Journal of Information Systems & Information Technology (JISIT)
ISSN: 2478-0677, Vol. 8 No.2, 2023, [Pages 45 - 72]

Predicting Progress: An In-Depth Study of Students' Academic Trajectories through Data Mining

Abdul Jabbar Mohamed Hasmy[1]
Mohideen Bawa Mohamed Irshad[2]
Department of Management and Information Technology[1 2]
South Eastern University of Sri Lanka[1 2]
hasmie@seu.ac.lk[1]
mbmirshad@seu.ac.lk[2]

Abstract

Accurately predicting students' academic trajectories is crucial for effective educational interventions. This research introduces a comprehensive predictive model for understanding and forecasting students' progress, leveraging advanced data mining techniques. Specifically, three powerful classifiers—Random Forest, Support Vector Machines (SVM), and Artificial Neural Network (ANN) are employed to explore the intricate dynamics of students' academic journeys.

The study places significant emphasis on uncovering patterns related to student absence days, lecturer involvement, and punctuality within the e-learning management system. This study examines the influence of Random Forest, SVM, and ANN on students' educational achievement by utilizing Random Forest, SVM, and ANN. The model proposed here incorporating these classifiers, demonstrates a substantial improvement of up to 12% to 18% in accuracy compared to models lacking these influential features.

This research contributes to the field by showcasing the effectiveness of Random Forest, SVM, and ANN in predicting academic trajectories, thereby facilitating targeted interventions and personalized strategies for student success. The findings underscore the importance of leveraging diverse classifiers to comprehensively capture the multifaceted aspects of students' academic progress.

***Keywords**: Educational Data Mining, Predictive Model, Random Forest, Support Vector Machines, Artificial Neural Network, Academic Trajectories.*

1. Introduction

In the dynamic landscape of data mining and Knowledge Discovery in Databases (KDD), the evolving field of Education Data Mining (EDM) has become pivotal in unraveling valuable insights from educational information systems. With an emphasis on platforms like Moodle, Blackboard, and other educational tools, EDM explores diverse sources such as course management systems, online learning platforms, and student registration systems. These systems, including the widely

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

used Moodle Learning Management System (LMS), play a central role in shaping students' educational trajectories from primary to higher education. [1] outlines Education Data Mining (EDM) as the analysis of data from educational institutions using Data Mining (DM) systems to address research challenges in education. The primary focus of EDM is to gain a deeper understanding of students and their learning environments. It involves the collection, storage, and interpretation of data related to students' studies and evaluations. Various methods, including Naïve Bayes, Decision Trees, Nearest Neighbor, Neural Networks, K-Regression, Correlation, etc., are employed in the EDM process. Within the realm of EDM, the objective goes beyond mere data analysis; it extends to the discovery of patterns that empower students to effectively manage their education while providing educational institutions with actionable insights. [2], in their notable analysis, leveraged several Machine Learning techniques to classify students, fostering enhanced academic performance.

The work of [3] reinforced the versatility of data mining techniques, demonstrating their utility in generating specific patterns, classifications, and predictions. In keeping with this trajectory, this research offers a student performance model, emphasizing on crucial variables such as lecturers' engagement and student absence for lectures. The dataset, rigorously collected from the Moodle LMS, the dataset comprises 450 records and containing 10 distinct attributes. Then, three data mining techniques were used, and the outcomes are assessed by employing several metrics.

This study adds depth by applying three prominent data mining algorithms—Random Forest, Support Vector Machines (SVM), and Artificial Neural Network (ANN). Through this exploration, we aim to contribute to the ongoing discourse on effective strategies for predicting and enhancing students' academic performance, aligning with the latest advancements in the field.

2. Literature Review

An institution of higher education aims to provide its students with a quality education and to improve their decision-making abilities. Analyzing academic data can provide insight into what factors affect learners' academic performance. The information gathered during the extraction process is valuable in assisting with decision making at the administrative level, as it provides a useful reference for decision makers. A number of advantages are associated with it, including the ability to improve the academic output of students, the ability to decrease negative performance, the capacity to effectively grasp student behavior, and the ability to enhance the educational process overall [4].

[5] developed a predictive model based on students' programming submissions, utilizing data-driven features to predict final exam grades. The explainable stacked ensemble model, outperforming various baseline models, incorporated the SHAP algorithm for transparent predictions. The analysis of SHAP results provided insights into students' problem-solving behavior and allowed profiling.

In the study conducted by [6], the focus was on exploring and analyzing fundamental student data within a 4-year degree program. The research aimed to address three key questions, primarily centering around the development of a classification model for early identification of student end-of-degree performance using readily available learning data.

[7] have developed a conceptual framework for attribute selection and forecasting student performance using ML models. ML is employed in the automated evaluation of students learning employing answers, simulations, educational assessment, etc. [8] have applied the K-Means clustering method and highlighted the possibilities of the clustering-aided strategy for forecasting student outcomes in higher education.

Artificial neural networks and other machine learning techniques are useful for classifying a range of educational outcomes, including student grade point averages, retention rates, and degree completion rates, among other factors [9].
In alignment with our research topic, we will concentrate on the most crucial category of features affecting students' academic performance, applying Random Forest, Support Vector Machines (SVM), and Artificial Neural Network (ANN) as our primary classifiers.

3. Data Pre-processing

The data for constructing the performance prediction model for students, aimed at anticipating academic outcomes and this dataset obtained from Bachelor of Science degree students at a state university, utilizing the Moodle Learning Management System (LMS), encompassing 450 student records with 10 distinct features.

3.1. Moodle Learning Management System (LMS)

The Moodle LMS serves as an innovative e-learning platform designed to engage learners, monitor progress, and deliver targeted outcomes. In the research conducted by Abhirami and [10], the key usability metrics highlighted for the Moodle Learning Management System (LMS) were efficiency, learnability, and satisfaction. Similarly, a study carried out by [11] also emphasized efficiency, effectiveness, and satisfaction as paramount usability metrics for the Moodle LMS.

Prior research endeavors have proposed diverse strategies aimed at enhancing the usability of the Moodle LMS. For instance, [11] recommended the implementation of an adaptive design, offering customization based on user needs and preferences. Additionally, findings from [12] suggested the provision of training and support for teachers to bolster their proficiency in effectively utilizing the Moodle LMS. Contrary to traditional methods such as books, PDFs, or PowerPoint presentations, Moodle LMS enables students to engage in fully interactive learning activities.

Following the collection of the dataset from 480 bachelor of science degree students at a state university through the Moodle LMS, the critical next step is data preprocessing. Real-world data often lacks completeness, containing inadequate attributes, missing values, and summarized data. To address these issues and enhance the quality of the dataset, preprocessing techniques are applied, encompassing data cleaning, transformation, and the selection and analysis of relevant features. This process is crucial for eliminating noise and handling outlier data, ensuring the dataset is well-prepared for subsequent analysis.

Table 1. The categories and features of the student dataset

Attribute	Description of Attribute	Category of Attribute
Computer Literacy Level	Student's ability to work with computer systems	**Academic Background**
Year of study	First Year, second year, third year and fourth year	
Semester	Semester I and Semester II.	
Course	BSc in IT, BSc in Physical Science, and BSc in Bio science	
Student Attendance for Lectures	No. of student available days in lecture hall	
Lecturer involvement	Lecturer interaction with students and lecturer ability to answer student queries.	**Participation of lectures on the whole teaching process**
Satisfaction of Lecturer	Lecturer's satisfaction on student's progress (Positive or Negative)	
Group Discussions	All of these features are designed to enhance the interaction between students and Moodle LMS.	**Behavioral Feature**
Resources visited by students		
Assignments submission by student		

COPYRIGHT © 2023 FACULTY OF MANAGEMENT AND COMMERCE, SOUTH EASTERN UNIVERSITY OF SRI LANKA (SEUSL), UNIVERSITY PARK, OLUVIL32360. SRI LANKA [RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

3.2 Data Pre-Processing and Data Cleaning

Data preprocessing involves employing techniques to convert unstructured data into a conventional format, facilitating acceptance and utilization by data mining algorithms.

Data cleaning is a crucial preprocessing procedure that addresses partial values and removes noisy and inconsistent data. In this study, the initial dataset comprised 450 records, of which 20 records contained missing values across different categories. Following the data cleaning process, the final dataset was refined to 430 records.

3.3 Data Transformation

In order to properly express class labels, data transformation is essential in transforming numerical quantities into nominal values for classification. Based on students' grades, Table 2 shows how the dataset is distributed into three different levels: highest level, medium level, and lowest level.

Table 2. Grouping according to numerical values.

Classes	
Value Interval	Class Label
0–59	Low Level
60–89	Medium Level
90–100	High Level

3.4 Feature Selection and Analysis

Research undertaken by [13] underscored the importance of feature selection as a crucial aspect of data preprocessing. This stage consists of selecting a subset of features from the dataset that are relevant and indispensable, which reduces the number of attributes included in the algorithm. This reduction is intended to improve the performance of the learning algorithm by eliminating redundant and irrelevant data, consequently enhancing data quality. Feature selection methods can be broadly categorized into two main types: (1) Wrapper-Based methods and (2) Filter-Based methods. [14] utilized both these methods, while [15] explored various feature ranking techniques.

To construct the student performance model, different feature scores are evaluated in order to identify the most significant features. An illustration of the top ranked features resulting from the filter-based evaluation process is provided in Figure 1 and student absence days ranked top, followed by lecturer involvement. It appears that a subset of important features has been selected while others have been left out. Consequently, the features considered in this research obtained the highest rank, underscoring the significant impact of student attendance and lecturer participation throughout the educational journey on academic performance.

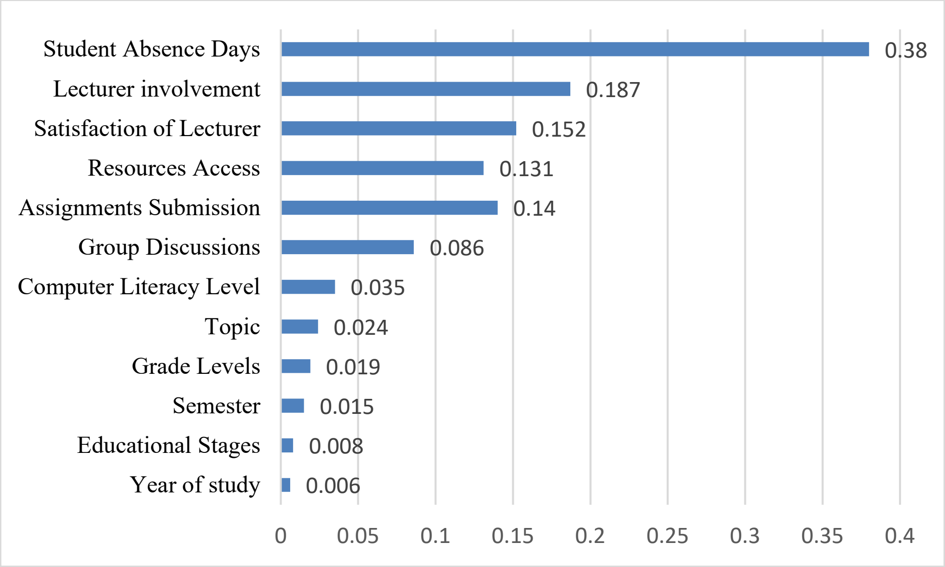


Figure 1. After using filter-based evaluation with gain ratio, highly ranked features were obtained

4. Methodology

In this paper, we introduce a framework for assessing students' performance utilizing three distinct classifiers, aimed at evaluating the subset of features influencing academic achievement. Figure 2 illustrates the key steps within this framework. The process initiates by collecting data from the Moodle Learning Management System, as detailed under the section dataset and data preprocessing.

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

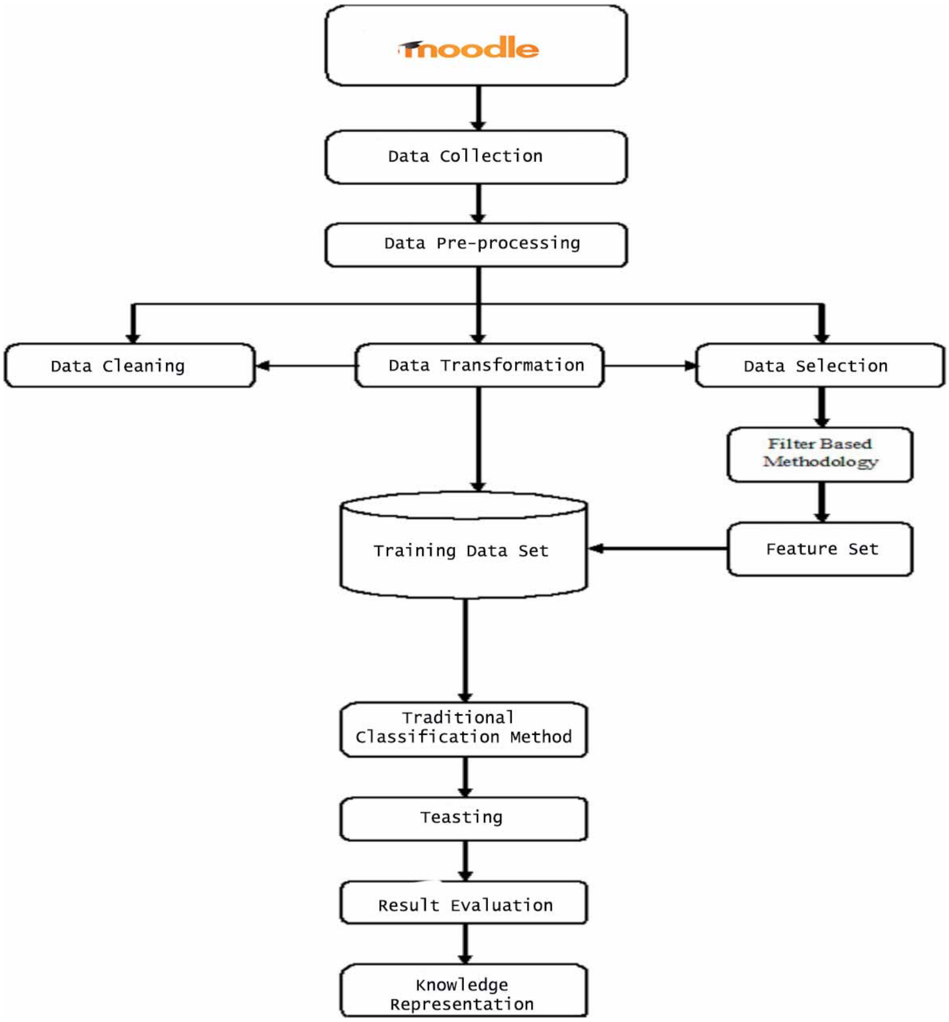


Figure 2. Stages of students' performance prediction model

The next stage then entails preparing the data by transforming the collected data into a more accessible manner. In the first step, duplicate and unnecessary data is removed from the dataset using the data cleaning method. In order to transform numerical values to nominal values, the dataset is divided as three class labels (medium level, low level, and high level) according to the cumulative grades of the students. As of right now, the dataset consists of 43 high-level students, 201 middle-level students, and 186 low-level kids.

In order to find the ideal collection of characteristics with the greatest scores, the method next moves on to feature selection and analysis. We used gain ratio-based

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

selection methods, which are a filter-based method of looking at different feature scores, as shown in Figure 1. Lastly, we provide a system that makes use of three classifiers: artificial neural networks (ANN), support vector machines (SVM), and random forests (RF). These categorization methods are used to identify characteristics that might affect students' academic performance.

4.1 Random Forest Classifier

Random Forest classifier employs an ensemble learning strategy, aggregating predictions from numerous decision trees [16], [17]. This approach involves evaluating probabilities for various attributes based on the training dataset for each class and leveraging these probabilities to classify new instances. In contrast to Naïve Bayes, the Random Forest classifier introduces a more intricate decision-making process, considering the collective input from a multitude of decision trees.

For instance, when assigning class probabilities, the random forest classifier may distribute probabilities differently. In a hypothetical scenario, it might allocate a probability of 0.25 for the middle level, 0.30 for the low level, and 0.45 for the high level. This variance in probability assignment reflects the inherent adaptability and robustness of the Random Forest algorithm, utilizing multiple decision trees to enhance predictive accuracy.

4.2 Support Vector Machine Classifier

This classification approach generates a hyperplane to identify things according to their classifications. The larger the distance between the hyperplane and the closest object, the smaller the generalization error of the SVM algorithm. SVM has been applied in various investigations, including those by [18]), [19], and [20]. In the present investigation, SVM is selected because to its adaptability for tiny datasets, and it is known to be quicker than other approaches, as highlighted by [21].

4.3 Artificial Neural Network (ANN)

ANN is a prominent approach in EDM and is supposed to replicate the organization of the human brain to solve complicated issues. It consists of a series of units that receive a weighted set of inputs and replies with an output.

Several study publications, such as those by [22], [23] and [24], have applied Artificial Neural Networks (ANN) to predict students' performance. In our investigation, we also decided for ANN owing to its capabilities to discover probable correlations among variables and its adeptness at learning from a restricted amount of instances. Furthermore, an earlier analysis found that ANN models outperformed classification strategies in properly classifying admitted candidates as accepted or not, as reported by [21].

The Artificial Neural Network (ANN) framework serves to derive patterns and address intricate prediction challenges. Structurally, it consists of an input layer, an output layer, and a hidden layer. The input layer receives input from the user, while the output layer sends the results back to the user. The intermediary layer, positioned between the input and output layers, connects neurons without direct interaction with the primary user application. This middle layer's neurons are intricately linked, contributing to the assessment of patterns and outcomes for knowledge representation.

5. Experiments and Results Evaluation

5.1. Environment Setup

The experimentation was conducted on a personal computer equipped with 8GB of RAM and an Intel Core processor (2.50 GHz). The Weka tool, known for its efficacy in classification algorithms [25], was chosen for its ability to analyze accuracy and prediction results. In this study, Weka was utilized to evaluate the proposed models, perform comparisons, and analyze outcomes. Various options, such as cross validation, supplied test set, percentage split and training set, were explored to test the effectiveness of the models.

The dataset was divided into a training set and a test set using a 10-fold cross-validation method. The dataset was randomly partitioned into 10 subsets. For each iteration, Weka tool employed one subset for testing and the remaining nine subsets for training. This process was repeated ten times, each time exchanging the testing subset with the next one. The final average success rate was then calculated.

5.2. Evaluation Measures

Four unique measures were used to assess the effectiveness of several categorization approaches, namely precision, accuracy, recall, and F-measure. Table 3 displays these derived metrics, showing a confusion matrix comprised of equations 1, 2, 3, and 4.

Table 3. Two classes of confusion matrix

		Predicted	
		Yes	No
Actual	Yes	TPV	FNV
	No	FPV	TNV

Positive responses are denoted by "Yes" while negative responses are denoted by "No". In this context, "TPV" represents true positive values, while "FPV" signifies false positive values. Correspondingly, "FNV" indicates false negative values, and "TNV" stands for true negative values. Accuracy is computed as the ratio of correct classifications to the total number of classifications. In statistics, recall refers to the proportion of instances that were correctly classified compared to the total number of unclassified and correctly classified cases. It is the proportion of instances that have been correctly classified relative to the total number of misclassified and correctly classified instances. Additionally, we incorporate the F-measure, a comprehensive metric combining precision and recall, serving as a robust indicator of their interplay.

Accuracy = True Positive + True Negative……….<1>

True Positive + False Negative + False Positive + True Negative

Recall = TruePositive……….<2>

True Positive + False Negative

Precision = True Postive………..<3>

True Positive +False Positive

F–measure = 2 Precision * Recall………. <4>

Precision + Recall

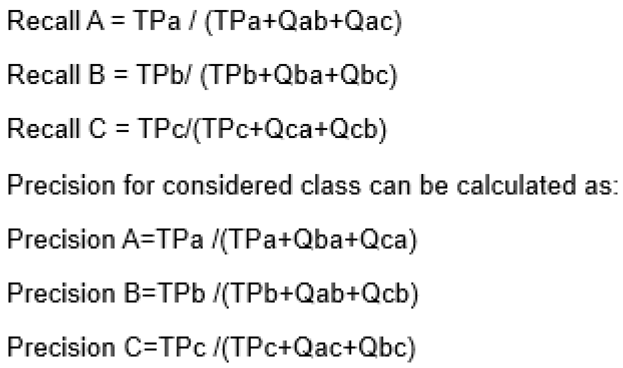
There are three classes namely Class A, Class B and Class C. Table 4 illustrates the categorization of confusion matrix based on the classes.

Table 4. More than two classes confusion matrix

	Predicted		
	A	B	C

Actual	A	TPa	Qab	Qac
	B	Qba	TPb	Qbc
	C	Qca	Qcb	TPc

For each row in a class, the cumulative false negative values represent the total value of all values, disregarding the actual positive value. The false positive values for every class, in contrast, are the sum of all values in the relevant column, which omits the real positive values. When all true negative values are added up for a particular class, it implies the sum of all true negative values across all rows and columns, thus eliminating both the column and rows associated with that class.



Recall A = TPa / (TPa+Qab+Qac)

Recall B = TPb/ (TPb+Qba+Qbc)

Recall C = TPc/(TPc+Qca+Qcb)

Precision for considered class can be calculated as:

Precision A=TPa /(TPa+Qba+Qca)

Precision B=TPb /(TPb+Qab+Qcb)

Precision C=TPc /(TPc+Qac+Qbc)

5.3. Results

Various outcomes are examined using three unique classification approaches applied on the student dataset to predict academic success. Confusion matrices for the Random Forest (RF), Support Vector Machines (SVM), and Artificial Neural Network (ANN) classifiers are shown in Tables 5, 6, and 7. These matrices constitute the basis for calculating metrics relating to Classes A, B, and C, as well as accuracy for the overall method.

Table 5. Confusion matrix for random forest classifier

		Predicted		
		A	**B**	**C**
Actual	**A**	133	16	42
	B	22	90	0

	C	24	2	101

$$\text{Accuracy}=\frac{133+90+101}{133+16+42+22+90+0+24+2+101}\text{x}100\%$$

$$=\frac{324}{430}\text{x}100\%$$

$$=\ 75.3\%$$

Recall for Class A, B and C:

Recall A=133/133+16+42=69.6, Recall B=90/90+22+0=80.3,

Recall C=101/101+24+2=79.5

Precision for Class A, B and C:

Precision A=133/133+22+24=74.3, Precision B=90/90+16+2=83.3

Precision C=101/101+42+0=70.6

F–measure for Class A, B and C:

$$\text{F–measure A} = \frac{2\ \text{Precision x Recall}}{\text{Precision+Recall}}$$

$$=\frac{2(74.3\ \text{x}\ 69.6)}{74.3+69.6}$$

$$=\ 71.8$$

F–measure B $= 81.8$

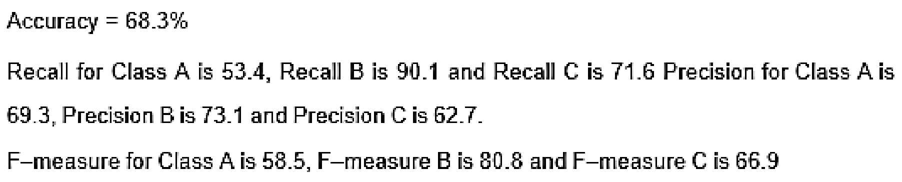
F–measure C $= 74.8$

In accordance with the aforementioned methodology, we extract the relevant data from the relevant tables in order to calculate the accuracy, precision, recall and F-measures for Support Vector Machines and Artificial Neural Networks.

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

Table 6. Confusion matrix for support vector machines classifier.

		Predicted		
		A	**B**	**C**
Actual	**A**	102	35	54
	B	11	101	0
	C	34	2	91



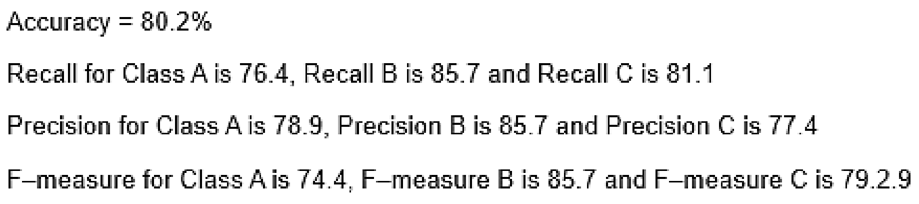
Accuracy = 68.3%

Recall for Class A is 53.4, Recall B is 90.1 and Recall C is 71.6 Precision for Class A is 69.3, Precision B is 73.1 and Precision C is 62.7.

F–measure for Class A is 58.5, F–measure B is 80.8 and F–measure C is 66.9

Table 7. Confusion matrix for artificial neural network.

		Predicted		
		A	**B**	**C**
Actual	**A**	146	16	29
	B	15	96	1
	C	24	0	103



Accuracy = 80.2%

Recall for Class A is 76.4, Recall B is 85.7 and Recall C is 81.1

Precision for Class A is 78.9, Precision B is 85.7 and Precision C is 77.4

F–measure for Class A is 74.4, F–measure B is 85.7 and F–measure C is 79.2.9

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

Table 8 displays outcomes employing 03 distinct data mining algorithms (RF, SVM, and ANN). Each algorithm produces two sets of classification results: one incorporating highly ranked features (RF), specifically for the days of absence of students and lecturer's involvement, and the other without these features (WRF). The previously mentioned details delineate the results achieved with highly ranked features. Comparable results without these features can be obtained using a similar approach. The tabulated data in Table 8 showcases superior classification outcomes with highly ranked features compared to results lacking these features. It is evident from this that student attendance and lecturers' involvement play a significant role in determining a student's academic success and success.

Table 8. Results of ranked features (RF & WRF)

Evaluation criteria	**RF**		**SVM**		**ANN**	
	RF	WRF	RF	WRF	RF	WRF
Accuracy	75.3	65.2	68.3	59.1	80.2	62.3

In reviewing Table 8, it becomes clear that ANN is outperforming other algorithms for classification. The ANN achieves an accuracy of 80.2% for RF and 62.3% for WRF. To elaborate, an accuracy of 80.2% indicates that 345 out of 430 students are accurately classified into the correct class labels—High, Medium, and Low—while 85 students were misclassified.

6. Conclusion

It is widely acknowledged that the students' academic performance is a cornerstone for the future success, and this has garnered significant attention from academic institutions around the world. E-learning management systems are becoming increasingly prevalent in the contemporary educational landscape. The substantial amount of data generated by these systems is causing many developed countries to transition to fully or partially automated systems. This latent knowledge and pattern can be used to derive meaningful insights, which will assist students in improving their academic performance.

A new student performance model is presented in this study, which incorporates novel feature categories linked to students' class attendance and lecturers' active involvement in the learning activities. A three-step classification algorithm is employed to assess the overall efficacy of the academic prediction framework: random forest, support vector machines, and artificial neural network. The findings underscore the substantial impact of these features on a student's academic success. The model demonstrates commendable accuracy when incorporating these feature categories, achieving a notable 12% to 18% improvement compared to results obtained by excluding such features.

Reference

[1] C. Lokpo, "Mining educational data to analyze students' performance: A case study of Mawuli School, Ho," 2020.

[2] S. Ghashout, Y. Gdura, and N. M. Drawil, "Early prediction of students' academic performance using artificial neural network: A case study in computer engineering department," in 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), pp. 40–45.

[3] G. Lampropoulos, "Educational data mining and learning analytics in the 21st century," in Encyclopedia of Data Science and Machine Learning, IGI Global, pp. 1642–1651, 2023.

[4] A. Chadha and V. Kumar, "An empirical study of the applications of data mining techniques in higher education," International Journal of Advanced Computer Science and Applications, vol. 2, no. 3, pp. 80–84, 2011.

[5] M. Hoq, P. Brusilovsky, and B. Akram, "Analysis of an explainable student performance prediction model in an introductory programming course," International Educational Data Mining Society, 2023.

[6] A. F. Meghji, N. A. Mahoto, Y. Asiri, H. Alshahrani, A. Sulaiman, and A. Shaikh, "Early detection of student degree-level academic performance using educational data mining," PeerJ Computer Science, vol. 9, p. e1294, 2023.

[7] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, "A conceptual framework to aid attribute selection in machine learning student performance prediction models," International Journal of Interactive Mobile Technologies, vol. 15, no. 15, 2021.

[8] O. Iatrellis, I. K. Savvas, P. Fitsilis, and V. C. Gerogiannis, "A two-phase machine learning approach for predicting student outcomes," Education and Information Technologies, vol. 26, no. 1, pp. 69–88, 2021.

[9] M. F. Musso, C. F. Rodríguez Hernández, and E. C. Cascallar, "Predicting key educational outcomes in academic trajectories: A machine-learning approach," 2020.

[10] K. Abhirami and M. K. Devi, "Student behavior modeling for an e-learning system offering personalized learning experiences," Computer Systems Science & Engineering, vol. 40, no. 3, 2022.

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]

[11] W. G. Alghabban and R. Hendley, "Perceived level of usability as an evaluation metric in adaptive e-learning: A case study with dyslexic children," SN Computer Science, vol. 3, no. 3, p. 238, 2022.

[12] L. Amhag, L. Hellström, and M. Stigmar, "Teacher educators' use of digital tools and needs for digital competence in higher education," Journal of Digital Learning in Teacher Education, vol. 35, no. 4, pp. 203–220, 2019.

[13] M. A. Umar and C. Zhanfang, "Effects of feature selection and normalization on network intrusion detection," Authorea Preprints, 2023.

[14] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and A. S. di Freca, "Comparing filter and wrapper approaches for feature selection in handwritten character recognition," Pattern Recognition Letters, vol. 168, pp. 39–46, 2023.

[15] R. Çekik and M. K. Kaya, "A new feature selection metric based on rough sets and information gain in text classification," Gazi University Journal of Science Part A: Engineering and Innovation, vol. 10, no. 4, pp. 472–486, 2023.

[16] S. Das, M. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," Expert Systems with Applications, vol. 213, p. 118914, 2023.

[17] A. H. S. Al Mamari, R. S. H. H. Al Ghafri, N. Aravind, R. Dhandapani, E. M. A. M. Al Hatali, and R. Pandian, "Experimental study and development of machine learning model using random forest classifier on shear strength prediction of RC beam with externally bonded GFRP composites," Asian Journal of Civil Engineering, vol. 24, no. 1, pp. 267–286, 2023.

[18] H. Li, "Support vector machine," in Machine Learning Methods, Singapore: Springer Nature Singapore, pp. 127–177, 2023.

[19] K. S. Gill, V. Anand, and R. Gupta, "Website classification through exploratory data analysis using naive Bayes, random forest, and support vector machine classifier," in 2023 3rd International Conference on Intelligent Technologies (CONIT), IEEE, pp. 1–5, June 2023.

[20] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, "Early segmentation of students according to their academic performance: A predictive modelling approach," Decision Support Systems, vol. 115, pp. 36–51, Nov. 2018.

[21] R. B. Crist, L. Manuel-Ignacio, M. J. López, et al., "Predicting students' final performance from participation in online discussion forums," Computers & Education, vol. 68, pp. 458–472, 2013.

[22] C. Huang, J. Zhou, J. Chen, J. Yang, K. Clawson, and Y. Peng, "A feature-weighted support vector machine and artificial neural network algorithm for academic course performance prediction," Neural Computing and Applications, vol. 35, no. 16, pp. 11517–11529, 2023.

[23] S. Hussain and M. Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," Annals of Data Science, vol. 10, no. 3, pp. 637–655, 2023.

[24] M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," in 2014 International Conference on Communication and Network Technologies, Sivakasi, India, pp. 113–118, 2014.

[25] U. Qamar and M. S. Raza, "Practical data science with WEKA," in Data Science Concepts and Techniques with Applications, Cham: Springer International Publishing, pp. 393–448, 2023.

[RECEIVED: 26TH SEPTEMBER, 2023; REVISED AND ACCEPTED: 27TH APRIL, 2024]