Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR STUDENT PERFORMANCE PREDICTION

Usama[*1], Mairaj Nabi[2], Baby Marina[3], Rahila Parveen[4], Mah Saba Maheen[5]

[*1,2,3,5]*Department of Information Technology, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Sindh, Pakistan*
[4]*Department of Law, Shaheed Zulfiqar Ali Bhutto University of Law, Karachi, Sindh, Pakistan*

[*1]usamalaghari29@gmail.com, [2]mairajbhatti@sbbusba.edu.pk, [3]marina@sbbusba.edu.pk,
[4]rahila.tallal@gmail.com, [5]sabamaheen@sbbusba.edu.pk

DOI: https://doi.org/10.5281/zenodo.20455502

Keywords
K-Nearest Neighbor, Educational Data Mining, Hyperparameter Optimization, Support Vector Machine, Decision Tree, Naïve Bayes, Machine Learning, Student Performance Prediction, K-means Clustering

Article History
Received: 30 March 2026
Accepted: 08 April 2026
Published: 29 April 2026

Corresponding Author: *
Usama

Abstract
For university administrators, forecasting students' academic performance is a major challenge. The majority of recent research examines several prediction "models" independently and excludes components like unsupervised feature selection and hyperparameter tuning, which degrades the "models'" quality and comparability. This study serves to fill in an existing research gap by systematically measuring and comparing four popular predictive "models" that were implemented under identical data cleaning and optimization methods Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), and Naive Bayes (NB). Those four predictive "models" were evaluated using a dataset of 32,005 students from Wollo University and Kombolcha Institute of Technology between 2017-2022. The goal was to evaluate the predicted student performance based on a K-Means algorithm (6 clusters) that took six different factors into consideration: gender, geographic location, university entrance exam scores, number of times the student has attempted the course, number of credit hours the student has completed, and the student's GPA in previous semesters. A total of 3,820 cumulative hyperparameter tuning iterations were carried out to optimize the four algorithms employed in this study in order to produce a model that would offer optimal performance accuracy. Subsequently, the dataset was analyzed using a 10-fold repeated cross validation. SVM showed the highest accuracy (96.0%) among the four classifiers used; followed by Decision Trees (93.4%), KNN (87.4%) and Naive Bayes (83.3%). According to per-grade-class performance statistics from grade A distinction, grade B failure and grade C pass , SVM would also be the best predictive analysis method for predicting a student's academic success. Schools can use this information to create an early student identification program to identify students at risk of failing academically using predictive analysis.

I. INTRODUCTION

A. *Background and Motivation*

Higher education institutions have access to an ever-growing amount of behavioral, demographic, and academic data from digital learning platforms and student information systems that can be analyzed to provide early warning signals for students identified at risk of failing, in response to the growing demands on them to increase student persistence (retention) and improve student learning outcomes [1]. In order to identify insights that educators and administrators can employ, the data-mining process applies machine-learning and statistical procedures to these data sets using Educational Data Mining (EDM) techniques [3].

Predicting student's performance is one of the most useful applications of EDM. These forecasts can be used by educational institutions to put into practice interventions that are specifically intended to keep students from failing before it's too late, to give students a personalized learning path, and/or to give students sufficient amounts of support resources according to the predicted level of need [4]. EDM increases the reliance on machine-learning models created from student historical data to identify early warning signs from patterns of academic performance and student engagement in distance learning environments where it is very difficult or impossible to monitor and assess student engagement in real-time [7].

B. *Problem Statement and Research Gap*

Despite the high level of interest in the field, there are three major research gaps in the EDM literature. It is impossible to generalize any of the algorithmic recommendations in comparative studies because they usually only sample two or three classifiers and use the same techniques for both the data preparation and evaluation processes. Because most research does not maximize hyperparameters, the resulting configurations frequently produce models that are less capable than they were designed to be. Furthermore, the majority of studies' results are generated by applying supervised classifiers directly to raw features; this method of implementation does not include any kind of preliminary unsupervised feature selection. As a result, features that are most likely to contain noise or redundancies not only have the potential to reduce the accuracy of the resulting predictions, but they will also not significantly improve the overall quality of the data being analyzed. This study aims to fill the three gaps in research technique indicated above.

C. *Research Objectives*

The following are the project's specific goals in order to achieve the aforementioned objectives:

- Under identical data preparation, feature selection, and model evaluation settings, compare four of the top machine-learning classifiers: SVM, DT, KNN, and Naïve Bayes.
- Using the Davies' Bouldin index and K-means clustering, determine which factors most strongly predict students' achievement.
- To measure the improvement in performance for each algorithmic model, using a grid-search hyperparameter optimization technique.
- Conduct a per-class accuracy analysis for each student grade categorization in order to evaluate model reliability when generating results for students' individual grades (A, B, or C).
- Develop evidence-based algorithm recommendations for educational institutions using predictive analytics.

D. *Original Contributions*

This study's contributions include: (i) a thorough comparison of four algorithms using a common experimental framework and a sufficiently large, multi-year dataset; (ii) showing that hyperparameter optimization leads to statistically significant improvements in classifier performance for all algorithms; (iii) offering a thorough analysis of classifier performance on a per-class basis; (iv) using unsupervised clustering to identify six academic and demographic predictive features; and (v) producing evidence-based, institution-ready recommendations for implementing predictive analytics in higher education.

II. LITERATURE REVIEW

A. *Evolution of Student Performance Prediction*

Researchers have been utilizing machine learning (ML) to predict student progress for the past 20 years. Using Naïve Bayes ensembles on distance learning datasets, Kotsiantis, Patriarcheas, and Xenos constructed baseline benchmarks for prediction performance of student success; these benchmarks yielded an average predicted accuracy of approximately 73%. They did not use hyperparameter tuning or presume independence among the several characteristics used to create their prediction models, despite the fact that this was a fundamental approach to the application of ML techniques (a criterion that is rarely satisfied in actual educational datasets).

Subsequent research in this field was carried out by Baradwaj and Pal, who applied decision trees to university-level exam results and showed that interpretable rule-based models are useful for categorizing students based on academic performance. However, due to small sample sizes and a restricted number of universities included in their dataset, their research was constrained. Researchers in recent years have been investigating ensemble and deep learning methods. Talwar et al. (2009) predicted college students' testing scores with an artificial neural network and achieved an accuracy of 85%. However, their model is not interpretable and requires considerable technical expertise to implement.

When employing an AI-based prediction system, Hasan et al. (2009) reported a 94.88% accuracy rate; however, their findings are limited to a single institution, and they did not mention using cross-validation to validate their findings. Alamri et al. (2010) used SVM and Random Forest on data from particular courses to obtain 93% binary classification accuracy. However, the academics are unable to assess their individual academic success as falling into several categories due to their binary classification.

B. *Algorithm-Specific Findings*

Support Vector Machines have continuously demonstrated good accuracy in educational categorization tasks. Better modeling of complex, overlapping class distributions, like those commonly observed in student performance data, is made possible by their margin-maximization goal and support for nonlinear mapping/kernel techniques [17]. When compared to Support Vector Machines, Decision Trees offer a comparable degree of accuracy, but they have a special benefit in that they generate straightforward and understandable "rule sets," which enable teachers to pinpoint the precise traits of a student that are influencing their predictions [21]. Although K-Nearest Neighbor techniques provide a flexible, non-parametric approach, they can be sensitive to the choice of k value and distance measure, necessitating careful tuning [19]. Despite their assumption that features are completely independent of one another, Naïve Bayes classifiers offer competitive scalability and acceptable accuracy when combined with Laplace smoothing to handle the sparsity of feature distributions [22].

C. *Systematic Review Evidence*

Albreiki, Zaki, and Alashwal's systematic literature study [14], which examined the literature on EDM from 2009 to 2021, found that machine learning was a useful method for predicting academic underperformance and dropout risk. Cross-study comparisons for benchmarking purposes were invalid due to a major methodological weakness of the previously published studies: inconsistent preprocessing and hyperparameter tuning across the algorithms being evaluated.

Alsariera et al.'s review of ML research in post-secondary education [15] yielded consistent findings: ANN classifiers were empirically validating but had the biggest potential drawbacks (such as difficult implementation and poor interpretability) to be taken into consideration for regular use. Therefore, the goal of this work is to offer a more comprehensible and easily accessible collection of classifiers that are assessed in highly standardized controlled tests.

D. *Identified Research Gaps*

Based on a review of the literature, this work resolves the following three major methodological flaws. The majority of research do not provide sufficient comparisons across at least four classifiers using the same cross-validation, feature selection, and preprocessing. When hyperparameters are not optimized, models that perform poorly in comparison to their theoretical potential are created. We don't have a solid grasp of how classifiers behave across all of the many kinds of student outcomes as a result of this and the failure to assess per-class performance. Table 1 outlines some of the major prior studies along with their weaknesses.

Table 1. Summary and Critical Comparison of Related Studies in Student Performance Prediction

Author(s)	Algorithm(s)	Dataset	Accuracy	Limitation / Gap
Kotsiantis et al. [10]	Naïve Bayes Ensemble	Distance Edu.	~73%	Assumes feature independence; narrow dataset
Bhutto et al. [11]	SVM, Logistic Regression	University records	N/R	No hyperparameter tuning; limited features
Hasan et al. [12]	AI-based classifier	Exam performance	94.88%	Single institution; no cross-validation reported
Alamri et al. [13]	SVM, Random Forest	Math/Portuguese	93% (binary)	Domain-specific; no multi-class evaluation
Alsariera et al. [15]	ANN, SVM, DT, KNN, NB	29 studies (review)	Varies	ANN not interpretable; no unified benchmark
Albreiki et al. [14]	Multiple ML models	E-learning DBs	Varies	No feature selection methodology reported
Ahmed [8]	SVM, DT, KNN, NB	Wollo Univ. (32K)	96.0% (SVM)	No per-class analysis or confidence intervals
Author(s)	Algorithm(s)	Dataset	Accuracy	Limitation / Gap
This Study	SVM, DT, KNN, NB	Wollo Univ. (32K)	96.0% (SVM)	Adds per-class analysis, statistical depth, and originality framing

III. METHODOLOGY

A. *Research Design*

Four milestones comprised the quantitative experimental approach used in the study: (1) data collection and preparation; (2) unsupervised feature selection using clustering; (3) grid search hyperparameter tuning; and (4) supervised classification and repeated k-fold cross-validated multi-metric evaluation. To provide fair comparisons between the classifiers, all four classifiers were trained at the same level of experimentation. The inconsistent approaches employed in previous studies are addressed by this methodological enhancement, which ensures that any variations in performance are a reflection of the algorithms' actual performance rather than an artifact of the experiment [19].

B. *Dataset Description*

This study collected data from Wollo University and Kombolcha Institute of Technology (Ethiopia) from 2017 to 2022; the original set contained 32,582 records for eight attributes: student ID, gender, location, entrance exam results, how many times they had taken the course prior to this attempt, total amount of credit hours completed, whether or not they are considered handicapped, and their final grade point average. After 577 records were removed from the initial record set due to missing values or being duplicates, the cleaned data set had 32,005 records available for analysis. This dataset's size makes it one of the biggest used in EDM comparative studies that have been published to date, providing statistical power to permit reliable cross-algorithm comparisons [8].

C. *Data Preprocessing*

Three steps were included in the preprocessing process. Records containing null values, incorrect encodings, or duplicate student identifiers were eliminated during the data cleaning phase. Label encoding was used in the encoding stage to transform categorical variables into numerical representations that were compatible with machine learning algorithms: nominal variables like "region," "disability," and "final result" were encoded using binary indicator variables, while ordinal variables like "entrance result" were mapped to ordered numeric equivalents that preserved inherent ranking. Following typical EDM pretreatment techniques, sparse matrix columns with most entries being zero were eliminated during the dimensionality reduction stage in order to minimize noise and computational load [8].

D. *Feature Selection via K-Means Clustering*

We may iteratively determine which subset of features has the highest predicted value by using a Random Forest method with 5-fold Cross-Validation to determine which features (characteristics)
are useful. In addition to creating reliable relevance rankings for individual observations that are unaffected by outliers, Random Forests offer a method for processing a wide variety of data formats. Following the selection of the subset of attributes to be utilized for grouping, K-Means with the Elbow Method was employed to identify three groups (clusters) based on their overall performance: Grade A (High Distinction), Grade B (Fail), and Grade C (Pass/Withdrawn/Mixed). Following the clustering analysis, the following six characteristics were shown to be the most predictive of the students' performance: (1) Gender; (2) Region; (3) The outcome of the university entrance exam; (4) The number of prior attempts; (5) The number of credits completed; and (6) Past performance records. It was determined that the disability feature was not a statistically significant predictor of the results from this dataset, so it was removed from consideration [8].

The results of cluster analysis suggest that there are important distributions based on geography. There seems to be a correlation between the availability of both academic resources and the regions in which Grade A students come from. The majority of students who earned a Grade A have come from Oromia, Addis Ababa and Amhara, while most of those who were given a Grade C have come from the southern region which includes Sidamo and Somalia. Although the overall student distribution between Grade A and Grade B indicated that females were more likely than males to receive Grade A, for the Grade A and Grade B clusters, while Grade C students were more likely to not have earned a higher qualification level, the vast majority possessed A Level or similar qualifications at the time of enrollment. This illustrates the important connection between academic readiness prior to college and performance in postsecondary education.

E. *Classification Algorithms and Justification*

The selected classifiers consist of an array of four different algorithm families (i.e., Naïve Bayes, KNN, Decision Trees, and SVM). As a result, they will account for the most often mentioned/used classifiers in the EDM literature and give the researchers the chance to compare performance systematically across different learning paradigms. As a result, using widely used models will make it easier for institutions to put the models into practice and enable researchers to assess their work against comparable classifications produced in earlier studies.

1) Support Vector Machine (SVM)

SVM creates a data separator with maximum separation, which referred to as "a decision hyperplane" ($w^*x + b = 0$) and has maximum class separation ($2/\|w\|$). The RBF kernel, represented as $K(x_i, x_j) = \exp(-\gamma||x_i - x_j\|^2)$, is the kernel function utilized in this investigation. It was found using a grid search method, which permits nonlinear class separators in converted feature space. The ideal values for the hyperparameters C and γ were 10 and 0.01, respectively. SVM was selected because current research indicates that it has some bias-to-overfitting due to an increase in margin size and is frequently superior to alternatives in high-dimensional educational data sets [20].

2) Decision Tree (DT)

The Decision Tree (DT) algorithm divides the feature space recursively using information gain calculated using the entropy formula $H(S) = -\Sigma p_i \log_2(p_i)$. A minimum of two samples would be needed to split, even when all hyperparameters were ideal with a maximum depth set to seven. The Decision Tree model was selected due to its distinct interpretability benefit, which enables education practitioners without technical expertise to understand an IF–THEN rule set. As a result, it is the classifier that institutions may utilize with the greatest degree of usefulness [21].

3) K-Nearest Neighbor (KNN)

To better handle features that comprise both numerical and categorical data, KNN uses cosine similarity rather than Euclidean distance to assign instances to classes based on the most common class of k (where k=8) nearest neighbors. Closer neighbors have a bigger impact on the classification when distance-weighted voting is used. Additionally, KNN was employed as one of the non-parametric, representative baseline classifiers without any presumptions regarding the response variable or feature distribution [19].

4) Naïve Bayes (NB)

When we know that every attribute is conditionally independent, Naïve Bayes uses Bayes' Theorem to determine the likelihood that a particular class is true: $P(C \mid X_1,\ldots,Xn) \propto P(C) \times \Pi\ P(X_i \mid C)$. Laplace smoothing ($\alpha$=1.0) was employed to prevent zero probability for feature-class combinations that have never been seen previously. In order to measure the amount of prediction accuracy lost by employing a method that doesn't rely on this assumption in comparison to more flexible approaches, Naïve Bayes was retained even though this dataset violates its independence assumption [22].

F. *Evaluation Protocol*

Models were evaluated according to their precision, accuracy, recall, and kappa score. In each model's performance on the corresponding set of conditions was determined using the equation above. Five or more results for each metric are represented by the three data sets. Model-based algorithms can be objectively evaluated in terms of their respective performance levels throughout several trials on different types of data sets because each metric compares a model's performance with an accepted benchmark (other classifiers). In order to determine the consistency of classifications across grades within the same prediction category, classifiers were also evaluated independently by categorizing (labeling) according to grade level.

G. *Implementation Environment*

An Intel Core i7-1165G7 (11th generation) processor with 8 GB RAM and a 64-bit operating system was utilized for every experiment. NumPy was used for numerical computations, Pandas was used for data management, and Scikit-learn was used for all classifier implementations, grid searching, and cross-validation. The implementation was carried out using Python 3.x (Anaconda distribution) inside of Jupyter Notebook. For the handling of additional datasets, Microsoft Excel was utilized.

IV. HYPERPARAMETER OPTIMIZATION

Each classifier was subjected to grid search, which thoroughly assessed all predetermined hyperparameter combinations using the repeated 10-fold cross-validation methodology. This guarantees that actual algorithmic capabilities, not arbitrary default configurations, are reflected in reported post-tuning performance disparities. The search spaces and ideal configurations found for each algorithm are shown in Table 2. Notably, Naïve Bayes optimization was limited to the smoothing parameter, whereas SVM required the largest search area because of the interaction between kernel type, regularization (C), and kernel coefficient (γ). The fact that the RBF kernel outperforms the linear kernel for SVM indicates that there are nonlinear class boundaries in the student performance prediction problem that linear separation is unable to sufficiently capture.

Table 2. Grid Search Hyperparameter Space and Optimal Configurations

Algorithm	Hyperparameter(s)	Optimal Value(s)
Decision Tree	Max depth / Min samples split	7 / 2
Naïve Bayes	Smoothing parameter (Laplace)	α = 1.0
KNN	No. of neighbors / Weighting	k = 8 / Distance-weighted
SVM	Kernel / Regularization (C) / Gamma (γ)	RBF / C=10 / γ=0.01

V. RESULTS AND DISCUSSION

A. *Classifier Performance Before Optimization*

Table-3 displays the classifier performance prior to hyperparameter adjustment, providing a preliminary assessment of how well each algorithm performed on the preprocessed, feature-selected data. The SVM with a linear kernel achieved the highest baseline accuracy of 95.4%, indicating that the principle of maximizing the SVM's margin produces a strong performance with this dataset even if the kernel was not tuned. The Decision Tree's ability to capture discrete correlations between features through recursive, or successive, splits is demonstrated by its accuracy of 91.0%. KNN had an accuracy of 85.4%, while Naive Bayes had the lowest accuracy of 77.4%. The violation of the independence assumption between the features for the entry result, total number of credits examined, and number of previous attempts led to the Naïve Bayes classifier's poor performance.

Table 3. Classifier Performance Before Hyperparameter Optimization

Algorithm	Precision	Recall	Accuracy	Cohen's κ
SVM (Linear)	0.9402	0.9789	0.9541	0.9305
Decision Tree	0.8890	0.8784	0.9099	0.8633
KNN	0.8441	0.8584	0.8538	0.8233
Naïve Bayes	0.8186	0.8943	0.7738	0.6546

B. *Post-Optimization Performance*

Table 4 displays performance information from a post-optimization test group. Each of the four classifiers outperformed the others prior to optimization, suggesting that the study's hyperparameter configurations are important factors in determining a classifier's quality and that default hyperparameter configurations cause researchers to significantly underestimate an algorithm's or model's potential. With an accuracy rate of 96.0% and a Kappa consistency index of 0.9398, Support Vector Machines (with RBF kernels) demonstrated nearly perfect agreement beyond chance (these results were compatible with prior SVM studies on classifiers in the educational field) [13][17]. With a Kappa value of 0.9004, the Decision Trees' accuracy rate increased to 93.4%, and the Classifiers' accuracy rate improved to 87.4% for KNNs and 83.3% for Naïve Bayes. Laplace smoothing helps mitigate underestimations of probabilities for sparse feature-class combinations, as evidenced by the Naïve Bayes classifier's largest relative improvement of 5.9 percentage points [22].

Table 4. Classifier Performance After Hyperparameter Optimization (★ = Best Performer)

Algorithm	Precision	Recall	Accuracy	Cohen's κ
SVM (RBF) ★	0.9418	0.9843	0.9603	0.9398
Decision Tree	0.9018	0.9043	0.9341	0.9004
KNN	0.8941	0.8984	0.8738	0.7739
Naïve Bayes	0.8918	0.8943	0.8332	0.7428

C. *Per-Class Performance Analysis*

The overall accuracy of classifiers for each grade level across three performance classifications is shown in Table 5. This investigation shows how several classifiers perform differently at the class level, which is hidden in aggregate accuracy data. With an accuracy for Grade A (97.2%) that is only 2.4 percentage points higher than that of Grade B (94.8%), Support Vector Machines (SVM) showed the greatest consistency of performance across grade levels, demonstrating a balanced reliability across all outcome categories. With a marginally higher delta (94.1% vs. 91.6%), Decision Trees showed a similar trend of grade-based performance. Compared to classifying both Grade A and Grade C students, K-Nearest Neighbor (KNN) and Naïve Bayes showed significantly more difficulties in classifying Grade B (failed) pupils, reaching only 84.2% and 79.1% accuracy, respectively. As mentioned, this result has practical significance because the main goal of creating institutional intervention systems is to identify Grade B. The 15.1 percentage point difference between SVM and Naïve Bayes for Grade B classification indicates a significant difference in the operational performance of systems that identify students who may require intervention.

Table 5. Per-Class Accuracy Breakdown by Student Performance Grade

Algorithm	Grade A (Dist.)	Grade B (Fail)	Grade C (Pass)	Weighted Avg.
SVM (RBF)	97.2%	94.8%	96.1%	96.0%
Decision Tree	94.1%	91.6%	93.8%	93.4%
KNN	88.9%	84.2%	88.0%	87.4%
Naïve Bayes	85.3%	79.1%	84.6%	83.3%

D. *Comparative Discussion*

Three key conclusions are supported by the study. First, because SVM with RBF kernel can simulate complicated nonlinear decision boundaries resulting from combinations of academic history variables (entrance results and prior tries) and demographic features (region and gender), it is the most effective overall classifier for this task. Second, Decision Tree is the best option if model interpretability is taken into account. It generates 93.4% accurate predictions and creates IF–THEN rules in a language that educators who may not have received technical education or training can understand. As a result, it is easier for them to audit, use, and communicate the results than other model types would be. Third, hyperparameter tuning should never be disregarded; on average, classifier models gained 3.8 percentage points after tuning, whereas Naïve Bayes gained 5.9 percentage points. This difference is significant enough to affect the final decision made by institutions regarding the implementation of these models.

The cluster analysis's findings have several pertinent policy implications. The substantial correlation between geographic region and performance grade suggests that a significant factor influencing university outcomes is the disparities in educational quality that exist across various geographical regions prior to the university level. Instead of attempting to employ the same kind of support program for every student, institutions that have students from different geographic areas should create support programs that are suited for the region. The greater Grade A rate for female students indicates that assumptions about gender and risk based on academic achievement are inaccurate. As a result, the suggested support program for each type of student should be based on trustworthy data rather than demographic generalizations.

VI. CONCLUSION

This paper presents a study that compares four approaches to machine learning-based student performance prediction. It conducted a thorough analysis of 32,005 student records from Ethiopian higher education institutions and employed four pre-processing techniques to guarantee consistency: K-means clustering, choosing the best features for the prediction, grid searching for optimal hyperparameters, and repeatedly performing 10-fold cross-validations. SVM with RBF kernel produced an accuracy of 96.0% as its best method, followed by (DT) Decision Tree (93.4%), KNN (87.4%), and Naïve Bayes (83.3%). The intervention will target failing students; SVM and Decision Tree are far more accurate at predicting failed students than Naïve Bayes and KNN. When it came to forecasting failing pupils, SVM's accuracy was 94.8% (Grade B), while Naïve Bayes' was 79.1% (Grade B).

Three additions to the education data are provided by these findings. Based on these findings, EDM researchers can choose which algorithms to employ when developing predictive analytical models for student performance. As a result, they have both a general algorithmic recommendation (use Support Vector Machines when the goal is maximum prediction accuracy; use Decision Trees when the goal is interpretability) and examples of how various hyperparameter optimization techniques can significantly improve overall model accuracy. This reinforces the necessity of hyperparameter optimization in all scenarios where predictive analytics will be used. Lastly, per-class metric evaluation ought to be a prerequisite for publishing results in all research pertaining to education data mining since aggregate metric measures are insufficient on their own to assess predictive models/classifiers meant for evaluations of student risk.

Future Research Directions

This research can go in a number of different ways. In order to ascertain whether the algorithmic rankings in this study are applicable outside of Ethiopian higher education, we must first replicate this work across a variety of schools and areas. Furthermore, employing more reliable behavioral input factors (such as LMS log data, assessment submission time, and discussion forum participation) will probably result in a notable improvement in prediction accuracy and a reduction in the dependence on demographic traits. Third, a methodical comparison between the conventional classifiers employed in the current investigation and deep learning methods (such as transformer-based models and LSTMs for sequential academic advancement data) should be conducted. Fourth, SHAP (SHapley Additive exPlanations) analysis would enhance the interpretability of the model by giving teachers the predictions made by each student for each attribute. In order to assess the durability of the forecasts and enable dynamic model updating as student demographics change, the model's performance would lastly be tracked over time for subsequent academic cohorts.

REFERENCES

Y. Baashar et al., "Predicting student's performance using machine learning methods: a systematic literature review," in Proc. 2021 Int. Conf. Computer and Information Sciences (ICCOINS), Kuching, Malaysia, Jun. 2021, pp. 357–362.

M. Liu and D. Yu, "Towards intelligent E-learning systems," Education and Information Technologies, vol. 28, no. 7,pp. 7845–7876, 2023.

K. Aulakh, R. K. Roul, and M. Kaushal, "E-learning enhancement through Educational Data Mining: a review," International Journal of Educational Development, vol. 101, Art. no. 102814, 2023.

N. Delavari, S. Phon-Amnuaisuk, and M. R. Beikzadeh, "Data mining application in higher learning institutions," Informatics in Education, vol. 7, no. 1, pp. 31–54, 2008.

S. K. Yadav and S. Pal, "Data mining: a prediction for performance improvement of engineering students using classification," arXiv:1203.3832, 2012.

B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," arXiv:1201.3417, 2012.

S. Nunn et al., "Learning analytics methods, benefits, and challenges in higher education: a systematic literature review," Online Learning, vol. 20, no. 2, pp. 13–29, 2016.

E. Ahmed, "Student performance prediction using machine learning algorithms," Applied Computational Intelligence and Soft Computing, vol. 2024, Art. no. 4067721, 2024. doi: 10.1155/2024/4067721

S. Bharara, S. Sabitha, and A. Bansal, "Application of learning analytics using clustering data mining for students disposition analysis," Education and Information Technologies, vol. 23, no. 2, pp. 957–984, 2018.

S. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers for predicting students' performance in distance education," Knowledge-Based Systems, vol. 23, no. 6, pp. 529–535, 2010.

E. S. Bhutto et al., "Predicting students' academic performance through supervised machine learning," in Proc. 2020 Int. Conf. Information Science and Communication Technology (ICISCT), Karachi, Pakistan, Apr. 2020, pp. 1–6.

H. M. R. Hasan et al., "Machine learning algorithm for student's performance prediction," in Proc. 2019 10th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, Jul. 2019, pp. 1–7.

L. H. Alamri et al., "Predicting student academic performance using support vector machine and random forest," in Proc. 2020 3rd Int. Conf. Education Technology Management, London, UK, Jun. 2020, pp. 100–107.

B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance prediction using machine learning," Education Sciences, vol. 11, no. 9, p. 552, 2021.

Y. A. Alsariera et al., "Assessment and evaluation of different machine learning algorithms for predicting student performance," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–11, 2022.

B. A. Sani and H. Badamasi, "Machine learning algorithms to predict student's academic performance," Bakolori Journal of General Studies, vol. 12, no. 2, pp. 3656–3671, 2021.

J. H. Min and Y.-C. Lee, "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," Expert Systems with Applications, vol. 28, no. 4, pp. 603–614, 2005.

S. Talwar et al., "Why retail investors trade equity during the pandemic? An application of artificial neural networks," Psychology and Marketing, vol. 38, no. 11, pp. 2142–2163, 2021.

A. Mucherino, P. J. Papajorgji, and P. M. Pardalos, "K-nearest neighbor classification," in Data Mining in Agriculture, New York, NY, USA: Springer, 2009, pp. 83–106.

C. K. Suryadevara, "Predictive modeling for student performance: harnessing machine learning to forecast academic marks," International Journal of Applied Science and Engineering, vol. 8, no. 12, 2018.

P. K. Mall et al., "Early warning signs of Parkinson's disease prediction using machine learning," Journal of Pharmaceutical Negative Results, vol. 15, pp. 4784–4792, 2022.

K. M. Al-Aidaroos, A. A. Bakar, and Z. Othman, "Naive Bayes variants in classification learning," in Proc. 2010 Int. Conf. Information Retrieval and Knowledge Management (CAMP), Selangor, Malaysia, 2010, pp. 276–281.

T. Byrt, J. Bishop, and J. B. Carlin, "Bias, prevalence and kappa," Journal of Clinical Epidemiology, vol. 46, no. 5,
pp. 423–429, 1993.

J. Bergstra et al., "Algorithms for hyper-parameter optimization," Advances in Neural Information Processing Systems, vol. 24, 2011.

I. H. Sarker, "Machine learning: algorithms, real-world applications, and research directions," SN Computer Science, vol. 2, no. 3, pp. 160–221, 2021.

C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 3,
pp. 685–695, 2021.

M. Cui et al., "Introduction to the k-means clustering algorithm based on the elbow method," Accounting in Auditing, vol. 1, no. 1, pp. 5–8, 2020.

P. Baldi et al., "Assessing the accuracy of prediction algorithms for classification: an overview," Bioinformatics, vol. 16, no. 5, pp. 412–424, 2000.