

Research Article

DÉVELOPPEMENT D'UN MODÈLE DE CLASSIFICATION POUR L'ÉVALUATION DES PERFORMANCES DES ENSEIGNANTS

* KADIATA MUKENGE Peter

Assistant à l'université pédagogique de Kananga, Democratic Republic of the Congo.

Received 16th April 2026; Accepted 17th May 2026; Published online 15th July 2026

RESUME

L'évaluation des performances des enseignants est un élément moteur du fonctionnement des écoles. Développer un modèle de classification contribue à la prise de décision et à l'accroissement de la qualité de l'enseignement. Notre étude consiste à concevoir un modèle qui permet d'améliorer l'évaluation des performances des enseignants des écoles secondaires. Pour y arriver nous avons récolté les fiches d'inspections et avons construit un dataset. Sur ce, nous avons utilisé la technique de clustering pour identifier les groupes des enseignants ayant le même profil. A l'aide des algorithmes Arbre de décision et Machine de vecteurs de support, nous avons déterminé les métriques tels que : la précision, recall, F1-score, accuracy et la matrice de confusion. L'algorithme un contre tous nous a permis de tracer les courbes ROC. Les différents métriques ont donné un pourcentage de plus 97% ce qui prouve que notre modèle présente une meilleure performance.

Mots clés: Système intelligent, Evaluation, Performances des enseignants.

INTRODUCTION GENERALE

La baisse sans cesse croissante de niveau que connaît notre enseignement et qui est décrite par les observateurs internes et externes de notre système éducatif est essentiellement due à la qualité des ressources humaines. En ce qui concerne particulièrement l'enseignant, sa responsabilité est d'une importance indéniable en sa qualité de premier opérateur pédagogique ; il joue le rôle de suprême arbitre : c'est sur son savoir, savoir-être et savoir faire que repose toute l'action de l'enseignement-apprentissage. C'est sur lui que sont jetées des fleurs quand les apprenants réussissent et des pierres en cas d'échec.

Le système cherche à améliorer l'efficacité des enseignants pour répondre aux besoins diversifiés des apprenants.

De nos jours l'intelligence artificielle est presque dans tous les domaines de la vie avec ses technologies qui lui permettent de traiter de grandes quantités de données. Ces outils peuvent offrir des analyses prédictives et des recommandations personnalisées, qui rendent facile et précise l'évaluation des enseignants. C'est dans ce contexte qu'entre ce travail de recherche intitulé « développement d'un modèle de classification pour l'évaluation des performances des enseignants »

Après une réflexion approfondie sur les défis actuels du système éducatif, notamment en ce qui concerne la qualité de l'enseignement au niveau secondaire. En tant que pédagogue de formation, il a été observé une baisse progressive du niveau des apprenants, souvent liée à plusieurs facteurs parmi lesquels figurent les performances des enseignants.

L'intégration des techniques d'intelligence artificielle dans le domaine éducatif est une opportunité majeure pour améliorer les processus d'évaluation en les rendant plus cohérents, précis et adaptés aux réalités de notre pays.

L'évaluation de la performance des enseignants est un processus d'importance capitale pour tout établissement qui préfère mesurer et améliorer la qualité de l'enseignement. Cependant, les méthodes traditionnelles présentent des imperfections qui entraînent une insatisfaction, une démotivation voire une injustice. Comment l'apprentissage automatique peut-il aider à améliorer l'évaluation des performances des enseignants

En vue d'améliorer la qualité et l'objectivité de l'évaluation des enseignants du secondaire, l'objectif principal de notre recherche consiste à concevoir un système intelligent qui permettra d'améliorer l'évaluation des performances des enseignants.

Cet objectif principal est accompagné des objectifs spécifiques énumérés ci-dessous:

- ✓ Identifier les critères clés qui influencent les performances des enseignants.
- ✓ Exploiter les données quantitatives et qualitatives pour obtenir une idée générale des performances des enseignants.

L'importance de cette étude est d'appliquer les méthodes d'apprentissage automatique dans le domaine éducatif en proposant un modèle d'analyse et de classification des performances des enseignants et offrir aux établissements scolaires et au corps inspectoral un outil d'aide à la prise de décision pour améliorer la qualité d'enseignement.

REVUE DE LA LITTÉRATURE

L'étude d'évaluation des performances d'enseignement fait l'objet de plusieurs travaux de recherche tant au niveau national qu'international. Cette section présente une synthèse des différentes contributions scientifiques qui ont abordé cette problématique sous différents aspects :

Antoine Murry (2016): Dans sa thèse intitulée « application d'un système probabiliste bayésien pour prédire la moyenne cumulative

*Auteur correspondant: KADIATA MUKENGE Peter,

Assistant à l'université pédagogique de Kananga, Democratic Republic of the Congo.

des étudiants à l'école polytechnique de Montréal » : L'idée de l'auteur est de tenter de prédire la moyenne cumulative des six dernières sessions obtenue par chaque étudiant, à partir de ses moyennes obtenues à chacune des six premières sessions, du nombre de crédits pris durant chaque session, et du département de l'étudiant. Il s'intéresse ainsi à tester si les informations disponibles et mesurables sont suffisantes pour fournir une prédiction informative, en dépit d'informations non disponibles, telles que des variables reflétant la psychologie, la vie personnelle ou associative de l'étudiant. Pour cela, il a développé un système basé sur les réseaux bayésiens, qui sont des modèles probabilistes graphiques, permettant ainsi d'estimer la distribution des prédictions possibles en fonction du pattern (forme) particulier d'un étudiant. Le système peut alors servir d'outil d'aide à la décision.

Taieb Khalil, B.(2021) : dans son mémoire intitulé « Etude comparative des travaux sur le traitement du langage naturel en utilisant le deep learning en milieu universitaire », il se concentre sur les émotions des étudiants et analyse leurs sentiments suivant une situation donnée en appliquant les techniques de traitement de langage naturel dans les systèmes informatiques (par ex : les chats bots) en se basant sur les algorithmes de deep learning pour avoir plus d'efficacité dans la détection et la déduction d'informations nécessaires au bon suivi des étudiants.

Gharbi N.(2024) : Dans son mémoire intitulé « un système intelligent à base des techniques de traitement du langage naturel pour le suivi pédagogique et psychologique des nouveaux étudiants universitaires », il conçoit un système intelligent qui permet de suivre les étudiants et leur permettre de solliciter de l'assistance en exprimant leurs interrogations à partir de leurs domiciles même en langage naturel. Il vise à classer les différentes formes de difficultés rencontrées par les étudiants, à proposer un moyen d'analyser les demandes exprimées dans le langage naturel, à prendre en compte le type d'assistance requise et à développer un moyen d'évaluer les contributions des assistants.

En dépit de ces contributions significatives, la plupart des travaux existants se concentrent principalement sur l'évaluation des apprenants ou sur des aspects partiels de l'évaluation des enseignants. Ainsi, notre recherche se distingue par la conception d'un modèle de classification appliqué spécifiquement au contexte des enseignants du secondaire, en utilisant des algorithmes tels que les arbres de décision et les machines à vecteurs de support, dans le but de proposer un système d'évaluation plus précis, objectif et adapté aux réalités locales.

MÉTHODES

Notre modèle est un modèle de classification multiclassé basé sur les données collectées à la sous division Kananga2, précisément au pool Secondaire ; 1500 fiches d'inspection sélectionnées et encodées dans la machine en utilisant le logiciel Microsoft Excel pour former un dataset.

Pour matérialiser notre étude, nous devons obligatoirement suivre une démarche qui se double en méthodes et techniques que nous décrivons ci-dessous :

Les méthodes ci-dessous ont été suivies tout au long de notre travail :

- ✓ **SMOTE** : Cette méthode de sur échantillonnage sera utilisée pour se concentrer sur les deux classes minoritaires à savoir

Elite et Médiocre, qui seront augmentées par la création d'instances "synthétiques" (Merwan Chelouah,2023)

- ✓ **ACP** : cette méthode sera utilisée pour étudier les liens entre les variables numériques en produisant les matrices de corrélation (Lan T. Jolliffe,2002).

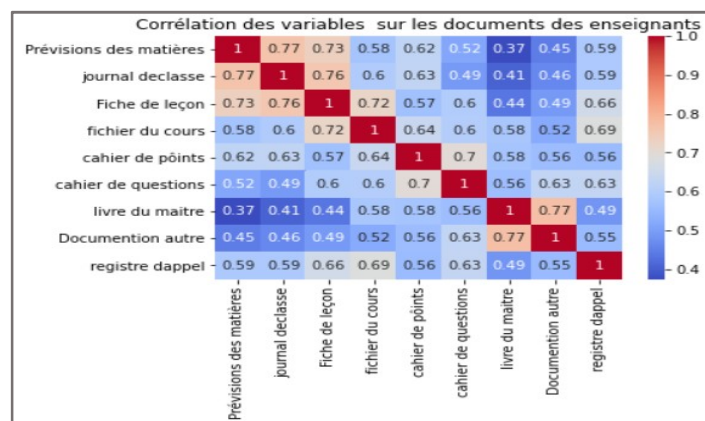
Il existe une multiplicité des techniques de recherche qu'on utilise pour rendre les méthodes fructueuses, ces techniques constituent des outils au service des méthodes pour la récolte et le traitement des données. Parmi ces techniques, nous allons recourir :

- ✓ Au Clustering: En utilisant le Kmeans pour identifier des groupes d'enseignants ayant des profils de performances semblables.
- ✓ A l'arbre de décision : pour mettre en place les catégories selon les performances (Excellent, Très bon, Bon, Assez bon, Médiocre) (Chloé-Agathe Azencott, 2018)
- ✓ A l'algorithme un contre tous : Qui utilise K classifieurs pour représenter les cinq courbes de ROC en discriminant une classe un contre toutes les autres. C'est-à-dire le modèle est entraîné pour chaque classe afin de la distinguer des autres et cela génère une courbe de ROC pour chaque classe. Pour arriver à générer une courbe de ROC, on considère une classe comme positive et toutes les autres négatives, on calcule les probabilités prédites pour cette classe ainsi de suite. (Fawcett, T., 2006)

RÉSULTATS ET DISCUSSIONS

Analyse bivariée

Corrélation des variables sur les documents des enseignants.



Source : nous-même

En étudiant cette matrice de corrélation nous constatons que : lorsque l'enseignant a ses prévisions de matières, il prépare ses leçons et remplit correctement son journal de classe. entre ces deux variables.

Les métriques

Tableau 4.1: Résultats sur les métriques

Class	Precision	Recall	F1-Score	Support
AB	0.92	0.95	0.93	74
B	0.94	0.96	0.95	85
Elite	1.00	0.99	0.99	98
M	1.00	0.97	0.98	93
TB	1.00	1.00	1.00	100
Accuracy			0.98	450
Macro Avg	0.97	0.97	0.97	450
Weighted Avg	0.98	0.98	0.98	450

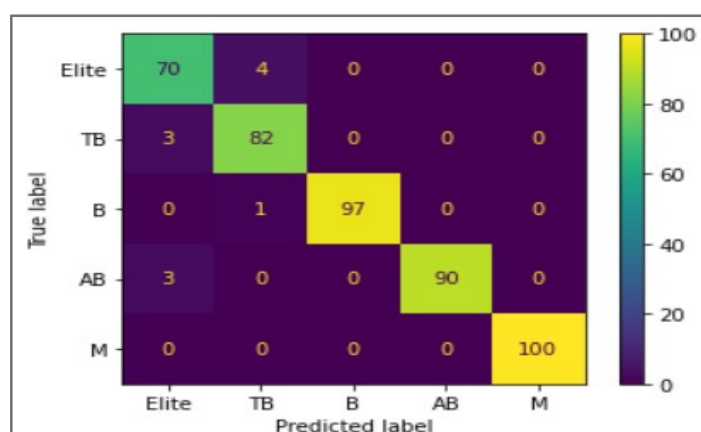
Source: nous-même

Dans le tableau ci-dessus, nous retrouvons trois paramètres synthétiques ci-dessous :

- ✓ Accuracy (Exactitude) (98%): ceci prouve à suffisance combien d'exemples de notre ensemble sont bien classés ou classés correctement. Pour déterminer ce paramètre nous divisons le nombre de prédictions correctes par le nombre total d'exemples.
- ✓ Macro Avg (Moyenne Macro) (97%): ce paramètre nous donne la moyenne des métriques pour chaque classe, donc il donne le poids de chaque classe. Alors avec 97%, nous pouvons conclure que notre modèle performe très bien de manière équilibrée sur les différentes classes et il n'y pas de faibles performances sur les classes minoritaires masquées par les classes majoritaires.
- ✓ Weight avg (Moyenne pondérée) (98%): ce dernier paramètre donne la moyenne pondérée par taille des classes. Avec ce pourcentage, nous pouvons dire que le modèle classe correctement une grande proportion des exemples, surtout ceux qui appartiennent aux classes majoritaires.

La Matrice de Confusion

Figure 4.11 : résultats sur matrice de confusion



Source: nous-même

Les vrais positifs : les valeurs de la diagonale (70,82, 97,90 et 100) représentent le nombre de bonnes classifications.

- Elite : 70 exemples correctement classés comme élite
- TB : 82 exemples correctement classés comme TB
- B : 97 exemples correctement classés comme B
- AB : 90 exemples correctement classés comme AB

- M : 100 exemples correctement classés comme M

Les faux positifs : les valeurs en dehors de la diagonale sur la même colonne indiquent les erreurs où d'autres classes ont été prédites à la place.

- Elite : 3 exemples de TB et 3 exemples de AB ont été classés comme Elite.
- TB : 4 exemples d'Elite et 1 exemple de B ont été classés comme TB.

Les faux négatifs : les valeurs de la diagonale sur la même ligne montrent les exemples qui ont été mal classés.

- Elite : 4 exemples ont été mal classés comme TB
- TB : 3 exemples ont été mal classés comme Elite
- B : 1 exemple a été classé comme TB
- AB : 3 exemples a été classé comme Elite

Concernant les autres éléments qui sont tous nuls, cela signifie que le modèle n'a commis aucune erreur de classification. il n'y a pas eu de faux positifs (instances d'autres classes prédites comme appartenant à une classe) ni de faux négatifs (instances d'une classe non reconnues).

Conclusion :

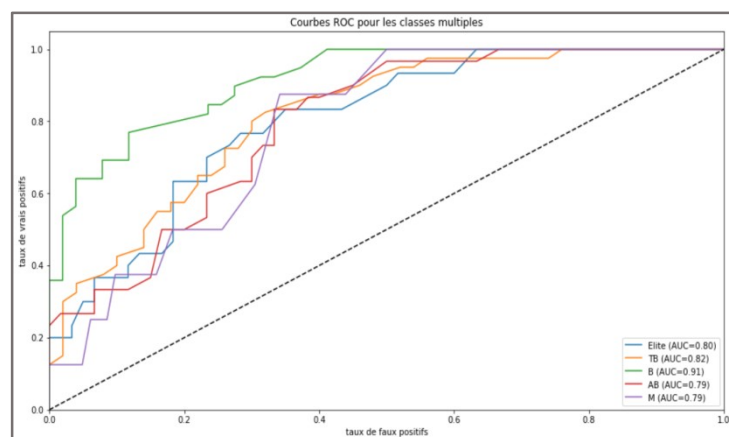
Précision : chaque classe a une bonne précision, surtout pour les classes B, AB et M qui n'ont pas d'erreurs.

Confusion entre classes : on observe quelques confusions, notamment entre Elite et TB.

Généralement : la matrice de confusion indique le modèle est plutôt performant, mais il pourrait être amélioré en réduisant les erreurs de certaines classes. Pour cela nous pouvons utiliser les techniques d'augmentations des données ou l'ajustement des hyper paramètres pour améliorer les performances

Courbes ROC

Figure 4.12 : Usage des courbes ROC



Source: nous-même

Dans la figure ci-dessus :

- Toutes les courbes sont plus proches du coin supérieur gauche cela indique une meilleure performance.
- D'après ces différents pourcentage, on se rend compte que: la classe B est excellente avec un bon pourcentage de discrimination, Elite et TB sont bonnes et enfin AB et M sont acceptables.

