

Prédiction du risque d'échec dans un MOOC : une approche basée sur l'analyse des séries temporelles multivariées

Anass El Ayady^{1,2}, Maxime Devanne², Germain Forestier², and Nour El Mawas¹

¹ CREM, Université de Lorraine, Metz, France

² IRIMAS, Université de Haute-Alsace, Mulhouse, France

Résumé Les MOOCs offrent un accès libre et gratuit à un large public, mais les taux de complétion demeurent faibles, souvent en raison du manque de personnalisation des contenus. Pour y remédier, il est essentiel d'évaluer et de prédire les performances des apprenants afin de leur fournir une rétroaction personnalisée. Les traces comportementales, comme les clics et événements, peuvent être analysées sous forme de séries temporelles pour anticiper leurs résultats. Ce travail compare des méthodes de classification de séries temporelles multivariées pour identifier les apprenants à risque d'échec à différents moments du cours (après 5, 10 semaines...). L'évaluation expérimentale, réalisée sur la base *Open University Learning Analytics Dataset* (OULAD), porte sur trois cours : deux en *Science, Technology, Engineering, and Mathematics* (STEM) et un en Sciences Humaines et Sociales (SHS). Les résultats préliminaires montrent que les approches évaluées semblent prometteuses pour prédire l'échec des apprenants dans les MOOCs. Les analyses suggèrent également que la précision des prédictions est influencée par le volume d'interactions enregistré, mettant en lumière l'importance de données comportementales riches et diversifiées.

Mots clés: MOOCs · Séries temporelles · Apprentissage automatique · Prédictions précoces · Personnalisation du contenu

Abstract MOOCs provide free and open access to a wide audience, but completion rates remain low, often due to a lack of personalized content. To address this issue, it is essential to evaluate and predict learners' performance in order to provide personalized feedback. Behavioral traces, such as clicks and events, can be analyzed as time series to anticipate their outcomes. This study compares multivariate time series classification methods to identify at-risk learners at different points in the course (after 5, 10 weeks, etc.). The experimental evaluation, conducted using the Open University Learning Analytics Dataset (OULAD), focuses on three courses : two in Science, Technology, Engineering, and Mathematics (STEM) and one in Social Sciences and Humanities (SHS). The preliminary results indicate that the evaluated approaches appear promising for predicting learner failure in MOOCs. The analyses also suggest that prediction accuracy is influenced by the volume of recorded interactions, highlighting the importance of rich and diverse behavioral data.

Keywords : MOOCs, Time series, Machine learning, Early predictions, Content personalization.

1 Introduction

Avec l'essor des technologies numériques, les MOOCs se sont imposés comme un outil clé de démocratisation de l'éducation. Ils permettent à des millions d'apprenants, de tous horizons, d'accéder à des contenus éducatifs variés. Pourtant, leur potentiel est souvent freiné par des taux d'abandon élevés qui varient généralement entre 0,7% et 52,1%, avec une médiane de 12,6% [5]. Cela souligne un défi persistant dans l'adaptation des cours aux besoins spécifiques des apprenants.

En parallèle, l'évolution des outils d'intelligence artificielle et des méthodes d'apprentissage automatique offre de nouvelles opportunités pour résoudre ces problématiques. Les approches modernes basées sur l'analyse des données permettent non seulement d'automatiser l'identification des facteurs de risque, mais également d'adopter une approche proactive pour engager les apprenants. En mobilisant des algorithmes sophistiqués, il devient possible de proposer des contenus adaptés et des stratégies pédagogiques personnalisées, maximisant ainsi l'expérience éducative.

L'intégration de méthodes d'analyse des données provenant des interactions des apprenants avec la plateforme pourrait constituer une réponse prometteuse à ce défi. Ces données, parfois négligées, représentent une véritable source d'informations sur les comportements et les trajectoires des apprenants au fil du temps. En particulier, les clics de souris peuvent être exploités pour modéliser des séries temporelles décrivant leurs interactions avec différentes activités pédagogiques. Ces séries permettent de suivre les rythmes d'apprentissage, d'identifier les tendances et de prédire les risques d'échec avant qu'ils ne deviennent critiques.

Ce travail se distingue par l'utilisation des séries temporelles multivariées, rarement exploitées dans ce contexte, et l'introduction du modèle *Fully Convolutional Networks* (FCN) pour la prédiction du risque d'échec dans les MOOCs. Par ailleurs, il propose une analyse différenciée selon le type de cours (STEM vs SHS), un aspect encore peu exploré dans la littérature. L'objectif est d'identifier les étudiants à risque dès les premières semaines afin de favoriser des interventions pédagogiques ciblées.

2 Cadre théorique

Les recherches sur la prédiction des performances des apprenants dans les environnements d'apprentissage en ligne s'appuient de plus en plus sur des approches issues de l'intelligence artificielle, de l'apprentissage automatique et de l'apprentissage profond. Le succès récent des modèles d'apprentissage profond motive leur adoption croissante pour exploiter les données disponibles, notamment dans des bases comme OULAD [9], reconnue pour sa richesse et son accessibilité.

Les modèles récurrents, tels que ceux utilisant des LSTM, se sont montrés efficaces pour capturer les dépendances temporelles dans les données issues d'OULAD. Par exemple, le modèle DOPP [7] a démontré son utilité pour identifier les apprenants à risque dès les premières semaines, ce qui le rend pertinent pour une intervention précoce. Une approche hybride combinant des couches denses et des LSTM a également été proposée pour la prédiction multiclassées, illustrant la capacité de ces modèles à traiter des tâches complexes. De plus, les modèles GRU et LSTM étudiés dans [3] et [1] confirment leur pertinence pour analyser les séquences temporelles et prédire les risques d'échec des apprenants, en particulier dans des contextes nécessitant des prédictions précoces et fiables.

D'autres méthodologies exploitent des relations complexes entre apprenants et cours. Une approche basée sur des graphes neuronaux, combinée à des modèles récurrents, représente ces relations sous forme de graphes de connaissances [6]. En utilisant des données de clics et des attributs relationnels temporels, cette méthode démontre son efficacité pour effectuer des prédictions précoces dans un contexte multi-relationnel. Ces travaux mettent en avant l'intérêt d'explorer des représentations graphiques pour améliorer la compréhension et la prédiction dans les systèmes éducatifs.

Les modèles basés sur des Transformers [8] se distinguent par leur capacité à capturer des dépendances séquentielles grâce à des mécanismes d'attention. Ces approches permettent de traiter des données normalisées et organisées de manière séquentielle pour des prédictions multiclassées. En exploitant des données agrégées à différentes granularités temporelles, les Transformers démontrent leur efficacité pour les tâches de classification séquentielle, surpassant souvent les modèles traditionnels comme les LSTM.

Des méthodes plus classiques exploitent les forêts de séries temporelles pour classer les apprenants en fonction de leurs comportements interactifs [2]. De manière similaire, des modèles d'apprentissage automatique plus simples, comme les forêts aléatoires, permettent des prédictions précoces en analysant une partie limitée des données disponibles. Ces travaux illustrent la robustesse des algorithmes classiques lorsqu'ils sont appliqués à des données bien structurées.

Comparativement, les modèles récurrents (LSTM, GRU) excellent dans l'analyse des données temporelles continues et les prédictions à différents moments d'un cours grâce à leur capacité à modéliser les séquences. Les approches par graphes et Transformers s'avèrent efficaces pour capturer des relations complexes entre apprenants et contenus. Les modèles classiques, tels que les forêts aléatoires, bien que robustes et simples, sont moins adaptés aux prédictions séquentielles ou aux contextes relationnels complexes.

Cependant, ces avancées ne prennent pas en compte l'influence des différences disciplinaires, notamment entre les cours en SHS et en STEM. Ces distinctions, pourtant essentielles, peuvent affecter la densité des données et les performances des modèles prédictifs. Notre étude comble cette lacune en analysant ces deux contextes distincts et en évaluant les approches pour chaque discipline.

Dans notre étude, nous avons sélectionné les modèles LSTM pour leur performance reconnue en analyse de séries temporelles, comme indiqué dans la littérature. Le modèle DOPP, basé sur LSTM, est réputé pour capturer efficacement les dépendances temporelles et fournir des prédictions précoces à partir des données de clics.

En complément, nous avons exploré le FCN, qui se distingue par sa capacité à extraire des motifs locaux et globaux dans les séries temporelles grâce à ses couches convolutives. Bien qu'ils soient moins présents dans les travaux de l'état de l'art sur les MOOCs, ils ont démontré leur efficacité dans la classification des séries temporelles dans d'autres domaines [4], ce qui justifie leur inclusion dans notre étude.

Pour compléter cette sélection, nous avons intégré des modèles classiques comme les k-plus-proches-voisins (KNN) et les perceptrons multicouches (MLP). Bien qu'ils ne soient pas couramment étudiés dans le domaine de la prédiction des performances dans les MOOCs, ces modèles offrent une base de comparaison pour évaluer la pertinence et la valeur ajoutée des approches avancées. L'objectif de cette étude comparative est d'évaluer différentes méthodes d'analyse des séries temporelles, dans des conditions cohérentes, afin d'identifier celles qui semblent les plus prometteuses pour prédire le risque d'échec dans les contextes ciblés. En particulier, les modèles avancés comme le FCN et le DOPP exploitent les dépendances temporelles et les séries multivariées, tandis que les KNN et les MLP permettent d'évaluer la performance de solutions plus simples. Ce choix équilibré entre approches classiques et avancées permet d'obtenir une vue d'ensemble complète et rigoureuse des méthodes disponibles pour ce contexte spécifique.

3 Méthodologie

Cette section détaille les approches méthodologiques adoptées pour analyser les séries temporelles des interactions des apprenants et prédire leur niveau de réussite. Nous décrivons d'abord le jeu de données utilisé, puis les étapes de préparation des données, les modèles évalués, ainsi que les méthodologies d'évaluation utilisées pour comparer leurs performances.

3.1 Description du jeu de données

Cette étude s'appuie sur le jeu de données public OULAD, qui regroupe les parcours de 32 593 étudiants répartis sur 22 cours en ligne, avec plus de 10 millions de clics enregistrés. Nous avons sélectionné trois cours fortement représentés (tableau I) : (BBB, SHS), (DDD, STEM) et (FFF, STEM), afin d'explorer l'influence du type disciplinaire sur les performances prédictives.

TABLEAU I – Statistiques des apprenants et des présentations pour chaque module

Code de module	Nombre de présentations	Nombre d'apprenants
AAA	2	748
BBB	4	7909
CCC	2	4434
DDD	4	6272
EEE	3	2934
FFF	4	7762
GGG	3	2534
Total	22	32593

Chaque cours est proposé deux fois par an (sessions de février "B" et d'octobre "J") (tableau II). Les sessions 2013B et 2013J sont utilisées pour l'entraînement, et les sessions 2014B et 2014J pour le test, ce qui garantit une évaluation sur des cohortes indépendantes dans le temps. Certains apprenants inscrits ne participent pas au cours, ce qui explique la différence entre le nombre total d'apprenants indiqué dans le tableau I et la somme des apprenants par période dans le tableau II.

La distribution des résultats finaux est la suivante : *Pass* (42,55%), *Withdrawn* (24,62%), *Fail* (23,94%), *Distinction* (8,89%). Pour la classification binaire, *Pass* et *Distinction* sont regroupés en classe positive, tandis que *Fail* et *Withdrawn* forment la classe négative. Une baseline simple basée sur la classe majoritaire atteint un F1-score de 0,58 (binaire) et 0,48 (multiclasse), ce qui reste significativement inférieur aux performances obtenues par les modèles proposés.

3.2 Préparation des données

Les interactions des étudiants avec les cours sont transformées en séries temporelles multivariées, où chaque étudiant est décrit par une séquence hebdomadaire de clics répartis par type d'activité (forum, quiz, ressource, etc.). Chaque activité constitue un *canal*, soit une dimension distincte de la série.

La structure obtenue est $(n_samples, n_weeks, n_activities)$, où $n_samples$ est le nombre d'apprenants, n_weeks la durée d'observation (variant de 5 à 40 semaines) et $n_activities$ le

TABLEAU II – Description des cours étudiés

Code	Type	Période	Nombre d'apprenants
BBB	SHS	2013B	1537
BBB	SHS	2013J	1870
BBB	SHS	2014B	1294
BBB	SHS	2014J	1921
DDD	STEM	2013B	1214
DDD	STEM	2013J	1768
DDD	STEM	2014B	1116
DDD	STEM	2014J	1647
FFF	STEM	2013B	1510
FFF	STEM	2013J	2098
FFF	STEM	2014B	1363
FFF	STEM	2014J	2121

nombre d'activités distinctes. Pour garantir l'homogénéité des canaux, une normalisation Min-Max est appliquée individuellement sur chaque canal.

Les étiquettes de performance sont assignées selon deux cadres : *binnaire* (Pass/Fail) et *multi-classe* (Distinction, Pass, Fail, Withdrawn), permettant d'évaluer les modèles à différents niveaux de granularité.

La figure 1 montre que les apprenants qui réussissent maintiennent une activité régulière tout au long du cours, tandis que la figure 2 illustre une baisse significative des interactions pour les apprenants en échec. Ces dynamiques confirment l'importance des traces d'interactions comme indicateurs précoces.

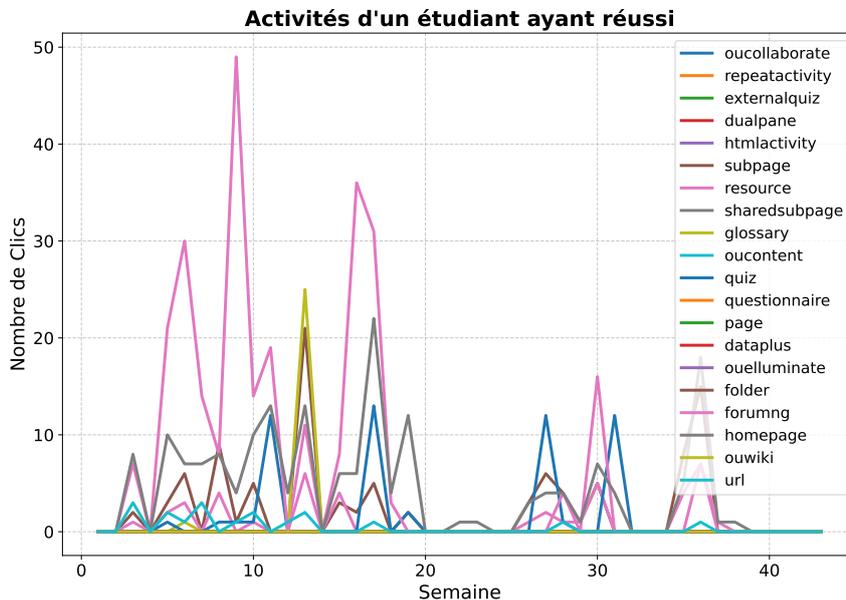


FIGURE 1 – Représentation des interactions hebdomadaires d'un apprenant ayant réussi. L'activité reste régulière tout au long du cours, indiquant un engagement constant.

3.3 Indicateur de performance

L'évaluation repose sur le *F1-score*, une métrique équilibrée combinant précision et rappel. Il est particulièrement adapté aux contextes de classification déséquilibrée, comme les MOOCs, où la simple précision globale (accuracy) peut être trompeuse.

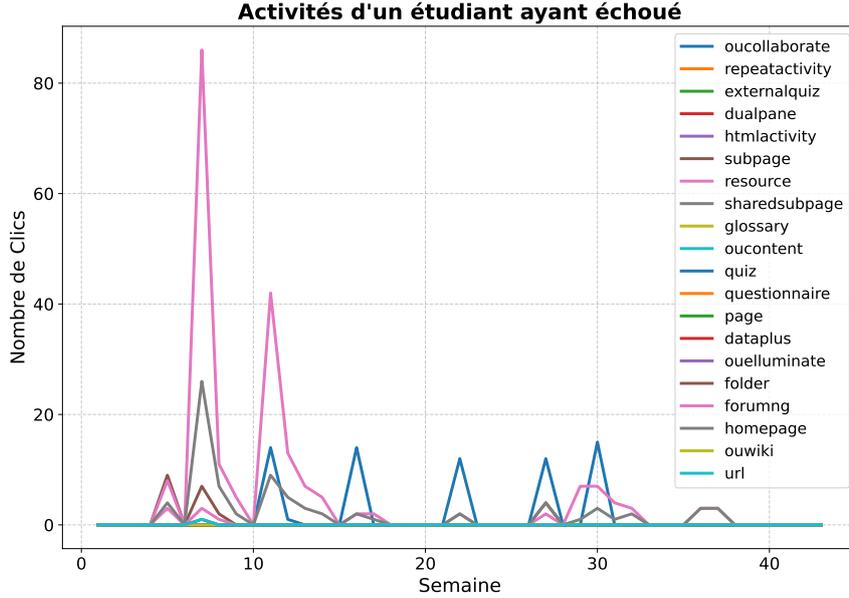


FIGURE 2 – Représentation des interactions hebdomadaires d’un apprenant échoué. Une baisse significative de l’activité est observée à partir de la 5e semaine, signalant un désengagement progressif.

Le F1-score est défini par :

$$F1 = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (1)$$

avec :

$$\text{Précision} = \frac{TP}{TP + FP}, \quad \text{Rappel} = \frac{TP}{TP + FN} \quad (2)$$

où TP , FP et FN désignent respectivement les vrais positifs, les faux positifs et les faux négatifs. Dans notre contexte, cela reflète la capacité du modèle à détecter correctement les apprenants à risque, tout en limitant les fausses alertes.

Les performances sont mesurées à différents moments du cours (après 5, 10, 15 semaines, etc.), afin d’évaluer la capacité des modèles à prédire tôt dans le parcours. Cette approche permet d’identifier les périodes critiques où les prédictions sont les plus fiables.

4 Résultats et Discussion

Les résultats obtenus pour la classification binaire et multiclassées révèlent des tendances similaires en termes de performances des modèles et de dépendance au type de cours et à la durée d’observation. Pour la **classification binaire** (la figure 3), les cours **STEM (DDD et FFF)** affichent des scores F1 significativement plus élevés que le cours **SHS (BBB)**, atteignant jusqu’à **94%** pour **FCN** après 40 semaines dans le cours FFF. Le cours **DDD**, bien qu’étant également un cours STEM, affiche des performances légèrement inférieures (86%). Cette différence s’explique par un volume et une régularité d’interactions inférieurs à ceux observés dans FFF, comme le confirment les statistiques d’usage. Ce résultat souligne la capacité du **FCN** à identifier des motifs temporels complexes grâce à ses convolutions, qui permettent d’extraire des relations locales et globales dans les séries temporelles tout en préservant leur structure séquentielle.

Cette supériorité des modèles pour les cours STEM s’explique sûrement par le volume élevé d’interactions et une variabilité temporelle plus marquée. Cela offre aux modèles, notamment **FCN** et **DOPP**, une meilleure capacité d’apprentissage des relations complexes entre les activités et les résultats des apprenants. À l’inverse, le cours SHS souffre d’une stagnation des performances, en particulier pour les séries temporelles courtes, à cause d’une faible densité d’interactions : en configuration binaire, le F1-score reste limité à **58%** après 5 semaines ; en multiclassées, il débute à **32%** après 5 semaines et plafonne à **54%** après 40 semaines.

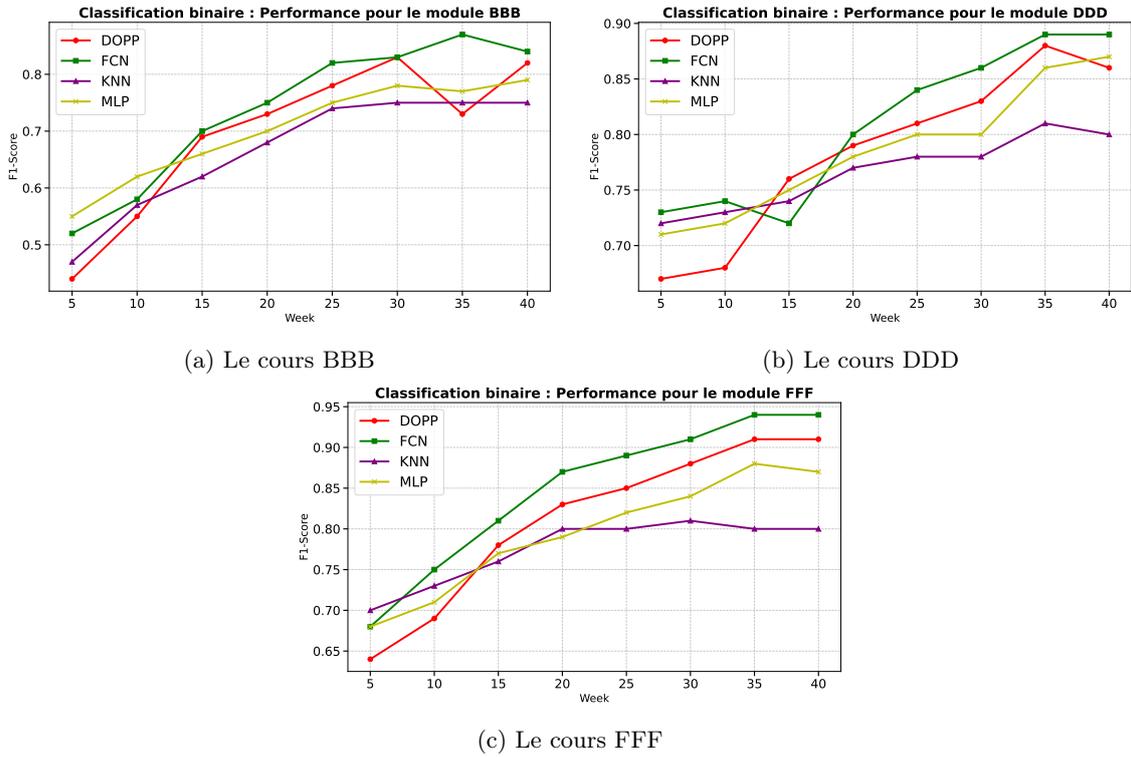


FIGURE 3 – Évaluation de performances : classification binaire

Pour la **classification multiclassés** (la figure 4), ces tendances persistent, mais les scores F1 sont globalement plus faibles, en particulier pour le cours **BBB**, où **FCN** plafonne autour de **55%** après 40 semaines. Cela illustre la difficulté des modèles à discriminer plusieurs classes dans un environnement où les données interactives sont limitées. En revanche, pour les cours **DDD** et **FFF**, les modèles montrent une progression continue, avec **FCN** atteignant **66%** et **68%** respectivement. Cette amélioration s'accroît avec l'augmentation de la période d'observation, confirmant que l'enrichissement des données temporelles est crucial pour des prédictions précises, mais elle souligne aussi le défi du *cold start* pour les semaines initiales. Les modèles classiques comme le **KNN** stagnent dans les deux configurations de classification, prouvant leur faible capacité à capturer des dynamiques temporelles complexes.

Ainsi, que ce soit pour la classification binaire ou multiclassés, les cours présentant une richesse et une diversité élevées d'interactions temporelles offrent un avantage notable pour la prédiction des performances. À l'inverse, les cours avec une faible densité d'interactions posent des défis importants. La performance supérieure des modèles neuronaux, comme **FCN** et **DOPP**, illustre leur capacité à apprendre efficacement des séries temporelles multivariées, même si des méthodes adaptées restent nécessaires pour améliorer les performances dans des contextes où les données sont limitées ou faiblement structurées, comme l'utilisation de modèles préentraînés ou d'apprentissage par transfert.

Dans les cours **SHS**, des stratégies complémentaires telles que l'intégration de données qualitatives, comme les réponses des apprenants dans les forums de discussion ou les commentaires sur les évaluations, ou d'autres indicateurs comportementaux, tels que le temps passé sur les activités, la participation aux sessions interactives ou l'utilisation des ressources pédagogiques, pourraient améliorer les performances des modèles.

La différence de performance observée entre les cours **SHS** et **STEM** peut être attribuée, en partie, à la conception pédagogique des cours. Une expérimentation complémentaire, illustrée dans la figure 5, montre que le cours **FFF** (**STEM**) présente un nombre unique d'activités et un total de clics nettement supérieurs à ceux du cours **BBB** (**SHS**). Cette différence suggère que les cours **STEM** bénéficient d'une conception favorisant une plus grande richesse et diversité d'interactions. Toutefois, cette observation ne peut être généralisée sans des analyses plus approfondies pour mieux comprendre l'impact des différences structurelles dans la conception pédagogique sur les performances des apprenants.

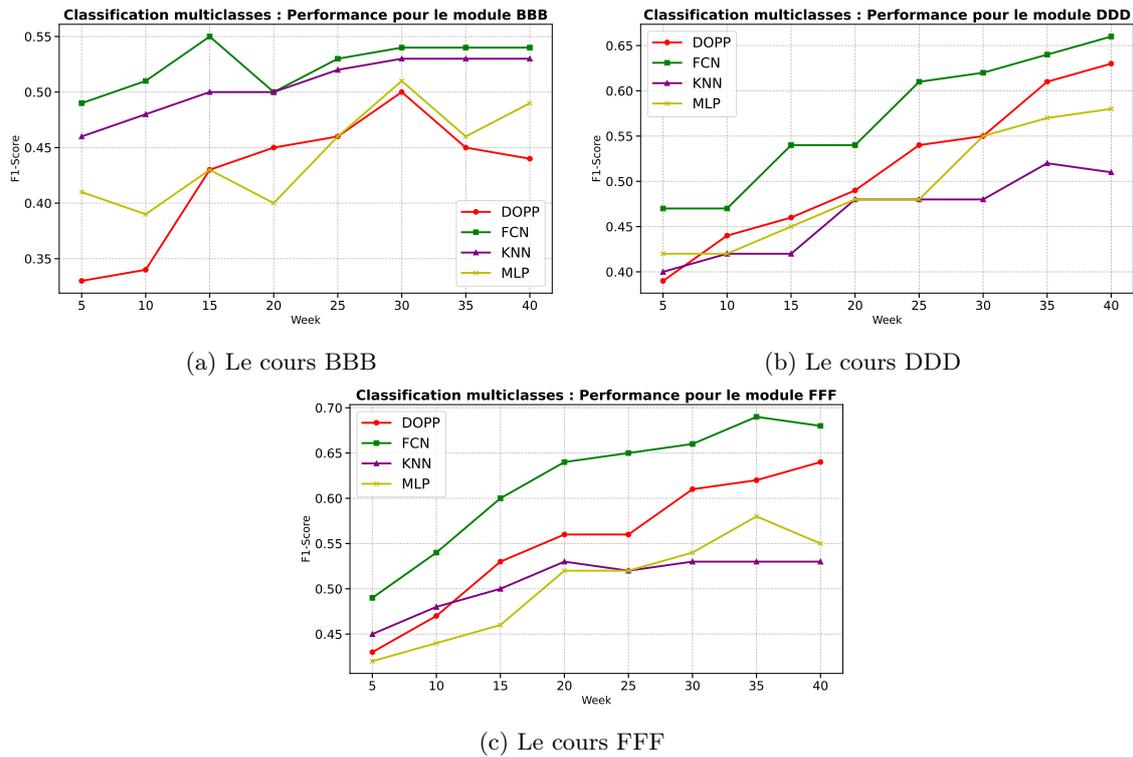


FIGURE 4 – Évaluation de performances : classification multiclasse

5 Conclusion

En conclusion, cette étude met en évidence l'intérêt de l'analyse des séries temporelles multivariées pour classer les apprenants selon leur performance attendue dans les MOOCs. Les résultats montrent que la richesse et la diversité des interactions hebdomadaires jouent un rôle clé dans la précision des prédictions. Les modèles profonds, en particulier le FCN introduit ici, permettent de capturer des motifs temporels complexes, dépassant les approches classiques.

Ces travaux s'inscrivent dans la continuité des recherches en EIAH intégrant l'apprentissage automatique, avec une volonté d'innovation dans la personnalisation des parcours. À terme, des outils comme des systèmes d'analyse automatisée ou des tableaux de bord pour enseignants pourraient faciliter l'identification précoce des apprenants en difficulté et soutenir des interventions pédagogiques ciblées.

Cependant, cette étude présente certaines limites. Les conclusions sont basées sur un seul jeu de données (OULAD) et un nombre restreint de cours. Bien que ce dataset soit riche et structuré, des validations sur d'autres plateformes et contextes éducatifs sont nécessaires pour confirmer la généralisabilité des résultats.

Le problème du *cold start*, marqué par une faible densité d'interactions durant les premières semaines, reste un défi. Nos essais avec des tailles de noyaux de FCN et la distance *Dynamique Time Warping* (DTW) n'ont pas permis d'améliorations significatives. L'intégration d'informations contextuelles (telles que les profils démographiques et la date d'inscription) pourrait atténuer cet effet en apportant des indices supplémentaires lorsque les données comportementales sont insuffisantes.

Enfin, l'intégration de données qualitatives issues de forums, d'évaluations ou de retours libres représente une piste prometteuse pour enrichir les profils d'apprenants. Leur usage soulève cependant des défis méthodologiques liés à leur nature interprétative et déclarative. Des approches hybrides combinant données quantitatives et qualitatives, renforcées par des techniques comme le transfert d'apprentissage [10], pourraient améliorer la robustesse des analyses et leur adaptation à une diversité de contextes éducatifs.

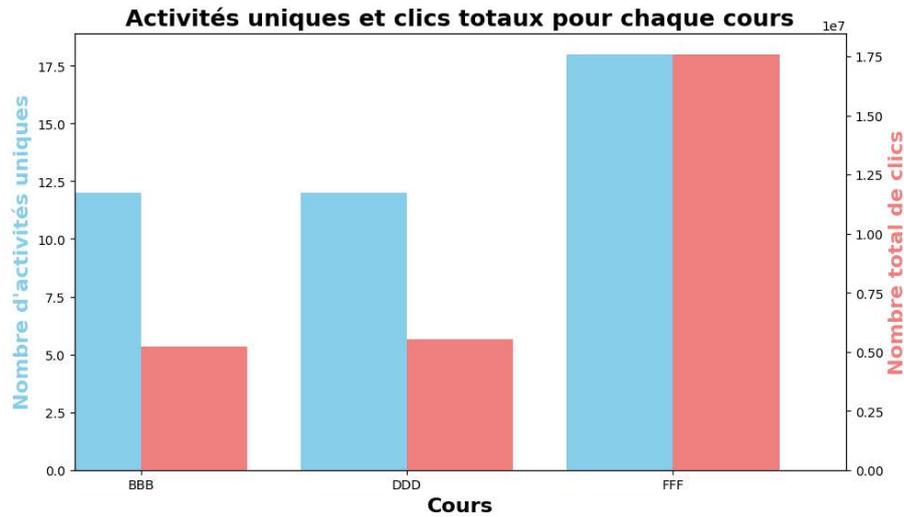


FIGURE 5 – Comparaison du nombre unique d’activités et du total de clics pour les cours SHS et STEM. Cette figure met en évidence que le cours STEM dispose de davantage d’activités et d’interactions, suggérant une conception pédagogique plus riche et variée.

Remerciements

Cette étude s’inscrit dans le projet **COPCOT** (ANR-22-CE38-0003), soutenu par l’Agence Nationale de la Recherche (**ANR**), visant à personnaliser les contenus éducatifs numériques pour faciliter l’acquisition de compétences en informatique. Nous remercions l’**ANR** pour son soutien financier et les partenaires du projet pour leurs contributions.

Références

1. Aljohani, N.R., Fayoumi, A., Hassan, S.U. : Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability* **11**(24), 7238 (2019)
2. Haiyang, L., Wang, Z., Benachour, P., Tubman, P. : A time series classification method for behaviour-based dropout prediction. In : 2018 IEEE 18th international conference on advanced learning technologies (ICALT). pp. 191–195. IEEE (2018)
3. He, Y., Chen, R., Li, X., Hao, C., Liu, S., Zhang, G., Jiang, B. : Online at-risk student identification using rnn-gru joint neural networks. *Information* **11**(10), 474 (2020)
4. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A. : Deep learning for time series classification : a review. *Data mining and knowledge discovery* **33**(4), 917–963 (2019)
5. Jordan, K. : Massive open online course completion rates revisited : Assessment, length and attrition. *International Review of Research in Open and Distributed Learning* **16**(3), 341–358 (2015)
6. Karimi, H., Derr, T., Huang, J., Tang, J. : Online academic course performance prediction using relational graph convolutional neural network. *International Educational Data Mining Society* (2020)
7. Karimi, H., Huang, J., Derr, T. : A deep model for predicting online course performance. *Association for the Advancement of Artificial Intelligence* pp. 1–6 (2020)
8. Kusumawardani, S.S., Alfarozi, S.A.I. : Transformer encoder model for sequential prediction of student performance based on their log activities. *IEEE Access* **11**, 18960–18971 (2023)
9. Kuzilek, J., Hlostá, M., Zdrahal, Z. : Open university learning analytics dataset. *Scientific data* **4**(1), 1–8 (2017)
10. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., Ragos, O. : Transfer learning from deep neural networks for predicting student performance. *Applied Sciences* **10**(6), 2145 (2020)