

Article

Analyse prédictive de l'engagement académique numérique des étudiants à partir d'indicateurs d'usage mobile et Web : une approche d'apprentissage automatique dans les établissements d'enseignement supérieur de Butembo (RDC)

Kavusa Kasoko

Domaine des Sciences Economiques et de Gestion, Université de l'Assomption au Congo, B.P. 104, Butembo, République Démocratique du Congo

Correspondant : jgkasoko@gmail.com

Citation : Kavusa, K. Analyse prédictive de l'engagement académique numérique des étudiants à partir d'indicateurs d'usage mobile et Web : une approche d'apprentissage automatique dans les établissements d'enseignement supérieur de Butembo (RDC). *Etincelle*. 2026, Vol. 26, no. 2. <https://doi.org/10.61532/rime262147>

Reçu: 01 juillet 2026

Accepté: 10 août 2026

Publié: 11 août 2026

Note de l'éditeur: Ishango-uac reste neutre en ce qui concerne les revendications juridictionnelles dans les cartes géographiques publiées et les affiliations institutionnelles des auteurs.



Copyright: © 2026 par l'auteur. Soumis pour une publication en libre accès selon les termes et conditions de la licence Creative Commons Attribution (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

Résumé : La transformation numérique de l'enseignement supérieur produit de nombreux indicateurs d'usage susceptibles d'éclairer les pratiques d'apprentissage. Cette étude examine la capacité d'indicateurs mobiles et Web déclarés par questionnaire à prédire l'engagement académique numérique de 396 étudiants issus de cinq établissements d'enseignement supérieur de Butembo. Un Indice d'Engagement Académique Numérique (IEAN), normalisé sur une échelle de 0 à 100 et organisé en trois niveaux, a servi de variable cible. Six classifieurs ont été comparés : régression logistique, arbre de décision, forêt aléatoire, machine à vecteurs de support, Gradient Boosting et Naïve Bayes gaussien. Sur 80 observations de test, le Gradient Boosting obtient la meilleure exactitude (83,75 %) et le meilleur F1-score macro (86,07 %). Les résultats indiquent une faisabilité prédictive, mais leur portée reste exploratoire en raison du caractère transversal et déclaratif des données, du rapport élevé entre le nombre de variables et la taille de l'échantillon, de l'absence de validation externe et d'un risque de recouvrement conceptuel entre certains prédicteurs et la construction de l'IEAN.

Mots-clés : apprentissage automatique, engagement académique numérique, usages mobiles, Web, Gradient Boosting, Learning Analytics, enseignement supérieur

1. Introduction

La généralisation des smartphones, des plateformes Web, des réseaux sociaux, des moteurs de recherche et, plus récemment, des outils d'intelligence artificielle générative transforme les modalités d'accès à l'information et les pratiques d'apprentissage dans l'enseignement supérieur. Dans le champ des Learning Analytics, les données relatives aux apprenants sont mobilisées pour décrire les comportements d'apprentissage, suivre la progression et soutenir les décisions pédagogiques (Baker & Inventado, 2014 ; Siemens, 2013 ; Romero & Ventura, 2020). Toutefois, la valeur de ces données dépend fortement de leur provenance, de leur qualité et de leur interprétation contextuelle.

L'engagement étudiant est un construit multidimensionnel qui ne se réduit pas à la fréquence d'une activité numérique. Dans l'enseignement supérieur, il résulte de l'interaction entre comportements, cognition, affect et contexte institutionnel (Kahu, 2013). Les environnements numériques rendent certaines dimensions comportementales plus observables, mais les clics, recherches, applications ou durées d'usage ne constituent que des indicateurs indirects. La revue systématique de Bergdahl et al.

(2024), fondée sur 159 études, montre d'ailleurs que les Learning Analytics privilégient largement les manifestations comportementales de l'engagement et saisissent moins bien ses dimensions cognitives et affectives.

Parallèlement, les méthodes d'apprentissage automatique occupent une place croissante dans l'analytique prédictive de l'enseignement supérieur. Elles sont utilisées pour anticiper la performance, le décrochage, la persévérance ou l'engagement, avec des résultats fortement dépendants du contexte, des variables et du protocole d'évaluation (Sghir et al., 2023 ; Chen et al., 2025). Ces travaux soulignent l'intérêt de comparer plusieurs familles d'algorithmes tout en contrôlant les risques de surapprentissage, de fuite d'information et de faible généralisabilité.

La présente recherche se situe dans un contexte encore peu documenté dans la littérature internationale : cinq établissements d'enseignement supérieur de Butembo, en République démocratique du Congo. Elle utilise des réponses à un questionnaire portant sur les caractéristiques sociodémographiques et académiques, les usages mobiles et Web et le recours à l'intelligence artificielle. L'objectif général est d'évaluer dans quelle mesure ces indicateurs d'usage permettent de discriminer des niveaux d'engagement académique numérique.

Trois questions structurent l'étude : (1) quelles caractéristiques décrivent les pratiques numériques des étudiants ? (2) dans quelle mesure ces caractéristiques permettent-elles de discriminer les niveaux d'IEAN ? et (3) quel algorithme offre les meilleures performances prédictives parmi les modèles comparés ? Les hypothèses sont formulées en termes prédictifs : H1, certaines caractéristiques d'usage contribuent à la discrimination des niveaux d'IEAN ; H2, la performance d'un modèle appris dépasse celle d'une règle de classification naïve ; H3, les performances diffèrent selon l'algorithme utilisé.

2. Revue de la littérature

2.1. Engagement étudiant et indicateurs numériques

L'engagement est généralement envisagé comme un processus dynamique et multidimensionnel. Le cadre de Kahu (2013) insiste sur l'articulation entre facteurs structurels, psychosociaux et éducatifs. Dans les environnements numériques, certaines manifestations comportementales — consultation de ressources, recherches en ligne, participation, production de contenus ou régularité d'usage — peuvent être observées ou déclarées. Leur interprétation doit néanmoins rester prudente : une activité numérique identique peut relever d'un objectif académique, social ou récréatif. Cette distinction est essentielle pour éviter de confondre intensité d'usage et engagement académique.

Bergdahl et al. (2024) montrent que l'opérationnalisation de l'engagement en Learning Analytics repose surtout sur des indicateurs comportementaux. Cette littérature justifie l'utilisation d'un indice numérique comme variable opérationnelle, à condition de documenter précisément sa construction et de ne pas l'assimiler à l'ensemble du construit d'engagement.

2.2. Learning Analytics et exploitation des données

Les Learning Analytics visent à analyser les données relatives aux apprenants et à leurs contextes afin de mieux comprendre et d'optimiser l'apprentissage (Siemens, 2013). L'expansion des environnements numériques a accru la disponibilité des données, mais également les enjeux de validité, de gouvernance et d'interprétation. Viberg et al. (2018) et Romero et Ventura (2020) soulignent le rôle des traces et indicateurs numériques dans la compréhension des trajectoires d'apprentissage, tandis que Paulsen et Lindsay (2024) montrent que les tableaux de bord évoluent progressivement

d'une logique centrée sur l'analytique vers une logique davantage orientée vers les processus d'apprentissage.

Stojanov et Daniel (2024) observent une croissance continue des usages du Big Data et de l'analytique dans l'enseignement supérieur, mais insistent sur la nécessité de mieux intégrer les résultats aux pratiques pédagogiques et aux processus institutionnels. Les prédictions n'ont donc de valeur éducative que si elles sont interprétables, contextualisées et associées à des interventions proportionnées.

2.3. Apprentissage automatique et prédiction en enseignement supérieur

L'apprentissage supervisé permet d'estimer une classe ou une valeur cible à partir d'un ensemble de caractéristiques. Les modèles linéaires, les arbres, les méthodes d'ensemble, les machines à vecteurs de support et les approches probabilistes présentent des biais inductifs différents (Kotsiantis, 2007 ; Sarker, 2021). La forêt aléatoire combine plusieurs arbres afin d'améliorer la généralisation (Breiman, 2001), tandis que le Gradient Boosting construit séquentiellement des apprenants faibles pour réduire l'erreur du modèle (Friedman, 2001). Dans l'enseignement supérieur, la comparaison de plusieurs familles de modèles est une pratique courante, mais les revues récentes recommandent des protocoles de validation plus rigoureux et plus transparents (Sghir et al., 2023 ; Chen et al., 2025).

2.4. Intelligence artificielle, engagement et éthique

L'essor de l'intelligence artificielle générative s'inscrit dans une évolution plus large de l'IA en éducation (Williamson & Eynon, 2020) et ajoute une dimension nouvelle aux pratiques numériques étudiantes. Lo et al. (2024) montrent que l'usage de ChatGPT peut être associé à des formes d'engagement comme de désengagement selon la tâche, la manière d'utiliser l'outil et l'encadrement pédagogique. L'usage d'un outil d'IA ne doit donc pas être interprété isolément comme un indicateur positif ou négatif d'engagement.

L'exploitation de données étudiantes par les Learning Analytics et l'IA soulève en outre des questions de confidentialité, de transparence, d'autonomie et de contrôle humain (Slade & Prinsloo, 2013 ; Alfredo et al., 2024). Bond et al. (2024) appellent à davantage de rigueur et d'éthique dans la recherche sur l'IA en enseignement supérieur. Les orientations de l'UNESCO insistent également sur une approche centrée sur l'humain, la protection des données et le développement d'un jugement critique face aux systèmes d'IA (Miao & Holmes, 2023 ; Miao et al., 2024).

2.5. Positionnement de l'étude

Cette étude combine des caractéristiques individuelles et académiques, des indicateurs d'usage mobile et Web et des variables relatives à l'IA afin de construire un indice opérationnel d'engagement académique numérique et de comparer six algorithmes de classification. Son apport principal réside dans l'exploration de ces relations dans le contexte universitaire de Butembo, au-delà des données issues exclusivement d'un LMS institutionnel. La Figure 1 résume le cadre conceptuel mobilisé.

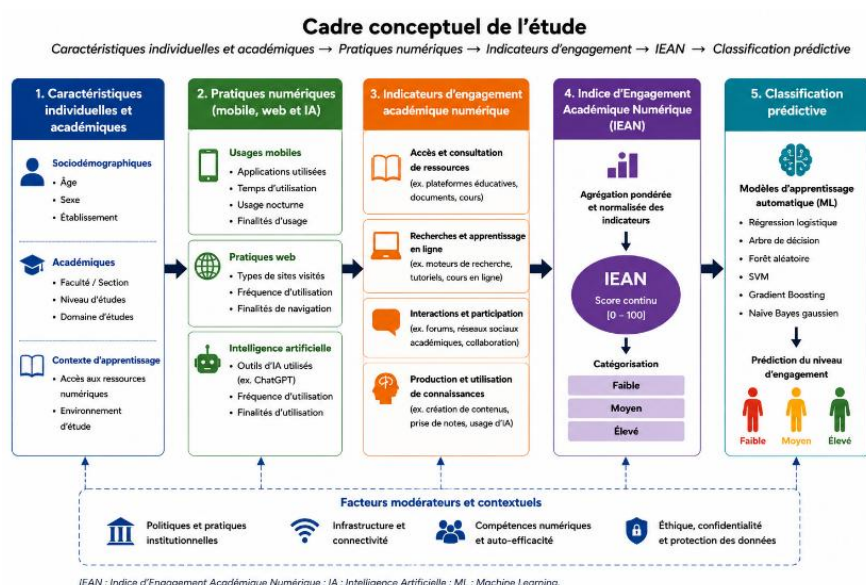


Figure 1 : Cadre conceptuel de l'étude.

3. Méthodologie

3.1. Cadre, population et échantillonnage

L'étude adopte une approche quantitative, transversale et prédictive. Elle a été menée auprès d'étudiants de cinq établissements d'enseignement supérieur de Butembo : FAB, ISP Muhangi, UAC, UOR et UCG. La taille théorique minimale a été estimée à l'aide de la formule de Cochran $n_0 = Z^2 p(1-p)/e^2$, conduisant à environ 374 observations. Quatre cents questionnaires ont été administrés et 396 réponses exploitables ont été retenues après contrôle et nettoyage. L'échantillonnage était non probabiliste, de convenance ; en conséquence, les résultats ne doivent pas être interprétés comme une estimation représentative de l'ensemble des étudiants de Butembo.

3.2. Variables et prétraitement

Le questionnaire couvrait l'âge, le sexe, l'établissement, la faculté ou section, le niveau d'études, les applications mobiles utilisées, le temps d'utilisation du smartphone, l'usage nocturne, les finalités d'usage, les pratiques Web et le recours à des outils d'intelligence artificielle. La variable « Commentaire » a été supprimée avant la modélisation. Les variables catégorielles ont été transformées par encodage et les variables numériques standardisées lorsque nécessaire. Le manuscrit source rapporte 901 variables explicatives après encodage pour 396 observations.

Le protocole indique que les variables directement dérivées de l'IEAN ont été exclues des prédicteurs afin de limiter la fuite d'information. Toutefois, le dictionnaire exact des variables exclues et la chaîne complète de prétraitement ne sont pas fournis. Ce point est important, car certaines variables d'usage Web ou de participation sont conceptuellement proches des dimensions utilisées pour construire l'IEAN. Une réplique doit donc publier la liste des variables sources de l'indice et des prédicteurs, et ajuster l'encodage et la standardisation uniquement sur les données d'apprentissage.

3.3. Construction de l'Indice d'Engagement Académique Numérique

Le cadre conceptuel fourni décrit l'IEAN comme une agrégation pondérée et normalisée sur une échelle de 0 à 100 de quatre dimensions : (1) accès et consultation de ressources ; (2) recherches et apprentissage en ligne ; (3) interactions et participation

académiques ; et (4) production et utilisation de connaissances. Le score est ensuite classé en trois niveaux : faible, moyen et élevé.

La version source ne précise pas les poids appliqués à chaque indicateur ni les seuils numériques exacts de catégorisation. En outre, la distribution observée (31 faibles, 164 moyens et 201 élevés) n'est pas compatible avec une partition en trois tertiles de tailles proches. La formulation « par tertiles » du manuscrit initial a donc été retirée. Pour assurer la reproductibilité, les poids et seuils utilisés doivent être documentés dans le protocole ou dans un matériel complémentaire.

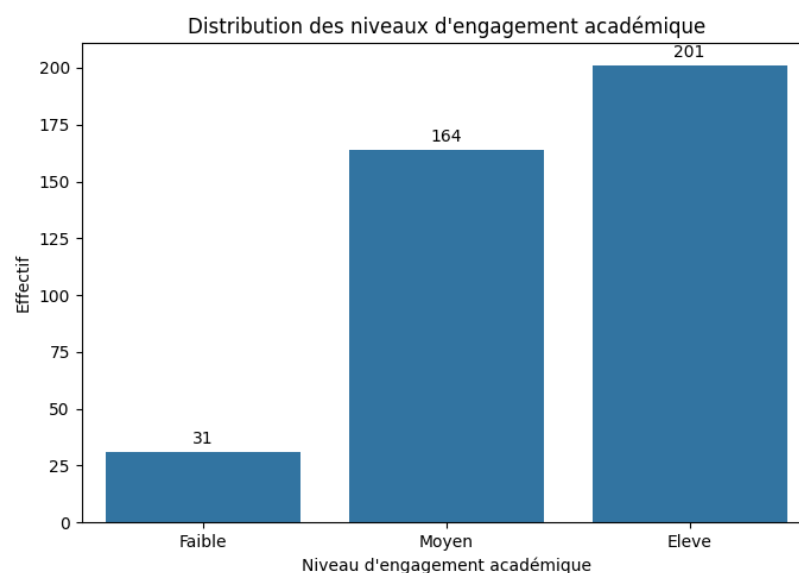


Figure 2 : Distribution des niveaux d'engagement académique numérique dans l'échantillon complet (N = 396).

3.4. Modélisation et évaluation

Six algorithmes ont été comparés : régression logistique, arbre de décision, forêt aléatoire, machine à vecteurs de support (SVM), Gradient Boosting et Naïve Bayes gaussien. Les traitements ont été réalisés sous Python avec pandas, NumPy et scikit-learn (Pedregosa et al., 2011) ; la graine aléatoire rapportée est 42. Les données ont été réparties en ensembles d'apprentissage (316 observations) et de test (80 observations), avec stratification selon le niveau d'IEAN. Une validation croisée stratifiée à cinq plis est également mentionnée pour apprécier la stabilité des modèles.

Les performances sont évaluées par l'exactitude (accuracy), la précision, le rappel et le F1-score macro. Le F1 macro est particulièrement utile ici, car il attribue le même poids à chacune des trois classes malgré leur distribution inégale. Le manuscrit source ne fournit pas les hyperparamètres détaillés, la stratégie éventuelle d'optimisation ni les résultats chiffrés de validation croisée. Ces éléments sont signalés comme limites de reproductibilité et doivent être conservés dans toute réplique ou validation externe.

4. Résultats

4.1. Structure des données et cohérence des effectifs

La distribution des classes a été vérifiée en croisant la Figure 2, la matrice de confusion et les mesures par classe. Ces éléments sont mutuellement cohérents et indiquent 31 étudiants de niveau faible, 164 de niveau moyen et 201 de niveau élevé. Avec

la séparation stratifiée 80/20, l'ensemble d'apprentissage comprend 25, 131 et 160 étudiants dans ces trois classes, tandis que l'ensemble test en comprend respectivement 6, 33 et 41. Ces effectifs remplacent une répartition contradictoire qui figurait dans le tableau initial.

Tableau 1 : Structure des données utilisées pour la modélisation.

Élément	Ensemble complet	Apprentissage	Test
Observations	396	316	80
Variables explicatives après encodage	901	901	901
Niveau faible	31	25	6
Niveau moyen	164	131	33
Niveau élevé	201	160	41
Total des classes	396	316	80

La classe « élevé » représente 50,8 % de l'échantillon complet et 51,25 % de l'échantillon test. Cette asymétrie rend nécessaire l'examen conjoint de l'exactitude et des métriques macro plutôt qu'une lecture fondée sur l'exactitude seule.

4.2. Comparaison des modèles

Les six modèles ont été évalués sur le même ensemble test. Le Gradient Boosting obtient les meilleures valeurs d'exactitude et de F1-score macro. La forêt aléatoire présente la deuxième meilleure valeur de F1 macro, tandis que le Naïve Bayes gaussien obtient l'exactitude la plus faible.

Tableau 2 : Performances des modèles sur l'échantillon test.

Modèle	Exactitude	Précision macro	Rappel macro	F1 macro
Gradient Boosting	83,75	84,19	88,45	86,07
Arbre de décision	78,75	72,77	79,86	75,36
Forêt aléatoire	78,75	85,13	75,12	78,97
SVM	77,50	84,59	70,15	74,29
Régression logistique	76,25	78,56	78,82	78,24
Naïve Bayes gaussien	68,75	79,07	68,81	72,01

Sur 80 observations de test, une exactitude de 83,75 % correspond à 67 classifications correctes. À titre de repère, un classifieur trivial prédisant systématiquement la classe majoritaire « élevé » obtiendrait 41/80, soit 51,25 % d'exactitude. Le gain observé est donc substantiel sur ce jeu de test, sans constituer pour autant une preuve de généralisabilité externe.

4.3. Évaluation détaillée du Gradient Boosting

Tableau 3 : Performances par classe du Gradient Boosting.

Classe	Précision	Rappel	F1-score
Élevé	0,87	0,80	0,84
Faible	0,86	1,00	0,92
Moyen	0,80	0,85	0,82
Moyenne macro	0,84	0,88	0,86

La matrice de confusion confirme ces résultats : 33 étudiants sur 41 de niveau élevé, 6 sur 6 de niveau faible et 28 sur 33 de niveau moyen sont correctement classés. Les erreurs concernent principalement des étudiants du niveau élevé classés comme moyens (7 cas) et des étudiants du niveau moyen classés comme élevés (5 cas).

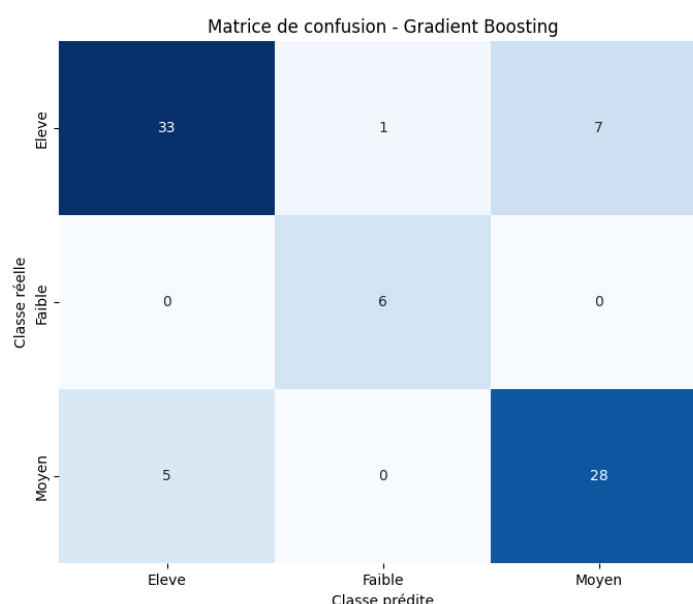


Figure 3 : Matrice de confusion du Gradient Boosting sur l'échantillon test (n = 80).

4.4. Importance des variables

L'analyse par permutation met en évidence plusieurs variables issues des usages Web et de certaines catégories d'applications ou du contexte académique. Les premières positions sont notamment occupées par des modalités techniques liées au type d'usage Web, suivies de variables de recherches Web et de plusieurs indicateurs catégoriels encodés. Les noms affichés dans la Figure 4 correspondent aux colonnes issues de l'encodage et doivent être interprétés à partir du dictionnaire des variables.

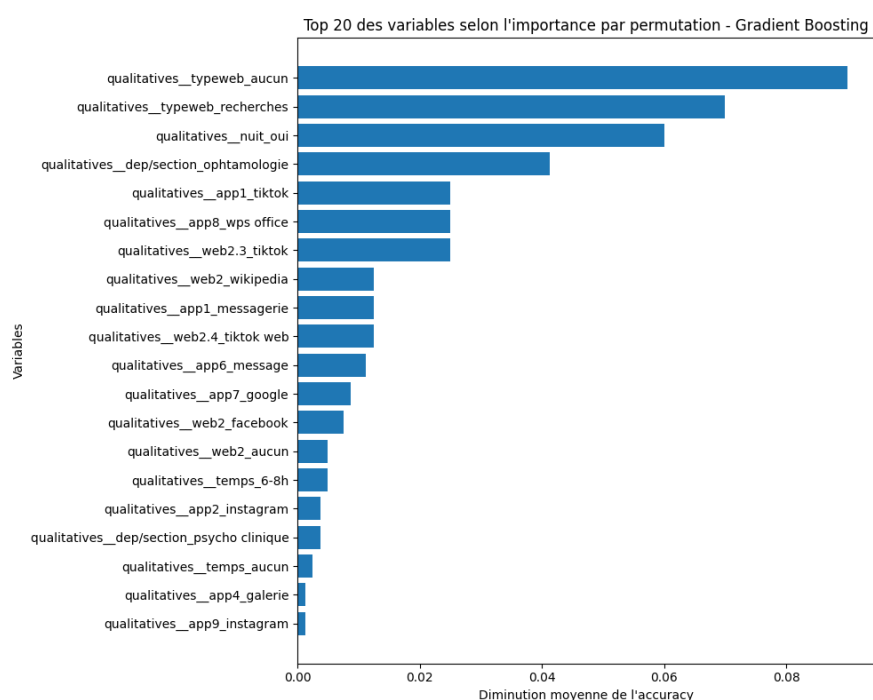


Figure 4 : Vingt variables présentant les plus fortes importances par permutation pour le Gradient Boosting.

Une importance prédictive ne démontre pas une relation causale. Elle ne signifie pas non plus qu'une variable est pédagogiquement souhaitable ou défavorable. La présence, parmi les variables les plus importantes, de modalités conceptuellement proches des dimensions de l'IEAN renforce la nécessité d'un audit de fuite d'information avant toute utilisation opérationnelle du modèle.

4.5. Examen des hypothèses

H1 reçoit un soutien prédictif : plusieurs caractéristiques d'usage contribuent effectivement à la discrimination des niveaux d'IEAN. Ce résultat ne doit pas être présenté comme une association « statistiquement significative », car aucun test d'inférence spécifique n'est rapporté pour cette hypothèse. H2 est soutenue sur l'échantillon test : le meilleur modèle dépasse nettement la règle de classe majoritaire (83,75 % contre 51,25 % d'exactitude). H3 est également soutenue descriptivement, les performances variant selon les algorithmes. Ces conclusions restent limitées au protocole et à l'échantillon étudiés.

5. Discussion

Le Gradient Boosting obtient la meilleure performance parmi les six algorithmes comparés, avec 83,75 % d'exactitude et 86,07 % de F1-score macro. Ce résultat est cohérent avec la capacité des méthodes de boosting à modéliser des relations non linéaires et des interactions complexes (Friedman, 2001). Il rejoint plus largement la littérature sur les Predictive Learning Analytics, où les méthodes d'ensemble figurent fréquemment parmi les approches efficaces, tout en rappelant que la supériorité d'un modèle dépend du jeu de données, des variables et du protocole d'évaluation (Sghir et al., 2023 ; Chen et al., 2025).

L'étude apporte un intérêt contextuel en explorant des indicateurs d'usage numérique dans cinq établissements de Butembo. Elle élargit ainsi l'analyse au-delà des seules traces d'un LMS. Cependant, les données utilisées sont principalement déclaratives : elles décrivent ce que les étudiants rapportent de leurs usages, et non des journaux numériques observés automatiquement. Le terme d'« empreinte numérique » doit donc être compris ici comme un ensemble d'indicateurs déclarés d'usage et non comme une télémétrie instrumentée.

La performance doit aussi être interprétée à la lumière de la construction de la cible. L'IEAN agrège des dimensions telles que les recherches en ligne, l'accès aux ressources, les interactions et la production de connaissances. Or certaines variables de même nature apparaissent également parmi les prédicteurs importants. Le manuscrit source affirme que les variables directement dérivées de l'IEAN ont été retirées, mais l'absence de dictionnaire complet empêche de vérifier une séparation stricte. Un recouvrement conceptuel, même sans duplication exacte de colonnes, peut faciliter artificiellement la prédiction. Une validation robuste doit donc reconstruire l'IEAN et les prédicteurs à partir d'ensembles de variables explicitement disjoints.

Un second enjeu concerne la dimension statistique. L'ensemble d'apprentissage contient 316 observations pour 901 variables après encodage. Ce rapport p/n élevé augmente le risque de surapprentissage, en particulier lorsque de nombreuses modalités catégorielles rares sont présentes. La validation croisée à cinq plis constitue une protection partielle, mais les résultats de cette validation ne sont pas rapportés. Des stratégies de réduction de dimension, de régularisation, de regroupement des modalités rares et de sélection de variables devraient être évaluées dans une réplique.

Enfin, les six modèles sont comparés sur le même ensemble test et le Gradient Boosting est ensuite retenu comme « modèle optimal ». Si ce choix a été effectué après observation des performances de test, l'estimation finale peut être optimiste. Un protocole plus solide utiliserait la validation croisée exclusivement dans l'ensemble d'apprentissage pour sélectionner modèle et hyperparamètres, puis un jeu de test verrouillé pour une seule évaluation finale, ou une validation croisée imbriquée. Des intervalles de confiance par bootstrap ou répétition des splits permettraient également de quantifier l'incertitude.

Sur le plan pédagogique, un modèle de ce type pourrait soutenir un dispositif d'alerte ou d'orientation, mais non une sanction automatique. Les Learning Analytics responsables exigent une supervision humaine, une explication des critères, une limitation des finalités, la protection des données et un mécanisme de contestation (Slade & Prinsloo, 2013 ; Alfredo et al., 2024 ; Bond et al., 2024). Ces exigences sont particulièrement importantes lorsque des données sur les usages numériques et l'IA sont mobilisées.

6. Limites et recommandations

Les principales limites sont les suivantes : échantillonnage de convenance limité à cinq établissements d'une même ville ; données transversales et largement auto-déclarées ; construction de l'IEAN insuffisamment documentée ; distribution déséquilibrée des classes ; 901 variables pour 316 observations d'apprentissage ; hyperparamètres et résultats de validation croisée non rapportés ; risque de recouvrement entre variables de la cible et prédicteurs ; sélection du meilleur modèle sur le même jeu de test ; absence d'intervalles d'incertitude et de validation externe.

Les travaux futurs devraient publier le questionnaire, le dictionnaire des variables, les poids et seuils de l'IEAN ainsi que le code de prétraitement ; utiliser une chaîne de traitement ajustée exclusivement sur les données d'apprentissage ; comparer des modèles régularisés et des méthodes de réduction de dimension ; sélectionner les modèles par validation croisée imbriquée ; rapporter la distribution des scores de validation, les intervalles de confiance et, si possible, des métriques complémentaires adaptées aux classes déséquilibrées ; et réaliser une validation temporelle ou externe sur d'autres cohortes universitaires.

Lorsque cela est légalement et éthiquement acceptable, les données déclaratives pourraient être complétées par des traces issues de plateformes pédagogiques, à condition de distinguer clairement les données servant à construire la cible de celles utilisées comme prédicteurs. Les modalités d'approbation éthique, de consentement, d'information des étudiants, de durée de conservation et de sécurisation des données doivent être explicitement documentées avant toute publication ou déploiement institutionnel.

7. Conclusion

Cette étude évalue la possibilité de prédire un indice d'engagement académique numérique à partir d'indicateurs déclarés d'usage mobile, Web et d'IA chez 396 étudiants de cinq établissements d'enseignement supérieur de Butembo. Parmi six algorithmes, le Gradient Boosting obtient la meilleure performance sur 80 observations de test, avec 83,75 % d'exactitude et 86,07 % de F1-score macro. La matrice de confusion et les métriques par classe sont cohérentes avec ces résultats.

L'intérêt principal du travail réside dans sa preuve de faisabilité et son contexte d'application. Les performances ne doivent toutefois pas être assimilées à une valida-

tion définitive. L'insuffisante documentation de l'IEAN, le nombre très élevé de variables par rapport à l'échantillon, la possibilité d'un recouvrement conceptuel entre cible et prédicteurs et l'absence de validation externe limitent la portée des conclusions. La priorité scientifique est donc de consolider la reproductibilité et la validation avant tout usage institutionnel.

Sous ces conditions, une approche de Learning Analytics centrée sur l'humain pourrait contribuer au suivi pédagogique en fournissant des signaux complémentaires à l'expertise des enseignants, sans se substituer à leur jugement ni aux échanges directs avec les étudiants.

Contributions : Conceptualisation, K.K.; méthodologie, K.K.; validation, K.K. ; Investigation, K.K.; ressources, K.K.; Traitement des données, K.K.; écrire le manuscrit, K.K.; visualisation, K.K.; supervision, K.K.; correction du manuscrit, K.K.

L'auteur a lu et approuvé la version publiée de ce manuscrit.

Sponsor financier : Cette recherche n'a reçu aucun soutien financier.

Disponibilité des données : Les données ne sont pas disponibles.

Remerciement : Non applicable.

Conflits d'intérêt : L'auteur déclare aucun conflit d'intérêt.

Références

1. Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
2. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4
3. Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: A systematic review. *International Journal of Educational Technology in Higher Education*, 21, 63. <https://doi.org/10.1186/s41239-024-00493-y>
4. Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, 4. <https://doi.org/10.1186/s41239-023-00436-z>
5. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
6. Chen, J., Zhou, X., Yao, J., & Tang, S. K. (2025). Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: A systematic literature review. *Asian Education and Development Studies*, 14(2), 205–240. <https://doi.org/10.1108/AEDS-08-2024-0166>
7. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
8. Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, 38(5), 758–773. <https://doi.org/10.1080/03075079.2011.598505>
9. Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268.
10. Lo, C. K., Hew, K. F., & Jong, M. S. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219, 105100. <https://doi.org/10.1016/j.compedu.2024.105100>
11. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
12. Miao, F., Shiohira, K., & Lao, N. (2024). AI competency framework for students. UNESCO.
13. Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review. *Education and Information Technologies*, 29, 14279–14308. <https://doi.org/10.1007/s10639-023-12401-4>

14. 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, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
15. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
16. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>
17. Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, 28, 8299–8333. <https://doi.org/10.1007/s10639-022-11536-0>
18. Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>
19. Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>
20. Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, 29, 5807–5831. <https://doi.org/10.1007/s10639-023-12033-8>
21. Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110. <https://doi.org/10.1016/j.chb.2018.07.027>
22. Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>