les limites du cadre incrémental proposé. Cette section discute les principales observations selon trois axes : performance prédictive, frugalité adaptative, et robustesse temporelle.

6.5.1 Performance prédictive et sensibilité aux décrocheurs

L'un des objectifs centraux du cadre proposé est de maintenir une sensibilité élevée à la classe des apprenants à risque. Les résultats confirment que les méthodes sans mécanisme de sélection (Weekly Incremental) perdent rapidement cette capacité, malgré une amélioration de l'exactitude globale. Cela confirme que, dans les jeux éducatifs déséquilibrés, l'exactitude ne constitue pas un indicateur suffisant, voire peut être trompeur, pour évaluer la qualité du modèle.

La combinaison du déclenchement et de la sélection dynamique (Trigger Incremental + DFS) permet de stabiliser la sensibilité à un niveau élevé (supérieur à 85 % sur les deux jeux), tout en maintenant une performance globale compétitive. Cette capacité à préserver la détection des cas minoritaires dans la durée constitue un atout majeur pour les systèmes de prédiction orientés vers l'intervention pédagogique.

6.5.2 Réduction du coût d'adaptation

L'autre contribution essentielle du cadre est sa capacité à limiter le nombre de réentraînements du modèle. Sur OULAD, la réduction du nombre de mises à jour est particulièrement significative : 8 mises à jour au lieu de 38, soit une économie de près de 80 % en coût d'apprentissage. Cette frugalité est obtenue sans recours à une logique fixe (par exemple, mise à jour toutes les n semaines), mais à travers une détection active des moments critiques dans la dynamique des données.

Dans le contexte des environnements virtuels d'apprentissage, où les flux de données sont continus et la capacité de calcul souvent limitée, cette approche offre un compromis efficace entre réactivité et maîtrise des ressources.

6.5.3 Robustesse face à la variabilité temporelle

Les résultats obtenus sur MoocCN montrent que le cadre est capable de fonctionner efficacement même sur des horizons temporels courts et instables. La stratégie de déclenchement s'adapte automatiquement à la structure du cours : elle déclenche deux mises à jour précoces (semaines 1 et 2) en réponse à des variations fortes, puis attend la dernière semaine (6) pour ajuster le modèle une dernière fois. Cela montre que le système n'est pas calibré sur une fréquence fixe, mais réagit aux signaux réels émis par les données.

Sur OULAD, le cadre identifie également les points de rupture pertinents (semaines 1-2, 4–6, 20, 33), qui correspondent à des changements significatifs dans la distribution ou la structure des interactions. Le modèle parvient ainsi à se maintenir à jour tout en évitant une sur-réaction aux fluctuations de faible amplitude.

6.5.4 Limites et perspectives

Malgré des résultats prometteurs, certaines limites propres au cadre incrémental proposé doivent être soulignées.

Tout d'abord, le mécanisme de déclenchement repose sur un ensemble de seuils fixés manuellement à partir d'une phase de validation exploratoire. Ce choix, bien que empirique et justifié dans le cadre de cette étude, limite la portée et la transférabilité du cadre à d'autres contextes. Une piste naturelle d'amélioration consisterait à automatiser l'ajustement de ces seuils, par exemple via une procédure d'optimisation adaptative ou de méta-apprentissage, permettant au système d'ajuster lui-même ses critères d'actualisation en fonction de l'évolution des données.

Ensuite, le cadre repose sur un classifieur unique, mis à jour de manière incrémentale au fil des semaines. Si cette stratégie s'est révélée efficace sur les cas étudiés, elle pourrait se montrer moins robuste face à des données fortement non stationnaires, ou dans des contextes multi-classes. Une extension envisageable serait d'intégrer des mécanismes de gestion de la dérive plus avancés, ou de combiner plusieurs modèles spécialisés selon les segments temporels ou les profils détectés.

Enfin, le cadre suppose que les données sont disponibles de manière séquentielle et régulièrement espacée dans le temps. Dans des contextes plus réalistes, les flux de données peuvent être irréguliers, bruités ou partiellement observés. Adapter l'apprentissage incrémental à de telles contraintes, tout en conservant des garanties de stabilité et de convergence, constitue une perspective importante pour la suite de ce travail.

Ces différentes limites soulignent les enjeux encore ouverts autour de l'apprentissage incrémental dans les environnements éducatifs, et mettent en évidence la nécessité de poursuivre l'exploration de solutions plus adaptatives et robustes.

6.6 Conclusion du chapitre

Ce chapitre a proposé un cadre d'apprentissage incrémental spécifiquement conçu pour les environnements éducatifs numériques, où les données sont générées progressivement au fil du temps, selon des rythmes souvent irréguliers. L'approche développée vise à répondre à un double enjeu : garantir une qualité prédictive élevée, notamment sur les apprenants à risque, tout en maîtrisant la charge computationnelle induite par la mise à jour continue des modèles.

Pour ce faire, le cadre repose sur deux principes complémentaires. D'une part, une *sélection dynamique des caractéristiques* (CI-DFS) est appliquée à chaque instant pour adapter l'espace de représentation aux données disponibles, en filtrant les caractéristiques redondantes ou peu informatives. D'autre part, un *mécanisme de déclenchement adaptatif* régule l'actualisation du modèle, en s'appuyant sur des signaux liés aux variations structurelles des données et aux pertes de performance prédictive. Le modèle n'est mis à jour que lorsque ces signaux dépassent certains seuils, évitant ainsi des réentraînements superflus.

L'expérimentation sur deux jeux de données contrastés, le MoocCN (6 semaines) et OULAD (38 semaines), a permis d'évaluer empiriquement la robustesse et la frugalité de l'approche. Les résultats montrent que la méthode *Trigger Incremental + DFS* parvient à maintenir une sensibilité élevée aux décrocheurs (jusqu'à 93 %), avec une exactitude comparable, voire supérieure, à celle des baselines entraînées systématiquement. Surtout, elle réduit drastiquement le nombre de mises à jour nécessaires : trois sur six semaines pour MoocCN, huit sur trente-huit pour OULAD. Cette réduction de la charge s'opère sans perte de performance, et souvent avec un gain en stabilité sur les indicateurs critiques.

Ces résultats valident le potentiel du cadre proposé pour des scénarios réels, où les modèles doivent s'adapter de manière continue mais raisonnée, en fonction de la dynamique propre à chaque parcours d'apprentissage. L'approche contribue à ouvrir un espace intermédiaire entre l'apprentissage en ligne strict (mise à jour à chaque observation) et les modèles statiques ou batch, en introduisant une forme de régulation intelligente de l'adaptation.

Ce chapitre constitue ainsi un socle méthodologique pour une gestion plus fine des flux de données éducatives, sensible à la variabilité temporelle et à la rareté des signaux utiles. Il ouvre la voie à l'intégration de mécanismes et de stratégies visant à optimiser l'adaptation des modèles tout en préservant l'efficacité computationnelle.

Conclusion générale

1 Retour sur la problématique et les objectifs

Cette thèse s'inscrit dans le champ de l'intelligence artificielle appliquée à l'éducation et s'attache à répondre à une question centrale : *faut-il améliorer un modèle en mobilisant toujours plus de données, ou concentrer l'effort sur les situations où il échoue à modéliser correctement ?*

L'objectif poursuivi est de concevoir une **approche ciblée et conditionnelle** pour améliorer la modélisation des apprenants dans les systèmes prédictifs, dans un contexte où la qualité et la quantité de données varient selon les profils et les environnements. Le cadre proposé vise à déclencher la mise à jour du modèle uniquement lorsque cela est nécessaire, en mobilisant sélectivement les données les plus pertinentes, tout en garantissant la **fiabilité et la frugalité** du processus d'apprentissage.

Face à ces limites, l'objectif poursuivi a été de proposer des méthodes permettant d'améliorer la **fiabilité**, la **frugalité** et la **capacité d'adaptation** des modèles prédictifs. Plus précisément, la thèse a visé à :

— Développer un cadre méthodologique de **prédiction frugale**, réduisant la consommation inutile de données tout en maintenant une performance élevée
— Introduire des mécanismes d'**estimation et d'exploitation de l'incertitude** pour diagnostiquer les erreurs et mieux calibrer la confiance des modèles
— Mettre en oeuvre une **sélection dynamique des caractéristiques**, permettant d'ajuster en continu les variables pertinentes au fil de l'évolution des traces d'apprentissage
— Concevoir des approches d'**apprentissage incrémental**, capables de mettre à jour les modèles en continu à partir des flux de données, sans perte d'information ni surcharge computationnelle.

2 Synthèse des contributions de la thèse

La thèse a apporté plusieurs contributions méthodologiques et expérimentales, organisées autour de quatre grands volets, correspondant aux chapitres principaux.

Un modèle frugal de prédiction (FEP)

Nous avons proposé le modèle *Frugal Early Prediction* (FEP), qui conditionne l'usage de données supplémentaires à une estimation d'incertitude. Contrairement aux approches multi-sources

systématiques, FEP exploite des données additionnelles uniquement lorsqu'elles s'avèrent nécessaires, réduisant ainsi la consommation totale de données de 27 % tout en améliorant l'accuracy de 7,3 % en moyenne. Cette contribution illustre que « plus de données » n'est pas toujours synonyme de « meilleures prédictions », et constitue le premier modèle explicitement conçu pour la frugalité en éducation.

Une métrique temporelle de l'incertitude (TWPUS)

La thèse a proposé le *Time-Weighted Persistent Uncertainty Score* (TWPUS), permettant de quantifier la persistance et l'évolution de l'incertitude dans le temps. À partir de ce signal, trois profils interprétables de trajectoires de confiance ont été identifiés : les *persistants incertains*, les *résolutifs* et les *émergents incertains*. Cette typologie, validée empiriquement, fournit un outil de diagnostic pour relier les erreurs du modèle à la structure des données disponibles, et offre une nouvelle manière de rendre les prédictions explicables en éducation.

Une sélection dynamique et sensible au déséquilibre (CI-DFS)

Nous avons introduit l'algorithme *Class-Imbalanced Dynamic Feature Selection* (CI-DFS), qui combine information mutuelle pondérée, seuils adaptatifs et gestion du drift temporel. Validé sur les jeux de données **OULAD** et **MoocCN**, CI-DFS améliore la détection des apprenants en décrochage avec une sensibilité moyenne de **0,58 sur OULAD** et **0,84 sur MoocCN**, soit un gain relatif de **+28%** et **+87%** par rapport à SAOLA [169]. Dans le même temps, le nombre de caractéristiques sélectionnées est réduit de plus de **70%** (en moyenne **204 contre 386** sur OULAD et **116 contre 358** sur MoocCN).

Sur le plan méthodologique, CI-DFS constitue une extension inédite des approches de sélection dynamique de caractéristiques, jusque-là peu adaptées aux contextes éducatifs déséquilibrés, en offrant un compromis robuste entre **sensibilité, frugalité et adaptabilité temporelle**.

Un déclenchement incrémental guidé par la sélection dynamique

Nous avons proposé un cadre d'apprentissage incrémental dans lequel la mise à jour du modèle n'est plus réalisée à intervalles fixes, mais déclenchée uniquement lorsque les signaux issus de la sélection dynamique des caractéristiques l'exigent. Ces signaux incluent des changements significatifs dans l'ensemble des caractéristiques retenues, une dérive de leur pertinence, ou une dégradation mesurable des performances.

Appliqué au jeu de données **OULAD**, ce cadre permet de **réduire de près de 40% le nombre de mises à jour nécessaires** tout en maintenant une **exactitude moyenne de 0,65** et une **sensibilité de 0,58**. Sur **MoocCN**, la même approche atteint une **exactitude de 0,69** et une **sensibilité de 0,84**, confirmant la capacité du modèle à s'adapter aux flux de données courts tout en préservant la performance. Cette stratégie d'actualisation conditionnelle limite les recalculs redondants et réduit la consommation de ressources, tout en garantissant une exactitude prédictive comparable à celle des mises à jour systématiques.

Ainsi, l'apprentissage incrémental guidé par la sélection dynamique s'inscrit pleinement dans une logique de **frugalité computationnelle** et d'**adaptation continue aux dynamiques éducatives**.

3 Limites de la recherche

Si les résultats obtenus sont encourageants, plusieurs limites doivent être soulignées :

— Dans le chapitre 4 consacré à la **prédiction frugale (FEP)**, les expérimentations ont porté sur un nombre restreint de sources additionnelles (principalement l'historique de cours), sans mobilisation de données plus diversifiées.

— Dans le même chapitre, l'évaluation de l'incertitude s'est limitée à des mesures probabilistes classiques (variance, entropie) appliquées à des modèles supervisés standards, sans recours à des approches bayésiennes ni à des architectures de réseaux profonds.

— Dans le chapitre 5, dédié à la **sélection dynamique des caractéristiques (CI-DFS)**, les tests ont été menés uniquement dans un cadre de classification binaire, avec des paramètres fixés de manière empirique, sans analyse de leur extension à d'autres tâches ni de leur ajustement automatique.

— Dans le chapitre 6, traitant du **déclenchement incrémental**, la méthode repose sur des seuils dont la détermination reste empirique et dépendante des jeux de données, ce qui limite la fiabilité et la généralisabilité du déclenchement.

— Enfin, la démarche adoptée est restée strictement **scientifique et méthodologique**, sans approfondissement des dimensions éthiques, pédagogiques ou organisationnelles de l'IA éducative, pourtant essentielles pour la mise en œuvre concrète de ces travaux.

4 Perspectives et pistes futures

Les travaux présentés ouvrent plusieurs prolongements possibles, que l'on peut regrouper en deux axes complémentaires.

Axe scientifique

Les perspectives scientifiques ouvertes par cette recherche s'articulent autour de trois axes complémentaires.

Un premier prolongement consiste à approfondir la question de l'incertitude dans les modèles prédictifs éducatifs. Au-delà de la simple mesure de confiance associée aux prédictions, il s'agira d'examiner des métriques de calibration probabiliste plus fines, afin d'évaluer la justesse des probabilités produites par le modèle. Une distinction explicite entre l'incertitude épistémique, liée au manque de connaissance du modèle, et l'incertitude aléatoire, inhérente à la variabilité des données, permettrait d'affiner l'interprétation des erreurs et de relier ces signaux aux approches d'explicabilité. Ces travaux visent à concevoir des modèles capables non seulement de prédire avec précision, mais aussi de justifier leurs décisions dans un cadre éducatif réel.

Un second axe de recherche porte sur le renforcement des stratégies d'apprentissage incrémental et frugal. L'enjeu principal consiste à intégrer une gestion plus fine de la dérive conceptuelle, c'est-à-dire des changements progressifs dans la relation entre les caractéristiques observées et les comportements d'apprentissage. Une telle évolution nécessiterait la détection automatique de ces dérives, suivie d'une réinitialisation partielle ou conditionnelle du modèle pour maintenir la

cohérence temporelle des prédictions. Par ailleurs, l'exploration de modèles légers et compressés, permettrait d'accroître encore la frugalité computationnelle, en réduisant les coûts d'entraînement tout en préservant la performance et l'interprétabilité.

Enfin, un troisième axe vise à élargir le champ expérimental au-delà des tâches de classification binaire. L'évaluation du cadre proposé sur des tâches de classification multi-classes, de régression continue ou d'apprentissage non supervisé constituerait une étape essentielle pour tester la généralisabilité et la transférabilité des méthodes développées. Une telle extension offrirait une compréhension plus fine du comportement des modèles dans des contextes éducatifs variés, allant de l'apprentissage formel et structuré à des environnements ouverts et auto-dirigés, où la dynamique d'adaptation et la robustesse des modèles prennent toute leur importance.

Axe applicatif

Sur le plan applicatif, plusieurs prolongements sont envisagés. Un premier enjeu consiste à exploiter les profils d'incertitude identifiés par le modèle pour guider la personnalisation des parcours d'apprentissage et adapter les interventions pédagogiques de manière ciblée. Ces profils pourraient permettre d'ajuster l'intensité et la nature des recommandations en fonction du degré de confiance du modèle, favorisant ainsi une prise de décision plus contextualisée et explicable.

Un second axe d'application porte sur l'évaluation de l'équité et de la fiabilité des prédictions, en analysant plus finement les cas de faux positifs et de faux négatifs. L'objectif est de détecter et de limiter les biais potentiels dans l'accompagnement des apprenants, en garantissant une qualité de service homogène, quels que soient le volume ou la nature des données disponibles pour chaque profil.

Enfin, le déploiement des approches proposées dans des environnements réels tels que des plateformes LMS ou des environnements virtuels d'apprentissage constitue une étape clé pour en mesurer la pertinence en conditions opérationnelles. Cette expérimentation permettra d'évaluer leur impact concret sur la réussite et la persévérance des apprenants, tout en fournissant un retour d'expérience précieux sur l'intégration de modèles frugaux et explicables dans des systèmes éducatifs à grande échelle.

En somme, cette thèse a proposé un ensemble de méthodes pour répondre aux défis liés à la variabilité et à la dynamique des données éducatives. En articulant frugalité, sélection dynamique des caractéristiques, estimation de l'incertitude et apprentissage incrémental déclenché, elle a posé les bases d'un cadre original visant à améliorer la fiabilité et la réactivité des modèles prédictifs dans les environnements d'apprentissage. Si certaines limites demeurent, les contributions apportées offrent des avancées à la fois méthodologiques et expérimentales, et ouvrent la voie à de futures recherches destinées à transformer ces approches en pratiques concrètes au service des apprenants.

Bibliographie

[1] Khan/tutoring-accuracy-dataset : This repository hosts the paper "LLM Based Math Tutoring : Challenges and Dataset", along with the accompanying dataset. It explores the performance and challenges of Large Language Models (LLMs) in math tutoring scenarios, providing a benchmark dataset for evaluating LLM accuracy in educational contexts.

[2] Youssef Abadade, Anas Temouden, Hatim Bamoumen, Nabil Benamar, Yousra Chtouki, and Abdelhakim Senhaji Hafid. A Comprehensive Survey on TinyML. *IEEE Access*, 11 :96892–96922, 2023.

[3] Mohammad Hassan Abbasi, Candidate ; Gholam, Ali Montazer, ; Fatemeh Ghorbani, Zahra Alipour, and Gholam Ali Montazer. Categorizing E-Learner Attributes in Personalized E-learning Environments : A Systematic Literature Review. *Z. Categorizing E-Learner Attributes in Personalized E-learning Environments : A Systematic Literature Review. Interdiscip J Virtual Learn Med Sci*, 12(1) :1–21, 2021.

[4] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning : Techniques, applications and challenges. *Information Fusion*, 76 :243–297, 12 2021.

[5] L M Abu-Zohair. Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(1), 2019.

[6] Abir Abyaa, Mohammed Khalidi Idrissi, and Samir Bennani. Learner modelling : systematic review of the literature from the last 5 years. *Educational Technology Research and Development*, 67(5) :1105–1143, 10 2019.

[7] Edmund F Agyemang, Joseph A. Mensah, Obu-Amoah Ampomah, Louis Agyekum, Justice Akuoko-Frimpong, Amma Quansah, and Oluwaferanmi M. Akinlosotu. Predicting Students' Academic Performance Via Machine Learning Algorithms : An Empirical Review and Practical Application. *Computer Engineering and Intelligent Systems*, 15(1) :86, 9 2024.

[8] Esmael Ahmed. Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, 2024, 2024.

[9] Noura Al Nuaimi and Mohammad M. Masud. Online streaming feature selection with incremental feature grouping. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 10(4), 7 2020.

[10] Ali Al-Zawqari, Dries Peumans, and Gerd Vandersteen. A flexible feature selection approach for predicting students' academic performance in online courses. *Computers and Education : Artificial Intelligence*, 3 :100103, 1 2022.

[11] Rahaf Alamri and Basma Alharbi. Explainable Student Performance Prediction Models : A Systematic Review. *IEEE Access*, 9 :33132–33143, 2021.

[12] Balqis Albreiki, Nazar Zaki, and Hany Alashwal. A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences 2021, Vol. 11, Page 552*, 11(9) :552, 9 2021.

[13] Essa Alhazmi and Abdullah Sheneamer. Early Predicting of Students Performance in Higher Education. *IEEE Access*, 11 :27579–27589, 2023.

[14] Bander Allogmany and Darsana Josyula. An approach to dealing with incremental concept drift in personalized learning systems. *Proceedings - 2022 IEEE 4th International Conference on Cognitive Machine Intelligence, CogMI 2022*, pages 139–147, 2022.

[15] Noura AlNuaimi, Mohammad Mehedy Masud, Mohamed Adel Serhani, and Nazar Zaki. Streaming feature selection algorithms for big data : A survey. *Applied Computing and Informatics*, 18(1-2) :113–135, 1 2022.

[16] Ebedia Hilda Am, Indriana Hidayah, and Sri Suning Kusumawardani. A Literature Review of Knowledge Tracing for Student Modeling : Research Trends, Models, Datasets, and Challenges. *Journal of Information Technology and Computer Science*, 6(2) :183–194, 10 2021.

[17] Maxime Amisse, Mélissa Faur, Lucie Gonard, and André Orcesi. Promouvoir des modèles d'intelligence artificielle frugale pour et par les politiques publiques. Technical report, 2024.

[18] Shuang An, Xingyu Guo, Changzhong Wang, Ge Guo, and Jianhua Dai. A soft neighborhood rough set model and its applications. *Information Sciences*, 624 :185–199, 5 2023.

[19] Vitor Cirilo Araujo Santos, Lucas Cardoso, and Ronnie Alves. The quest for the reliability of machine learning models in binary classification on tabular data. *Scientific Reports*, 13(1) :1–11, 12 2023.

[20] Soufiane Ardchir, Amina Rachik, Youssef Ouassit, Reda Moulouki, and Mohamed Azzouazi. An Efficient Feature Selection Approach for MOOCs Dropout Prediction. pages 587–599, 2022.

[21] ArgaLudovic, BélorgeyFrançois, BraudArnaud, CarbouRomain, CharbonniaudNathalie, ColomesCatherine, Delphin-PoulatLionel, ExcoffierDavid, FauchéChristel, GeorgeThomas, GuyardFrédéric, HassanThomas, LampinQuentin, LemaireVincent, NodetPierre, PiotrowskiPawel, SapiejewskiKrzysztof, Sirvent-HienEmilie, and TosicTamara. Frugal AI : Introduction, Concepts, Development and Open Questions. *ACM SIGKDD Explorations Newsletter*, 27(1) :72–111, 7 2025.

[22] Amal Asselman, Mohamed Khaldi, and Souhaib Aammou. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, 31(6) :3360–3379, 2023.

[23] Fiddin Yusfida A'La and Berliana Kusuma Riasti. Optimizing Machine Learning Models for Predicting User Engagement in Online Learning Platforms : The Role of Feature Selection. In *2024 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 157–161. IEEE, 10 2024.

[24] Ryan S Baker. Challenges for the Future of Educational Data Mining : The Baker Learning Analytics Prizes. Technical report, 2019.

[25] Ryan S J D Baker. Mining Data for Student Models. Technical report, 2010.

[26] V. Balachandar and K. Venkatesh. A multi-dimensional student performance prediction model (MSPP) : An advanced framework for accurate academic classification and analysis. *MethodsX*, 14 :103148, 6 2025.

[27] David Bañeres, M. Elena Rodríguez-González, Ana Elena Guerrero-Roldán, and Pau Cortadas. An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, 20(1), 12 2023.

[28] Firas Bayram, Bestoun S. Ahmed, and Andreas Kassler. From concept drift to model degradation : An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245 :108632, 6 2022.

[29] A Ben Soussia, C Labba, A Roussanaly, and A Boyer. Time-dependent metrics to assess performance prediction systems. *International Journal of Information and Learning Technology*, 39(5) :451–465, 2022.

[30] Amal Ben Soussia, Célina Treuillier, Azim Roussanaly, and Anne Boyer. Learning Profiles to Assess Edu-cational Prediction Systems. pages 41–52, 2022.

[31] Johannes Berens, Kerstin Schneider, Simon Görtz, Simon Oster, and Julian Burghoff. Early Detection of Students at Risk-Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods †. Technical Report 3, 2019.

[32] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight Uncertainty in Neural Networks. *32nd International Conference on Machine Learning, ICML 2015*, 2 :1613–1622, 5 2015.

[33] Carlo Blundo, Vincenzo Loia, and Francesco Orciuoli. A Time-Aware Approach for MOOC Dropout Prediction Based on Rule Induction and Sequential Three-Way Decisions. *IEEE Access*, 11 :113189–113198, 2023.

[34] Marcelo C. Borba, Aparecida Santana de Souza Chiari, and Helber Rangel Formiga Leite de Almeida. Interactions in virtual learning environments : new roles for digital technology. *Educational Studies in Mathematics*, 98(3) :269–286, 7 2018.

[35] Scott Bulfin, Luci Pangrazio, and Neil Selwyn. Making 'MOOCs' : The construction of a new digital higher education within news media discourse. *The International Review of Research in Open and Distributed Learning*, 15(5) :290–305, 10 2014.

[36] Vanessa Morgado Madeira Caldeira, Vera Lucia Kochen, Iracema Cristina Fernandes, Pollyana Saluci Esquincalha Martins, and Cleonice Lucimar Ribeiro Nunes. DE GIZ E LOUSA AO METAVERSO : A REVOLUÇÃO DIGITAL NA EDUCAÇÃO. *ARACÊ*, 7(3) :11499–11513, 3 2025.

[37] Laura Caprara and Cataldo Caprara. Effects of virtual learning environments : A scoping review of literature. *Education and Information Technologies*, 27(3) :3683–3722, 4 2022.

[38] Gabriella Casalino, Giovanna Castellano, and Corrado Mencar. Incremental and Adaptive Fuzzy Clustering for Virtual Learning Environments Data Analysis. In *Proceedings of the International Conference on Information Visualisation*, volume 2019-July, pages 382–387. Institute of Electrical and Electronics Engineers Inc., 7 2019.

[39] Wilson Chango, Rebeca Cerezo, and Cristóbal Romero. Predicting academic performance of university students from multi-sources data in blended learning. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 12 2019.

[40] Edward Chen, Han Bao, and Nam Dinh. Evaluating the reliability of machine-learning-based predictions used in nuclear power plant instrumentation and control systems. *Reliability Engineering & System Safety*, 250 :110266, 10 2024.

[41] Jing Chen, Bei Fang, Hao Zhang, and Xia Xue. A systematic review for MOOC dropout prediction from the perspective of machine learning. *Interactive Learning Environments*, 32(5) :1642–1655, 5 2024.

[42] Yanfei Chen and Sanmin Liu. A novel learning method for feature evolvable streams. *Evolving Systems 2024 15 :5*, 15(5) :1733–1751, 5 2024.

[43] Youngduck Choi, Youngnam Lee, Junghyun Cho, Jineon Baek, Byungsoo Kim, Jonghun Kim, Younghoon Cha, Dongmin Shin, Chanhyuk Bae, Hyeongu Kim, Joon Kim, and Jeonghee Heo. Ednet : A large-scale hierarchical dataset in education. In *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2020)*, pages 522–528. International Educational Data Mining Society, 2020.

[44] Stephane Cholet and Emmanuel Biabiany. A Framework for Frugal Supervised Learning with Incremental Neural Networks. *Applied Sciences*, 13(9) :5489, 4 2023.

[45] Konstantina Chrysafiadi and Maria Virvou. Student modeling approaches : A literature review for the last decade, 9 2013.

[46] K T Chui, D C L Fung, M D Lytras, and T M Lam. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior*, 107 :105584, 2020.

[47] Jae Young Chung and Sunbok Lee. Dropout early warning systems for high school students using machine learning. *Child Youth Serv Rev*, 96 :346–353, 1 2019.

[48] Anthony Corso, David Karamadian, Romeo Valentin, Mary Cooper, and Mykel J. Kochenderfer. A Holistic Assessment of the Reliability of Machine Learning Systems. 7 2023.

[49] E B Costa, B Fonseca, M A Santana, F F de Araújo, and J Rego. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73 :247–256, 2017.

[50] National Research Council. How People Learn : Brain, Mind, Experience, and School : Expanded Edition. 8 1999.

[51] Thomas M. Cover and Joy A. Thomas. Elements of Information Theory. *Elements of Information Theory*, pages 1–748, 4 2005.

[52] Ian Covert, Wei Qiu, Mingyu Lu, Nayoon Kim, Nathan White, and Su-In Lee. Learning to Maximize Mutual Information for Dynamic Feature Selection. 1 2023.

[53] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey : Defying forgetting in classification tasks. 2021.

[54] Oriane Dermy, Azim Roussanaly, and Anne Boyer. Using Behavioral Primitives to Model Students' Digital Behavior. *Procedia Computer Science*, 207 :2444–2453, 1 2022.

[55] Sebastien Deschamps and Hichem Sahbi. Frugal Reinforcement-based Active Learning. 12 2022.

[56] S. Gayathri Devi and M. Sabrigiriraj. Swarm intelligent based online feature selection (OFS) and weighted entropy frequent pattern mining (WEFPM) algorithm for big data analysis. *Cluster Computing*, 22 :11791–11803, 9 2019.

[57] Pierre Dillenbourg and John Self. A Framework for Learner Modelling. *Interactive Learning Environments*, 2(2) :111–137, 6 1992.

[58] Zhangling Duan, Tianci Li, Xingyu Wu, Zhaolong Ling, Jingye Yang, and Zhaohong Jia. Fair Streaming Feature Selection. 6 2024.

[59] Elizabeth L. Eisenstein. The Printing Press as an Agent of Change. *The Printing Press as an Agent of Change*, 9 1980.

[60] Emrah Emirtekin. A MODEL PROPOSAL WITH DEEP LEARNING : ON STUDENT SUCCESS. *Journal of Modern Technology and Engineering*, pages 130–139, 8 2024.

[61] M Evchenko, J Vanschoren, H H Hoos, M Schoenauer, and M Sebag. Frugal Machine Learning, 2021.

[62] Christian Fachola, Agustín Tornaría, Paola Bermolen, Germán Capdehourat, Lorena Etcheverry, and María Inés Fariello. Federated Learning for Data Analytics in Education. *Data 2023, Vol. 8, Page 43*, 8(2) :43, 2 2023.

[63] Miaomiao Fei and Dit-Yan Yeung. Temporal models for predicting student dropout in massive open online courses. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 256–263. IEEE, 2015.

[64] Wenzheng Feng, Jie Tang, and Tracy Xiao Liu. Understanding Dropouts in MOOCs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01) :517–524, 7 2019.

[65] Yaoyao Fiona Zhao, Jiarui Xie, and Lijun Sun. On the Data Quality and Imbalance in Machine Learning-based Design and Manufacturing—A Systematic Review. *Engineering*, 45 :105–131, 2 2025.

[66] Zenun Kastrati Fisnik Dalipi, Ali Shariq Imran. *MOOC Dropout Prediction Using Machine Learning Techniques : Review and Research Challenges*. IEEE, 2018.

[67] Regol Florence, Schwinn Leo, Sprague Kyle, Coates Mark, and Markovich Thomas. When to retrain a machine learning model. 5 2025.

[68] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning, 6 2016.

[69] María Gómez Gallego, Alfonso Palazón Perez de los Cobos, and Juan Cándido Gómez Gallego. Identifying Students at Risk to Academic Dropout in Higher Education. *Education Sciences 2021, Vol. 11, Page 427*, 11(8) :427, 8 2021.

[70] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A Survey of Uncertainty in Deep Neural Networks. *Artificial Intelligence Review*, 56 :1513–1589, 7 2021.

[71] Katerina Georgouli. Virtual learning environments - An overview. *Proceedings - 2011 Panhellenic Conference on Informatics, PCI 2011*, pages 63–67, 2011.

[72] Ilie Gligorea, Marius Cioca, Romana Oancea, Andra Teodora Gorski, Hortensia Gorski, and Paul Tudorache. Adaptive Learning Using Artificial Intelligence in e-Learning : A Literature Review. *Education Sciences 2023, Vol. 13, Page 1216*, 13(12) :1216, 12 2023.

[73] Jibing Gong, Yao Wan, Ye Liu, Xuewen Li, Yi Zhao, Cheng Wang, Yuting Lin, Xiaohan Fang, Wenzheng Feng, Jingyi Zhang, and Jie Tang. Reinforced MOOCs Concept Recommendation in Heterogeneous Information Networks. *ACM International Conference Proceeding Series*, 37(4) :349–354, 5 2023.

[74] Thomas Grote, Konstantin Genin, and Emily Sullivan. Reliability in Machine Learning. *Philosophy Compass*, 19(5) :e12974, 5 2024.

[75] Sachini Gunasekara and Mirka Saarela. Explainability in Educational Data Mining and Learning Analytics : An Umbrella Review. pages 887–892, 2024.

[76] Suraya Hashim, Muhd Khaizer Omar, Habibah Ab Jalil, and Nurfadhlina Bt Mohd Sharef. Trends on Technologies and Artificial Intelligence in Education for Personalized Learning : Systematic Literature Review. *Article in International Journal of Academic Research in Progressive Education and Development*, 2022.

[77] Jiangpeng He, Runyu Mao, Zeman Shao, and Fengqing Zhu. Incremental Learning In Online Scenario. 2020.

[78] Yuhang He, Yingjie Chen, Yuhan Jin, Songlin Dong, Xing Wei, and Yihong Gong. Dynamic Feature Space Self-Organization for Online Task-Free Class Incremental Learning. 2024.

[79] Andrew D. Ho, Justin Reich, Sergiy O. Nesterko, Daniel T. Seaton, Tommy Mullaney, James Waldo, and Isaac Chuang. Harvardx and mitx : The first year of open online courses. Technical report, HarvardX and MITx Working Paper No. 1, 2014.

[80] Danial Hooshyar, Gustav Šír, Yeongwook Yang, Eve Kikas, Raija Hämäläinen, Tommi Kärkkäinen, Dragan Gašević, and Roger Azevedo. Towards responsible AI for education : Hybrid human-AI to confront the Elephant in the room. 4 2025.

[81] Hao Huang, Lihjen Jew, and Dandan Qi. Take a MOOC and then drop : A systematic review of MOOC engagement pattern and dropout factor, 4 2023.

[82] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning : an introduction to concepts and methods. *Machine Learning*, 110(3) :457–506, 3 2021.

[83] Sergio Iván Ramírez. Learner Models for MOOC in a Lifelong Learning Context : A Systematic Literature Review. 2021.

[84] Fatema Tuz Johora, Md Nahid Hasan, Aditya Rajbongshi, Md Ashrafuzzaman, and Farzana Akter. An explainable AI-based approach for predicting undergraduate students academic performance. *Array*, 26 :100384, 7 2025.

[85] Isam Kareem Thajeel, Khairulmizam Samsudin, Shaiful Jahari Hashim, and Fazirulhisyam Hashim. Dynamic feature selection model for adaptive cross site scripting attack detection using developed multi-agent deep Q learning model. 2023.

[86] Marzieh Karimi-Haghighi, Carlos Castillo, Davinia Hernández-Leo, and Veronica Moreno Oliver. Predicting Early Dropout : Calibration and Algorithmic Fairness Considerations. 2021.

[87] Judy Kay, Kathryn Bartimote, Kirsty Kitto, Bob Kummerfeld, Danny Liu, and Peter Reimann. Enhancing learning by Open Learner Model (OLM) driven data design. *Computers and Education : Artificial Intelligence*, 3 :100069, 1 2022.

[88] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(13) :3521–3526, 1 2017.

[89] Kenneth R. Koedinger, John C. Stamper, Elizabeth A. McLaughlin, and Theodore Nixon. Algebra i 2005–2006 data set from cognitive tutor. Pittsburgh Science of Learning Center DataShop (PSLC), 2010. Dataset used in the KDD Cup 2010 Educational Data Mining Challenge.

[90] Maria Kurday and Gergana Vladova. Learning analytics and educational data mining applications : bibliometric and ChatGPT-based analysis of research publications from 2014 to 2023. *Journal of Computers in Education*, pages 1–36, 6 2025.

[91] J Kuzilek, M Hlosta, and Z Zdrahal. Open University Learning Analytics dataset. *Scientific Data*, 4(1), 2017.

[92] Chahrazed Labba and Anne Boyer. When and How to Update Online Analytical Models for Predicting Students Performance ? page 10, 2022.

[93] Chahrazed Labba and Anne Boyer. Towards An Online Incremental Approach to Predict Students Performance. 5 2024.

[94] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Advances in Neural Information Processing Systems*, 2017-December :6403–6414, 12 2016.

[95] Yassamine Lala Bouali, Olfa Ben Ahmed, Abbas Bradai, and Smaine Mazouzi. Frugal AI : The New Era of Sustainable Deep Learning. 11 2024.

[96] Justin Leo and Jugal Kalita. Survey of continuous deep learning methods and techniques used for incremental learning. *Neurocomputing*, 582 :127545, 5 2024.

[97] Kewen Li, Mingxiao Yu, Lu Liu, Timing Li, and Jiannan Zhai. Feature Selection Method Based on Weighted Mutual Information for Imbalanced Data. *https ://-doi.org/10.1142/S0218194018500341*, 28(8) :1177–1194, 8 2018.

[98] Na Li, Xiaojun Zhang, Maria Limniou, and Youmin Xi. Meaning-making in virtual learning environment enabled educational innovations : a 13-year longitudinal case study. *Interactive Learning Environments*, 32(1) :168–182, 1 2024.

[99] Shu Li. The Historical Evolution of Educational Technology : from the Printing Press to Online Education. Technical report.

[100] Yu Li and Yusheng Cheng. Streaming feature selection for multi-label data with dynamic sliding windows and feature repulsion loss. *Entropy*, 21(12), 12 2019.

[101] Jing Liang, Ryan Hare, Tianyu Chang, Fangli Xu, Ying Tang, Fei Yue Wang, Shimeng Peng, and Mingyu Lei. Student Modeling and Analysis in Adaptive Instructional Systems. *IEEE Access*, 10 :59359–59372, 2022.

[102] Chin-Yang Lin, Chung-Ho Wu, Chang-Han Yeh, Shih-Han Yen, Cheng Sun, and Yu-Lun Liu. FrugalNeRF : Fast Convergence for Extreme Few-shot Novel View Synthesis without Learned Priors, 2025.

[103] Helai Liu, Mao Mao, Xia Li, and Jia Gao. Model interpretability on private-safe oriented student dropout prediction. *PLOS ONE*, 20(3) :e0317726, 3 2025.

[104] Aarón López-García, Olga Blasco-Blasco, Marina Liern-García, and Sandra E. Parada-Rico. Early detection of students' failure using Machine Learning techniques. *Operations Research Perspectives*, 11, 12 2023.

[105] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under Concept Drift : A Review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12) :2346–2363, 4 2020.

[106] Simone Magistri, Tomaso Trinci, Albin Soutif-Cormerais, Joost Van De Weijer, and Andrew D Bagdanov. ELASTIC FEATURE CONSOLIDATION FOR COLD START EXEMPLAR-FREE INCREMENTAL LEARNING. 2024.

[107] Ananth Mahadevan and Michael Mathioudakis. Cost-aware retraining for machine learning. *Knowledge-Based Systems*, 293 :111610, 6 2024.

[108] Saleem Malik, S. Gopal Krishna Patro, Chandrakanta Mahanty, Rashmi Hegde, Quadri Noorulhasan Naveed, Ayodele Lasisi, Abdulrajak Buradi, Addisu Frinjo Emma, and Naoufel Kraiem. Advancing educational data mining for enhanced student performance prediction : a fusion of feature selection algorithms and classification techniques with dynamic feature ensemble evolution. *Scientific Reports*, 15(1) :1–28, 12 2025.

[109] Shengzhong Mao, Chaoli Zhang, Yichi Song, Jindong Wang, Xiao-Jun Zeng, Zenglin Xu, Senior Member, and Qingsong Wen. Time Series Analysis for Education : Methods, Applications, and Future Directions. 8 2024.

[110] Jacob Montiel, Max Halford, Saulo Martiello Mastelini, Geoffrey Bolmier, Raphael Sourty, Robin Vaysse, Adil Zouitine, Heitor Murilo Gomes, Jesse Read, Talel Abdessalem, et al. River : machine learning for streaming data in python. 2021.

[111] Amir Moslemi, Anna Briskina, Zubeka Dang, and Jason Li. A survey on knowledge distillation : Recent advancements. *Machine Learning with Applications*, 18 :100605, 12 2024.

[112] Ahmed A. Mubarak, Han Cao, and Ibrahim M. Hezam. Deep analytic model for student dropout prediction in massive open online courses. *Computers & Electrical Engineering*, 93 :107271, 7 2021.

[113] Fatemeh Nargesian, Horst Samulowitz, Udayan Khurana, Elias B Khalil, and Deepak Turaga. Learning Feature Engineering for Classification. 2017.

[114] Bach Hoai Nguyen, Bing Xue, and Mengjie Zhang. A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54, 5 2020.

[115] Fan Ouyang, Mian Wu, Luyi Zheng, Liyin Zhang, and Pengcheng Jiao. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20(1), 12 2023.

[116] Ozcan Ozyurt, Hacer Ozyurt, and Deepti Mishra. Uncovering the Educational Data Mining Landscape and Future Perspective : A Comprehensive Analysis. *IEEE Access*, 11 :120192–120208, 2023.

[117] Prasadini Padmasiri and Sanvitha Kasthuriarachchi. Interpretable Prediction of Student Dropout Using Explainable AI Models. *Proceedings - International Research Conference on Smart Computing and Systems Engineering, SCSE 2024*, 2024.

[118] Leonid Perlovsky, Marie-Claude Bonniot-Cabanac, and Michel Cabanac. Curiosity and Pleasure. 10 2010.

[119] Thomas W. Price, Yihuan Dong, and Dzenan Lipovac. The csedm 2019 data challenge : Predicting student performance in a block-based programming environment. In *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019)*. International Educational Data Mining Society, 2019.

[120] Dukundane Gashongore Prince and Savong Bou. Fast Relevant Feature Selection Based on Markov Boundary Learning from Data Stream for Classification Tasks. *Journal of Information Processing*, 31 :321–331, 2023.

[121] S. Priya, T. Ankit, and D. Divyansh. Student performance prediction using machine learning. *Advances in Parallel Computing*, 39 :167–174, 2021.

[122] Carlos Puerto-Santana, Pedro Larrañaga, and Concha Bielza. Feature subset selection in data-stream environments using asymmetric hidden Markov models and novelty detection. *Neurocomputing*, 554, 10 2023.

[123] 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.

[124] Maryam Rahmaninia and Parham Moradi. OSFSMI : Online stream feature selection method based on mutual information. *Applied Soft Computing Journal*, 68 :733–746, 7 2018.

[125] Markson Rebelo Marcolino, Thiago Reis Porto, Tiago Thompsen Primo, Rafael Targino, Vinicius Ramos, Emanuel Marques Queiroga, Roberto Munoz, and Cristian Cechinel. Student dropout prediction through machine learning optimization : insights from moodle log data. *Scientific Reports*, 15(1), 12 2025.

[126] D. Renuka Devi and S. Sasikala. Online Feature Selection (OFS) with Accelerated Bat Algorithm (ABA) and Ensemble Incremental Deep Multiple Layer Perceptron (EIDMLP) for big data streams. *Journal of Big Data*, 6(1), 12 2019.

[127] Daeyoung Roh, Donghee Han, Daehee Kim, Keejun Han, and Mun Yong Yi. SIG-Net : GNN based dropout prediction in MOOCs using Student Interaction Graph. *Proceedings of the ACM Symposium on Applied Computing*, pages 29–37, 4 2024.

[128] Cristobal Romero and Sebastian Ventura. Educational data mining and learning analytics : An updated survey. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 10(3), 5 2020.

[129] Daniela Rotelli, Anna Monreale, and Riccardo Guidotti. Uncovering Student Temporal Learning Patterns. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13450 LNCS :340–353, 2022.

[130] J Samuelsen, W Chen, and B Wasson. Integrating multiple data sources for learning analytics— review of literature. *Research and Practice in Technology Enhanced Learning*, 14(1), 2019.

[131] Dario Sansone. Beyond Early Warning Indicators : High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics*, 81(2) :456–485, 4 2019.

[132] Erik Schaffernicht and Horst Michael Gross. Weighted mutual information for feature selection. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6792 LNCS(PART 2) :181–188, 2011.

[133] Tegjyot Singh Sethi and Mehmed Kantardzic. On the Reliable Detection of Concept Drift from Streaming Unlabeled Data. *Expert Systems with Applications*, 82 :77–99, 3 2017.

[134] Mamta Shah, Ryan S. Baker, Peter Granville, and Kimberly Sharp. Examining Student Engagement in Online Learning Platforms for Promoting Exam Readiness and Success in Undergraduate Nursing Education. In *L@S 2024 - Proceedings of the 11th ACM Conference on Learning @ Scale*, pages 443–445. Association for Computing Machinery, Inc, 7 2024.

[135] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3) :379–423, 1948.

[136] Raghubir Singh and Sukhpal Singh Gill. Edge AI : A survey. *Internet of Things and Cyber-Physical Systems*, 3 :71–92, 1 2023.

[137] John Stamper, Alexandru Niculescu-Mizil, Steve Ritter, Geoff Gordon, and Kenneth Koedinger. Kdd cup 2010 : Educational data mining challenge. In *Proceedings of the KDD Cup 2010 Competition*. ACM, 2010. Data from the ASSISTments platform.

[138] S. M.F.D. Syed Mustapha. Predictive Analysis of Students' Learning Performance Using Data Mining Techniques : A Comparative Study of Feature Selection Methods. *Applied System Innovation*, 6(5), 10 2023.

[139] Pierre Tchounikine and André Tricot. Environnements informatiques et apprentissages humains. *Informatique et sciences cognitives*, 2011.

[140] Dipti Theng and Kishor K. Bhoyar. Feature selection techniques for machine learning : a survey of more than two decades of research, 3 2024.

[141] Arthur Thuy and Dries F. Benoit. Explainability through uncertainty : Trustworthy decision-making with neural networks. *European Journal of Operational Research*, 317(2) :330–340, 9 2024.

[142] N Tomasevic, N Gvozdenovic, and S Vranes. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143 :103676, 2020.

[143] Célina Treuillier and Anne Boyer. A New Way to Characterize Learning Datasets. 2022.

[144] Christos Troussas, Konstantina Chrysafiadi, and Maria Virvou. Personalized tutoring through a stereotype student model incorporating a hybrid learning style instrument. *Education and Information Technologies*, 26(2) :2295–2307, 3 2021.

[145] Maria Tsiakmaki, Georgios Kostopoulos, Sotiris Kotsiantis, and Omiros Ragos. Transfer learning from deep neural networks for predicting student performance. *Applied Sciences (Switzerland)*, 10(6), 3 2020.

[146] Hristos Tyralis and Georgia Papacharalampous. A review of predictive uncertainty estimation with machine learning. *Artificial Intelligence Review*, 57(4), 4 2024.

[147] Gido M. van de Ven, Tinne Tuytelaars, and Andreas S. Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4(12) :1185–1197, 12 2022.

[148] Mrzieh Vatandoustmohammadieh, Mohammad Mahdi Mohajeri, Ali Keramati, and Majid Nili Ahmadabadi. AI-powered Digital Framework for Personalized Economical Quality Learning at Scale. 2024.

[149] Jorge R. Vergara and Pablo A. Estévez. A Review of Feature Selection Methods Based on Mutual Information. *Neural Computing and Applications*, 24(1) :175–186, 9 2015.

[150] Alice Villar and Carolina Robledo Velini de Andrade. Supervised machine learning algorithms for predicting student dropout and academic success : a comparative study. *Discover Artificial Intelligence*, 4(1) :1–24, 12 2024.

[151] T Volarić, H Ljubić, M Dominković, G Martinović, and R Rozić. Data Augmentation with GAN to Improve the Prediction of at-risk Students in a Virtual Learning Environment. *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, pages 260–265, 2023.

[152] H Waheed, S.-U. Hassan, N R Aljohani, J Hardman, S Alelyani, and R Nawaz. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104 :106189, 2020.

[153] Chengliang Wang, Xiaojiao Chen, Teng Yu, Yidan Liu, and Yuhui Jing. Education reform and change driven by digital technology : a bibliometric study from a global perspective. *Humanities and Social Sciences Communications*, 11(1) :1–17, 12 2024.

[154] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A Comprehensive Survey of Continual Learning : Theory, Method and Application. 2024.

[155] Niantai Wang, Haoran Liu, Liyue Zhang, Yanbin Cai, and Qianrui Shi. Loose-to-strict Markov blanket learning algorithm for feature selection. *Knowledge-Based Systems*, 283, 1 2024.

[156] X Wang, B Guo, and Y Shen. Predicting the at-risk Online Students Based on the Click Data Distribution Characteristics. *Scientific Programming*, 2022 :1–12, 2022.

[157] Xiyu Wang, Maeda Yukiko, and Hua Hua Chang. Development and techniques in learner model in adaptive e-learning system : A systematic review. *Computers & Education*, 225 :105184, 2 2025.

[158] Zhenyi Wang and Heng Huang. Model Sensitivity Aware Continual Learning.

[159] Jiun Yu Wu, Chen Hsuan Liao, Chin Chung Tsai, and Oi Man Kwok. Using learning analytics with temporal modeling to uncover the interplay of before-class video viewing engagement, motivation, and performance in an active learning context. *Computers and Education*, 212, 4 2024.

[160] Siyu Wu, Hong Qian, B O Jiang, Wei Zhang, Yang Cao, Jiajun Cui, Runze Li, Bo Jiang, and Wei 2024 Zhang. A Comprehensive Exploration of Personalized Learning in Smart Education : From Student Modeling to Personalized Recommendations. *Journal of the ACM*, 37(111), 1 2024.

[161] Xingyu Wu, Bingbing Jiang, Xiangyu Wang, Taiyu Ban, and Huanhuan Chen. Feature Selection in the Data Stream Based on Incremental Markov Boundary Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10) :6740–6754, 10 2023.

[162] Xingyu Wu, Bingbing Jiang, Tianhao Wu, and Huanhuan Chen. Practical Markov Boundary Learning without Strong Assumptions. Technical report, 2023.

[163] Yao Xie and Rebecca Willett. Online logistic regression on manifolds. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3367–3371, 10 2013.

[164] Ruohao Xu, Mengmeng Li, Zhongliang Yang, Lifang Yang, Kangjia Qiao, and Zhigang Shang. Dynamic feature selection algorithm based on Q-learning mechanism. *Applied Intelligence*, 51(10) :7233–7244, 10 2021.

[165] Taro Yamada. Frugal Approaches to Scalable and Green AI. 7 2024.

[166] Tengyu Yin, Hongmei Chen, Zhong Yuan, Tianrui Li, and Keyu Liu. Noise-resistant multilabel fuzzy neighborhood rough sets for feature subset selection. *Information Sciences*, 621 :200–226, 4 2023.

[167] W. J. Youden. Index for rating diagnostic tests. *Cancer*, 3(1) :32–35, 1 1950.

[168] Mourdi Youssef, Sadgal Mohammed, El Kabtane Hamada, and Berrada Fathi Wafaa. A predictive approach based on efficient feature selection and learning algorithms' competition : Case of learners' dropout in MOOCs. *Education and Information Technologies*, 24(6) :3591–3618, 11 2019.

[169] Kui Yu, Xindong Wu, Wei Ding, and Jian Pei. Towards Scalable and Accurate Online Feature Selection for Big Data. In *Proceedings - IEEE International Conference on Data Mining, ICDM*, volume 2015-January, pages 660–669. Institute of Electrical and Electronics Engineers Inc., 1 2014.

[170] Lei Yu and Huan Liu. Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research*, 5(Oct) :1205–1224, 2004.

[171] Shanshan Yu, Yiran Cai, Baicheng Pan, and Man Fai Leung. Semi-Supervised Feature Selection of Educational Data Mining for Student Performance Analysis. *Electronics (Switzerland)*, 13(3), 2 2024.

[172] Lanqin Yuan, Bernhard Pfahringer, and Jean Paul Barddal. Iterative subset selection for feature drifting data streams. In *Proceedings of the ACM Symposium on Applied Computing*, pages 510–517. Association for Computing Machinery, 4 2018.

[173] Ezzatul Akmal Kamaru Zaman, Azlinah Mohamed, and Azlin Ahmad. Feature selection for online streaming high-dimensional data : A state-of-the-art review, 9 2022.

[174] Xiaoge Zhang and Indranil Bose. Reliability estimation for individual predictions in machine learning systems : A model reliability-based approach. *Decision Support Systems*, 186 :114305, 11 2024.

[175] Yingbin Zhang and Luc Paquette. Sequential pattern mining in educational data : The application context, potential, strengths, and limitations. pages 219–254, 2 2023.

[176] Peng Zhou, Xuegang Hu, Peipei Li, and Xindong Wu. OFS-Density : A novel online streaming feature selection method. *Pattern Recognition*, 86 :48–61, 2 2019.

[177] Peng Zhou, Yunyun Zhang, Peipei Li, and Xindong Wu. General assembly framework for online streaming feature selection via Rough Set models. *Expert Systems with Applications*, 204, 10 2022.

[178] Peng Zhou, Shu Zhao, Yuanting Yan, and Xindong Wu. Online Scalable Streaming Feature Selection via Dynamic Decision. *ACM Transactions on Knowledge Discovery from Data*, 16(5), 10 2022.

[179] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1) :43–76, 1 2021.

[180] Philip Zmushko, Aleksandr Beznosikov, Martin Takáč, and Samuel Horváth. FRUGAL : Memory-Efficient Optimization by Reducing State Overhead for Scalable Training. 267, 11 2024.