Développement de détecteurs précoces de l'attrition des étudiants et du Wheel Spinning à l'aide de l'apprentissage profond (Deep Learning)

Anthony F. Botelho, Ashvini Varatharaj, Thanaporn Patikorn, Diana Doherty, Seth A. Adjei, et Joseph E. Beck

*Résumé - L'*utilisation accrue de plateformes d'apprentissage informatisées et d'outils en ligne dans les salles de classe offre de nouvelles possibilités non seulement d'étudier les concepts sous- jacents impliqués dans le processus d'apprentissage, mais aussi d'utiliser ces informations pour identifier et aider les élèves en difficulté. De nombreuses plates-formes d'apprentissage, en particulier celles qui pilotent ou complètent l'enseignement, ne sont en mesure d'apporter une aide qu'aux étudiants qui interagissent avec le système. Dans cette optique, la persévérance des étudiants apparaît comme une construction d'apprentissage importante contribuant à la réussite des étudiants dans l'apprentissage de nouvelles matières. À l'inverse, une persistance élevée n'est pas toujours productive pour les élèves, lorsque la pratique supplémentaire ne les aide pas à atteindre un état de maîtrise de la matière. Dans cet article, nous appliquons une méthodologie d'apprentissage par transfert utilisant l'apprentissage profond et les techniques de modélisation traditionnelles pour étudier les représentations élevées et faibles de la persistance improductive. Nous nous concentrons sur deux problèmes importants dans les domaines de l'exploration de données éducatives et de l'analyse de l'apprenant, représentant une faible persistance, caractérisée par l'" abandon " de l'étudiant, et une persistance élevée improductive, opérationnalisée par le " WHEEL SPINNING " de l'étudiant, dans le but de mieux comprendre la relation entre ces mesures de persistance improductive (c'est-à-dire l'abandon et le WHEEL SPINNING) et de développer des détecteurs précoces de ces comportements. Nous constatons que les modèles développés pour détecter l'abandon au sein de l'affectation et entre les affectations et le WHEEL SPINNING sont capables d'apprendre des ensembles de caractéristiques qui se généralisent pour prédire l'autre. Nous observons en outre la performance de ces modèles à chaque occasion d'apprentissage dans les affectations des étudiants afin d'identifier le moment où des interventions peuvent être déployées pour aider au mieux les étudiants qui sont susceptibles de faire preuve d'une persistance improductive.

*Termes de l'index :* détection *précoce*, WHEEL SPINNING, Stopout, apprentissage par transfert, apprentissage profond, persistance.

données nécessaires pour mieux comprendre ces processus d'apprentissage, tout en offrant simultanément une plate-forme par laquelle cette recherche peut être mise en œuvre et déployée pour améliorer les expériences d'apprentissage des élèves. Comme c'est le cas pour de nombreuses plateformes d'apprentissage, si ce n'est toutes, en particulier celles qui visent à piloter ou à compléter l'enseignement dispensé par les enseignants, elles ne sont en mesure de fournir une aide qu'aux étudiants qui interagissent avec le système ; c'est pour cette même raison que les tuteurs humains emploient souvent une série de techniques pour maintenir l'engagement des étudiants et encourager leur persévérance lorsqu'ils abordent des contenus difficiles [1]. Cela renforce la nécessité de mieux comprendre la persistance de l'élève au cours du processus d'apprentissage afin de développer de meilleurs détecteurs

d'élèves en difficulté.

1. INTRODUCTION

T

L'utilisation d'environnements d'apprentissage numériques dans les écoles a donné lieu à de nouvelles possibilités d'étudier les constructions d'apprentissage influentes des élèves, à la fois longitudinalement et à des niveaux de granularité fins. Les environnements

d'apprentissage numériques sont apparus pour tirer parti de ces possibilités, en fournissant aux chercheurs les outils et les

et ensuite développer des interventions pour promouvoir des stratégies d'apprentissage productives.

Lorsqu'ils abordent un contenu difficile, il est essentiel que les élèves fassent preuve d'une grande persistance en travaillant sur un nombre suffisant de problèmes pratiques afin d'apprendre la matière avec succès. Ainsi, le concept de persistance joue un rôle important dans la réussite des élèves, comme l'ont montré les recherches sur le courage [2], la persévérance [3] et l'échec productif [4]. Les élèves qui ne parviennent pas à terminer leur travail après un petit nombre de problèmes seulement, définis dans le présent document comme des élèves qui font preuve d'abandon, ratent des occasions d'apprendre la matière difficile par une pratique supplémentaire ; c'est particulièrement le cas lorsque les élèves font preuve d'abandon au début d'un devoir, par exemple dans les premiers problèmes.

Bien que la présence de la persistance soit essentielle pour que les étudiants puissent surmonter les obstacles à l'apprentissage, il existe des cas où une persistance élevée peut être improductive. Cet aspect négatif d'une persistance élevée et improductive a été opérationnalisé dans des travaux antérieurs par un comportement connu sous le nom de "WHEEL SPINNING" [5]. Le WHEEL SPINNING décrit le

cas d'un étudiant qui persiste dans une tâche d'apprentissage particulière mais qui est incapable d'atteindre un état de maîtrise dans un délai raisonnable.

L'abandon et le WHEEL SPINNING représentent tous deux des exemples improductifs de persévérance des étudiants ; dans un cas, l'arrêt représente les étudiants qui ne font pas preuve d'assez de persévérance pour réussir, tandis que le WHEEL SPINNING représente une persévérance trop importante pour laquelle l'étudiant aurait probablement intérêt à s'arrêter et à chercher une aide supplémentaire auprès d'un instructeur ou d'un tuteur. Pour cette raison, nous définissons l’abandon et le WHEEL SPINNING comme des mesures mutuellement exclusives au sein d'un même devoir. Comme les travaux précédents ont défini le comportement de WHEEL SPINNING comme le fait pour un étudiant d'atteindre le dixième problème, ou opportunité d'apprentissage, d'un devoir basé sur la maîtrise (discuté plus en détail dans la section 3), les étudiants ne sont considérés comme ayant arrêté un devoir que s'ils l'ont fait avant le dixième problème ; il est important de souligner cette définition car chaque mesure représente ainsi ce que nous considérons comme un comportement d'apprentissage improductif.

Il est important de pouvoir détecter quand les élèves sont susceptibles de présenter un comportement d'arrêt ou de WHEEL SPINNING afin de développer des interventions pour promouvoir la persistance quand elle est probablement bénéfique pour les élèves et pour suggérer une aide supplémentaire quand cette persistance a peu de chances de mener au succès. Cependant, à la lumière de cette importance, déployer une intervention une fois que l'arrêt est détecté n'a probablement pas beaucoup d'impact car l'étudiant a déjà cessé d'interagir avec le système, et de même, dans le cas du WHEEL SPINNING, déployer une intervention au moment de la détection est probablement trop tard car

l'étudiant a déjà...

a perdu du temps et des efforts (et est peut-être devenu frustré). C'est en gardant ces scénarios à l'esprit qu'il devient impératif de déployer de telles interventions de manière préventive en anticipant un tel comportement et en s'attaquant aux causes potentielles de l'arrêt et du comportement de WHEEL SPINNING avant que l'élève ne présente des formes improductives de persistance élevée et faible. Comme nous le verrons plus en détail dans la section " Contexte ", les récentes applications de l'apprentissage profond dans le contexte de l'éducation ont donné des résultats prometteurs, ce qui soutient l'exploration de tels modèles pour la tâche de développer des détecteurs précoces de ces comportements d'étudiants.

L'objectif de ce travail est d'explorer la détection précoce de la persistance improductive telle qu'elle est opérationnalisée par le WHEEL SPINNING et le stopout. En utilisant des techniques d'apprentissage automatique, y compris l'application de l'apprentissage profond en conjonction avec des méthodes d'apprentissage par transfert de modèles et de résultats, nous explorons la relation entre les prédicteurs appris du WHEEL SPINNING et de l'abandon, à la fois au sein d'une affectation et entre les affectations. Avec cet objectif en tête, nous cherchons à répondre aux questions de recherche suivantes :

1. Comment les modèles d'apprentissage profond temporel se comparent-ils aux méthodes traditionnelles dans la tâche de prédire le comportement de WHEEL SPINNING et d'arrêt à la fois dans et entre les affectations ?
2. Les prédicteurs appris de chaque comportement de WHEEL SPINNING et de stopout sont-ils également prédictifs de l'autre comportement respectif (par exemple, les prédicteurs de WHEEL SPINNING sont- ils également prédictifs de stopout ainsi que l'inverse) ?
3. Comment la récence influe-t-elle sur la performance des modèles prédisant le Wheel Spinning et l'arrêt de la course, dans le cadre d'une même affectation ou d'une affectation différente ?

Ce travail se concentre sur l'exploration de la relation entre les représentations de la persistance improductive des étudiants dans le but de développer des détecteurs précoces de ces comportements. La section suivante décrira d'abord les travaux existants qui ont déjà étudié les comportements d'attrition des étudiants et de WHEEL SPINNING en plus des applications précédentes de l'apprentissage profond dans le contexte de l'éducation. Nous décrirons ensuite la source et les attributs des données utilisées dans ce travail avant de détailler la méthodologie appliquée et les analyses menées pour étudier ces comportements d'étudiants. Les résultats de ces analyses seront ensuite discutés, avec un accent particulier sur la détection précoce de chaque comportement d'arrêt et de Wheel Spinning au sein de l'affectation et entre les affectations. Enfin, nous discuterons des travaux futurs potentiels, soulignerons les contributions de ce travail et discuterons des conclusions finales des analyses effectuées.

1. CONTEXTE
2. *ROTATION DE LA ROUE*

Plusieurs travaux antérieurs ont exploré et tenté de modéliser le comportement de l'étudiant en matière de

WHEEL SPINNING dans plusieurs plates-formes, y compris le tuteur cognitif [6] et les ASSISTANCES [5] [7], tandis que d'autres travaux ont exploré des politiques pour aider à prévenir le WHEEL SPINNING [8]. Comme décrit dans la section Introduction, le WHEEL SPINNING est le comportement d'un étudiant qui fait preuve d'une grande persistance dans une tâche d'apprentissage, mais qui est incapable d'obtenir une compréhension suffisante du matériel d'apprentissage. Le terme "WHEEL SPINNING" est analogue à celui d'une voiture bloquée dans la neige.

ou de la boue ; malgré les efforts déployés pour se déplacer, les roues tournent sans aller nulle part.

Dans ce travail, nous utiliserons la définition du WHEEL SPINNING donnée dans le travail de Beck et Gong [5], qui consiste à ne pas atteindre la maîtrise après avoir vu dix opportunités d'apprentissage. C'est pour cette raison que les travaux antérieurs sur l'observation du WHEEL SPINNING ont porté sur les interactions des étudiants avec les devoirs basés sur la maîtrise. Les devoirs basés sur la maîtrise, par opposition aux devoirs traditionnels qui exigent des élèves qu'ils répondent à toutes les questions assignées, exigent plutôt des élèves qu'ils démontrent un niveau suffisant de compréhension, ou de maîtrise, du matériel assigné afin de compléter le devoir. Dans le cas d'ASSISTments, ce seuil de compréhension, par défaut, exige des étudiants qu'ils répondent simplement à trois problèmes consécutifs correctement à la première tentative sans utiliser l'aide fournie par l'ordinateur.

Les tentatives précédentes de modélisation du WHEEL SPINNING ont observé l'activité de l'élève sur des devoirs basés sur la maîtrise au niveau du problème afin de prédire si l'élève va éventuellement faire une roue dans ce devoir [7]. Le modèle a été entraîné sur des caractéristiques générées par des experts décrivant chaque problème et les actions récentes de l'élève pour estimer la probabilité qu'un élève fasse une roue dans le devoir en cours. Nous supposons qu'un tel modèle est susceptible d'être plus performant sur les problèmes ultérieurs d'un devoir que sur les problèmes antérieurs, mais les travaux précédents ont fait état d'une performance moyenne du modèle pour toutes les opportunités, ou problèmes.

Le présent article tente, en partie, de s'appuyer sur ces travaux antérieurs pour construire des modèles permettant de prédire le WHEEL SPINNING à l'aide d'une granularité plus fine des données (par exemple au niveau de l'action), d'observer le comportement du WHEEL SPINNING (ainsi que le stopout qui sera décrit ci-après) sur des périodes plus longues (par exemple sur plusieurs affectations), et d'observer comment la performance du modèle change sur des problèmes consécutifs.

1. *Attrition et abandon des étudiants*

L'attrition des étudiants, plus communément appelée abandon, a fait l'objet d'une grande attention ces dernières années en tant que problème d'éducation, en grande partie en raison de son importance dans les environnements numériques tels que les cours en ligne ouverts et massifs (MOOC) [9] [10] [11] [12] [13]. Dans ces systèmes, il a été observé qu'une grande partie des étudiants ne terminent pas leurs cours ; ce comportement est appelé abandon. Les enquêtes ont montré que les raisons de la faible persistance dans les MOOC sont multiples et varient d'un apprenant à l'autre. Par exemple, certains peuvent abandonner en raison de connaissances de base insuffisantes ou de la difficulté du contenu, mais d'autres peuvent être interrompus en raison de la gestion du temps ou de l'emploi du temps, ou simplement arrêter de revenir parce qu'ils ont appris tout ce qu'ils voulaient savoir [14]. L'attrition des étudiants au sein des MOOC a également été étudiée précédemment par le biais du

développement d'un modèle d'apprentissage profond, appelé "GritNet", qui s'est avéré plus performant que les méthodes de base existantes [15] et même transféré d'un cours à l'autre [16]. Bien que ces domaines aient, comme décrit, reçu une grande attention, les caractéristiques de la persistance et les raisons de l'attrition dans les MOOC diffèrent grandement de celles observées dans les classes de la maternelle à la 12e année, car la plupart des étudiants ne présentent pas d'abandon de la même manière.

L'abandon n'est pas courant dans le contexte traditionnel des classes de la maternelle à la 12e année (c'est-à-dire l'enseignement obligatoire), car l'assiduité et l'obtention de diplômes sont souvent imposées et encouragées par les parents. Au lieu de cela, l'attrition et la faible persistance des étudiants sont observées sous la forme d'étudiants qui ne terminent pas certaines tâches d'apprentissage ; nous appelons ce comportement "abandon". La principale différence entre l'abandon et le décrochage est que lorsqu'un étudiant abandonne, il est toujours dans le cours et peut choisir de terminer les tâches suivantes, alors que les apprenants sont définis comme abandonnant lorsqu'ils ne reviennent pas pour terminer le cours.

Dans de nombreux cas, l'abandon des étudiants au niveau des devoirs empêche les étudiants d'apprendre suffisamment la matière et peut ensuite entraîner des difficultés supplémentaires lors de l'apprentissage des compétences post- requises (voir par exemple [17]), mais introduit également une série d'autres problèmes relatifs au développement et au déploiement d'interventions d'apprentissage efficaces. Étant donné que les élèves présentant un comportement d'arrêt cessent d'interagir avec l'environnement d'apprentissage, aucune aide ne peut leur être apportée par le biais de la plateforme, l'aide de l'élève dépendant alors uniquement de sources externes, telles que l'enseignant. Les données manquantes ou incomplètes sur les étudiants causées par l'attrition rendent difficile l'étude du processus d'apprentissage (car aucune donnée ne peut être enregistrée pour les étudiants qui n'interagissent pas avec le système), la mesure de l'efficacité des interventions par le biais d'essais contrôlés randomisés [18] et, comme la cause de l'abandon est souvent difficile à identifier, le développement d'interventions efficaces pour soutenir une persistance plus productive. Pour ces raisons, il est important d'élaborer des modèles permettant d'identifier de manière préventive les élèves susceptibles d'avoir un comportement d'abandon, afin de mieux comprendre les signes précurseurs de ce comportement et de développer des interventions pour le prévenir.

1. *L'apprentissage profond dans les contextes éducatifs*

L'utilisation de méthodes d'apprentissage profond dans le contexte de l'analyse de l'éducation et de l'apprentissage a conduit à un nombre croissant de recherches axées sur une meilleure modélisation du comportement et des performances des étudiants. Dans ce domaine, un grand nombre de ces travaux ont commencé à utiliser les réseaux de neurones récurrents (RNN) [19], pour leur capacité à modéliser des modèles temporels complexes de comportements d'étudiants. Ces modèles se sont révélés très prometteurs dans des travaux récents de modélisation des connaissances et des performances à court terme des étudiants [20] [21] [22], de prédiction de l'obtention du diplôme [15] et des performances en temps réel [16] dans les MOOC, de détection de l'état affectif des étudiants [23] et de prédiction des résultats à long terme [24] [25].

Malgré les performances élevées souvent rapportées de ces modèles lorsqu'ils sont appliqués à leurs tâches respectives dans l'éducation, le grand nombre de paramètres appris et les structures complexes des modèles les rendent souvent difficiles à interpréter. Bien que cette difficulté s'applique aux

paramètres appris du modèle, cela ne signifie pas que les estimations produites par les modèles sont également ininterprétables et peuvent être utilisées pour explorer le comportement de l'étudiant au fil du temps à des niveaux de granularité fins (voir par exemple [26]). Quelque chose d'aussi simple que d'observer les estimations elles-mêmes, ou même la performance du modèle, au fil du temps, peut conduire à une meilleure compréhension des comportements modélisés ainsi que du moment où il est préférable d'intervenir.

La grande complexité des structures des réseaux profonds permet au modèle d'apprendre des incorporations de caractéristiques riches, soit explicitement.

(par exemple, [27]) ou implicitement (par exemple, [25]), qui décrivent mieux les données afin de faire des estimations de modèle mieux informées. De cette façon, ces modèles permettent également l'application de l'apprentissage par transfert.

28] pour mieux comprendre la relation entre les résultats d'intérêt en fournissant les moyens d'observer comment les caractéristiques apprises se généralisent à travers les tâches de prédiction.

1. DATASET

Les données utilisées dans le cadre de ce travail sont constituées d'étudiants travaillant avec ASSISTments au cours de l'année scolaire 2016-2017. AS- SISTments est une plateforme d'apprentissage en ligne qui fournit aux enseignants les outils nécessaires pour assigner des devoirs en classe ou à domicile pour lesquels les étudiants reçoivent un retour immédiat sur la correction [29]. Lors de l'exécution de chaque devoir, de nombreux problèmes fournissent aux élèves une aide facultative à la demande fournie par l'ordinateur ; des indices, dont le nombre peut aller de 0 à plusieurs, fournissent aux élèves un message pédagogique, tandis que l'échafaudage, lorsqu'il est disponible, décompose le problème en petites étapes à résoudre. En plus de cela, le système fournit pour chaque problème un indice "de sortie" qui fournit aux étudiants la réponse correcte si l'étudiant n'est pas en mesure de résoudre le problème, car les étudiants ne sont pas autorisés à passer aux problèmes suivants tant que la réponse correcte n'a pas été entrée dans le système ASSIST- ments.

ASSISTments est utilisé quotidiennement par plusieurs milliers d'étudiants distincts, dont la plupart sont en 6ème- 8ème année et résolvent principalement des contenus mathématiques, fournissant un ensemble de données d'une échelle et d'une variation suffisantes pour appliquer les méthodes d'apprentissage profond qui nécessitent souvent de telles données. Bien que la majorité des élèves soient d'âge moyen, l'ensemble de données est composé de tous les utilisateurs du système pendant l'année scolaire susmentionnée. Les données sont filtrées pour n'inclure que l'interaction des étudiants avec les devoirs basés sur la maîtrise, appelés "skill builders" dans le système, où le seuil d'achèvement est désigné pour exiger simplement que les étudiants répondent correctement à trois problèmes consécutifs sans l'utilisation d'une aide fournie par l'ordinateur (c.-à-d., sans conseils, échafaudage ou conseils de bas de page). Reconnaissant que le WHEEL SPINNING est un comportement d'apprentissage indésirable, le système met en œuvre une " limite quotidienne ", arrêtant les élèves sur le devoir de construction de compétences pour la journée si le seuil d'achèvement n'est pas atteint au dixième problème (sauf dans le cas où l'élève est sur le point d'atteindre le seuil au dixième problème ou directement après) ; le système donne à l'élève l'instruction de chercher une aide supplémentaire et de revenir au devoir le jour suivant.

Étant donné que les enseignants qui utilisent le système assignent un éventail de contenus, aussi bien ceux qui sont mis à disposition par le système que ceux qui sont construits

par les élèves eux-mêmes, nous incluons les données des devoirs de construction de compétences pour lesquels au moins 10 élèves ont commencé le devoir et le taux d'achèvement global est d'au moins 70 %. Ces limitations permettent d'éliminer les cas aberrants tels que les classes types et les devoirs supplémentaires facultatifs que l'enseignant n'exige pas de tous les élèves. Ces cas aberrants sont exclus car nous estimons que l'attrition due à de tels facteurs ne constitue pas un abandon tel que nous l'avons défini dans cette tâche (par exemple, une faible persistance improductive).

|  |  |
| --- | --- |
| Nom de la fonction | Description |
| Type d'action | Encodage à un coup de l'action (tentative, demande d'aide, etc.) |
| Nombre de tentatives | Le nombre de tentatives effectuées jusqu'à l'action en cours |
| Nombre d'indices | Le nombre d'indices demandés jusqu'à l'action en cours. |
| Nombre de problèmes | Le nombre de problèmes vus jusqu'à l'action en cours |
| Probabilité d'action | La probabilité de l'action actuelle compte tenu du problème |
| Probabilité d'action compte tenu du nombre d'actions | La probabilité de l'action actuelle étant donné à la fois le problème et le nombre d'actions prises dans le problème. |
| Probabilité de réponse | Lors d'une tentative, la probabilité qu'un élève réponde par la réponse spécifique étant donné le problème. |
| Probabilité de réponse compte tenu du nombre d'actions | Lors d'une tentative, la probabilité qu'un élève réponde par la réponse spécifique étant donné le problème et le nombre d'actions effectuées dans le problème. |
| Probabilité logarithmique  cumulée de la réponse | La probabilité logarithmique cumulative qu'un étudiant réponde par la réponse spécifique au problème. |
| Temps normalisé pris | Le temps écoulé depuis la dernière action, avec un score z pour le type d'action et le problème. |
| Avant-dernier indice utilisé | Si l'avant-dernier indice a été vu avant l'action actuelle. |
| Indice de fond utilisé | Si l'élève a vu le dernier indice (contenant la réponse) avant l'action en cours. |
| Correctness | Correct ou incorrect si l'action courante est une tentative, ou une non-tentative (comme un encodage à 3 valeurs à un coup). |
| 3 actions précédentes | Encodage à un coup décrivant les trois actions précédentes, à l'exclusion de l'action actuelle. |
| Actions actuelles et antérieures 2 | Encodage à un coup décrivant les trois actions précédentes effectuées, y compris l'action actuelle (actuelle et précédente 2) |

TABLEAU I : Description des caractéristiques générées au niveau de l'action.

1. *Caractéristiques*

Les données consistent en des données de niveau d'action enregistrées par le système, décrivant un niveau fin d'interaction avec le contenu. Ainsi, chaque ligne des données décrit une action unique effectuée par un élève concernant la réponse à un problème, ou les tentatives, ainsi que la demande d'indices dans le système, en plus des mesures liées au temps, de la probabilité de chaque réponse (par exemple, l'identification des mauvaises réponses courantes) et des informations de récence (par exemple, les actions précédentes). À partir des 15 caractéristiques générées, un codage à un coup a été appliqué à toutes les caractéristiques catégorielles, ce qui a donné un total de 86 caractéristiques à utiliser comme entrée dans nos modèles. Une brève description de chacune de ces caractéristiques est fournie dans le tableau I.

1. *Étiquettes de dérapage et d'arrêt des roues*

Les étiquettes WHEEL SPINNING et stopout sont appliquées aux données en suivant largement les définitions précédentes de ces comportements, avec toutefois un petit nombre d'exceptions de cas limites qui sont détaillées ici pour éviter toute ambiguïté. Comme nous avons émis l'hypothèse que le WHEEL SPINNING et le stopout sont, respectivement, des représentations d'une persistance improductive élevée et faible, comme souligné dans l'introduction, nous avons défini ces comportements comme mutuellement exclusifs. Le

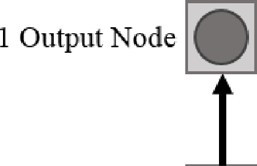
WHEEL SPINNING se produit lorsque les élèves n'ont pas atteint un seuil de compréhension suffisant à la dixième opportunité d'apprentissage ; nous reconnaissons que ce seuil de dix problèmes pour définir le comportement WHEEL SPINNING est plutôt arbitraire (et mériterait peut-être d'être affiné dans des travaux futurs), mais il est utilisé ici par souci de cohérence avec les travaux antérieurs étudiant la roue.

comportement de filature. Comme nous l'avons souligné dans l'introduction, nous considérons qu'il y a abandon uniquement si l'élève ne parvient pas à terminer le travail et abandonne avant le dixième problème. L'attitude manifestée après le dixième problème n'est pas considérée comme un comportement d'abandon, mais plutôt comme un comportement de wheel spinning (puisque le dixième problème a été atteint sans avoir terminé le travail de maîtrise). De cette façon, tout élève ayant dix problèmes ou plus, à moins que l'achèvement n'ait été atteint précisément au dixième item, est étiqueté comme ayant présenté un comportement de WHEEL SPINNING.

Les étiquettes de chaque arrêt et de chaque WHEEL SPINNING sont représentées comme des valeurs binaires distinctes et, bien que calculées au niveau de l'affectation de l'étudiant, elles sont appliquées à chaque ligne de l'ensemble de données. De cette façon, tous les modèles présentés dans cet article prédisent le WHEEL SPINNING et l'abandon à chaque action entreprise par un étudiant, de manière similaire aux estimations au niveau du problème observées dans des travaux antérieurs [5] [7]. Bien que nous ne nous attendions pas à ce que ces modèles aient le même niveau de précision pour toutes les actions, ce niveau de prédiction permettra d'étudier ces performances dans le temps.

|  |  |
| --- | --- |
| Nombre d'étudiants distincts | 12,714 |
| Nombre d'affectations d'étudiants | 123,539 |
| Nombre de rangées (actions) | 1,055,588 |
| Pourcentage d'affectations avec WS | 4.85% |
| Pourcentage d'affectations avec STOPOUT | 4.72% |

TABLEAU II : Les descriptifs notables de l'ensemble de données.



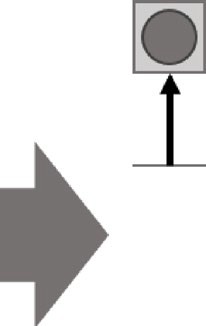
10 CRe c1m en t -

Nœuds cachés - 86 Entrée Nade s-

- •

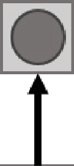
·•

# •-•



.e-+I

# - •



- -

--

•-• •

un réseau **LSTM (**Long-Short Tenn Memory) [30], en plus d'un modèle traditionnel d'arbre de décision et de régression logistique. Les travaux précédents axés sur la prédiction du comportement de WHEEL SPINNING ont utilisé une approche de régression logistique utilisant un large ensemble de caractéristiques techniques [5] [7]. Bien qu'un ensemble de caractéristiques techniques soit également utilisé dans ce travail, les modèles précédents du WHEEL SPINNING ont tenté de modéliser au niveau du problème et ont inclus un plus grand ensemble de caractéristiques contextuelles qui décrivent les performances antérieures pour chaque composante de connaissance, ou compétence, dans l'affectation ; l'ensemble de caractéristiques que nous utilisons ici permet d'obtenir des

résultats plus précis.

dn'iovbesaeurvders léelsèvr**e**psrésentations de chaque comportement au

et les travaux futurs peuvent certainement s'étendre sur ce point

Action des étudiants *t0*

Action de l'étudiant *t1*

s n1dent Action *t2*

pour inclure des caractéristiques plus contextuelles, basées sur le contenu.

Pour chacune des quatre étiquettes appliquées à l'ensemble de données, un modèle distinct de régression logistique, d'arbre de décision et de LSTM est entraîné.

Fig. 1 : Représentation simplifiée de la structure du modèle LSTM, illustrant la manière dont les informations circulent à partir des étapes précédentes pour informer chaque estimation du modèle.

Dans le cadre de ce travail, quatre étiquettes sont appliquées aux données correspondant aux indicateurs intra et inter- affectations. En d'autres termes, un indicateur de WS et de SO au sein de l'affectation (si l'élève présente chaque comportement sur l'affectation actuelle sur laquelle il travaille) est appliqué en plus des indicateurs de WS et SO sur l'affectation suivante. Dans les deux cas, une étiquette est appliquée à chaque ligne des données, correspondant à nouveau à une action unique de l'élève. De cette façon, le comportement de l'étudiant lors de la prochaine affectation sera prédit à partir, par exemple, de la première action de l'affectation précédente, puis de la deuxième action, et ainsi de suite. De même, étant donné qu'il n'y a pas d'indication du sujet du devoir suivant, les modèles de représentations des comportements de WS et de SO d'un devoir à l'autre capturent de manière inhérente les représentations de ces comportements au niveau de l'étudiant (par exemple, sans tenir compte du contenu).

L'ensemble de données résultant, tel que décrit par le tableau II, contient plus de

100 000 devoirs d'étudiants provenant de plus de 12 000 étudiants, ce qui donne lieu à environ 1 million d'actions à utiliser par nos modèles.

intégrées qui sont prédictives d'un résultat (c.-à-d. la rotation de la roue) afin d'observer ensuite comment ces caractéristiques se généralisent pour prédire l'autre résultat (c.-à-d. l'abandon). Cette section détaille les modèles utilisés pour atteindre cet objectif ainsi que l'ensemble des méthodes appliquées pour répondre à nos questions de recherche décrites dans l'introduction.

1. *Modèles de construction du WS et de SO*

Afin de prédire le comportement du WHEEL SPINNING et du stopout au sein de l'affectation et entre les affectations, nous utilisons un type de RNN appelé

1. MÉTHODOLOGIE

Les méthodes utilisées dans cet article visent à répondre aux questions de recherche décrites dans la section d'introduction, centrées sur l'application d'un mode d'apprentissage profond] en conjonction avec l'apprentissage par transfert pour prédire les représentations de la persistance improductive à l'intérieur et entre les affectations. De cette façon, nous développons un modèle d'apprentissage profond récurrent comme moyen d'apprentissage d'un ensemble riche de caractéristiques

pour prédire l'étiquette respective. Pour tous les modèles formés dans ce travail, nous évaluons chacun d'entre eux en utilisant une validation croisée stratifiée de 10 fois au niveau de l'étudiant (en utilisant les mêmes plis dans tous les modèles pour des comparaisons équitables). Étant donné le grand déséquilibre des étiquettes stopout et WHEEL SPINNING (car la plupart des étudiants ne présentent pas un tel comportement par affectation), nous avons stratifié chaque pli en regroupant d'abord les étudiants en fonction du pourcentage d'affectations dans lesquelles chacun présentait un comportement WHEEL SPINNING et stopout, puis en pliant chaque groupe en 10 plis égaux.

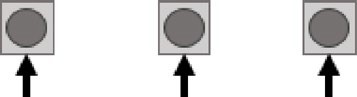
Dans le cas des modèles plus traditionnels d'arbre de décision et de régression logistique, les caractéristiques brutes sont présentées en entrée du modèle, chaque action étant fournie comme un échantillon d'entraînement indépendant ; là encore, le résultat est prédit à chaque action de l'élève dans le système. La performance résultante de chaque modèle est ensuite calculée sur tous les échantillons de chaque pli et la moyenne sur les 10 plis. Les modèles traditionnels ont été implémentés à l'aide de la bibliothèque Scikit-Learn [31] en Python en utilisant les hyperparamètres par défaut, à l'exception de la profondeur maximale de l'arbre de décision qui a été restreinte à 3 niveaux pour éviter un surajustement potentiel ; ces paramètres ont été utilisés pour tous les modèles logistiques et d'arbre de décision décrits dans ce travail.

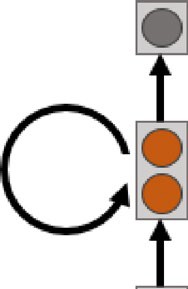
Le modèle LSTM, cependant, en tant que modèle temporel, diffère

légèrement en ce qui concerne la manière dont les échantillons sont présentés au modèle en entrée pendant la procédure de formation. Dans ce cas, les échantillons sont regroupés par affectation des élèves, chaque échantillon représentant une série d'actions entreprises par un élève dans le cadre de chaque affectation. La série complète d'actions d'affectation est présentée au modèle et une série d'estimations (de longueur égale à l'entrée) est produite. De cette façon, le modèle est entraîné comme un modèle de séquence à séquence avec une longueur de séquence dynamique, mais finie (car les étudiants ont complété un nombre variable de problèmes). Le modèle tente d'apprendre les relations temporelles au sein de chaque affectation d'étudiant pour mieux informer ses estimations, mais produit toujours le même nombre de sorties que les modèles traditionnels. De même, comme certaines des caractéristiques représentent une activité récente, la comparaison des modèles aidera à révéler les aspects de ces relations temporelles ; la comparaison des modèles LSTM et traditionnels, par exemple, révélera si l'utilisation de l'historique des performances des étudiants à plus long terme conduit à une meilleure performance du modèle.

Le modèle LSTM a été développé à l'aide de Tensorflow.

[32] en Python avec une structure à 3 couches ; la couche d'entrée comprenait 86 nœuds correspondant à chacune des couches disponibles.

,.. --,



••



**1 1**

**1 1**

**1 1**

1 ..

**1 1**

**1 1**

**1**.. J **1**

!



•• ••

**LSTM**

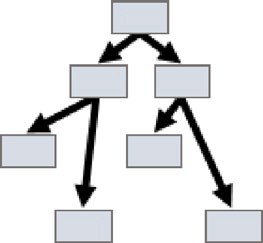
■■■■■■■■■■■■■■■



Fig. 2 : Un exemple visuel de la procédure d'apprentissage par transfert. La couche cachée du mode **LSTM** formé ! est utilisée comme entrée pour former un arbre de décision et une régression logistique afin de prédire le comportement de chaque roue en rotation et de chaque arrêt.

**Régressio n logistîque**

**[i]**



**Arbre de décision**

qui a ensuite été introduite dans une couche cachée de 10 nœuds LSTM, avant de passer à une couche de sortie composée de

1 nœud de sortie auquel est appliquée une fonction d'activation sigmoïde. Un réglage minimal des hyperparamètres a été effectué pour ce réseau afin de réduire les risques de donner un avantage injuste au modèle ; par souci de reproductibilité, le modèle a utilisé une fonction de mise à jour d'Adam [33], une fonction de coût d'entropie croisée, une taille de pas de 0,001, une taille de lot de 32, et a utilisé 20 % de l'ensemble d'apprentissage comme ensemble de validation pour déterminer quand arrêter l'apprentissage du modèle.

1. *Apprentissage par transfert*

Une fois que chacun des modèles a été construit et évalué pour prédire le comportement de rotation de la roue et d'abandon au sein de l'affectation et entre les affectations, nous appliquons une approche d'apprentissage par transfert pour étudier la relation entre ces constructions. Nous avons émis l'hypothèse que la rotation de la roue et le comportement d'abandon sont deux mesures extrêmes de la persistance improductive. En utilisant l'apprentissage par transfert, nous pouvons tester cette hypothèse, que les deux mesures sont étroitement liées, en observant comment les prédicteurs d'un comportement se transfèrent pour prédire l'autre comportement.

Pour cette tâche, nous utilisons le modèle LSTM comme base de la méthode d'apprentissage par transfert. En tant que réseau récurrent, la structure permet au modèle d'apprendre un riche ensemble de caractéristiques qui tentent d'utiliser les relations temporelles complexes dans les données pour faire des estimations mieux informées à chaque étape temporelle ; ce riche ensemble de caractéristiques est stocké dans la couche cachée du réseau et, bien que non directement interprétable, cet ensemble de caractéristiques est appris pendant le

processus d'apprentissage du modèle. Ce développement de caractéristiques intégrées est bien étudié dans d'autres modèles d'apprentissage profond, tels que ceux utilisés pour le traitement des images [34] [35]. Le mode LSTM !, bien qu'il ne permette pas d'identifier les lignes et les formes comme c'est le cas dans les modèles d'apprentissage profond.

Le modèle LSTM est entraîné comme un modèle séquence à séquence (c'est-à-dire plusieurs à plusieurs), ce qui permet d'extraire un ensemble de caractéristiques pour chaque étape temporelle et de les présenter ensuite comme entrée dans un modèle séparé. Le modèle LSTM est formé comme un modèle de séquence à séquence (c'est-à-dire many-to-many), ce qui permet d'extraire un ensemble de caractéristiques pour chaque étape temporelle et de les présenter ensuite comme entrée dans un modèle séparé. C'est de cette façon que le transfert se produit, lorsque le LSTM apprend un ensemble de caractéristiques dans sa couche cachée qui sont ensuite transférées à un autre modèle qui observe une tâche de prédiction différente. Par exemple, comme il y a 10 nœuds dans la couche cachée du **LSTM,** le modèle apprend 10 caractéristiques à partir de la séquence précédente de caractéristiques au niveau de l'action (voir le tableau **1)** qui distinguent les étiquettes positives et négatives de la variable dépendante (c'est-à-dire soit l'arrêt, soit la rotation de la roue)

; les 10 caractéristiques sont ensuite extraites pour chaque pas de temps et utilisées comme entrée pour l'arbre de décision ou le mode de régression logistique ! Une représentation simplifiée de ce processus est illustrée à la figure 2. Les modèles de régression logistique et d'arbre de décision sont ensuite entraînés à prédire l'arrêt ou le WHEEL SPINNING à chaque pas de temps (c'est-à-dire à chaque action de l'élève), en utilisant les caractéristiques transférées du mode LSTM !

Avec cette méthodologie, quatre ensembles de modèles d'apprentissage par transfert sont comparés pour chaque étiquette d'affectation interne et externe de WHEEL SPINNING et de stopout. Ces quatre ensembles comparent différentes combinaisons de caractéristiques, acquises par la formation du modèle LSTM pour prédire le comportement WHEEL SPINNING ou stopout, et chaque résultat. Tout d'abord, les caractéristiques apprises par le modèle LSTM pour prédire le comportement de WHEEL SPINNING au sein de l'affectation, appelées ci-après " caractéristiques de WHEEL SPINNING ", sont présentées à un arbre de décision mode ! et à une régression logistique pour prédire le comportement de WHEEL SPINNING au sein de l'affectation

; cette tâche nous permet d'identifier d'abord toute différence potentielle de performance causée par le transfert de mode !

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DT | | LR | | LSTM | |
| Caractéristiques | AUC | RMSE | AUC | RMSE | AUC | RMSE |
| Brut | 0.847 | 0.327 | 0.511 | 0.437 | 0.887 | 0.313 |
| LSTM - FILAGE DES ROUES | 0.87 | 0.318 | 0.887 | 0.313 | —- | —- |
| LSTM - Stopout | 0.679 | 0.388 | 0.708 | 0.39 | —- | —- |

Modèle de classe majoritaire RMSE : 0.482

TABLEAU III : Prévision de l'effacement de la roue dans l'affectation actuelle

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DT | | LR | | LSTM | |
| Caractéristiques | AUC | RMSE | AUC | RMSE | AUC | RMSE |
| Brut | 0.706 | 0.224 | 0.46 | 0.275 | 0.759 | 0.223 |
| LSTM - FILAGE DES ROUES | 0.71 | 0.224 | 0.683 | 0.226 | —- | —- |
| LSTM - Stopout | 0.747 | 0.223 | 0.757 | 0.222 | —- | —- |

Modèle de classe majoritaire RMSE : 0.234

TABLEAU IV : Prévision de l'arrêt dans l'affectation actuelle

comme couche de sortie du LSTM). Ensuite, les caractéristiques du WHEEL SPINNING sont à nouveau présentées à un arbre de décision différent et à un modèle de régression logistique qui sont ensuite entraînés pour prédire l'abandon au sein de l'affectation. Le troisième ensemble de modèles observe ensuite, à l'inverse, comment les caractéristiques d'arrêt, apprises par le modèle LSTM entraîné à prédire l'arrêt au sein de l'affectation, sont transférées à un arbre de décision et à un modèle de régression logistique pour prédire à nouveau l'arrêt au sein de l'affectation. Enfin, le quatrième ensemble de modèles utilise les caractéristiques d'arrêt dans un arbre de décision et une régression logistique pour prédire le WHEEL SPINNING. Il est important de préciser que ce travail ne tente pas de faire des comparaisons entre les caractéristiques intra-affectation transférées pour prédire les résultats de l'affectation suivante.

est la plus élevée. Par conséquent, la mesure tient compte de la rareté de la classe positive. La valeur est comprise entre 0 et 1, les valeurs les plus élevées indiquant un meilleur ajustement du modèle. Les valeurs proches de 0,5 indiquent que le modèle fonctionne de manière similaire au hasard.

1. RÉSULTATS
2. *Métriques*

Nous comparons les résultats à l'aide de deux mesures principales, l'AUC et la RMSE, ainsi que le rappel, dans le cas de l'observation des performances du modèle dans le temps. Il y a plusieurs avantages à utiliser cette gamme particulière de mesures pour évaluer chaque modèle, en particulier dans le cas de la modélisation du WHEEL SPINNING et du stopout où il y a un grand déséquilibre entre les étiquettes (la plupart des étudiants ne présentent pas de tels comportements). Dans de tels cas de déséquilibre, les modèles de classe majoritaire ont tendance à apparaître comme performants même si aucune distinction entre les classes n'est apprise. Pour éviter que les modèles formés ne produisent une erreur faible en biaisant leurs estimations vers la classe majoritaire, nous utilisons l'AUC pour évaluer l'ajustement du modèle.

L'utilisation de l'AUC évalue la façon dont un modèle distingue les échantillons positifs des échantillons négatifs ; étant donné une instance de la classe positive et de la classe négative, l'AUC peut être considérée comme la probabilité que la classe positive soit celle dont l'estimation de probabilité

Alors que l'AUC évalue la capacité du modèle à distinguer les classes, la RMSE identifie la distance de chaque estimation (en termes d'erreur) par rapport à l'étiquette réelle

; la métrique est calculée en utilisant la probabilité à valeur continue de chaque classe telle que produite par le modèle et en la comparant à l'étiquette de la vérité terrain. De cette façon, le modèle pénalise l'indécision du modèle. Par exemple, si, pour un ensemble d'étiquettes positives et négatives, le modèle produit toutes les estimations de 0,1 et 0,09 respectivement, l'AUC indiquerait un ajustement parfait du modèle tandis que le RMSE serait comparativement faible (car l'erreur sur les instances positives est très élevée). Cette métrique ne tient toutefois pas compte du biais de la classe majoritaire et doit donc être comparée à la valeur RMSE d'un modèle de classe majoritaire. La valeur de la RMSE est comprise entre 0 et 1 dans ce cas (car toutes les estimations sont comprises dans cette plage et les étiquettes sont des valeurs binaires), les valeurs les plus faibles indiquant une meilleure performance du modèle.

Enfin, nous indiquerons également une valeur de rappel lorsque nous observerons la performance des modèles d'assignation suivante WHEEL SPINNING et d'assignation suivante stopout au fil du temps. Le rappel, en tant que mesure de la précision de l'étiquette positive (pour tous les cas positifs, combien le modèle en a-t-il identifié avec succès), aide à identifier la performance du modèle dans l'identification des cas positifs de WHEEL SPINNING et de stopout. Ceci est particulièrement important, encore une fois, en raison du grand déséquilibre, car il fournit un moyen d'évaluer la capacité des modèles à identifier les cas de comportement d'arrêt et de WHEEL SPINNING. L'inconvénient de cette mesure est qu'elle nécessite la définition d'un seuil d'arrondi, et comme il est probable que les estimations soient biaisées vers la classe majoritaire, un seuil d'arrondi de la moyenne de sortie du modèle est utilisé plutôt que l'utilisation plus traditionnelle de 0,5 ; en d'autres termes, les valeurs supérieures à la moyenne sont arrondies vers le haut pour identifier un cas positif de WHEEL SPINNING ou d'arrêt et les estimations inférieures à la moyenne sont arrondies vers le bas pour identifier un cas négatif de l'une ou l'autre mesure. La valeur du rappel est également délimitée entre 0 et 1, les valeurs les plus élevées indiquant une meilleure performance du modèle.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DT | | LR | | LSTM | |
| Caractéristiques | AUC | RMSE | AUC | RMSE | AUC | RMSE |
| Brut | 0.581 | 0.238 | 0.539 | 0.273 | 0.600 | 0.251 |
| LSTM - Prochaine affectation PISTAGE DE ROUE | 0.595 | 0.250 | 0.601 | 0.250 | - — | - — |
| LSTM - Arrêt de l'assignation suivante | 0.570 | 0.251 | 0.569 | 0.251 | - — | - — |

Modèle de classe majoritaire RMSE : 0.246

TABLEAU V : Prévision de l'essorage de la roue dans la prochaine mission

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DT | | LR | | LSTM | |
| Caractéristiques | AUC | RMSE | AUC | RMSE | AUC | RMSE |
| Brut | 0.545 | 0.209 | 0.492 | 0.25 | 0.557 | 0.221 |
| LSTM - Prochaine affectation PISTAGE DE ROUE | 0.547 | 0.221 | 0.548 | 0.221 | - — | - — |
| LSTM - Arrêt de l'assignation suivante | 0.553 | 0.221 | 0.557 | 0.221 | - — | - — |

Modèle de classe majoritaire RMSE : 0.215

TABLEAU VI : Prévision de l'arrêt dans la prochaine mission

1. *Performance du modèle*

Nos résultats sont enregistrés de telle sorte que chacun des tableaux III à VI enregistre les résultats d'une variable de résultat. Le tableau III décrit les différents modèles qui ont été construits pour prédire si un étudiant va faire une roue dans le devoir en cours. Le premier modèle a été construit en utilisant les caractéristiques brutes (c'est-à-dire les caractéristiques originales de l'ensemble de données telles qu'elles sont énumérées dans le tableau I). Nous constatons que le modèle LSTM est le plus performant, avec une AUC de 0,887 et une RMSE de 1,5.

0.313. Il est ensuite suivi par le modèle d'arbre de décision avec une AUC de 0,847 et une RMSE de 0,327. Le modèle de régression logistique n'est pas performant avec une faible AUC de 0,511, à peine meilleure que la performance aléatoire. Les deuxième et troisième modèles du tableau III sont construits en utilisant l'apprentissage par transfert, où nous utilisons la couche cachée apprise du LSTM formé pour prédire le WHEEL SPINNING. Ses caractéristiques apprises sont utilisées comme entrée pour les modèles d'arbre de décision et de régression logistique. Cette expérience démontre la capacité des caractéristiques apprises à être transférées entre les modèles et à être généralisées à de nouveaux résultats. Le résultat principal est que les deux modèles s'améliorent lorsqu'ils sont formés à l'aide des caractéristiques découvertes par le LSTM : les arbres de décision voient une légère amélioration de l'AUC et du RMSE, tandis que la régression logistique s'améliore considérablement. Nous voyons que le modèle LSTM- Régression logistique avec une AUC de 0,887 (RMSE 0,313) est plus performant que le modèle LSTM-Arbre de décision avec une AUC de 0,87 et RMSE 0,318. Nous pouvons observer que le transfert du modèle LSTM sur le modèle de régression logistique aboutit à la même AUC du LSTM avec des caractéristiques brutes, ce qui n'est pas surprenant puisque la couche de sortie du LSTM est essentiellement un modèle de régression logistique. Le dernier modèle est un autre modèle d'apprentissage de transfert, dans lequel le LSTM qui a été

construit pour prédire l'arrêt dans l'affectation actuelle est utilisé pour transférer ses caractéristiques apprises à un arbre de décision et un modèle de régression logistique pour prédire le WHEEL SPINNING. Les résultats sont mitigés, l'arbre de décision présentant peu d'avantages par rapport à l'utilisation des caractéristiques brutes, tandis que la régression logistique est plus performante que les caractéristiques brutes. Il est intéressant de constater que la régression logistique s'est améliorée même lorsque des caractéristiques extraites pour

une activité d'apprentissage différente. Nous observons que ces deux modèles ont donné de bons résultats avec une AUC de 0,679 et une RMSE de 0,388 dans le cas du modèle LSTM-arbre de décision et une AUC de 0,708 et une RMSE de 0,39 pour le modèle LSTM-régression logistique.

Comme dans le tableau III, le tableau IV enregistre la performance du modèle pour prédire si un étudiant va abandonner dans le devoir en cours. L'ordre est similaire à celui du tableau III : dans la première ligne, les caractéristiques originales ont été utilisées pour ajuster l'arbre de décision, la régression logistique et le modèle LSTM. Le modèle LSTM semble être le plus performant avec une AUC de 0,759 et une RMSE de 0,223, suivi par les modèles d'arbre de décision et de régression logistique avec des AUC de 0,706 et 0,46, respectivement. Le deuxième modèle est le premier des modèles d'apprentissage par transfert visant à prédire l'abandon au sein de l'assignation. Les caractéristiques apprises par le LSTM pour prédire le WHEEL SPINNING ont été transférées en entrée d'un arbre de décision et d'un modèle de régression logistique pour prédire l'abandon. Malgré l'utilisation de caractéristiques apprises pour une tâche de prédiction différente, l'arbre de décision et la régression logistique ont montré des performances améliorées par rapport à l'utilisation des caractéristiques brutes. Pour les arbres de décision, l'avantage est léger avec une augmentation triviale de l'AUC. Cependant, la régression logistique a montré un gain de performance important avec une AUC passant de 0,46 à 0,683 et une RMSE passant de 0,275 à 0,226. Enfin, en utilisant le modèle LSTM qui a été construit pour prédire l'arrêt en utilisant les caractéristiques brutes, nous avons transféré ses caractéristiques apprises comme entrée à un arbre de décision et à un modèle de régression logistique pour prédire la même étiquette d'arrêt. Les deux modèles montrent une amélioration par rapport à l'utilisation des caractéristiques d'arrêt du LSTM. L'arbre de décision et la régression logistique présentent tous deux des gains de performance notables en AUC, avec des gains plus faibles en RMSE.

Le tableau V décrit les résultats du modèle construit pour prédire si un élève va faire une roue dans la *prochaine* affectation, plutôt que dans l'affectation actuelle. En utilisant les caractéristiques originales, le LSTM présente une AUC de 0,600 (RMSE 0,251), suivi de l'arbre de décision avec une AUC de 0,581 (RMSE 0,238), puis de la régression logistique avec une AUC de 0,539 (RMSE 0,273). Le site

La deuxième ligne décrit la performance des modèles d'apprentissage par transfert du LSTM construit pour prédire le WHEEL SPINNING dans la prochaine affectation. Ce modèle LSTM - arbre de décision avait une AUC de 0,595 (RMSE 0,250) tandis que le modèle LSTM - régression logistique présentait une AUC de 0,601 (RMSE 0,250). De même, le LSTM construit pour prédire le prochain arrêt d'affectation est utilisé pour construire des modèles d'apprentissage par transfert avec l'arbre de décision et la régression logistique pour la tâche de prédiction de la prochaine affectation WHEEL SPINNING. Ces modèles ont abouti à une AUC de 0,570 (RMSE 0,251) pour le modèle d'arbre de décision transféré et à une AUC de 0,570 pour le modèle de régression logistique.

0,569 (RMSE 0,251) pour le modèle de régression logistique. Une fois encore, nous observons le schéma général selon lequel les caractéristiques apprises donnent une meilleure précision que les caractéristiques brutes. Pour la régression logistique, même les caractéristiques construites pour un arrêt réussissent à surpasser les caractéristiques brutes, bien que ce résultat ne soit pas valable pour l'arbre de décision.

Le tableau VI décrit les modèles construits pour prédire si un étudiant va abandonner dans le prochain devoir. Suivant la structure similaire des tableaux précédents, les caractéristiques originales ont été utilisées pour construire un arbre de décision, un modèle de régression logistique et un modèle LSTM. Les résultats sont loin d'être aussi forts que les prédictions à court terme pour le devoir en cours, mais ils sont tout de même supérieurs au hasard, ce qui souligne peut-être la difficulté d'identifier ce comportement dès le devoir précédent sans informations contextuelles sur le contenu du devoir suivant. Le LSTM semble à nouveau être le plus performant des trois avec une AUC pas si élevée de 0,557 (RMSE 0,221). Il est suivi par l'arbre de décision avec une AUC de 0,545 (RMSE 0,209) et le modèle de régression logistique avec une AUC inférieure au hasard de 0,492 (RMSE 0,25). Après les caractéristiques brutes, nous utilisons ce qui a été appris par le modèle LSTM construit pour prédire le WHEEL SPINNING dans la prochaine affectation pour transférer son apprentissage à un arbre de décision et à un modèle de régression logistique. Ces modèles ont donné des AUC de 0,547 (RMSE 0,221) et 0,548 (RMSE 0,221),

respectivement. Nous observons qu'il y a peu de différences entre les deux modèles. Ensuite, nous utilisons le modèle LSTM formé pour prédire le prochain arrêt d'affectation pour transférer son apprentissage à un modèle d'arbre de décision et de régression logistique pour prédire le même label d'arrêt d'affectation, ce qui donne des AUC de

0,553 (RMSE 0,221) et 0,557 (RMSE 0,221) respectivement.

Il est important de répéter que chaque modèle prédit l'étiquette respective à chaque pas de temps. En d'autres termes, chaque comportement est prédit à chaque action de l'élève. C'est probablement pour cette raison que certains modèles présentent des valeurs d'AUC proches du hasard ; la faible performance du modèle de régression logistique dans le tableau III, par exemple, et la performance élevée de l'arbre de décision, suggère que les étiquettes positives et négatives du comportement ne sont pas linéairement séparables en utilisant uniquement les caractéristiques brutes et qu'elles nécessitent plus d'informations (telles que les caractéristiques temporelles

fournies par le LSTM) afin de présenter une meilleure performance.

1. *Observer la performance des modèles par opportunité*

En plus d'observer la performance du modèle en moyenne sur toutes les estimations, nous observons également comment la performance du modèle change à chaque opportunité d'apprentissage, ou problème, lors de la prédiction de chaque mesure de résultat. En observant la performance de ces modèles à chaque opportunité d'apprentissage, nous pouvons commencer à

déterminer à quel moment de l'affectation précédente nous sommes susceptibles de détecter des indicateurs de persistance improductive à l'avenir ; cela peut ensuite aider à

1) identifier les causes ou les facteurs potentiels qui peuvent être corrélés avec la persistance improductive future et 2) commencer à comprendre non seulement quand mais aussi quel type d'intervention peut être déployé pour soutenir les comportements d'apprentissage productifs.

Comme les données sont représentées sous la forme d'une série d'actions d'élèves, nous prenons d'abord la performance moyenne du modèle dans chaque problème d'élève et nous traçons cette performance sur les dix premiers problèmes des devoirs d'élèves, comme le montre la figure 3. Comme le nombre d'étudiants présents à chaque occasion change en raison du comportement des étudiants qui abandonnent ou qui terminent effectivement le devoir, il est important d'inclure également des intervalles de confiance car chaque valeur sera mesurée de manière moins précise à chaque occasion suivante. Dans le cas de la RMSE, cet intervalle de confiance est calculé en calculant la racine carrée des limites supérieure et inférieure des erreurs standard calculées à partir des erreurs au carré entre les estimations à chaque occasion. Dans le cas du rappel, les limites de confiance sont calculées en utilisant un intervalle de score de Wilson [36] pour la valeur de rappel calculée à chaque opportunité. Les limites de confiance pour l'AUC sont calculées à l'aide de pROC [37], un package R open source.

Nous représentons la performance du modèle pour chaque rotation de roue et arrêt de l'affectation suivante, estimée à l'aide du modèle LSTM sans apprentissage par transfert, dans les Figures 4 et 6 respectivement ; nous les comparons ensuite à la performance du modèle pour chaque rotation de roue et arrêt de l'affectation suivante, représentée dans les Figures 3 et 5 respectivement. Il est important de souligner, comme cela a été décrit dans la section sur les métriques, que des valeurs RMSE plus faibles indiquent une meilleure performance du modèle, tandis que des valeurs de rappel et d'AUC plus élevées sont indicatives d'une meilleure performance du modèle ; de cette façon, bien que le RMSE et le rappel, par exemple, présentent une tendance générale à la hausse sur chaque opportunité ultérieure, les métriques sont contradictoires dans leur tendance de performance du modèle. Ce cas particulier observé dans la figure 4 suggère donc que, bien que le modèle soit capable d'identifier correctement un plus grand nombre d'élèves susceptibles de faire la roue à la fin du devoir précédent, le modèle est moins précis dans sa capacité à le faire. Ceci est confirmé par la diminution de l'AUC observée dans cette figure, où le modèle est susceptible de mal étiqueter les élèves qui ne font pas la roue lors du prochain devoir.

Lors de la prédiction de la prochaine affectation WHEEL SPINNING, comme l'illustre la figure 4, la RMSE du modèle est la plus faible sur les trois premières occasions d'affectation des élèves. Cela n'est pas très surprenant car, puisque le seuil d'achèvement des devoirs est de répondre correctement à trois problèmes consécutifs, un grand nombre d'étudiants répondront probablement correctement aux trois premiers problèmes et achèveront effectivement le devoir. De tels étudiants, bien que dépendant certainement du contenu, sont probablement moins susceptibles de faire preuve de

WHEEL SPINNING lors de futurs devoirs que les étudiants présentant des difficultés au début du devoir ; les étudiants qui n'apprennent pas efficacement le matériel sont susceptibles de lutter pour apprendre des compétences ultérieures qui peuvent nécessiter la maîtrise du contenu antérieur. La performance du modèle, en termes de RMSE, diminue ensuite régulièrement après la troisième opportunité, car il est probable que les estimations soient biaisées vers la valeur la plus basse.

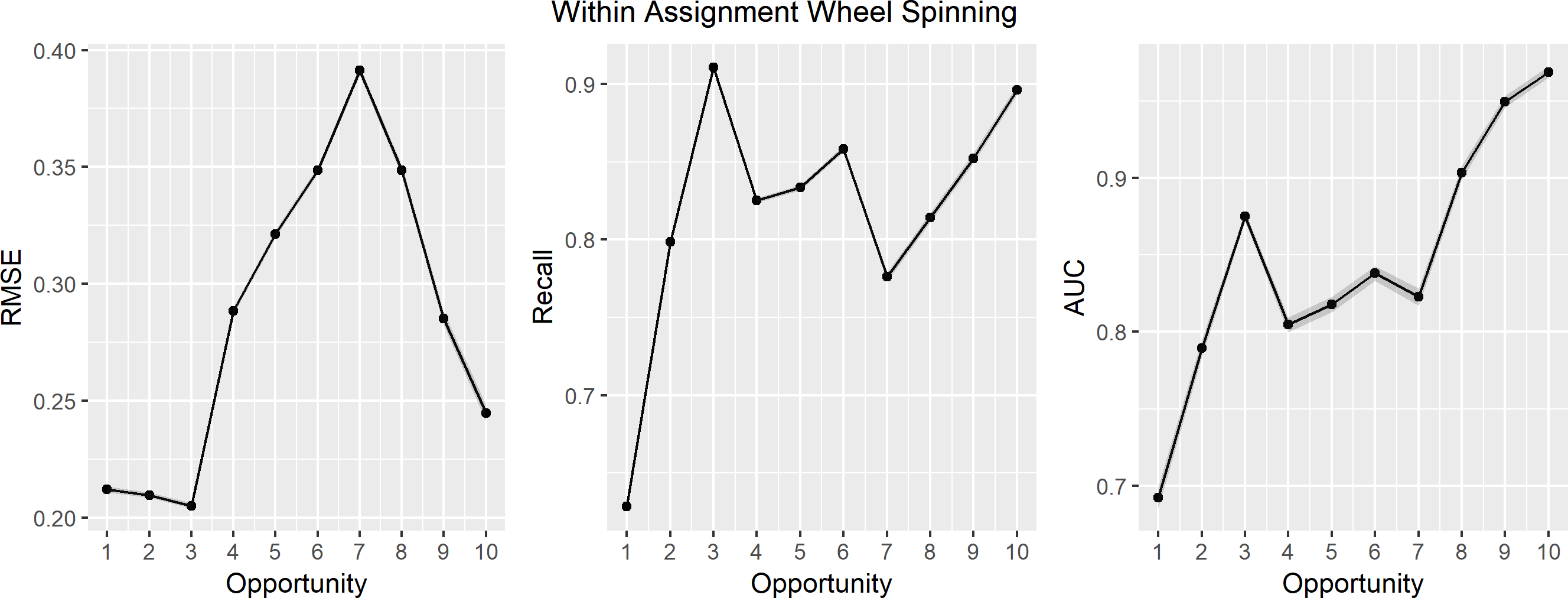


Fig. 3 : La performance du modèle LSTM dans la prédiction de la rotation de la roue à l'intérieur de l'affectation par opportunité.

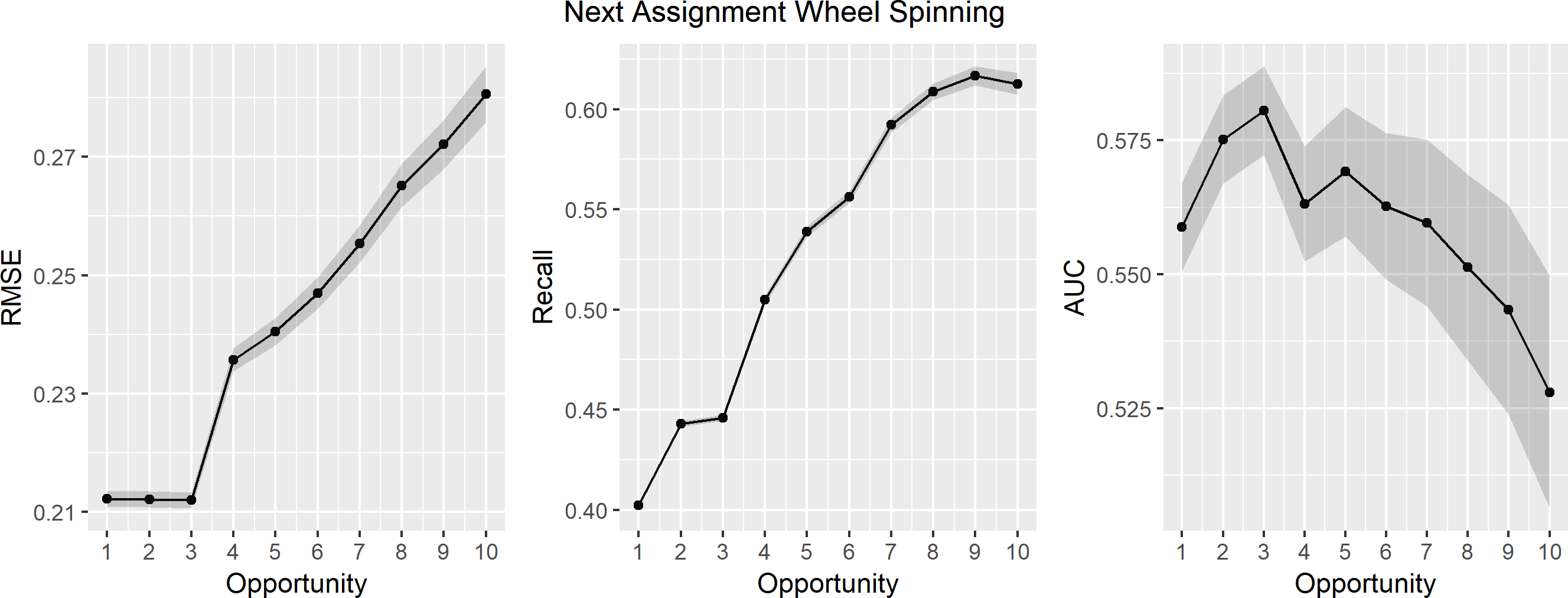


Fig. 4 : La performance du modèle LSTM dans la prédiction de la prochaine affectation WHEEL SPINNING par opportunité.

classe majoritaire. Cependant, en ce qui concerne le rappel et l'AUC, le modèle s'améliore régulièrement avec chaque opportunité suivante, ce qui suggère que, bien que peut-être biaisé vers la classe majoritaire, le modèle est capable d'identifier plus efficacement les futurs cas de comportement de WHEEL SPINNING au fur et à mesure que les élèves restent dans l'affectation. Le rappel du modèle semble se stabiliser vers la fin de l'exercice.

Les résultats suggèrent qu'à la fin de l'exercice, le modèle est capable d'identifier 60 % des élèves qui font du " WHEEL SPINNING " pour l'exercice suivant (sans même savoir quel sera le contenu de l'exercice). On peut supposer que le modèle identifie simplement les cas où les élèves qui font preuve de MOUVEMENT dans le devoir en cours sont plus susceptibles de le faire dans les devoirs suivants, d'autant plus que les élèves qui restent dans le devoir à la dixième occasion font du MOUVEMENT (sauf s'ils ont terminé le dixième item selon notre définition du comportement).

Dans un sens, cela suggère que, de manière peu surprenante, une intervention visant à prévenir le WHEEL

SPINNING lors d'une affectation ultérieure est susceptible d'avoir le plus d'impact au premier signe de comportement potentiel de WHEEL SPINNING lors de l'affectation actuelle.

d'affectation. Dans le cas de nos résultats, cela semble être autour de la troisième opportunité d'apprentissage, comme l'illustrent le rappel et la métrique de la figure 3. Dans cette figure, la troisième opportunité présente à la fois le rappel le plus élevé, ce qui suggère que le modèle est capable d'identifier les cas où le WHEEL SPINNING est présent à la fin de l'affectation, et le RMSE le plus faible, ce qui, même avec un biais de classe majoritaire, est l'opportunité où toutes les métriques s'accordent généralement pour présenter de bonnes performances du modèle.

La performance du modèle LSTM dans la prédiction de l'arrêt de l'as- signation suivante, telle que décrite dans la figure 6, illustre une tendance similaire à celle du modèle WHEEL SPINNING. Bien que présentant une variation sensiblement plus élevée, la RMSE du modèle d'arrêt est la plus faible au cours des trois premières opportunités d'apprentissage et augmente régulièrement lors des opportunités suivantes. Le rappel présente à nouveau une tendance contradictoire, avec les pires performances au cours des trois premières occasions, puis une augmentation substantielle des performances après la troisième occasion, identifiant correctement environ 59% des étudiants qui abandonnent lors du prochain devoir. Par les larges bornes de confiance de l'AUC, cependant,

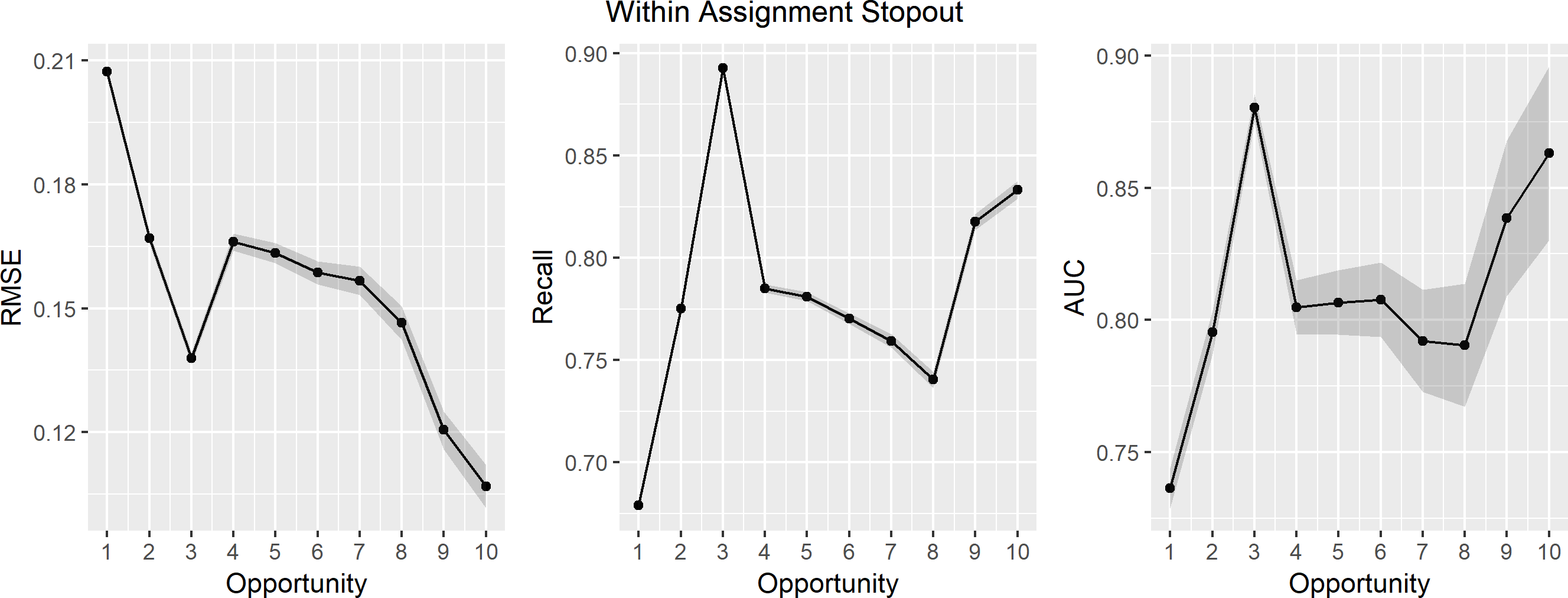


Fig. 5 : La performance du modèle LSTM dans la prédiction de l'arrêt de la mission par opportunité.

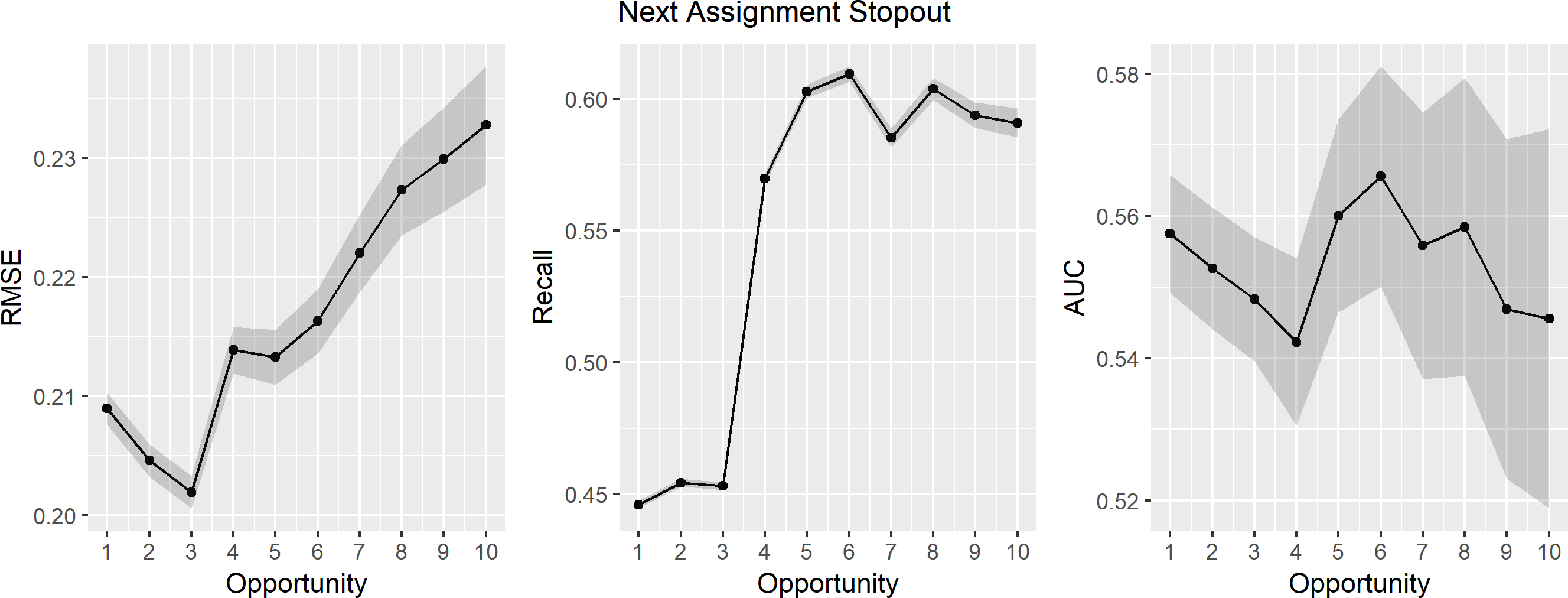


Fig. 6 : La performance du modèle LSTM dans la prédiction de l'arrêt de la prochaine affectation par opportunité.

il semblerait que, à l'instar de l'AUC du modèle WHEEL SPINNING de l'affectation suivante illustré à la figure 4, le modèle ait des difficultés à distinguer les étudiants susceptibles de présenter chacun de ces comportements à l'avenir.

En observant la performance de ce modèle au sein de l'affectation dans la Figure 5, on peut toutefois identifier une autre tendance intéressante. Comme pour le modèle WHEEL SPINNING, les métriques semblent concorder en termes de meilleures performances du modèle à la troisième occasion. Cependant, le RMSE s'améliore régulièrement et les mesures du rappel et de l'AUC diminuent un peu régulièrement après ce point. Cette tendance presque inverse de celle observée pour les performances du modèle WHEEL SPINNING suggère, bien que cela ne soit pas surprenant, que le modèle est incapable de distinguer les étudiants susceptibles d'abandonner et de persister lors des opportunités ultérieures ; selon notre définition, l'abandon ne peut se produire que lors des dix premières opportunités, mais les étudiants présents lors des opportunités ultérieures font preuve de persistance, ce

qui peut être difficile pour le modèle d'identifier le moment où l'abandon se produira dans de tels cas.

1. TRAVAIL FUTUR

Bien que ce travail fasse progresser la compréhension de l'apprentissage par transfert dans la compréhension des performances éducatives, il existe plusieurs questions de suivi intéressantes. Premièrement, nous avons constaté que la régression logistique bénéficie généralement de l'apprentissage par transfert, alors que les résultats pour les arbres de décision étaient plus mitigés. Cette tendance est- elle générale ou particulière à notre ensemble de données et de caractéristiques ? De même, comment se comporteraient d'autres classificateurs tels que les forêts aléatoires ou les arbres de décision ? Bénéficieraient-ils ou non des caractéristiques construites ? La première étape de l'exploration de l'apprentissage par transfert est utile, mais le domaine doit mieux comprendre dans quelles circonstances les caractéristiques seront transférées à de nouveaux apprenants.

Le deuxième domaine d'investigation concerne les différents avantages que les apprenants de transfert obtiennent. Lorsque les caractéristiques étaient alignées avec la tâche, par exemple les caractéristiques d'arrêt pour prédire l'arrêt, les arbres de décision et la régression logistique ont tous deux montré des avantages. Cependant, lorsque les caractéristiques étaient moins alignées, par exemple, les caractéristiques de WHEEL SPINNING utilisées pour prédire l'arrêt, les résultats étaient plus mitigés. Il y a plusieurs questions à poser dans ce domaine.

Premièrement, quelle est l'étendue de l'application des ensembles de caractéristiques apprises ? Montreraient-ils une amélioration par rapport aux caractéristiques brutes prédisant des tâches moins connexes, comme l'affect de l'apprenant ? Deuxièmement, est-il possible d'entraîner un réseau neuronal avec des sorties multiples pour l'encourager à apprendre des caractéristiques plus largement applicables (par exemple, par l'apprentissage multi-tâches [38]) ? De cette façon, un domaine de recherche important pourrait consister à former des réseaux sur une variété de sorties et à utiliser les caractéristiques apprises pour une variété de nouveaux sujets de recherche. En retirant les humains de la génération des caractéristiques, on risque d'obtenir des caractéristiques moins interprétables, mais on pourrait aussi obtenir des modèles plus précis et de nouvelles caractéristiques que nous n'avons pas encore découvertes manuellement.

Le dernier domaine qui nous semble digne d'intérêt est la compréhension de la forte baisse de performance entre la prédiction de l'ensemble de problèmes actuel WHEEL SPINNING et stopout, et la prédiction de l'ensemble de problèmes suivant WHEEL SPINNING et stopout. Une partie de la baisse de performance est fondamentale pour toute tâche de prédiction : les prédictions à plus long terme sont plus incertaines que celles concernant les événements à court terme. Dans quelle mesure la baisse de performance est-elle une limitation fondamentale, et dans quelle mesure est-elle due au fait qu'il n'y a pas autant d'antériorité dans les prédictions à long terme ? Est-il possible d'augmenter la précision des ensembles de problèmes ultérieurs jusqu'à une AUC de 0,7 en améliorant la construction des caractéristiques ou le choix des modèles, ou existe-t-il des limites fondamentales à la précision de la prédiction des performances des étudiants ?

1. CONTRIBUTIONS ET CONCLUSIONS

Cet article apporte deux contributions en ce qui concerne l'apprentissage par transfert. Premièrement, nous avons constaté que dans certains cas, l'apprentissage par transfert fonctionne mieux que les caractéristiques originales. Nous avons été surpris de constater que les caractéristiques apprises par la machine, conçues pour fonctionner avec un réseau neuronal, étaient applicables à un arbre de décision. Étant donné les formes de modèles identiques, il était moins surprenant que les caractéristiques améliorent les performances des modèles de régression logistique. La deuxième contribution est que l'apprentissage par transfert fonctionne (parfois) pour des tâches non identiques. En utilisant les caractéristiques LSTM-stopout pour pré-diriger WHEEL SPINNING, et vice versa, les performances se sont améliorées pour les modèles de régression logistique et parfois pour les modèles d'arbre de décision. Cette constatation démontre qu'il est possible de construire automatiquement des caractéristiques applicables à de nouvelles tâches de prédiction.

Cet article apporte également des contributions en ce qui concerne la prédiction d'événements à plus long terme. Les travaux antérieurs sur la modélisation des élèves se sont concentrés sur les événements immédiats, comme la prédiction des performances de l'élève sur le problème en

cours. Les travaux ultérieurs ont allongé l'intervalle de prédiction pour voir comment un élève se comporterait sur un ensemble de problèmes, qui était composé de nombreux problèmes. Ce travail augmente l'intervalle temporel pour prédire les performances d'un élève sur le prochain ensemble de problèmes. À bien des égards, ce travail représente une augmentation plus importante que le passage d'un problème actuel à un ensemble de problèmes actuel, car dans les deux cas, le modèle prédictif dispose d'informations sur les performances de l'élève pour cette compétence. Pour prédire le prochain ensemble de problèmes, le modèle n'est pas sûr de la manière dont l'élève va se comporter dans cette compétence. La tâche de prédiction est donc comparativement plus difficile.

En conclusion, cet article se concentre sur la mise en place d'un système d'alerte précoce permettant de prédire quels élèves auront des difficultés. Fournir de l'aide et des ressources d'apprentissage supplémentaires aux étudiants qui ont des difficultés d'apprentissage fait partie intégrante de tout système d'apprentissage.

Il est essentiel d'identifier les élèves qui vont avoir des difficultés pour les aider ; plus tôt nous savons si un élève va faire une roue ou s'arrêter, mieux nous pouvons lui apporter l'aide appropriée. Il vaut mieux prévenir que guérir, de même qu'il est préférable d'empêcher l'élève de faire un tête-à- queue ou de s'arrêter que de lui fournir des remèdes plus tard. D'après nos résultats, nous pouvons dire que nos modèles sont bons pour identifier le comportement d'abandon et de WHEEL SPINNING à partir des actions des étudiants dans l'affectation actuelle. Nos modèles nous permettent de comprendre la persistance des étudiants sous la forme du WHEEL SPINNING et du stopout. En utilisant ces concepts, nous pouvons essayer de faire persister les élèves plus longtemps s'ils ne persistent pas assez longtemps. Ou nous pouvons les empêcher de persévérer si nous constatons qu'ils ont des difficultés depuis longtemps. Nous pouvons utiliser ces modèles pour intervenir à un stade précoce de l'affectation, par exemple lorsque le modèle détecte le comportement après une action effectuée par l'élève. Si le modèle prédit que l'élève va faire la roue, nous pouvons arrêter de lui proposer d'autres problèmes pour la journée. Au lieu de cela, nous pourrions diriger l'élève vers une ressource d'apprentissage telle que des notes de cours ou une vidéo. De la même manière, si le modèle prédit qu'un élève va s'arrêter, nous pouvons essayer de réduire la difficulté des problèmes afin que l'élève prenne confiance dans la résolution des problèmes au lieu de s'arrêter. En utilisant les détecteurs de comportements pour le prochain devoir, nous détectons les étudiants vulnérables un devoir plus tôt.

ACCUSÉ DE RÉCEPTION

Nous remercions les multiples subventions actuelles de la NSF (ACI-1440753, DRL-1252297, DRL-1109483, DRL- 1316736, DGE- 1535428, OAC-1724889,OAC-1636782 & DRL-1031398), le US Dept. of Ed (IES R305A180401, R305A120125 & R305C100024 et GAANN), et l'ONR.

RÉFÉRENCES

1. R. D. Roscoe et M. T. Chi, " Understanding tutor learning : Knowledge- building and knowledge-telling in peer tutors explanations and ques- tions", *Review of Educational Research*, vol. 77, no. 4, pp. 534-574, 2007.
2. A. L. Duckworth, C. Peterson, M. D. Matthews, et D. R. Kelly, "Grit : perseverance and passion for long-term goals". *Journal of personality and social psychology*, vol. 92, no. 6, p. 1087, 2007.
3. C. Peterson, M. E. Seligman *et al.* , *Character strengths and virtues : A handbook and classification*. Oxford University Press, 2004, vol. 1.
4. M. Kapur, " Productive failure ", *Cognition and instruction*, vol. 26, no. 3, pp. 379-424, 2008.
5. J. E. Beck et Y. Gong, "Wheel-spinning : Students who fail to master a skill ", dans *International Conference on Artificial Intelligence in Educa- tion*. Springer, 2013, p. 431-440.
6. N. Matsuda, S. Chandrasekaran, et J. C. Stamper, "How quickly can WHEEL SPINNING be detected ?" dans *EDM*, 2016, pp. 607-608.
7. Y. Gong et J. E. Beck, " Towards detecting wheel-spinning : Future failure in mastery learning ", in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 2015, p. 67-74.
8. T. Ka¨ser, S. Klingler, et M. Gross, "When to stop? : towards universal instructional policies", in *Proceedings of the Sixth International Confer- ence on Learning Analytics & Knowledge.* ACM, 2016, p. 289-298.
9. D. S. Chaplot, E. Rhim, et J. Kim, " Predicting student attrition in moocs using sentiment analysis and neural networks ", in *AIED Workshops*, 2015.
10. W. Xing, X. Chen, J. Stein et M. Marcinkowski, " Temporal predi-

cation of dropouts in moocs : Reaching the low hanging fruit through stacking generalization ", *Computers in Human Behavior*, vol. 58, pp. 119-129, 2016.

1. D. Yang, T. Sinha, D. Adamson, et C. P. Rose', "Turn on, tune in, drop out : Anticipating student dropouts in massive open online courses ", dans *Proceedings of the 2013 NIPS Data-driven education workshop*, vol. 11, 2013, p. 14.
2. C. P. Rose', R. Carlson, D. Yang, M. Wen, L. Resnick, P. Goldman, et

J. Sherer, "Social factors that contribute to attrition in moocs," in *Pro- ceedings of the first ACM conference on Learning@ scale conference*. ACM, 2014, p. 197-198.

1. A. Lamb, J. Smilack, A. Ho et J. Reich, "Addressing common analytic challenges to randomized experiments in moocs : Attrition and zero- inflation," in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 2015, p. 21-30.
2. R. F. Kizilcec et S. Halawa, " Attrition and achievement gaps in online learning ", dans *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 2015, pp. 57-66.
3. B.-H. Kim, E. Vizitei, et V. Ganapathi, "Gritnet : Student performance prediction with deep learning ", *arXiv preprint arXiv:1804.07405*, 2018.
4. --, " Gritnet 2 : prédiction en temps réel des performances des étudiants avec adaptation au domaine ", *arXiv preprint arXiv:1809.06686*, 2018.
5. A. Botelho, H. Wan, et N. Heffernan, " The prediction of student first response using prerequisite skills ", dans *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 2015, p. 39-45.
6. J. Heckman, N. Hohmann, J. Smith, et M. Khoo, "Substitution and dropout bias in social experiments : A study of an influential social experiment", *The Quarterly Journal of Economics*, vol. 115, no. 2, pp. 651-694, 2000.
7. R. J. Williams et D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural computation*, vol. 1, no. 2, pp. 270-280, 1989.
8. C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, et

J. Sohl-Dickstein, " Deep knowledge tracing ", dans *Advances in Neural Information Processing Systems*, 2015, pp. 505-513.

1. M. Khajah, R. V. Lindsey et M. C. Mozer, " How deep is knowledge tracing ? ", dans *Proceedings of the 9th International Conference on Educa- tional Data Mining*, 2016, pp. 94-101.
2. X. Xiong, S. Zhao, E. Van Inwegen et J. Beck, " Going deeper with deep knowledge tracing ", dans *Proceedings of the 9th International Conference on Educational Data Mining*, 2016, pp. 545-550.
3. A. F. Botelho, R. S. Baker et N. T. Heffernan, " Improving sensor-free affect detection using deep learning ", dans *International Conference on Artificial Intelligence in Education*. Springer, 2017, p. 40-51.
4. A. C. Sales, A. Botelho, T. Patikorn et N. T. Heffernan, " Using big data to sharpen design-based inference in a/b tests ", dans *Proceedings of the 11th International Conference on Educational Data Mining*, 2018, pp. 479-486,.
5. C.-k. Yeung, Z. Lin, K. Yang, et D.-y. Yeung, "Incorporating features learned by an enhanced deep knowledge tracing model for stem/non- stem job prediction", *arXiv preprint arXiv:1806.03256*, 2018.
6. A. F. Botelho, R. S. Baker, J. Ocumpaugh et N. T. Heffernan, " Studying affect dynamics and chronometry using sensor-free detectors ", dans *Proceedings of the 11th International Conference on Educational Data Mining*, 2018, p. 157-166.
7. L. Zhang, X. Xiong, S. Zhao, A. Botelho, et N. T. Heffernan, "Incor- porating rich features into deep knowledge tracing," in *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*. ACM, 2017, p. 169-172.
8. L. Y. Pratt, "Discriminability-based transfer between neural networks," in *Advances in neural information processing systems*, 1993, pp. 204- 211.
9. N. T. Heffernan et C. L. Heffernan, "The assistments ecosystem : Building a platform that brings scientists and teachers together for min- imally invasive research on human learning and teaching ", *International Journal of Artificial Intelligence in Education*, vol. 24, no. 4, pp. 470- 497, 2014.
10. S. Hochreiter et J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
11. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,

O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vander- plas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, et E. Duch- esnay, "Scikit-learn : Machine learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.

1. M. Abadi, A. Agarwal, P. Barham, ..., et X. Zheng, "TensorFlow : Large-scale machine learning on heterogeneous systems," 2015, logiciel disponible sur tensorflow.org. [En ligne]. Disponible à l'[adresse suivante](http://www.tensorflow.org/)

[:](http://www.tensorflow.org/) https[://www.](http://www.tensorflow.org/)ten[sorflow.org/](http://www.tensorflow.org/)

1. D. P. Kingma et J. Ba, "Adam : A method for stochastic optimization,"

*Préprint arXiv arXiv:1412.6980*, 2014.

1. A. Krizhevsky, I. Sutskever, et G. E. Hinton, " Imagenet classification with deep convolutional neural networks ", dans *Advances in neural infor- mation processing systems*, 2012, pp. 1097-1105.
2. D. Erhan, Y. Bengio, A. Courville, et P. Vincent, " Visualizing higher- layer features of a deep network ", *Université de Montréal*, vol. 1341, no. 3, p. 1, 2009.
3. E. B. Wilson, "Probable inference, the law of succession, and statistical inference", *Journal of the American Statistical Association*, vol. 22, no 158, pp. 209-212, 1927.
4. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, et

M. Muer¨l, "proc : an open-source package for r and s+ to analyze and compare roc curves," *BMC bioinformatics*, vol. 12, no. 1, p. 77, 2011.

1. R. Caruana, "Multitask learning", *Machine learning*, vol. 28, no. 1, pp. 41-75, 1997.

Anthony F. Botelho Anthony est un étudiant en doctorat en sciences et technologies de l'apprentissage au Worcester Polytechnic Institute et travaille en tant que membre de l'équipe ASSISTments pour aider à développer et à effectuer des recherches au sein de la plateforme d'apprentissage en ligne.

Ashvini Varatharaj est étudiante en maîtrise au département d'informatique du Worcester Poly- technic Institute. Elle est assistante de recherche dans le laboratoire ASSISTments et élève de Neil Heffernan.

Thanaporn Patikorn est un étudiant en doctorat de l'Institut polytechnique de Worcester, spécialisé dans l'informatique. Il est assistant de recherche dans le laboratoire ASSISTments et élève de Neil Heffernan.

Diana Doherty a obtenu son diplôme de premier cycle en informatique à l'Institut polytechnique de Worcester à l'automne 2018.

Seth A. Adjei est professeur assistant invité à la Northern Kentucky University. Ses recherches visent à trouver des moyens d'améliorer l'apprentissage des élèves et d'aider les élèves peu performants à bénéficier d'interventions qui améliorent leurs résultats, notamment en mathématiques.

Joseph E. Beck est professeur adjoint de sciences informatiques au WPI. Ses recherches portent sur l'exploration de données éducatives, une nouvelle discipline qui développe des techniques d'analyse de grands ensembles de données éducatives pour faire des découvertes qui amélioreront l'enseignement et l'apprentissage. Son travail est centré sur l'estimation de l'impact des tuteurs sur l'apprentissage.