Development of machine learning algorithms in student performance classification based on online learning activities

Muhammad Aqif Hadi Alias, Mohd Azri Abdul Aziz, Najidah Hambali, Mohd Nasir Taib
School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Selangor, Malaysia

Article Info

Article history:

Received Jun 6, 2024
Revised Jul 19, 2024
Accepted Aug 6, 2024

Keywords:

Classification algorithms
Feature selection
K-nearest neighbors
Logistic regression
Multi-layer perceptron
Student performance
Synthetic minority oversampling technique

ABSTRACT

The field of educational data mining has gained significant traction for its pivotal role in assessing students' academic achievements. However, to ensure the compatibility of algorithms with the selected dataset, it is imperative for a comprehensive analysis of the algorithms to be done. This study delved into the development of machine learning algorithms utilizing students' online learning activities to effectively classify their academic performance. In the data cleaning stage, we employed VarianceThreshold for discarding features that have all zeros. Feature selection and oversampling techniques were integrated into the data preprocessing, using information gain to facilitate efficient feature selection and synthetic minority oversampling technique (SMOTE) to address class imbalance. In the classification phase, three supervised machine learning algorithms: k-nearest neighbors (KNN), multi-layer perceptron (MLP), and logistic regression (LR) were implemented, with 3-fold cross-validation to enhance robustness. Classifiers' performance underwent refinement through hyperparameter tuning via GridSearchCV. Evaluation metrics, encompassing accuracy, precision, recall, and F1-score, were meticulously measured for each classifier. Notably, the study revealed that both MLP and LR achieved impeccable scores of 100% across all metrics, while KNN exhibited a noticeable performance boost after using hyperparameter tuning.

Corresponding Author:

Mohd Azri Abdul Aziz
School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM)
Selangor, Malaysia
Email: azriaziz@uitm.edu.my

1. INTRODUCTION

The performance of students in educational institutions has garnered increasing attention in which a substantial number of institutions have recognized this as a pivotal determinant in enhancing both the overall quality of the institutions and the educational outcomes of their students [1]–[3]. Identifying at-risk students early in the course offers us the capacity to implement interventions and initiatives to improve their academic performance [4]–[10]. Consequently, in the pursuit of a deeper comprehension of the learning process and the environmental factors influencing it, the field of educational data mining has gained notable momentum. This discipline assumes a critical role in the classification of students' academic achievements [11], [12]. The application of artificial intelligence in education, particularly machine learning, has increased, with the technology expected to give effective approaches to enhance education in general in the near future [13]. Intelligent m-learning systems have recently gained traction as a method of offering more effective education and flexible learning that is tailored to each student's learning ability [14]. The early attempts to enable such systems, for creating tools to help students and learning in a conventional or online context, through the use of machine learning techniques focused on anticipating student achievement in terms of grades attained [15].

Despite the importance of data preprocessing procedures, classification models must be well-developed to provide more accurate classification performance, considering the suitability of the algorithms with the selected dataset. Thus, expanding the research in this area will provide further insights for improving student performance classification in terms of the capability of classification algorithms as well as the aspects that may contribute to their success.

Machine learning plays a pivotal role in educational data mining by focusing on the prediction of students' academic performance to enhance the quality of learning. Within this context, our emphasis will be on supervised machine learning methods like k-nearest neighbors (KNN), artificial neural networks (ANN), and logistic regression (LR). Notably, various researchers have conducted studies to evaluate student performance in the learning process through the application of supervised machine learning techniques.

The supervised machine learning technique known as KNN algorithm is utilized to estimate the likelihood of a data point being categorized into one of two groups based on feature similarities. Previous studies [16], [17] employed multiple classification algorithms to predict student academic performance efficiently. In particular, a comparative analysis revealed that the genetic algorithms (GA) feature selection approach, in conjunction with the KNN algorithm, achieved the highest accuracy of 91.37% [16]. Furthermore, a comprehensive evaluation involving diverse feature selection techniques and classification algorithms demonstrated that the minimum-redundancy-maximum-relevance (mRMR) feature selection approach, combined with the KNN classifier and a selection of 10 features, yielded an accuracy of 91.12% [17].

In the study by Wafi *et al.* [11], the authors employed the modified k-nearest neighbor (M-KNN) algorithm to classify students' academic performance. The primary objective of this investigation was to compare the performance of M-KNN against the conventional KNN method. The evaluation criterion employed in this research focused on the accuracy of classification. Remarkably, the application of genetic algorithms was found to enhance the accuracy of M-KNN, achieving an accuracy rate of 82.6%, whereas the traditional KNN method yielded a lower accuracy of 73.6%.

The KNN algorithm represents a classification method analyzed in studies [18], [19], and its performance was evaluated in comparison to other classifiers, including support vector machine (SVM), naïve Bayes (NB), decision tree (DT), and discriminant analysis (DISC). Various feature selection methods were employed to assess the effectiveness of KNN, and it emerged as a significant factor in both studies. Ajibade *et al.* [18] observed that the KNN classifier outperformed other classifiers in the context of student data, based on the derived findings. Meanwhile, Abdelkader *et al.* [19] conducted an evaluation, measuring subset quality with varying cardinalities in terms of prediction accuracy and the number of selected features for 11 wrapper-based feature selection (FS) algorithms, utilizing KNN and SVM as baseline classifiers. In terms of exploration and exploitation abilities (fitness), the sequential forward selection (SFO) method with KNN and SVM demonstrated higher performance by discovering a subset of only four features out of 20.

Artificial neural networks are increasingly popular in various fields, including education. Artificial neural networks consist of interconnected artificial neurons, each assigned individual weights, and are typically organized into three layers: the input layer, hidden layer, and output layer. Throughout the learning process, these networks modify their structure based on both internal and external input. The multilayer perceptron (MLP) is a popular feed-forward neural network that sends data from the input layer to the neurons in the output layer. Both research [20], [21] employed the MLP model within the ANN framework. Notably, these studies reported a substantial positive impact on accuracy, achieving an impressive accuracy rate of 93% [20].

Numerous studies consistently identify ANN as a prominent and effective classifier, exhibiting superior performance when compared to alternative methods [17], [22]–[26]. In one comprehensive study, Dafid and Ermatita [23] delved into the classification stage by evaluating five distinct classifiers, namely DT, KNN, ANN, NB, and SVM. Their goal was to find the most effective classifier after considering various feature selection methods. Their analysis clearly showed that ANN and DT were the top classifiers. These two classifiers not only demonstrated higher accuracy but also exhibited remarkable precision, recall, and F1-score, outperforming their counterparts [23], [26]. Additionally, Amrieh *et al.* [22] reported that the ANN model surpassed other data mining approaches, including the NB classifier and DT. The study found that the accuracy rate of ANN was 73.8% when it used behavioral features and 55.6% when it used non-behavioral features.

Certain studies present an alternative perspective on the performance of ANN when coupled with various feature selection techniques. For instance, Punlumjeak and Rachburee [17] revealed that the GA feature selection method achieved the highest accuracy, reaching 90.6% when combined with the ANN classifier. Before the application of feature selection methods, the ANN correctly classified about 78.3% of instances. After implementing the information gain (IG) method, this rate improved to 79.375%, corresponding to 376 instances and 381 instances correctly classified, respectively [25].

In some studies, the development of ANN models involved the significant application of the backpropagation algorithm [27], [28] and cross-validation [20], [29]. The backpropagation algorithm played a crucial role in establishing effective connections between neurons by adjusting the connection weights to create a well-structured neural network. Results indicated that employing the MLP yielded more accurate results than decision tree, with accuracy percentages spanning from 42% to 97% [27]. In addition, Tomasevic *et al.* [28] achieved the highest overall precision in their study by combining the backpropagation algorithm with cross-validation. They employed an artificial neural network, feeding it with student engagement data and past performance records while testing various numbers of hidden layers. Cross-validation within the MLP model enabled successful dataset predictions, including 223 out of 524 students in one scenario and 83 out of 178 students in another when employing percentage splits [20]. Additionally, cross-validation proved valuable for fine-tuning the models during the training process, using 5-fold cross-validation to determine the optimal parameter values [29].

In a study conducted by Abbasi *et al.* [30], a neural network was constructed with five layers, including three hidden layers. The study's simulations explored three specific scenarios by varying the number of hidden layers (2, 3, and 4 layers). The findings consistently pointed to optimal performance when employing three hidden layers, effectively addressing concerns related to overfitting or underfitting. Imdad *et al.* [21] identified the best configuration for classification using an ANN. They discovered that the configurations with two hidden layers, a momentum value of 0.2, and a learning rate of 0.3 was the most effective. With these configurations, yielding a 100% accuracy rate, fewer errors per epoch, and a reduction in time and errors.

In another study [29], grid search and randomized search techniques were employed to identify the optimal hyperparameter values for three classifiers: ANN, SVM, and random forest (RF). The performance of the ANN model dramatically increased after this fine-tuning procedure. Accuracy went from 90.94% to 92.00%, recall climbed from 94.41% to 95.76%, precision improved from 88.29% to 89.07%, and the F1-score increased from 91.25% to 92.29%. Meanwhile, in the context of predicting students' chances of graduating from a tertiary institution, Olalekan *et al.* [31] found that the ANN outperformed the Bayes theorem in terms of accuracy. They observed that as the number of hidden layers in the ANN increased, so did the accuracy. Notably, using four hidden layers yielded the highest performance, achieving an impressive accuracy rate of 99.97% on the training dataset.

Binary classification, achieved through the logistic regression method, provides a means to predict the likelihood of an outcome for a category-dependent variable. The logistic function allows the dependent variable to be classified into one of two possible outcomes by converting a linear combination of independent factors into a probability score between 0 and 1. Two studies [32], [33] underscore the significance of logistic regression in comparison to various machine learning algorithms when predicting student performance. These algorithms include linear discriminant analysis (LDA), KNN, SVM, DT, RF, NB, and LR.

In the context of categorizing students as “High risk” or “Low risk,” Ramaswami *et al.* [33] determined that the logistic regression model exhibited the highest F1-score among various classifiers. Alraddadi *et al.* [32] observed that LR and LDA outperformed other classifiers based on the area under the curve (AUC) value while in a study focused on maintaining the integrity of online student assessments [34], several machine learning algorithms, including RF, LR, SVM, KNN, and NB were applied alongside two feature engineering techniques: mutual information (MI) and analysis of variance (ANOVA). Based on the performance of the classifiers using the top five features chosen by MI, the results showed that LR was the second-best classifier, with an accuracy of 82% and F-score of 72% top of form.

In summary, this work will evaluate prior research that also considered data preprocessing approaches and present several classification algorithms. This is the format for the remainder of the paper: The methods employed are presented in section 2. Results from the experiment are discussed in section 3, and the study's overall process is wrapped up in section 4.

2. METHOD

In this research, we aimed to develop and assess several supervised machine learning algorithms for predicting students' academic performance based on their online learning activities. Specifically, we employed three algorithms: KNN, ANN, and LR as classifiers to predict students' performance. These classifiers were subjected to performance analysis and comparison. The dataset for this study comprised 102 students enrolled in the School of Electrical Engineering, College of Engineering at Universiti Teknologi MARA (UiTM), Malaysia. The input data was extracted from Online Learning Center for ECE431 [35], which represented students' online learning activities. The primary objective of this study was to predict the students' academic performance for the semester and identify the most significant predictive model. A flowchart of the entire research process is depicted in Figure 1.

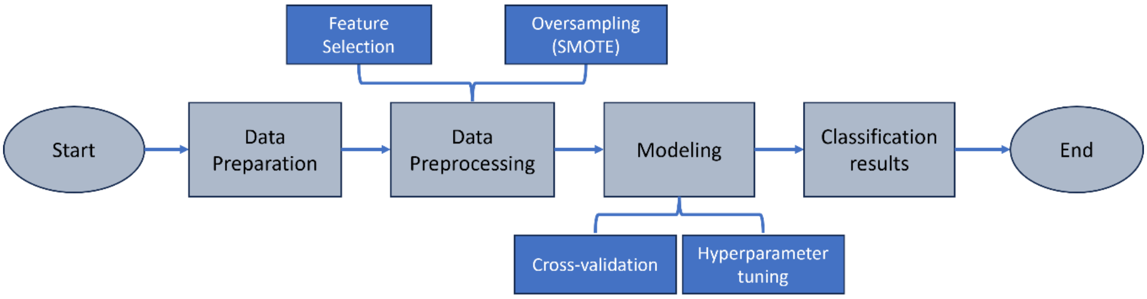

Figure 1. Flowchart of the process

2.1. Data preparation

In our study, we aimed to categorize data that pertains to three key aspects of online learning activities: notes, exercises, and tutorials. The target variable in our dataset was the students' grades, which were categorized as pass or fail, based on their course grade point average (GPA). The electronic and electrical engineering program students of Universiti Teknologi MARA (UiTM), Malaysia, specifically in the ECE431 programming course, were analyzed based on their online learning activities which then will be structured as a dataset over the course of one semester. An outline of the students' online learning activities is shown in Table 1. Students' access to the notes on the online learning platform are represented by Features1 through Feature5, their attempts to complete exercises on the online learning platform are represented by Features6 through Feature9, and the patterns of students' tutorial responses in the online learning platform are captured by Features10 through Feature12. Our dataset comprised approximately 102 samples, consisting of early-semester students in the electronic/electrical engineering program, all of whom were assessed based on these three categories, reflecting their efforts to improve their academic performance.

Table 1. Details of the online learning activities

Category	Description	Label
Notes	Students’ access to the notes before the class for the upcoming lesson begins	Feature1
	Students’ access to the notes after the class lesson has started	Feature2
	Students’ access to the notes after the first class has ended	Feature3
	Students’ access to the notes after all classes have ended	Feature4
	Length of notes left by the students	Feature5
Exercises	Students do the exercise before the class for the upcoming lesson begins	Feature6
	Students do the exercise after the class lesson has started	Feature7
	Students do the exercise after the first class has ended	Feature8
	Students do the exercise after all classes have ended	Feature9
Tutorials	Students get 3 questions and above correct	Feature10
	Students answer all tutorial questions	Feature11
	Students get wrong answer for the questions before the questions of correct answer	Feature12

2.2. Data preprocessing

This section presents the early phase of classification process which is known as data preprocessing. This step is crucial in any classification so that the model can be predicted accurately. Due to its simplicity and time efficiency, we utilized the feature selection technique, IG, to identify the most crucial features. These selected features will be utilized in the subsequent classification stage as a new subset of data whereby it will be split into 70:30 ratio of training and testing sets.

2.2.1. Information gain

In the realm of decision tree algorithms, particularly within the domain of feature selection, the term “information gain” assumes significance. This metric plays a pivotal role in assessing the relevance or importance of a feature when classifying or predicting a target variable. The foundational principle of information gain lies in the concept of entropy, which is integral to the associated formula. This formula compares the information gain of each independent feature to the information gain of the dependent feature in order to quantify the reduction in entropy. The selection process then favors the feature that exhibits the highest information gain [36]:

– Entropy (H(S)): Entropy measures the degree of disorder or impurity in a collection of data. Using the formula (1), it is computed within the framework of a dataset that has numerous classes:

$$H(S) = -\sum_{i=1}^{n} p_i log_2(p_i) \tag{1}$$

where n is the number of classes and p_i is the proportion of samples in a certain class i.

- Entropy of a feature (H(A)): Feature A's entropy in relation to a target variable S is calculated as the weighted sum of the entropies of the subsets created by dividing the data based on the values of the feature A (2):

$$H(A) = \sum_{v \in values(A)} \frac{|S_v|}{|S|} \times H(S_v) \tag{2}$$

where $values(A)$ represents the set of all potential values for feature A, $|S_v|$ denotes the number of samples in which feature A has a value of v. $|S|$ describes the total number of samples, and $H(S_v)$ denotes the entropy of the subset corresponding to value v of feature A.

- Information gain (IG), as determined in (3), is the amount of entropy that is reduced when data is divided according to a particular attribute.

$$IG(A) = H(S) - H(A) \tag{3}$$

where $H(S)$ signifies the entropy of the initial dataset and $H(A)$ represents the entropy of the dataset following the split based on feature A.

2.2.2. Class imbalance

Class imbalance is one of the issues that has been pointed out in this study since it may imply bias towards the dataset used. To address the issue of class imbalance depending on the target variable, the dataset will be resampled using the synthetic minority over-sampling technique (SMOTE). It functions by generating new synthetic instances based on the minority class up to the majority class. Therefore, the resampled data will be used in the next step which is basically for the training purpose.

2.3. Modelling

This section focuses on developing several machine learning algorithms—specifically KNN, ANN, and LR—and evaluating their performance in classifying student performance using particular assessment measures. The procedure starts with dataset preprocessing, which involves filter-based feature selection with Information Gain. This stage guarantees that only the most informative features are utilized in the subsequent modeling procedure. Following feature selection, GridSearchCV was used to tune hyperparameters with threefold cross-validation. This method systematically explores a preset grid of hyperparameters to improve the model's performance and robustness. Cross-validation helps to reduce overfitting by validating the model on several subsets of the data.

Furthermore, the issue of class imbalance in the target variable was addressed using SMOTE. SMOTE creates new synthetic instances for the minority class, corresponding to the number of instances in the majority class throughout each fold. This technique ensures that the model is trained on a well-balanced dataset, which is critical for achieving consistent and accurate results. By merging these processes, the study attempts to provide a comprehensive evaluation of the selected machine learning algorithms for classifying students' academic performance.

2.3.1. K-nearest neighbors

K-nearest neighbors is a straightforward and extensively used supervised machine learning technique for classification and regression applications. It is a sort of instance-based learning in which the model predicts based on the similarity of a new data point to its k-nearest neighbors in the training data. In prediction phase, when a new data point is provided for prediction, the algorithm computes the distances (*e.g.*, Euclidean distance) between it and all the points in the training dataset and the k-nearest data points (neighbors) were identified based on the estimated distances. In classification tasks, the algorithm predicts the most common class among the k-nearest neighbors for the new data point. Below are the key equations involved in the KNN algorithm:

- Distance calculation: A distance measure is used to locate the nearest neighbors where a popular option is the Euclidean distance. In a d-dimensional space, with two data points P and Q:

$$Euclidean\ distance: \sqrt{\sum_{i=1}^{d}(P_i - Q_i)^2} \tag{4}$$

- Classification: In a classification task, the new data point's class label is selected by majority vote among its k-nearest neighbors. Let C_i be the class label of the i-th neighbour:

$$Predicted\ class: \hat{y} = \arg max_j \sum_{i=1}^{K} I(C_i = j) \tag{5}$$

where I is the indicator function.

2.3.2. Artificial neural network

The structure and functionality of biological neural networks found in the human brain served as the model's inspiration for the artificial neural network. Fundamentally, ANN comprises interlinked nodes, typically arranged into layers. Each node, known as a neuron or perceptron, conducts elementary computations. Weights, symbolizing connection strength, are allocated to connections between nodes. Throughout training, the network fine-tunes these weights to enhance its efficiency for a particular objective as illustrated in Figure 2, the input layer, hidden layer, and output layer make up the minimum of three layers in the most common feed-forward neural network, the MLP. To make it function, information is sent from the input layer to the output layer's neurons. Below are several steps involved in forward propagation (6)-(8):

- Weighted input to a neuron (for a hidden layer or output layer):

$$z_j = \sum_{i=1}^{n} w_{ij} \cdot x_i + b_j \tag{6}$$

where z_j is the weighted input to neuron j, w_{ij} denotes the weight linking neuron i to j, x_i denotes the input from neuron i, and b_j denotes the bias for neuron j.

- Activation function (for a hidden layer or output layer):

$$a_j = f(z_j) \tag{7}$$

where $f(\cdot)$ is the activation function (typically, sigmoid, tanh, or ReLU) and a_j is the output (activation) of neuron j.

- MLP output:

$$\hat{y} = f\left(\sum_{j=1}^{m} w_{oj} \cdot a_j + b_o\right) \tag{8}$$

where a_j denotes the activations of the output layer neurons, b_o denotes the bias for the output layer, and $\hat{y}$ represents the expected output. The weights of w_{oj} connect the output layer neurons to the final output.

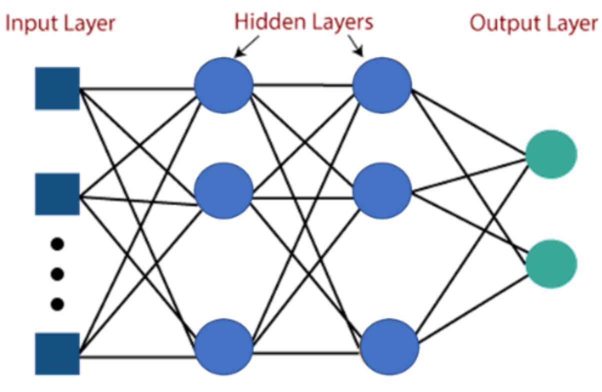

Figure 2. Structure of MLP

2.3.3. Logistic regression

Logistic regression is a statistical approach for binary classification tasks in which the target variable has only two possible outcomes which are either 0 or 1. The output of logistic regression is processed using the logistic (sigmoid) function to guarantee that it falls inside the range of 0 to 1. The logistic function is described as (9):

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{9}$$

where:
- σ(z) denotes the logistic function and z represents the linear combination of feature values and their corresponding weights (parameters).
- z is the linear combination of feature values and their corresponding weights (parameters).

The linear combination of z is calculated as:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{10}$$

where:
- β_0 indicates the intercept or bias term.
- β_1, β_2,…, β_n denote the coefficients associated with each feature.
- x_1, x_2, ..., x_n represent the values of each feature.

By integrating the logistic function with the linear combination, we obtain the probability of the output belonging to the positive class as shown in (11).

$$P(Y = 1 \mid X) = \sigma(z) = \frac{1}{1+e^{-z}} \tag{11}$$

2.4. Hyperparameter tuning

GridSearchCV is a methodical technique in machine learning experimentation, aiming to enhance model performance by exploring a predefined hyperparameter grid and selecting the set yielding optimal results. This involves constructing a hyperparameter grid that outlines the potential values or configurations for key parameters in the considered algorithms. This grid serves as the parameter space for GridSearchCV. Using the provided hyperparameter grid and three machine learning models, GridSearchCV systematically conducts cross-validated training, evaluating model performance based on relevant metrics like accuracy. It meticulously assesses various hyperparameter sets to pinpoint the one that achieves the best outcomes.

2.5. Evaluation metrics

Metrics from the confusion matrix will be used to assess the performance of each feature selection combination in conjunction with three algorithms. This includes examining important performance metrics such as accuracy, precision, recall, and F1-score, which provide a comprehensive picture of how well each model classifies students' performance. By comparing these metrics, we could determine the efficacy of each method and feature selection combination. The measurements used to evaluate the mode are listed in (12) to (15):

$$Accuracy = \frac{True\ positive + True\ negative}{Total\ samples} \tag{12}$$

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{13}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{14}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

3. RESULTS AND DISCUSSION

Data from the ECE431 subject code's online learning platform at Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia, was used in this study. The dataset included 102 early semester samples of students studying electrical and electronic engineering. To eliminate misinformation and noise, data cleaning was performed early in the data preparation process, and one feature was discarded since it had only zeros. Following that, feature selection was taken part in obtaining the most influenced features wherein a filter-based feature selection was used which is known as IG. A particular subset of the data will be generated based on the chosen features following the extraction of pertinent features. Following that, this subset was split into training and testing sets in a 70:30 ratio, respectively. For the training dataset, it will undergo resampling technique which in this case we used oversampling technique known as SMOTE to solve the class imbalance issue within the target variable. In the modeling part, we used three supervised machine learning algorithms which are KNN, MLP and LR, with the use of 3 folds cross validation to classify students' performance. Hyperparameter tuning was integrated into the modeling process for all three algorithms using GridSearchCV in order to obtain the most performing hyperparameters for each classifier.

3.1. Filter-based feature selection

Table 2 provides information about how each characteristic affects the target variable. The values, which range from 0 to 0.253611, show how much a characteristic depends on another feature to predict the target variable. A smaller number denotes less significance and less impact on the result. The importance values of features 1, 3, 7, 8, and 9, which show how students access the notes (features 1 and 3) and students do the exercises (features 7, 8, and 9) are zero, indicating that they have little effect on the target variable. Consequently, these features can be safely excluded from the analysis, being essentially independent of the outcome. This focused feature selection is vital for optimizing model performance and improving analysis efficiency.

Table 2. Table of features' importance for information gain

Features	Importance values
Feature12	0.253611
Feature11	0.229787
Feature2	0.225398
Feature10	0.078053
Feature6	0.023170
Feature4	0.001085
Feature1	0
Feature3	0
Feature7	0
Feature8	0
Feature9	0

3.2. Hyperparameters selection

In Table 3, the optimal hyperparameters for each classifier are presented, which were obtained by applying GridSearchCV wherein a set of hyperparameters was formulated for each model to ascertain the most effective combinations. We considered the time computation complexity and the feasibility of combining certain hyperparameters where such hyperparameters cannot be tested with all types of hyperparameters such as the other 'solver' of LR like *'sag'* and *'lbfgs'* cannot be matched or paired up with certain type of 'penalty' which is 'L1'. Thus, we only considered those hyperparameters that could be paired up including the following: for KNN (algorithm: $'ball\_tree', 'kd\_tree', 'brute'$ and n_neighbors: value of 1-10), for MLP (activation: $'identity', 'tanh', 'relu'$; learning_rate: $'constant', 'adaptive'$; solver: $'lbfgs', 'adam'$; hidden_layer_sizes: value of 10-100), and for LR (penalty: $'l1'$, $'l2'$; C: value of 1-10; solver: $'liblinear', 'saga'$).

Table 3. Hyperparameters used for the three algorithms

Models	List of hyperparameters	Values	Best selected values
KNN	algorithm	ball_tree, kd_tree, brute	ball_tree
	n_neighbors	1-10	1
MLP	activation	identity, tanh, relu	identity
	learning_rate	constant, adaptive	constant
	solver	lbfgs, adam	lbfgs
	hidden_layer_sizes	10-100	10
LR	C	1-10	1
	penalty	L1, L2	L1
	solver	liblinear, saga	liblinear

3.3. Model evaluation

The presented tables illustrate the classification performance of three machine learning algorithms—KNN, MLP, and LR where Table 4 represents the models' performance without using GridSearchCV while Table 5 explains the models' performance with the inclusion of GridSearchCV. The accuracy values for MLP and LR are 100% for both tables, which suggests that these models perfectly predict the students' performance categories, in this case, $'pass'$ or $'fail.'$ The precision values for all three algorithms are also at 100%, indicating that when they predict a student to pass, they are almost always correct. Theoretically, it aligns with the models' ability to precisely identify positive instances, which is particularly important when assessing student performance.

The performance evaluation of KNN, MLP, and LR reveals notable insights. For MLP and LR, the F1-scores are 100% for both tables, reflecting a perfect balance between precision and recall. This demonstrates their exceptional ability to provide a high degree of accuracy while ensuring no 'Pass' students are missed. In the case of KNN, the F1-scores are 80.0% in Table 4 and 88.9% in Table 5, showing a slight increase by using GridSearchCV. This emphasizes the positive impact of hyperparameter tuning on KNN's predictive accuracy. The systematic implementation of GridSearchCV significantly enhances overall model performance. Notable improvements in accuracy metrics across all algorithms in Table 5 underscore the importance of hyperparameter optimization in refining models for better predictive accuracy.

Table 4. Model performance without using GridSearchCV

	KNN	MLP	LR
Accuracy (%)	67.7	100.0	100.0
Precision (%)	100	100.0	100.0
Recall (%)	66.7	100.0	100.0
F1-score (%)	80.0	100.0	100.0

Table 5. Model performance using GridSearchCV

	KNN	MLP	LR
Accuracy (%)	80.6	100	100
Precision (%)	100	100	100
Recall (%)	80	100	100
F1-score (%)	88.9	100	100

4. CONCLUSION

In conclusion, the experimental investigation detailed in this study has provided valuable insights into the development of machine learning algorithms for classifying student performance based on their online learning activities. The comprehensive analysis of machine learning algorithms has yielded significant results, contributing to our understanding of student performance classification. The findings presented in this research underscore the importance of selecting significant features and highlighted the effectiveness of implementing hyperparameter tuning on the algorithms. As found in the experiment, two models which are MLP and LR obtained 100% F1-score, reflecting a perfect balance between precision and recall which demonstrates their exceptional ability to provide a high degree of accuracy while ensuring no 'Pass' students are missed while KNN recorded a slight increase in F1-score from 80% to 88.9% by employing hyperparameter tuning. The implications of these findings extend beyond the confines of this study and offer valuable insights for educational purposes. Overall, this research contributes to the ongoing discourse in student performance classification and underscores the importance of continued investigation for more understanding.

ACKNOWLEDGEMENT

The authors would like to express gratitude and acknowledge the Universiti Teknologi MARA (UiTM) for funding under the Journal Support Fund (JSF).

REFERENCES

[1] Y. Widyaningsih, N. Fitriani, and D. Sarwinda, "A semi-supervised learning approach for predicting student's performance: First-year," in *2019 12th International Conference on Information and Communication Technology and System (ICTS)*, Jul. 2019, pp. 291–295, doi: 10.1109/ICTS.2019.8850950.

[2] P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Materials Today: Proceedings*, vol. 47, pp. 5260–5267, 2021, doi: 10.1016/j.matpr.2021.05.646.

[3] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Computers and Electrical Engineering*, vol. 89, Jan. 2021, doi: 10.1016/j.compeleceng.2020.106903.

[4] M. Riestra-González, M. del P. Paule-Ruíz, and F. Ortin, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Computers and Education*, vol. 163, Apr. 2021, doi: 10.1016/j.compedu.2020.104108.

[5] V. Christou *et al.*, "Performance and early drop prediction for higher education students using machine learning," *Expert Systems with Applications*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120079.

[6] M. Nachouki, E. A. Mohamed, R. Mehdi, and M. Abou Naaj, "Student course grade prediction using the random forest algorithm: Analysis of predictors' importance," *Trends in Neuroscience and Education*, vol. 33, 2023, doi: 10.1016/j.tine.2023.100214.

[7] L. Vives *et al.*, "Prediction of students' academic performance in the programming fundamentals course using long short-term memory neural networks," *IEEE Access*, vol. 4, pp. 1–17, 2024, doi: 10.1109/ACCESS.2024.3350169.

[8] W. Qiu, A. W. H. Khong, S. Supraja, and W. Tang, "A dual-mode grade prediction architecture for identifying at-risk students," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 803–814, 2024, doi: 10.1109/TLT.2023.3333029.

[9] M. Adnan *et al.*, "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *IEEE Access*, vol. 9, pp. 7519–7539, 2021, doi: 10.1109/ACCESS.2021.3049446.

[10] R. Z. Pek, S. T. Ozyer, T. Elhage, T. Ozyer, and R. Alhajj, "The role of machine learning in identifying students at-risk and minimizing failure," *IEEE Access*, vol. 11, pp. 1224–1243, 2023, doi: 10.1109/ACCESS.2022.3232984.

[11] M. Wafi, U. Faruq, and A. A. Supianto, "Automatic feature selection for modified k-nearest neighbor to predict student's academic performance," *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*, pp. 44–48, Sep. 2019, doi: 10.1109/SIET48054.2019.8986074.

[12] V. Mhetre and M. Nagar, "Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA," in *Proceedings of the International Conference on Computing Methodologies and Communication*, 2017, pp. 475–479, doi: 10.1109/ICCMC.2017.8282735.

[13] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," *Computers and Education: Artificial Intelligence*, vol. 2, 2021, doi: 10.1016/j.caeai.2021.100018.

[14] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent M-learning environments," *Computers and Education: Artificial Intelligence*, vol. 2, 2021, doi: 10.1016/j.caeai.2021.100035.

[15] A. Polyzou and G. Karypis, "Feature extraction for next-term prediction of poor student performance," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 237–248, 2019, doi: 10.1109/TLT.2019.2913358.

[16] M. R. Ahmed, S. T. I. Tahid, N. A. Mitu, P. Kundu, and S. Yeasmin, "A comprehensive analysis on undergraduate student academic performance using feature selection techniques on classification algorithms," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225341.

[17] W. Punlumjeak and N. Rachburee, "A comparative study of feature selection techniques for classify student performance," in *2015 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering*, 2015, pp. 425–429, doi: 10.1109/ICITEED.2015.7408984.

[18] S. S. M. Ajibade, N. B. Ahmad, and S. M. Shamsuddin, "An heuristic feature selection algorithm to evaluate academic performance of students," *2019 IEEE 10th Control and System Graduate Research Colloquium*, pp. 110–114, 2019, doi: 10.1109/ICSGRC.2019.8837067.

[19] H. E. Abdelkader, A. G. Gad, A. A. Abohany, and S. E. Sorour, "An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the COVID-19," *IEEE Access*, vol. 10, pp. 6286–6303, 2022, doi: 10.1109/ACCESS.2022.3143035.

[20] M. Sivasakthi, "Classification and prediction based data mining algorithms to predict students' introductory programming performance," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Nov. 2017, pp. 346–350, doi: 10.1109/ICICI.2017.8365371.

[21] U. Imdad, W. Ahmad, M. Asif, and A. Ishtiaq, "Classification of students results using KNN and ANN," in *13th International Conference on Emerging Technologies*, 2017, pp. 1–6, doi: 10.1109/ICET.2017.8281651.

[22] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Preprocessing and analyzing educational data set using X-API for improving student's performance," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov. 2015, pp. 1–5, doi: 10.1109/AEECT.2015.7360581.

[23] Dafid and Ermatita, "Filter-based feature selection method for predicting students' academic performance," *2022 International Conference on Data Science and Its Applications, ICoDSA 2022*, pp. 309–314, 2022, doi: 10.1109/ICoDSA55874.2022.9862883.

[24] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," in *2015 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering*, 2015, pp. 420–424, doi: 10.1109/ICITEED.2015.7408983.

[25] L. Rahman, N. A. Setiawan, and A. E. Permanasari, "Feature selection methods in improving accuracy of classifying students' academic performance," in *2nd International Conferences on Information Technology, Information Systems and Electrical Engineering*, 2017, no. 1, pp. 267–271, doi: 10.1109/ICITISEE.2017.8285509.

[26] E. Al Nagi and N. Al-Madi, "Predicting students performance in online courses using classification techniques," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, Oct. 2020, pp. 51–58, doi: 10.1109/IDSTA50958.2020.9264113.

[27] Y. S. Alsalman, N. Khamees Abu Halemah, E. S. AlNagi, and W. Salameh, "Using decision tree and artificial neural network to predict students academic performance," in *2019 10th International Conference on Information and Communication Systems (ICICS)*, Jun. 2019, pp. 104–109, doi: 10.1109/IACS.2019.8809106.

[28] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers and Education*, vol. 143, Jan. 2020, doi: 10.1016/j.compedu.2019.103676.

[29] S. Alwarthan, N. Aslam, and I. U. Khan, "An explainable model for identifying at-risk students at higher education," *IEEE Access*, vol. 10, no. September, pp. 107649–107668, 2022, doi: 10.1109/ACCESS.2022.3211070.

[30] F. Abbasi, M. Naderan, and S. E. Alavi, "Anomaly detection in internet of things using feature selection and classification based on logistic regression and artificial neural network on N-BaIoT dataset," in *2021 5th International Conference on Internet of Things and Applications (IoT)*, May 2021, pp. 1–7, doi: 10.1109/IoT52625.2021.9469605.

[31] A. M. Olalekan, O. S. Egwuche, and S. O. Olatunji, "Performance evaluation of machine learning techniques for prediction of graduating students in tertiary institution," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Mar. 2020, pp. 1–7, doi: 10.1109/ICMCECS47690.2020.240888.

[32] S. Alraddadi, S. Alseady, and S. Almotiri, "Prediction of students academic performance utilizing hybrid teaching-learning based feature selection and machine learning models," 2021, doi: 10.1109/WIDSTAIF52235.2021.9430248.

[33] G. S. Ramaswami, T. Susnjak, A. Mathrani, and R. Umer, "Predicting students final academic performance using feature selection approaches," in *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Dec. 2020, pp. 1–5, doi: 10.1109/CSDE50874.2020.9411605.

[34] M. Garg and A. Goel, "Preserving integrity in online assessment using feature engineering and machine learning," *Expert Systems with Applications*, vol. 225, 2023, doi: 10.1016/j.eswa.2023.120111.

[35] M. A. A. Aziz, "Online Learning Center for ECE431," *azriaziz.uitm.edu.my*. Accessed: Mar. 01, 2024. [Online] Available: https://azriaziz.uitm.edu.my/onlinelearning/download/onlineactivities.csv

[36] K. Sixhaxa, A. Jadhav, and R. Ajoodha, "Predicting students performance in exams using machine learning techniques," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2022, pp. 635–640, doi: 10.1109/Confluence52989.2022.9734218.

BIOGRAPHIES OF AUTHORS

Muhammad Aqif Hadi Alias received the B.Eng. degree in electronic engineering from Universiti Teknologi MARA (UiTM), Malaysia, in 2022. Currently, he is a master's degree student at the School of Electrical Engineering, Universiti Teknologi MARA. His research interest includes the area of classification which specifically focuses on students' performance by obtaining their online learning activities. He can be contacted at email: aqifhadi7@gmail.com.

Mohd Azri Abdul Aziz is a senior lecturer at the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Malaysia. He received the B.Eng. (Hons) degree in electronic engineering from University of Manchester Institute of Science and Technology (UMIST), UK. M.Sc. degree in electrical engineering and Ph.D. degree in electrical engineering, both from Universiti Teknologi MARA, Malaysia. He is also a fellow at Innovative Electromobility Research Lab (ITEM) UiTM, primarily researching on autonomous vehicles. His research area includes pattern classification, object detection, and internet of things-based application. Currently, he is supervising and co-supervising 5 master's and 6 Ph.D. students. He has authored and co-authored 8 journals and 16 conference proceedings with 5 H-index and 73 citations. He can be contacted at email: azriaziz@uitm.edu.my.

Najidah Hambali completed her PhD in advanced process control at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM). She received bachelor of engineering (B.Eng.Hons.) in electronics engineering from Universiti Sains Malaysia (USM) in 2004. In 2005, she joined Universiti Malaysia Pahang (UMP) as a tutor. After received her master's engineering science (M.Eng.Sc.) in systems and control from the University of New South Wales (UNSW), Sydney, Australia in 2006, she continued working in UMP as a lecturer until 2010. In January 2011, she joined UiTM and later in October 2015, she started her PhD study in UiTM. Currently, she is a senior lecturer in the Centre of System Engineering Studies (CSES), School of Electrical Engineering, College of Engineering UiTM, Shah Alam. Her current research interests are control system and nonlinear modelling for process control. She can be contacted at email: najidah@uitm.edu.my.

Mohd Nasir Taib is a professor of control and instrumentation, Universiti Teknologi MARA, Selangor, Malaysia. His area of expertise are control system and signal processing. He is a former director of Malaysia Institute of Transport, Malaysia. He is also head of the Advanced Signal Processing Research Group (ASPRG), UiTM Shah Alam. He has published more than 500 research and academic journals around the world and won many gold awards in innovations that he participated in. He can be contacted at: dr.nasir@uitm.edu.my.