

Consultez les discussions, les statistiques et les profils des auteurs de cette publication à l'adresse [suivante : https://www.researchgate.net/publication/342675876.](https://www.researchgate.net/publication/342675876_Early_Detection_of_Wheel-Spinning_in_ASSISTments?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_2&_esc=publicationCoverPdf)

[Détection précoce du patinage des roues dans les assistances](https://www.researchgate.net/publication/342675876_Early_Detection_of_Wheel-Spinning_in_ASSISTments?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_3&_esc=publicationCoverPdf)

**Chapitre** - juin 2020

DOI : 10.1007/978-3-030-52237-7\_46

CITATION

1

LECTURES

32



Abonnez-vous à DeepL Pro pour éditer ce document.

Visitez [www.DeepL.com/Pro](http://www.DeepL.com/Pro) pour en savoir plus.

**3 auteurs**, dont :

[Ryan Baker](https://www.researchgate.net/profile/Ryan-Baker-2?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_5&_esc=publicationCoverPdf)

[Université de Pennsylvanie](https://www.researchgate.net/institution/University_of_Pennsylvania?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_6&_esc=publicationCoverPdf)

**428** PUBLICATIONS **11 619** CITATIONS

[VOIR LE](https://www.researchgate.net/profile/Ryan-Baker-2?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_7&_esc=publicationCoverPdf)

**Certains des auteurs de cette publication travaillent également sur ces projets connexes :**



Métacognition des élèves (par exemple, confiance dans les études) [Voir leprojet](https://www.researchgate.net/project/Students-metacognition-eg-academic-confidence?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_9&_esc=publicationCoverPdf)

Évaluation analytique de la présence cognitive par l'apprentissage [Voir le projet](https://www.researchgate.net/project/Learning-analytics-assessment-of-cognitive-presence?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_9&_esc=publicationCoverPdf)

Tout le contenu de cette page a été téléchargé par [RyanBaker](https://www.researchgate.net/profile/Ryan-Baker-2?enrichId=rgreq-76bdf7c7042581e12712c0112f86a070-XXX&enrichSource=Y292ZXJQYWdlOzM0MjY3NTg3NjtBUzo5MTE0MDQyOTkyMTg5NTFAMTU5NDMwNzEyNjg4MQ%3D%3D&el=1_x_10&_esc=publicationCoverPdf) le 09 juillet 2020.

L'utilisateur a demandé l'amélioration du fichier téléchargé.

**Détection précoce du WHEEL SPINNING dans les assistances**

Yeyu Wang1 [0000-0003-1978-5453], Shimin Kai2 et Ryan Shaun Baker3 [0000-0002-3051-3232].

1 Université du Wisconsin-Madison, Madison, WI 53706, USA

[ywang2466@wisc.edu](mailto:ywang2466@wisc.edu)

2 Université de Columbia, New York, NY 10027, USA

3 Université de Pennsylvanie, Philadelphie, PA 19104, USA

**Résumé.** La persistance est un trait crucial pour les apprenants. Cependant, un problème commun à l'apprentissage par la maîtrise est que la persistance n'est pas toujours productive, une construction appelée WHEEL SPINNING. Dans cet article, nous nous appuyons sur des travaux antérieurs pour développer des détecteurs de WHEEL SPINNING dans le système d'apprentissage ASSISTments qui distinguent la non-persistance, la persistance productive et le WHEEL SPINNING. Pour comprendre la rapidité avec laquelle nous pouvons détecter chaque état, nous utilisons des données provenant de différents nombres d'opportunités de pratique et nous comparons la performance du modèle sur les paires d'étudiants et d'ensembles de problèmes. Nous constatons qu'un modèle construit à l'aide des données des neuf premières opportunités de pratique surpasse les modèles utilisant moins de données de pratique. Cependant, il est possible de différencier les élèves qui finiront par faire du WHEEL SPINNING qui persisteront de manière productive en utilisant uniquement les données des trois premières opportunités. Il est possible de différencier le WHEEL SPINNING de la non-persistance dès les cinq premières occasions, et la non- persistance peut être différenciée de la persistance productive dès les sept premières occasions. Ces résultats montrent qu'une différenciation précoce entre le WHEEL SPINNING et la persistance productive est possible. Ces détecteurs s'appuient sur les demandes d'indices, la justesse des occasions antérieures et la quantité de pratique et de temps consacré à la compétence. L'identification des caractéristiques prédictives permet de comprendre l'impact des comportements internes au système sur le WHEEL SPINNING et de guider la conception du système.

**Mots clés :** WHEEL SPINNING *-* Persistance *-* Arbre de décision *-* Détection précoce *-* Système de tutorat intelligent.

# Introduction

## Persistance et non-persistance dans l'apprentissage

Ces dernières années, la recherche s'est concentrée sur le développement de compétences non cognitives pour améliorer l'apprentissage des élèves, telles que la résilience et la persistance pendant l'apprentissage. La persistance est définie comme la capacité de maintenir une action ou de terminer une tâche, quelle que soit l'inclinaison de la personne à l'égard de cette tâche [[5,7].](#_bookmark7) Des études récentes ont montré que la persévérance dans les contextes éducatifs est associée à la réussite scolaire [[3,19], à la](#_bookmark19) créativité [[20]](#_bookmark20) et à des résultats scolaires à long terme tels que la poursuite de la scolarité [21].

et les revenus futurs [[6,19].](#_bookmark19) Cependant, la persistance n'est pas toujours positive. Des chercheurs de l'Université d'Ottawa [[2]](#_bookmark5) ont affirmé qu'une certaine persistance peut être improductive, ou WHEEL SPINNING, définie comme le fait de passer trop de temps à se battre sans atteindre la maîtrise. La définition du WHEEL SPINNING varie selon les études et les textes d'apprentissage. [2] a](#_bookmark5) défini le WHEEL SPINNING comme le fait de ne pas atteindre la maîtrise même après avoir tenté de résoudre 10 problèmes ou plus dans un ensemble de problèmes ; [[14]](#_bookmark14) a fait appel à deux évaluateurs humains pour coder les comportements de WHEEL SPINNING de manière qualitative sur la base d'un manuel de codage, avec un Kappa de Cohen de 0,9. D'autre part, [[10] a](#_bookmark10) défini le WHEEL SPINNING comme le fait de tenter plus de 10 problèmes mais de ne pas réussir à obtenir trois réponses correctes consécutives ou de ne pas démontrer la rétention ultérieure de l'habileté.

La non-persistance, ou le fait d'abandonner la tâche d'apprentissage en cours sans maîtriser les connaissances requises, a également été documentée dans plusieurs environnements d'apprentissage assistés par ordinateur. Par exemple, dans le jeu éducatif Physics Play-ground, la non-persistance a été définie comme le fait de quitter le niveau sans avoir réussi à résoudre le problème en utilisant les connaissances en physique [[11,](#_bookmark11)[12].](#_bookmark12) Dans le système d'apprentissage ASSISTments, [[4]](#_bookmark6) s'est penché sur les comportements de non- persistance dans lesquels les étudiants quittent la série de problèmes sans avoir atteint la maîtrise d'une compétence, en faisant la différence entre l'abandon immédiat et l'abandon après avoir tenté quelques problèmes. Dans le même système d'apprentissage, [[10] a](#_bookmark10) défini la non-persistance comme le fait de tenter moins de dix problèmes pour une compétence, mais n'a pas considéré la détection de la non-persistance dans son travail.

## Détection de la persistance dans l'apprentissage

La détection des comportements de WHEEL SPINNING est importante pour identifier les élèves qui peuvent avoir besoin d'un soutien supplémentaire pendant une tâche d'apprentissage. Comme la persistance est généralement définie par le nombre d'occasions de pratiquer une tâche d'apprentissage, certaines approches de modélisation ou de détection du WHEEL SPINNING ont été conçues pour ne fonctionner qu'une fois que le système a recueilli des données sur les élèves pour un nombre suffisamment important d'occasions de pratiquer. Par exemple, [[2], la](#_bookmark5) première étude sur le WHEEL SPINNING, indique que le WHEEL SPINNING pouvait être détecté dès la huitième occasion de pratique dans le système ASSISTments par un modèle de régression logistique. Une étude de suivi a affiné ce modèle et a permis de détecter le WHEEL SPINNING dès la septième occasion d'entraînement [[8].](#_bookmark8) D'autres méthodes d'apprentissage automatique, telles que les réseaux neuronaux [[14], le](#_bookmark14) gradient boosting [[17]](#_bookmark17) et la forêt aléatoire [18], ont également été utilisées.

[[22]](#_bookmark22) ont également été utilisés pour permettre la détection du WHEEL SPINNING à des stades plus précoces de la pratique. Plus particulièrement, [[4]](#_bookmark6) a été en mesure d'identifier les étudiants WHEEL SPINNING à leur troisième occasion, en appliquant des réseaux neuronaux récurrents à mémoire à long terme. Bien que ces études se déroulent toutes dans un environnement ITS, des travaux ont également été réalisés sur la détection du WHEEL SPINNING dans les jeux éducatifs. [16] ont](#_bookmark16) construit un modèle permettant de détecter le WHEEL SPINNING sur la base des caractéristiques développées au cours des 5 premières minutes, des 10 premières minutes et des 15 premières minutes de jeu, et [[15]](#_bookmark15) ont construit un modèle permettant de différencier le WHEEL SPINNING de la persistance productive dans une séquence de jeux de mathématiques.

En examinant ces travaux antérieurs, nous constatons que pour qu'un détecteur

de WHEEL SPINNING soit pratique pour une utilisation en temps réel, il y a deux critères importants à considérer. Premièrement, un détecteur doit être capable de différencier le WHEEL SPINNING des autres types de véhicules.

Les enseignants et les concepteurs de systèmes devraient être en mesure de détecter le plus tôt possible les états de persistance (qu'il s'agisse de persistance réussie ou de non-persistance non réussie) ainsi que de persistance productive. Grâce à la détection précoce de ces états, les enseignants et les concepteurs de systèmes peuvent avoir plus de possibilités de créer des interventions pour améliorer l'expérience d'apprentissage des élèves qui risquent de persister de manière improductive ou d'abandonner prématurément sans avoir terminé une tâche d'apprentissage. Deuxièmement, les prédictions basées sur des modèles interprétables, tels que les arbres de décision, offriront aux enseignants et aux concepteurs de systèmes un aperçu plus utile des facteurs qui influencent la persistance et le WHEEL SPINNING Les travaux antérieurs n'ont pas encore répondu pleinement à ces deux critères. Actuellement, la plupart des détecteurs ne prennent en compte que la prédiction binaire, soit en éliminant les cas de non-persistance [[17],](#_bookmark17) soit en traitant tous les cas qui ne sont pas du WHEEL SPINNING comme étant acceptables [[8]. En même temps, l](#_bookmark8)es efforts récents pour améliorer la prédiction de l'essorage en utilisant le boosting de gradient ou les réseaux neuronaux ont amélioré la vitesse et la qualité de la prédiction au prix de l'interprétabilité, ce qui pose un défi aux chercheurs en éducation pour découvrir et comprendre l'impact des comportements d'apprentissage sur l'essorage.

Dans cet article, nous tentons de répondre à chacune de ces limitations. Nous

1) concevons des détecteurs multi-classes qui distinguent les trois catégories d'états mentionnées ci-dessus - non persistance, persistance productive et persistance improductive (WHEEL SPINNING) - afin de saisir et de comparer les comportements spécifiques qui diffèrent à la fois entre les étudiants persistants et non persistants et entre les étudiants persistants et non persistants. 2) explorer le nombre minimum d'occasions de pratique qui pourraient être utilisées avec une précision raisonnable pour détecter les différents états de persistance dans ces conditions, et en déduire des caractéristiques spécifiques qui peuvent être traduites en interventions pratiques. Ce faisant, afin de comparer nos résultats avec les travaux précédents sur les détecteurs binaires de WHEEL SPINNING, prédisant le WHEEL SPINNING par rapport au non-WHEEL SPINNING [[2,4,17],](#_bookmark17) nous construisons également des modèles pour faire des comparaisons par paire pour deux classes sur les trois. En outre, nous résumerons les caractéristiques prédictives utilisées dans les modèles basés sur différentes possibilités de pratique, afin de promouvoir une meilleure compréhension du WHEEL SPINNING. Nous concluons en discutant de l'impact possible des caractéristiques sur la persistance et la persistance improductive dans l'apprentissage.

# Méthodes

## AIDES

ASSISTments est une plateforme d'apprentissage en ligne gratuite qui fournit un retour d'information immédiat aux élèves et une évaluation formative des performances des élèves aux enseignants [[9].](#_bookmark9) Dans le système ASSISTments, les Skill Builders sont un type d'ensemble de problèmes mathématiques où les élèves s'exercent à résoudre des problèmes générés de manière aléatoire, basés sur des modèles existants et correspondant à la même compétence [[9].](#_bookmark9) Dans un générateur de compétences, les élèves ne peuvent pas passer au problème suivant tant qu'ils n'ont pas donné la bonne réponse. Des conseils sont disponibles pour les aider à résoudre les problèmes. Pour chaque problème, les élèves peuvent faire plusieurs tentatives et demander plusieurs conseils. En général, il y a deux ou trois niveaux de conseils par problème, suivis d'un conseil en bas de page qui

fournit la bonne réponse.

réponse finale. Les élèves doivent répondre correctement à trois questions consécutives pour compléter un ensemble de problèmes. Ils passent ensuite un test à une seule question après un certain temps - généralement une semaine plus tard, mais les enseignants peuvent le configurer - avec un espacement progressivement croissant entre les réévaluations. Ce test comprend un élément sélectionné de manière aléatoire à partir d'un modèle dans l'ensemble de problèmes complété, et est fourni par le système de réévaluation et de réapprentissage automatique (ARRS) [[21]. L'](#_bookmark21)objectif principal de l'ARRS est d'évaluer la rétention d'une compétence par un élève au fil du temps. Si l'étudiant ne répond pas correctement à cette question, et échoue donc dans la rétention des compétences, il se verra attribuer l'ensemble de problèmes Skill Builder correspondant pour réapprendre les matériaux.

## Collecte de données et génération d'étiquettes

Notre ensemble de données de recherche est l'ensemble de données ASSISTments Skill Builders de l'année scolaire 2014-2015, accessible au public, qui comprend 26 522 élèves qui ont tenté 1 088 ensembles de problèmes Skill Builder sur une année. Chaque enregistrement de l'ensemble de données représente une paire élève-ensemble de problèmes, qui comprend les données du journal lorsqu'un apprenant s'exerce à un ensemble de problèmes Skill Builder. Cet ensemble de données a été choisi en raison de son utilisation dans des recherches antérieures sur le WHEEL SPINNING et la persistance (c'est-à-dire [[10]).](#_bookmark10) Nous avons ensuite construit huit nouveaux ensembles de données : *first-3*, *first-4*, . . . *first-9*, et *first-10* (*first-1* et *first-2* n'ont pas été générés, car il n'y avait pas assez de données pour déduire le WHEEL SPINNING dans les travaux précédents). Chaque ligne de l'un de ces ensembles de données *first-x* présente des données agrégées sur l'apprentissage d'un élève dans un certain ensemble de problèmes (c'est-à-dire une paire élève-ensemble de problèmes), où *x* est le nombre seuil de problèmes sur lequel les données sont agrégées. Par exemple, *first-3* ne contient que des données sur les 3 premiers problèmes que l'élève a tenté dans chaque ensemble de problèmes, tandis que *first-4*, *first-5* et *first-6* contiennent des données sur les 4, 5 et 6 premiers problèmes respectivement. Il faut noter que, pour un ensemble de problèmes, un élève qui n'a tenté que 3 problèmes sera inclus dans *first-3* mais pas dans *first-4* à *first-10*, alors qu'un élève qui a complété 10 problèmes sera inclus dans chaque ensemble de données de *first-3* à *first-10*. De manière plus générale, le nombre de paires élèves/ensemble de problèmes diminue lorsque *x* augmente, car il y a moins d'élèves qui tentent plus de problèmes.

**Tableau 1.** Critères de non-persistance (NP), de persistance productive (PP) et de filage au rouet (WS) dans le système Skill Builder.

|  |  |  |  |
| --- | --- | --- | --- |
| Définition | Trois Corrections d'affilée (Maîtrise)  à partir du 10ème problème | PremierARRS  Test | Dix ou plus  Problèmes |
| NP | Tout | Tout | Non |
| PP | Oui | Adopté | Oui |
| WS | Non  Oui | Tout  Non | Oui |

Ensuite, nous avons étiqueté chaque ligne de la paire élève/ensemble de problèmes comme étant soit une persistance productive (PP), soit une persistance en roue libre (WS), soit une non-persistance (NP), en fonction des

critères suivants

définitions opérationnelles dans [[10]](#_bookmark10) (tableau [1).](#_bookmark0) Si un élève a fait moins de 10 problèmes dans une série de problèmes, la paire élève-série de problèmes correspondante est étiquetée NP. Sinon, la paire est étiquetée PP si l'élève a atteint la maîtrise (c'est-à-dire s'il a obtenu trois réponses correctes d'affilée et a réussi le test ARRS) ou WS dans le cas contraire. Bien que nos définitions concernent le test ARRS, certains élèves n'ont pas été affectés à ce test même après avoir obtenu trois réponses correctes d'affilée parce que les enseignants ont désactivé la fonction ARRS. Ces cas, qui représentent 211 612 paires sur les 287 093 paires initiales élèves/ensemble de problèmes, ont été considérés comme hors sujet et retirés de l'analyse. Parmi les paires restantes de problèmes d'élèves, 6 855 ont été classées comme WS et 2 093 comme PP ; ces paires sont présentes dans chaque ensemble de données premier-x mais prennent des valeurs de caractéristiques différentes en fonction de *x*. Le nombre de paires NP dans les ensembles de données du *premier-3* au *premier-10* est respectivement de 51866, 33197, 26983, 12663, 7833, 4290, 1900 et 0. Comme indiqué précédemment, il y a moins d'enregistrements NP lorsque *x* augmente ; l'ensemble de données *first-10*, en particulier, n'a pas

d'enregistrements NP.

parce que les étudiants qui ont atteint le 10ème problème ont été considérés comme persistants.

## Ingénierie des caractéristiques et apprentissage automatique

Nous nous sommes basés sur l'ensemble des fonctionnalités développées par [[1],](#_bookmark4) qui consiste en des actions et des attributs de l'élève au sein de la plateforme ASSISTments Skill Builder qui fournit des informations sur la persistance et l'apprentissage de l'élève. Plus précisément, nous avons inclus 25 caractéristiques de base liées à l'utilisation d'indices par les élèves, au nombre d'occasions de s'exercer à un ensemble de problèmes, au nombre d'occasions d'acquérir des compétences et au délai entre les actions des élèves. Comme dans [[10],](#_bookmark10) nous avons calculé les valeurs respectives de la somme, du minimum, du maximum, de la moyenne et de l'écart-type de ces attributs de base pour chaque séquence d'élèves et avons généré 125 caractéristiques basées sur 25 caractéristiques de base. Ensuite, nous avons construit un ensemble de modèles pour distinguer les NP, PP et WS. Chaque modèle est basé sur l'un des ensembles de données de la *première année*. Ce processus se compose de trois étapes principales :

**Fractionnement des données.** Nous avons effectué un fractionnement de chaque ensemble de données de *première instance en* un ensemble de formation-validation (90% des étudiants) et un ensemble de test (10% des étudiants).

**Sélection des caractéristiques.** Pour chaque valeur de *x*, nous avons effectué une sélection de caractéristiques en boucle externe sur l'ensemble train- validation. Cette routine commence avec un ensemble de caractéristiques vide et, à chaque étape, sélectionne la caractéristique qui génère la meilleure performance, selon le résultat de la validation croisée. Pour réduire la suradaptation, nous avons fixé le nombre maximal de caractéristiques à 20 et imposé une condition d'arrêt précoce : si la prochaine caractéristique candidate ne produit pas une amélioration de la performance de plus de 0,001, la routine s'arrête.

**Évaluation du modèle.** Nous avons construit un modèle basé sur les caractéristiques de l'étape précédente, et l'avons entraîné sur l'ensemble train- validé. Nous avons ensuite évalué le modèle sur l'ensemble de test en fonction de l'AUC macro-moyenne et de l'AUC par paire entre NP-WS, PP-WS et NP-PP. De cette façon, nous nous sommes assurés qu'aucune donnée n'était utilisée à la fois pour la sélection des caractéristiques et l'évaluation du modèle, ce qui aurait biaisé les résultats.

Dans les étapes ci-dessus, notre mesure de performance est l'AUC validée par recoupement 10 fois. En raison d'un déséquilibre entre les classes WS, NP et PP, nous avons suréchantillonné les données de formation en ajoutant de manière aléatoire des copies d'enregistrements des classes minoritaires. Pour mesurer la qualité du modèle, nous avons adopté la macro-moyenne de l'AUC pour les modèles multi-classes.

prédiction. Enfin, pour comparer nos résultats avec ceux du détecteur binaire [de](#_bookmark10) [[10],](#_bookmark10) qui différencie les états persistants productifs des états persistants WHEEL SPINNING, nous avons choisi l'implémentation de l'arbre de décision de [[18].](#_bookmark18) Nous avons utilisé l'entropie comme critère de division, fixé la profondeur maximale de l'arbre à 12 et le nombre minimal d'instances par feuille à 2. Bien que ces hyperparamètres puissent être réglés individuellement pour chaque modèle de jeu de données afin d'obtenir de meilleures performances, notre objectif est d'utiliser le même processus de construction de modèle pour les huit jeux de données afin de comparer leurs performances ainsi que les caractéristiques saillantes de chacun, et d'éviter le surajustement associé au réglage des hyperparamètres.

Parmi les 125 caractéristiques, certaines ont été calculées sur la base des actions des élèves dans un certain nombre de problèmes antérieurs. *Past8BottomOut* et *Past8HelpRequest*, par exemple, se réfèrent au nombre d'indices de sortie et de demandes d'aide faites au cours des 8 derniers problèmes. Nous avons supprimé ces caractéristiques du processus de sélection des caractéristiques sur les ensembles de données où elles ne sont pas applicables

- pour cet exemple, les ensembles de données *first-3* à *first-8*, qui ne comprennent pas de données sur les compétences des élèves pour plus de 8 problèmes.

# Résultats

## Résultats de la sélection des caractéristiques

En appliquant l'algorithme de sélection avancée des caractéristiques, nous avons identifié les ensembles de caractéristiques qui maximisent les performances du modèle pour chaque ensemble de données *first-x*. Parmi les huit modèles d'arbre de décision, six ont des caractéristiques de nœud racine liées à l'utilisation d'indices, telles que la moyenne (*first-3*, *first-7*, *first-8* ) et la somme (*first-4*, *first-6* ) du nombre total d'indices utilisés, et la moyenne de l'indice le plus bas demandé dans les huit dernières occasions (*first-10* ). Les nœuds racines des deux autres modèles sont des facteurs temporels, tels que la somme (*first-5* ) et la moyenne (*first-9* ) de la durée depuis la dernière fois que l'élève a pratiqué la compétence. Bien que chaque ensemble de données ait son propre ensemble de caractéristiques, nous avons observé qu'il y avait des caractéristiques communes à tous les ensembles de données. Pour mieux représenter ces points communs, nous avons résumé toutes les caractéristiques sélectionnées en sept catégories, qui comprennent le type de question dans l'ensemble de problèmes, les comportements de demande d'aide, l'utilisation d'indices, l'étayage, le nombre d'opportunités, la quantité de pratique et de temps, et le nombre d'opportunités échouées. Dans le tableau [2,](#_bookmark2) nous avons énuméré trois exemples de caractéristiques

les catégories sélectionnées avec leurs descriptions [4.](#_bookmark1) La liste des numéros après chaque

caractéristique

indique pour quels modèles de l'ensemble de données *first-x* il a été sélectionné.

Sur la base de la sélection avancée, tous les modèles, des *trois premiers* aux *dix premiers,* comprennent des caractéristiques liées aux comportements de demande de conseils (*HintTotal* ). De plus, les caractéristiques liées à *HintTotal* sont sélectionnées pour le nœud racine de cinq modèles, ce qui indique que les caractéristiques liées aux demandes de conseils jouent un rôle crucial dans la prédiction des WS, NP et PP. D'autres caractéristiques, comme le nombre d'occasions de s'exercer et le temps imparti, ainsi que le nombre de tentatives erronées faites lors des problèmes précédents, sont présentes dans les

différents modèles. Nous discuterons des implications de ces résultats dans la section discussion.

4 Le tableau complet des caractéristiques sélectionnées pour chaque modèle peut être consulté à l'adresse [https://github.](https://github.com/yeyuw215/AIED_WS_2020/blob/master/FullTable2.pdf) [com/yeyuw215/AIEDWS2020/blob/master/FullTable2.pdf.](https://github.com/yeyuw215/AIED_WS_2020/blob/master/FullTable2.pdf)

**Tableau 2****.** Exemples de caractéristiques, catégories et descriptions sélectionnées.

|  |  |
| --- | --- |
| Caractéristi ques  Catégories | Caractéristiques et descriptions |
| Indice | * **HintTotal (3,4,5,6,7,8,9,10)** : Le nombre total de demandes d'indices. * **Past8BottomOut (9,10)** : Le nombre de demandes d'allusion à la sortie par le bas dans les 8 dernières tentatives. |
| Quantité de pratique et de temps | * **TimeBetweenProblems (5,7,8,9)** : La durée du temps entre les problèmes liés à la compétence. * **Temps pris (3,4,5,6)** : Le temps passé à résoudre le problème en cours. * **TotalSkillOpportunities (5,6,7,8,9)** : Le nombre total de problèmes tentés qui sont liés à la compétence dans l'ensemble de problèmes actuel. * **TotalTimeOnSkill (3,4)** : Le temps total passé sur la   compétence dans le système. |
| Mauva is compt e | * **TotalPastWrongCount (3,4,9)** : Le nombre total d'at- incorrect les tentatives faites sur les problèmes de la série de problèmes en cours. * **TotalPercentPastWrong (4,5)** : Le pourcentage d'at- tentes incorrectes faites sur les problèmes de la série de problèmes actuelle. * **Past5WrongCount (9)** : Le nombre de tentatives effectuées qui étaient   incorrect lors des 5 dernières tentatives. |

**Tableau 3.** Caractéristiques sélectionnées pour chaque modèle de l'ensemble de données *first-x*. La caractéristique de la racine désigne la caractéristique du nœud racine de chaque modèle d'arbre de décision. "*m"* indique que la caractéristique est agrégée en tant que moyenne ; *"s"* indique que la caractéristique est agrégée en tant que somme.

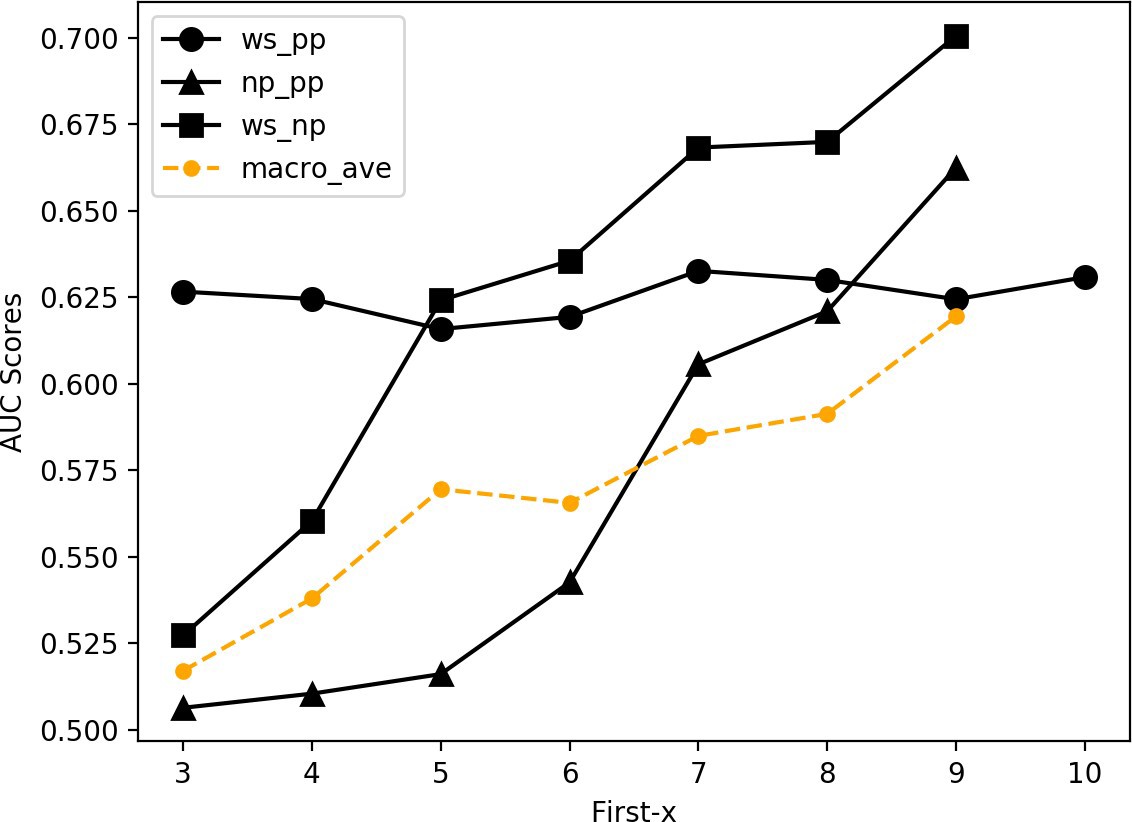
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **premier-3premier-4premier-5premier-6** | | | | | |
| de | Caractéri  stiques | 10 | 12 | 11 | 6 |
| Raci ne | Fonction nalité | m HintTotal | s HintTotal | s TimeBtwProb | s HintTotal |
|  |  | **premier-7** | **premier-8** | **premier-9** | **premier-10** |
| de | Caractéri stiques | 11 | 8 | 10 | 5 |
| Raci ne | Fonction nalité | m HintTotal | m HintTotal | m TimeBtwProb | m P8BottomOut |

## Performance du modèle pour les ensembles de données *"first-x".*

Pour tous les ensembles de données de *première instance,* nous avons appliqué la même procédure de sélection des caractéristiques et d'évaluation des modèles. Afin d'identifier à quel point nous pouvons prédire rapidement le WHEEL SPINNING, nous avons calculé la macro-moyenne de l'AUC (pour les multiples classes de PP, NP et WS) comme mesure de la qualité et nous avons comparé l'amélioration résultant de la prise en compte d'un plus grand nombre de problèmes, ou de possibilités de pratique. Nous avons également calculé l'AUC par paire pour les prédictions WS-PP, NP-PP et WS-NP, afin de comprendre dans quelle mesure le modèle peut différencier des paires d'états spécifiques.

Lorsque l'on prend en compte de plus en plus d'opportunités de pratique, les scores AUC macro-moyens du détecteur multi-classes augmentent progressivement (voir Fig. [1).](#_bookmark3) Le modèle incluant les données des 9 premières occasions a la meilleure performance, avec une AUC macro-moyenne de 0,62. Pour le contraste NP-PP et WS-NP, l'AUC montre une augmentation avec plus de

données d'opportunité de pratique. Pour la prédiction contrastant NP et PP, l'inclusion des données des 9 premières occasions de pratique a la meilleure performance avec une AUC macro-moyenne de 0,62.



**Fig. 1.** Scores de l'AUC pour différents ensembles de données de premier ordre.

L'inclusion des données des sept premières occasions de pratique entraîne la plus forte augmentation de l'AUC, qui passe de 0,54 (*première-6*) à 0,61 (*première- 7*). De même, le score de l'AUC du contraste entre WS et NP a augmenté le plus après l'inclusion des données des 5 premières opportunités de pratique, de

0.56 (*first-4* ) à 0.62 (*first-5* ). Cependant, l'AUC du détecteur WS-PP fluctue autour de 0,625 et ne montre aucune tendance à la hausse de la première-3 à la première-10.

# Discussion

## Résultats de la sélection des caractéristiques

Nous avons observé que les caractéristiques liées aux hint étaient présentes dans tous les modèles de l'ensemble de données ainsi qu'au nœud racine de cinq modèles, ce qui indique que ces caractéristiques ont le plus grand pouvoir prédictif. Ce résultat est cohérent avec les études précédentes. Par exemple, [[8] a](#_bookmark8) identifié des caractéristiques impliquant des indices pour prédire le WHEEL SPINNING, telles que l'utilisation d'indices, le nombre d'occasions de pratique précédentes avec des demandes d'indices et le fait que les étudiants aient demandé au moins cinq indices. Un autre résultat de notre modèle est l'effet des demandes d'allusions en bas de page pour prédire le WS. Le nombre moyen de demandes d'indices par le bas pour les huit dernières occasions d'entraînement est sélectionné comme nœud racine pour prédire les WS et PP (*premier-10*), ce qui indique que l'indice par le bas est un prédicteur fort pour prédire les WS par rapport aux PP. [2]](#_bookmark5) ont également fait une constatation similaire : après la quatrième occasion de s'entraîner, la demande d'allusion par le bas est positivement associée à la rotation de la roue. [10] ont](#_bookmark10) également rapporté que l'utilisation intensive d'indices de sortie par le bas est associée au WHEEL SPINNING.

Une autre catégorie de caractéristiques fortement liées à la détection du WHEEL SPINNING est l'exactitude des opportunités de pratique précédentes (*TotalPastWrongCount*, *To- talPercentPastWrong*, et *Past5WrongCount* ). Ce résultat est également cohérent avec les études précédentes [[8,15,](#_bookmark15)[22].](#_bookmark22) En particulier, [[22] a](#_bookmark22) comparé la dé- tection de WHEEL SPINNING à travers différents tuteurs, algorithmes et caractéristiques. Ils ont constaté qu'un modèle de régression logis- tique avec une seule caractéristique, le pourcentage de réponses correctes, permettait d'atteindre les résultats suivants

La précision est moindre mais comparable à celle d'autres modèles multi- fonctions construits à l'aide de la forêt domaniale, ce qui indique que la correction est un prédicteur fort pour la prédiction du WHEEL SPINNING. De plus, selon [[8], le](#_bookmark8) nombre de réponses incorrectes précédentes sur la même compétence a une relation positive avec le WHEEL SPINNING. Dans un jeu d'apprentissage des mathématiques, [[15] a](#_bookmark15) également constaté que la connaissance préalable mesurée par le taux d'absence et la non compétence des compétences est fortement liée au WHEEL SPINNING.

Les caractéristiques liées à la quantité de pratique et au temps (*TimeTaken*, *TotalTime- OnSkill*, *TotalSkillOpportunities*, *TimeBetweenProblems*) sont sélectionnées dans tous les modèles *first-x*. Pour les modèles comprenant moins d'occasions de pratiquer la compétence (*premier-3* à *premier-6*), le *temps pris* et le *temps total sur la compétence* sont prédictifs de la rotation de la roue. Cependant, pour les modèles avec plus de données accumulées (*first-5* à *first- 9*

), les caractéristiques sont passées de la durée du temps (*TotalTimeOnSkill* ) aux mesures du nombre d'opportunités (*TotalSkillOpportunities* ). Les auteurs de l'étude [[2] ont](#_bookmark5) également constaté que le temps de réponse est plus prédictif lors des premières occasions d'entraînement. Pour les réponses ultérieures, une réponse rapide peut indiquer soit la maîtrise d'une compétence, soit une manipulation du système, ce qui rend la signification du temps de réponse ambiguë.

## Performance des modèles de prédiction multi-classes et par paires

D'après la figure [1, l](#_bookmark3)a performance de la prédiction multi-classes augmente à mesure que nous incluons des données provenant d'un plus grand nombre de possibilités de pratique. En incluant les données des 9 premières occasions de pratique, l'AUC macro-moyenne atteint 0,62. À notre connaissance, il s'agit de la première étude explorant la détection intégrée de la non persistance, du WHEEL SPINNING et de la persistance productive, qui prolonge la recherche précédente sur le détecteur de WS utilisant un classificateur à arbre de décision [[10].](#_bookmark10) Par conséquent, elle pourrait être utilisée comme base de référence pour évaluer les performances du modèle dans des travaux futurs.

En différenciant le WHEEL SPINNING (WS) de la persistance productive (PP), nous avons constaté que les valeurs de l'AUC de la performance du modèle fluctuent autour de 0,625 entre les 3 premiers et les 10 premiers ensembles de données, ce qui implique que notre modèle prédictif est stable et capable de différencier les étudiants à risque de WHEEL SPINNING des étudiants qui sont persistants de manière productive dès la troisième opportunité de pratique. Cette constatation peut sembler contredire des études antérieures qui montrent que les modèles s'améliorent avec le nombre de données [[8, 14, 22].](#_bookmark22) Cette différence entre les études peut être due à la différence de définition de la maîtrise dans les différentes études. Dans les études antérieures, le critère de maîtrise productive est défini sur la base de la performance dans le système, comme trois corrections d'affilée [[8].](#_bookmark8) Cependant, la définition plus stricte de la persistance productive dans notre étude exige que les étudiants ne se contentent pas de répondre aux critères de maîtrise de " trois corrections à la suite ", mais qu'ils réussissent également le test ARRS différé pour démontrer la rétention de l'apprentissage [[10]. Il est](#_bookmark10) possible qu'une définition de la mas- trie basée sur un apprentissage robuste, une barre plus haute que la simple réalisation de trois réponses correctes à la suite, soit plus facile à détecter de manière précoce. Cependant, une conclusion contrastée a été obtenue par [[22],](#_bookmark22) qui a obtenu une prédiction plus précise et une détection plus précoce en utilisant un critère de maîtrise plus généreux que trois réponses correctes à la suite.

Dans nos modèles générés pour différencier le WHEEL SPINNING (WS) de

la non-persistance (NP), nous avons observé que si la performance du modèle

augmente avec le nombre d'opportunités de pratique, l'augmentation de la valeur AUC est la plus élevée be-

entre la 4e et la 5e occasion de s'exercer. Cela implique que nos détecteurs peuvent être capables de différencier WS de NP avec suffisamment de précision à la cinquième occasion de pratique. [4]](#_bookmark6) ont examiné la performance des réseaux de mémoire à long et à court terme pour prédire le WHEEL SPINNING et la non- persistance sur les ASSISTANCES en fonction du nombre d'occasions de s'exercer fournies à l'algorithme. Ils ont constaté que la troisième occasion pourrait être le moment le plus précoce pour prédire à la fois le WS et la NP, un point plus précoce que celui observé dans notre étude. Nos détecteurs nécessitent donc plus de données que [[4].](#_bookmark6) Cependant, nous sommes en mesure d'interpréter les caractéristiques de notre modèle sur la base des structures d'arbre de décision pour en tirer des enseignements plus généraux. Ce compromis entre la performance du modèle et l'interprétabilité est également présent dans d'autres domaines de l'analyse de l'apprentissage, comme la modélisation des composants de la connaissance [[13].](#_bookmark13)

# Conclusion

Dans cette étude, nous explorons le potentiel de détection précoce du WHEEL SPINNING, de la persistance productive et de la non-persistance dans les ASSISTments. En construisant des modèles d'arbre de décision et en observant l'évolution des performances du modèle au fur et à mesure de l'agrégation des données sur un plus grand nombre d'opportunités de pratique, nous avons constaté que le modèle basé sur neuf opportunités de pratique donne les meilleures performances ; le modèle basé sur les trois premières opportunités de pratique permet la détection précoce du WHEEL SPINNING par rapport à la persistance productive, les cinq premières opportunités de pratique sont suffisantes pour différencier le WHEEL SPINNING de la non-persistance, et les sept premières opportunités de pratique sont suffisantes pour différencier la persistance productive de la non-persistance. En raison de l'interprétabilité des modèles d'arbre de décision, nous avons examiné les caractéristiques communes à tous les modèles et les caractéristiques du nœud racine de chacun. Les caractéristiques prédictives, telles que l'utilisation d'indices et d'indices de sortie, la correction et la quantité de temps et d'opportunités sur la pratique précédente, nous donnent un aperçu des facteurs qui pourraient conduire au WHEEL SPINNING.

Un autre domaine potentiel de travail futur, l'intervention personnalisée basée sur les caractéristiques prédictives pourrait être intégrée dans les systèmes d'apprentissage existants pour mieux optimiser l'apprentissage des étudiants. Puisque les caractéristiques prédictives du WHEEL SPINNING sont au moins quelque peu cohérentes entre les études et les ensembles de données (voir la discussion ci-dessus), cela pourrait nous aider à concevoir de futurs systèmes de tutorat intelligents qui s'adapteraient mieux à la possibilité de WHEEL SPINNING dans les premières étapes de l'apprentissage. Un tel système pourrait encourager les étudiants à utiliser les indications de bas de page lors des premières occasions de pratique, si nécessaire ; ensuite, le système pourrait limiter la disponibilité des indications de bas de page dans les occasions de pratique ultérieures. De cette façon, le système pourrait tirer parti de ce que nous savons du WHEEL SPINNING pour nous aider à le prévenir.

**Remerciements**. Nous remercions Neil Heffernan et l'équipe ASSISTments, ainsi que la subvention NSF #DRL-1535340.

# Références

1. Baker, R.S., Goldstein, A.B., Heffernan, N.T. : Detecting learning moment-by- moment. International Journal of Artificial Intelligence in Education **21**(1-2), 5-25 (2011)
2. Beck, J.E., Gong, Y. : WHEEL SPINNING : Les étudiants qui ne parviennent pas à maîtriser une compétence. In : In- ternational conference on artificial intelligence in education. pp. 431–440. Springer (2013)
3. Borghans, L., Meijers, H., Ter Weel, B. : The role of noncognitive skills in explaining cognitive test scores. Economic inquiry **46**(1), 2-12 (2008)
4. Botelho, A.F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S.A., Beck, J.E. : Développer des détecteurs précoces de l'attrition des étudiants et du WHEEL SPINNING en utilisant l'apprentissage profond. IEEE Transactions on Learning Technologies **12**(2), 158-170 (2019).
5. Cloninger, C.R., Svrakic, D.M., Przybeck, T.R. : A psychobiological model of tem- perament and character. Archives of general psychiatry **50**(12), 975-990 (1993)
6. Deke, J., Haimson, J. : Valuing student competencies : Lesquelles permettent de prédire la réussite scolaire et les revenus au niveau postsecondaire, et pour qui ? rapport final. Mathe- matica Policy Research, Inc. (2006)
7. Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R. : Grit : persévérance et passion pour des objectifs à long terme. Journal of personality and social psychology **92**(6), 1087 (2007)
8. Gong, Y., Beck, J.E. : Towards detecting WHEEL SPINNING : Future failure in mastery learning. In : Proceedings of the second (2015) ACM conference on learning@ scale. pp. 67-74 (2015).
9. Heffernan, N.T., Heffernan, C.L. : The assistments ecosystem : Construction d'une plateforme qui rassemble les scientifiques et les enseignants pour une recherche peu invasive sur l'apprentissage et l'enseignement humain. Journal international de l'intelligence artificielle dans l'éducation **24**(4), 470-497 (2014).
10. Kai, S., Almeda, M.V., Baker, R.S., Heffernan, C., Heffernan, N. : Modélisation par arbre de décision de l'essorage de la roue et de la persistance productive chez les constructeurs de compétences. JEDM- Journal of Educational Data Mining **10**(1), 36-71 (2018).
11. Karumbaiah, S., Baker, R.S., Barany, A., Shute, V. : Using epistemic networks with automated codes to understand why players quit levels in a learning game. In : Conférence internationale sur l'ethnographie quantitative, p. 106-116. Springer (2019)
12. Karumbaiah, S., Baker, R.S., Shute, V. : prédire l'abandon chez les étudiants qui jouent à un jeu d'apprentissage. Société internationale de fouille de données éducatives (2018).
13. Liu, R., McLaughlin, E.A., Koedinger, K.R. : Interpréter la découverte de modèles et tester la généralisation à un nouvel ensemble de données. In : Educational Data Mining 2014. Citeseer (2014)
14. Matsuda, N., Chandrasekaran, S., Stamper, J.C. : A quelle vitesse peut-on détecter le WHEEL SPINNING ? In : EDM. pp. 607-608 (2016)
15. Owen, V.E., Roy, M.H., Thai, K., Burnett, V., Jacobs, D., Keylor, E., Baker, R.S. : Détecter le WHEEL SPINNING et la persistance productive dans les jeux éducatifs. Société internationale de fouille de données éducatives (2019)
16. Palaoag, T.D., Rodrigo, M.M.T., Andres, J.M.L., Andres, J.M.A.L., Beck, J.E. : WHEEL SPINNING in a game-based learning environment for physics. In : Conférence internationale sur les systèmes de tutorat intelligents, pp. 234-239. Springer (2016)
17. Park, S., Matsuda, N. : Predicting students' unproductive failure on intelligent tu- tors in adaptive online courseware. In : Actes du sixième symposium annuel des utilisateurs de GIFT, vol. 6, p. 131. Laboratoire de recherche de l'armée américaine (2018)
18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. : Scikit-learn : Machine learning in Python. Journal of Machine Learning Research **12**, 2825-2830 (2011)
19. Poropat, A.E. : Une méta-analyse du modèle à cinq facteurs de la personnalité et des performances scolaires. Psychological bulletin **135**(2), 322 (2009)
20. Prabhu, V., Sutton, C., Sauser, W. : Creativity and certain personality traits : Un- derstanding the mediating effect of intrinsic motivation. Creativity Research Jour- nal **20**(1), 53-66 (2008)
21. Wang, Y., Heffernan, N.T. : Towards modeling forgetting and relearning in its : Analyse préliminaire des données arrs. In : EDM. p. 352 (2011)
22. Zhang, C., Huang, Y., Wang, J., Lu, D., Fang, W., Stamper, J., Fancsali, S., Holstein, K., Aleven, V. : Détection précoce du WHEEL SPINNING : comparaison entre les tuteurs, les modèles, les caractéristiques et les opérationnalisations. Soumission des bénéficiaires de subventions (2019)

[Voir les statistiques de publication](https://www.researchgate.net/publication/342675876)