VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2023, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 11, Number 4, October-December 2023 pp:01-14

An Ensemble Modeling Approach to Enhance Grade Prediction in Academic Engineering Programming Courses

Khalid Mahboob *, **Sarfaraz Abdul Sattar Natha** , **Syed Saood Zia** , **Priha Bhatti** , **Abeer Javed Syed** , **Samra Mehmood**

Department of Software Engineering, Sir Syed University of Engineering Technology, Karachi, Pakistan

Keywords: *Ensemble, Programming Courses, Prediction, Machine Learning*

Journal Info:
Submitted:
September 26, 2023
Accepted:
October 30, 2023
Published:
November 13, 2023

Abstract

Predicting the future academic grades of students can play a pivotal role in enhancing their performance in specific courses, consequently yielding a positive impact on their prospective academic, professional, and personal achievements, as well as on society at large. The field of programming is rapidly gaining prominence as an essential profession spanning multiple domains, marked by abundant opportunities and financial rewards. To cater to the diverse interests of students, the recommended curriculum structure for engineering programs in computing adeptly combines theoretical knowledge with practical programming skills. This approach ensures that students acquire a comprehensive understanding of programming courses, allowing them to choose the path that aligns best with their envisioned careers as programmers This research endeavors to introduce ensemble prediction techniques aimed at identifying students who exhibit the potential for advancement, or conversely, those who may not excel in four university-level programming courses. The outcomes of this study are presented alongside valuable performance assessment metrics for five ensemble methodologies, namely AdaBoost, Bagging, Random Forest, Stacking, and Voting. This evaluation employs a 10-fold cross-validation methodology and incorporates the Principal Component Analysis (PCA) for feature ranking. The results unequivocally demonstrate that both the Stacking and Random Forest ensemble approaches have attained the highest level of accuracy when applied to two distinct datasets.

***Correspondence author email address:** kmahboob@ssuet.edu.pk
DOI: 10.21015/vtse.v11i4.1641

1 Introduction

Ensuring the nation's economic prosperity is a key government priority when formulating educational policies, with a particular emphasis on making higher-level engineering education more affordable [1]. Universities play a pivotal role in the education system, where they engage in rigorous research and development within a competitive environment. These institutions rely on the generation, accumulation, and dissemination of knowledge as fundamental prerequisites [2]. Universities often require access to historical and current data to inform decision-making by university administrators and address various issues [3]. Given that academic performance prediction is a common challenge in education, AI-based techniques are regularly employed for tasks related to forecasting student success. These analyses have demonstrated their ability to predict student outcomes and identify potential dropouts across a wide range of educational technology contexts. Predictive efforts often center on formal courses like university courses or structured online programs [4, 5]. Consequently, the availability of diverse pedagogical data is essential for applying various techniques aimed at enhancing the learning process. After analyzing students' academic data, this data is employed to monitor pedagogical progress, facilitate improvements, predict and enhance student retention early on, and investigate the likelihood of student failures or mistakes within a learning system [6]. The task of predicting and enhancing student learning has grown increasingly intricate, to elevate students' grades and support educational institutions in implementing various teaching methodologies. Ensemble learning, which incorporates both quantitative and qualitative data analysis, plays a central role in this endeavor. It involves effectively utilizing these techniques, whether a student seeks to enhance their learning skills through e-learning or an instructor needs to assess a student's academic records to evaluate their learning performance [7]. Ensemble approaches have been widely employed in education across different levels. Various parameters have been utilized to evaluate and monitor student pedagogical progress using ensemble techniques at different levels, including predicting the mastery of specific knowledge or skills at the system level, forecasting a student's ability to pass a course or obtain a degree at the course or degree level, or predicting their grades [8]. Research has suggested that one of the reasons for high dropout rates is inadequate career choices and a lack of personal interest. However, there has been limited investigation into the success of students' chosen career paths, particularly the factors influencing these choices. Consequently, parents and students have limited guidance in identifying suitable career options and course selections that align with their future goals and academic performance [9, 10].

Thus, this research endeavor is underpinned by a comprehensive set of motivations. It seeks to elevate academic achievement by employing accurately forecasting student grades in the context of demanding engineering programming courses, thereby providing a tailor-made educational experience. This foresight equips educational institutions with the tools to optimize resource allocation, adapt curricula to target prevalent learning challenges, and ultimately reinforce student retention within engineering programs. In a competitive educational environment, precise grade prediction has the potential to attract high-achieving students, thereby conferring a distinct advantage on the institutions [10].

Furthermore, this study aligns with the prevailing trend of data-driven education, demonstrating the potential of analytics and machine learning in enhancing educational outcomes. Consequently, it not only contributes to the enhancement of engineering programming courses but also adds to the broader conversation surrounding data-informed decision-making in education, fostering a culture of ongoing refinement and interdisciplinary insights.

2 Research Objectives

Predicting academic success at the university level is particularly important among the numerous prediction purposes. This research varies from other techniques to outcome prediction that has been proposed. The following are the primary research objectives:

- They are allowing the educational institution(s) to identify students who are (or are not) likely to show interest in programming courses in earlier semesters.
- They are identifying students who are at risk and providing appropriate guidance and specialized assistance to help them decrease their risk rates of failure and develop their interest in programming courses.
- They are identifying high-achieving students in programming subjects and assisting them in advancing their careers, particularly in the programming field.
- Furthermore, the research seeks to uncover the attainments in courses as key features using Principal Component Analysis (PCA) that might be used to predict the grades in programming courses.
- The predicted outcomes were checked for the accuracy of the ensemble models.
- . Finally, the research compares many ensemble models to determine which one gives the highest accuracy.

3 Related Work

In recent years, there has been a growing trend in the utilization of data mining and machine learning techniques to forecast the academic performance of students across various educational settings. A study, referenced as [11], employed five different data mining methods, including Extra Tree (ET), Linear Discriminant Analysis (LDA), Passive Aggressive Classifier (PAC), Radius Neighbor Classifier (RNC), and Support Vector Machine (SVM), to construct a practical framework aimed at enhancing the accuracy of predicting students' performance. Subsequently, the study compared the outcomes generated by these five algorithms to determine the most effective one. Notably, SVM emerged as the algorithm with the highest accuracy in predicting students' performance, offering valuable insights into identifying academically inactive students among newly admitted ones. Another piece of research, cited as [12], proposed a framework to guide students in making informed choices throughout their academic journey. This framework predicted students' GPAs in core disciplines based on their high school averages and IELTS scores, leveraging course-related data. Various machine learning algorithms were applied to train and test the data, ultimately forecasting optimal courses for different semesters. The study revealed that for Engineering and Business courses, the performance measures yielded Root Mean Square Errors (RMSE) of 0.469 and 0.45, respectively. However, it should be noted that this study employed a more limited dataset, resulting in decreased accuracy. In [13], researchers employed three distinct decision tree algorithms—Hoeffding Tree, J48, and REPTree—to predict university results and aid students in improving their performance by identifying areas of weakness. The dataset for university students encompassed a wide range of information, including assignment marks, attendance, class test marks, extracurricular activities, gender, tuition frequency, programming skills, and preceding semester Grade Point Averages (GPA). The study concluded that the J48 algorithm outperformed the others in terms of accuracy when applied to this specific dataset. To address the issue of student dropouts and provide support for retention, [14] utilized the Random Forest approach to identify attributes of first-year undergraduate students at risk of dropping out. A five-fold cross-validation model was implemented for more accurate dropout predictions, focusing solely on first-year programming courses. The model demonstrated an 81% accuracy rate, with potential for enhancement through the inclusion of additional characteristics to predict dropout rates over subsequent academic years. In [15], the study explored various ensemble prediction methods such as AdaBoost, Bagging, Gradient Boosting, Random Forest, and XGBoost to enhance student performance predictions. A combination of 13 features was used in training ensemble techniques for divination. Results indicated that both sampling and ensemble techniques improved the accuracy of student performance prediction, particularly in addressing multi-class imbalanced problems in educational data. The study identified ROS, a hybrid method with AdaBoost, as the most effective technique, and the SMOTEENN technique using ensemble classifiers showed considerable potential for enhancing student performance prediction models. A study detailed in [16] aimed to create a prediction model to assist computer science students in achieving optimal learning outcomes. The study employed 15 classification algorithms from diverse categories and found that the Naive Bayes and Hoeffding Tree algorithms produced the best results, achieving an accuracy of 91%. Additionally, the study discovered that certain courses could serve as early indicators of final GPA, with students receiving either an F or an A, thus predicting a Pass or Excellent final GPA. In [17], an ensemble model with two-layer stacking was proposed for predicting and analyzing student performance in academic competitions. The model categorized student performances into two symmetrical categorical classes and integrated three machine learning algorithms—AdaBoost, Random Forest, and Support Vector Machine (SVM)—at the first level. These algorithms were combined through logistic regression via stacking to enhance accuracy. Feature analysis revealed that candidates' competitive attitudes and GPAs played pivotal roles in predicting competition outcomes, with academic experience being deemed unnecessary. Lastly, [18] conducted experiments with various tree classification methods, accompanied by ensemble techniques, to identify the most effective classifier based on forecast accuracy. Results indicated that average grades in Mathematics were significant predictors for Mathematics and Engineering programs, while middle grades in English were influential in predicting outcomes for Non-Engineering programs. Bagging with the J48 tree classifier achieved the highest accuracy among the classifiers, while forest tree classifiers excelled in dropout precision and graduation recall. Combining these tree-training algorithms with an ensemble approach resulted in marginal improvements in classification performance.

4 Research Methodology

Ensemble algorithms are machine learning algorithms that combine predictions from multiple models into a single prediction. When compared to a single model, this method offers improved predictive performance. This section uses ensemble algorithms to predict the student's grades of two consecutive academic batches, i.e., batch 2016 and batch 2017, comprising 340 graduate students' instances belonging to computer engineering technology. Four programming course marks and gender as indicators have been used as arbitrating metrics to assess student's academic grades, as shown in Table 1. The acronyms for the course names are C&P.F (Computer and Programing Fundamentals), OOP (Object Oriented Programming), DS&A (Data Structures and Algorithms), and ADV.OOP (Advanced Object-Oriented Programming). In this study, we have used ensemble techniques with the Weka tool for predictions as one of the reasons for selecting Weka for applied machine learning is the wide range of ensemble machine learning algorithms available [19, 32]. The following is a breakdown of the research approach used by:

- Collecting data, data pre-processing, and features ranking.
- Applying ensemble algorithms and analyzing performance.
- Calculating accuracy, precisions, recall, and F1-score with 10-fold cross-validation.

Table 1. List of Variables and their Values in a Dataset

Variables	Values
Gender	{Male, Female}
CE-101 (C&P.F)	{0-100}
CE-102 (OOP)	{0-100}
CE-205 (DS.&A)	{0-100}
CE-206 (ADV.OOP)	{0-100}
Grade	A {90-100} Excellent A- {85-89} Very Good B+ {80-84} Good B {70-79} Above Average C+ {60-69} Average C {50-59} Satisfactory F {Less than 50} Fail

4.1 A. Data Collection, Pre-processing, and Visualizations

Computer engineering students' academic dataset used in this research is collected from a renowned and large private sector university Sir Syed University of Engineering & Technology (SSUET), located in the urban city of Karachi. Then, data pre-processing is applied to both datasets to clean and validate the data.

Data pre-processing is the process of transforming unstructured data into well-formed data sets that may be used in data mining techniques. When it comes to raw data, it's common to be incomplete and structured incorrectly. The quality of data preparation is directly linked to the success of any data analytics effort. The pre-processing procedure includes both data validation and data imputation. Data validation aims to see if the data is complete and accurate. Machine learning (ML) methods require data preparation to guarantee that large datasets are prepared so that the information they contain can be handled and evaluated by learning algorithms [20].

The descriptive statistics and their values obtained at the pre-processing stage are listed below in Table 2 belonging to batch 2016 and Table 3 belonging to batch 2017, respectively. For numeric variables, a four-number summary is supplied, including the minimum values, the maximum values, the mean values, the values of standard deviation for each programming course and grade, and the total number of instances of male and female.

4.1.1 Histogram

The most typical purpose of histograms is to show how a set of data looks when seen as a whole. Histograms quickly reveal whether a dataset's values are concentrated around a limited number of ranges or are more evenly distributed. A histogram shows the frequency distribution of data attributes within a class interval and measures their maximum, minimum, and median values.

Here, the Weka automatically generates the histogram visualizations at the pre-processing stage. The histogram generated at the pre-processing step demonstrates in Figures 1 and 2 that the grade variable yielded 26 cases of grade A with Excellent progress (visualized in a cyan bar) and just 1 case of grade A with Excellent progress (pictured in a light green bar), whereas the 48 cases of grade F with the Fail progress (visualized in a red bar) and the 60 cases of grade F with the Fail progress (pictured in a red bar), and so on.

Table 2. Summary of pre-processing results on batch 2016

Variables	**Pre-processing Results**						
Gender	Male	Female					
	140	30					
CE-101 (C&P.F)	Min	Max	Mean	StdDev			
	0	92	65.124	12.395			
CE-102 (OOP)	Min	Max	Mean	StdDev			
	0	96	55.929	24.744			
CE-205 (DS.&A)	Min	Max	Mean	StdDev			
	0	94	51.588	29.316			
CE-206 (ADV.OOP)	Min	Max	Mean	StdDev			
	0	100	52.835	30.122			
Grade	A	A-	B+	B	C+	C	F
	26	6	18	29	37	6	48

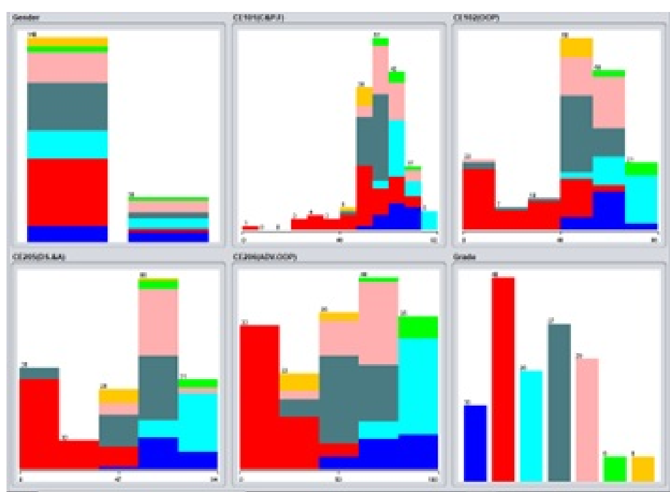

Figure 1. Representation of Histograms of Batch 2016

4.1.2 Box and Whisker Plot

A box and whisker plot is a graphical representation of variance in data collection. A histogram analysis is usu-

Table 3. Summary of pre-processing results on batch 2017

Variables	Pre-processing Results						
Gender	Male	Female					
	148	22					
CE-101 (C&P.F)	Min	Max	Mean	StdDev			
	0	94	65.141	15.153			
CE-102 (OOP)	Min	Max	Mean	StdDev			
	0	94	54.859	23.526			
CE-205 (DS.&A)	Min	Max	Mean	StdDev			
	0	98	47.759	31.406			
CE-206 (ADV.OOP)	Min	Max	Mean	StdDev			
	0	95	44.935	29.446			
Grade	A	A-	B+	B	C+	C	F
	1	5	12	32	36	24	60

ally sufficient, but a box and whisker plot can give more information while also allowing various sets of data to be shown in the same graph. A rectangle is created to represent the second and third quartiles, generally with a vertical line to indicate the median value, in a fundamental manner of depicting statistical data on a plot. Horizontal lines on either side of the rectangle represent the lower and upper quartiles.

The box and whisker plots are generated to compute the statistical values of median, range, and interquartile range (IQR) of all four programming courses separately, as shown in Figures 3 and 4. Figure 3 depicts the median is 66, the range is 92, and IQR is 12 for the course C&P.F, the median is 62, the range is 96, and IQR is 21for the course OOP, the median is 61, the range is 94, and IQR is 38 for the course DS.&A, the median is 59, the range is 100, and

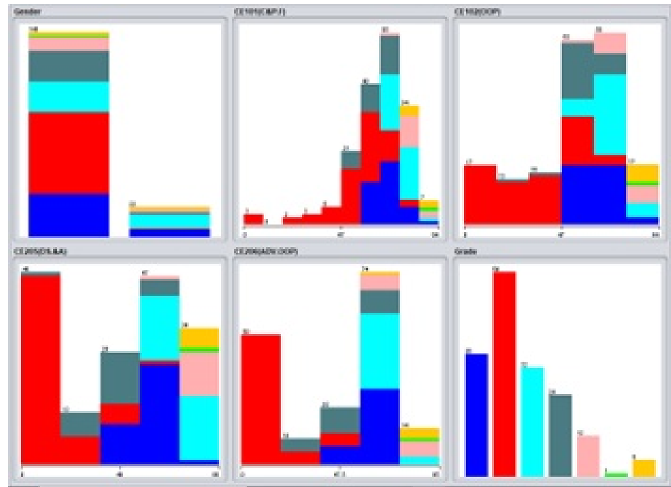

Figure 2. Representation of Histograms of Batch 2017

the IQR is 47 for the course ADV.OOP is respectively belonging to batch 2016. Figure 4 depicts the median is 67, the range is 94, and IQR is 18 for the course C&P.F, the median is 61, the range is 94, and IQR is 30 for the course OOP, the median is 56, the range is 98, and IQR is 61 for the course DS.&A, the median is 59, the range is 95, and the IQR is 59 for the course ADV.OOP is respectively belonging to batch 2017.

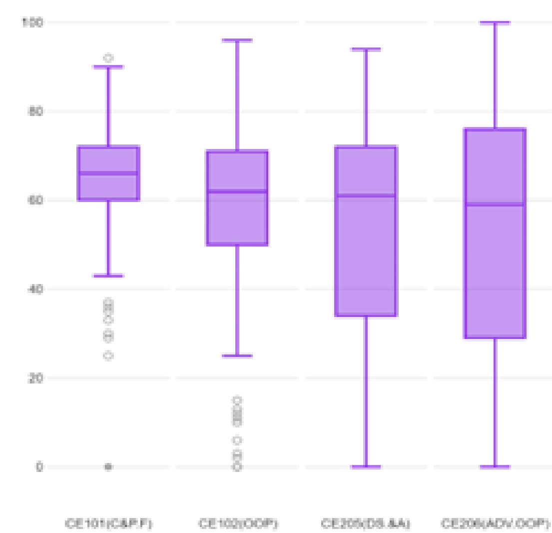

Figure 3. Representation of Box and Whisker Plot for programming courses of batch 2016

4.2 B. Ensemble Techniques Description

We employed the feature ranking approach, i.e., Principal Component Analysis (PCA), and ensemble learning algorithms using a cross-validation approach in this study below:

4.2.1 1. Cross-Validation

Cross-validation is a method through which the model is trained on a subset of the dataset and then assessed on the other subset. This approach divides the data set into k subsets (also known as folds) and then trains all of the subsets while leaving one (k-1) subset to assess the trained model. We iterate k times with a distinct subset allocated for testing purposes in this procedure.

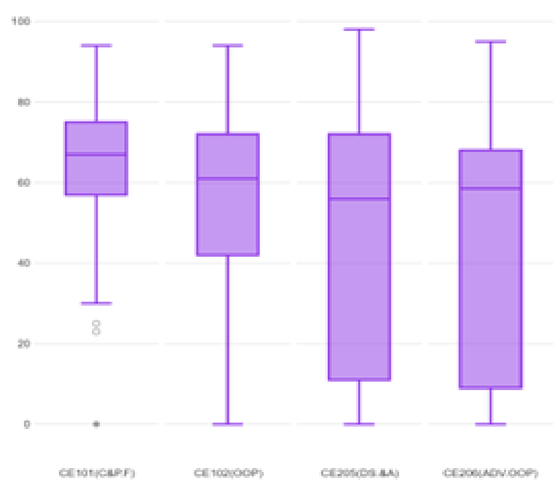

Figure 4. Representation of Box and Whisker Plot for programming courses of batch 2017

Cross-validation is a very effective technique. It allows us to make better use of our data and provides us with a lot more information regarding the performance of our algorithms. Furthermore, determine the model's hyperparameters or which parameters will result in the lowest test error [25]. 2. Principal Component Analysis (PCA) Principal Component Analysis (PCA) is a method for decreasing the dimensionality of the datasets, maximizing interpretability while minimizing information loss. It accomplishes this by generating new uncorrelated variables that systematically optimize variance. We can compress data using PCA by making the machine learning algorithms "efficient" and the data set smaller. A simpler predictive model with fewer input variables can better predict new data [21]. 3. AdaBoost / Boosting AdaBoost or Boosting is a machine learning method that may improve the performance of any other machine learning technique. It works well with weak learners. On a classification task, these models reach accuracy slightly above random chance. Decision trees with one level are the most suitable and commonly used method with AdaBoost or Boosting [6, 15]. 4. Bootstrap Aggregation (Bagging) Bagging, also known as bootstrap aggregation, is a machine learning technique that evolves machine learning models via ensemble learning. When dealing with a noisy dataset, it is frequently used to minimize variance. Bagging is the process of selecting a random sample of data from a training set with replacement—that is, the individual data points can be sampled many times [6, 15].

5. Random Forest Random forest is a classification technique that uses numerous decision trees to classify data. Creating each tree employs bagging and feature randomization to generate an uncorrelated forest of trees whose collective prediction is more accurate than that of any single tree [15, 28]. 6. Stacking Stacking is an essential extension of Voting Ensembles that may be applied to classification and regression tasks. Stacking determines another model to learn how to optimally combine the predictions from the sub-models in conjunction with picking numerous sub-models. Because a meta-model is used to integrate sub-model forecasts optimally, this method is also known as blending [22]. 7. Voting Voting is one of the most basic and influential ensemble algorithms. It may be applied to problems involving classification and regression. Voting works by splitting the population into two or more sub-models. Each sub-model provides predictions that are merged in some fashion, such as by calculating the mean or mode of the predictions, enabling each sub-model to vote on the outcome [23].

4.3 C. Proposed Research Model

The following are the elements of the proposed research model:

- The student's academic datasets in computer engineering are gathered from SSUET.
- The dataset was normalized after missing values were removed at the pre-processing stage.
- The Histograms and Box and Whisker Plots are used to represent data frequency distribution.
- The Principal Component Analysis (PCA) feature ranking approach is used for essential features ranking in the dataset.
- For stronger correlations among variables, the Correlation Heatmap is used.
- For improved prediction (Accuracy, Precision, Recall, and F1-Score), the AdaBoost Ensemble approach (DecisionTable) is used.
- For improved prediction (Accuracy, Precision, Recall, and F1-Score), the Bagging Ensemble approach (J48 DecisionTree) is used.
- . For improved prediction (Accuracy, Precision, Recall, and F1-Score), the Random Forest Ensemble approach is used.
- For improved prediction (Accuracy, Precision, Recall, and F1-Score), the Stacking Ensemble approach (AttributeSelectedClassifier as a Meta classifier) is used.
- For a final prediction, Stacking and Voting Ensemble approaches are used.

5 Results and Analysis

This section presents the results and analysis using five ensemble techniques, i.e., AdaBoost, Bagging, Random Forest, Stacking, and Voting using 10-fold cross-validation with the PCA method for features ranking [24, 30].

5.1 A. Performance Metrics Analysis and Comparison

Classifier performance is the most critical metric for any predictive model, especially when designed to predict student performance. Consequently, performance metrics are crucial in predicting performance [11]. We have expected the grades of graduate students in four programming courses enrolled in computer engineering technology with 340 instances from two consecutive academic batches, i.e., batch 2016 and batch 2017. Table 4 lists the accuracy, precision, recall, and F1-measure values for each course using ensemble classifiers. The proportion of all correct predictions divided by the total number of instances is used to estimate a classifier's accuracy. If the classifier's accuracy is deemed acceptable mathematically, the classifier may categorize subsequent data tuples for which the class label is unknown. Precision is determined as the number of true positive values divided by the total number of true positive and false positive values in a two-class unbalanced classification problem. An outcome ranges from 0.0 (no precision) to 1.0 (perfect precision). The recall is a statistic that measures how many correct positive predictions were produced out of all possible positive predictions. The F1-score or F1-measure measures a test's accuracy in binary classification statistical analysis. Precision and recall are used to compute it. The formulas of the valuation metrics are given in Table 5.

- **TP = True Positive** = A true positive is an event in which the model predicts the positive class adequately.
- **TN = True Negative** = A true negative is an event in which the model predicts the negative class adequately.
- **FP = False Positive** = A false positive occurs when the model mispredicts the positive class.
- **FN = False Negative** = A false negative occurs when the model mispredicts the negative class.

Figure 5 depicts the results of a comparative evaluation of four programming courses taught in engineering technologies across two consecutive batches in terms of data classification accuracy using five ensemble classifiers for the class Grade. Table 4 and Figure 5 show that Stacking on batch 2017 and Random Forest on batch 2016 have the highest accuracies, respectively.

Figures 6 and 7 illustrate the students' attainments of marks in four different programming courses of

Table 4. Ensemble models performance metrics analysis and comparison

Batch	Ensemble Model	Accuracy	Precision	Recall	F1-Measure
5*2016	AdaBoost / Boosting	77.64%	0.775	0.776	0.770
	Bootstrap Aggregation (Bagging)	79.41%	0.798	0.794	0.792
	Random Forest	84.70%	0.829	0.847	0.837
	Stacking	81.76%	0.806	0.818	0.799
	Voting	81.17%	0.790	0.812	0.786
5*2017	AdaBoost / Boosting	76.47%	0.355	0.765	0.355
	Bootstrap Aggregation (Bagging)	79.41%	0.369	0.794	0.375
	Random Forest	75.88%	0.761	0.759	0.759
	Stacking	84.11%	0.462	0.841	0.464
	Voting	81.17%	0.384	0.812	0.385

Table 5. Computational formula summary for prediction

Metrics	Formula
Accuracy	$\frac{TN+TP}{TN+FN+FP+TP}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-Measure	$\frac{2TP}{2TP+FP+FN}$

both batches 2016 and 2017, respectively. The visualization in Figure 6 shows that the students attained the highest marks for the target variable Grade in a course CE-206 (ADV.OOP). Whereas, Figure 7 visualization shows that the students achieved the highest marks for the target variable Grade in a course CE-205 (DS.&A).

5.2 B. Ranking of Features using Principal Component Analysis (PCA)

PCA is used for feature ranking in this study, as mentioned earlier. When utilizing PCA as a feature extraction or ranking method, the primary concept is to choose variables based on the magnitude of their coefficients (from most significant to lowest in absolute values). The volume of the associated values in the eigenvectors reflects the relevance of each attribute (larger magnitude — more importance). When using machine learning algorithms in analysis, feature reduction is an efficient strategy to enhance classification performance. Figure 8 shows that the top three ranked features that emerged in this study on applying PCA are Gender with the value of Female, CE-101 (C&P.F), and CE-102 (OOP) for both batches.

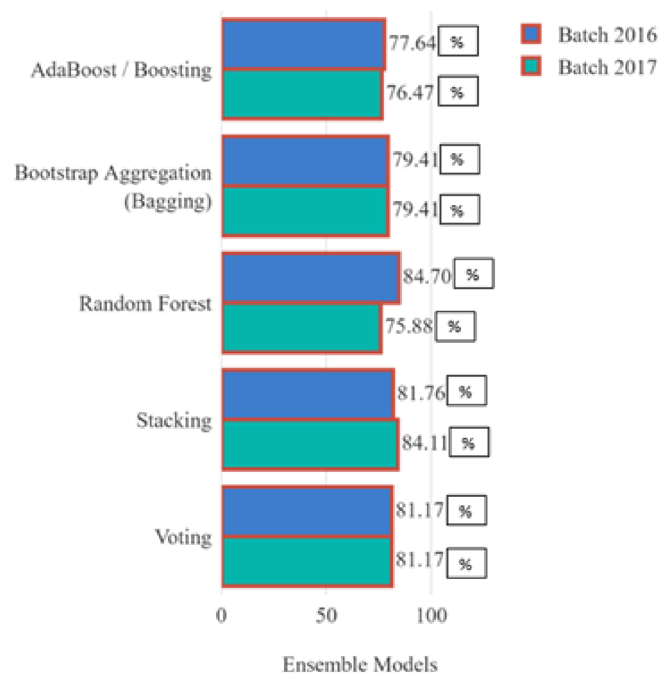

Figure 5. Computational representation of Ensemble Classification Accuracies

5.3 C. Correlation Analysis between Courses, Gender, and Grade Distribution

Now, we discuss the correlation matrix of both batches used to predict student success by analyzing grades

Figure 6. Batch 2016 course attainments visualization

Figure 7. Batch 2017 course attainments visualization

predicted through the marks attained in programming courses. The first step is to obtain a correlation matrix to understand the data and determine the marks distribution for the target variable Grade in each course. The combined distribution of grades can be observed using a correlation matrix between pairs of courses. We next calculate the Pearson correlation coefficient to see whether there is any correlation between the programming courses and gender. A more excellent correlation value indicates that students who perform well in one subject are also likely to achieve well in another course. The correlation coefficient determined in the first stage may be used to rank courses in relevance; the more robust the correlation, the better the course grade in predicting success. However, correlation is only one metric; a better method would be to rate courses based on various metrics and then average the results. Different ways

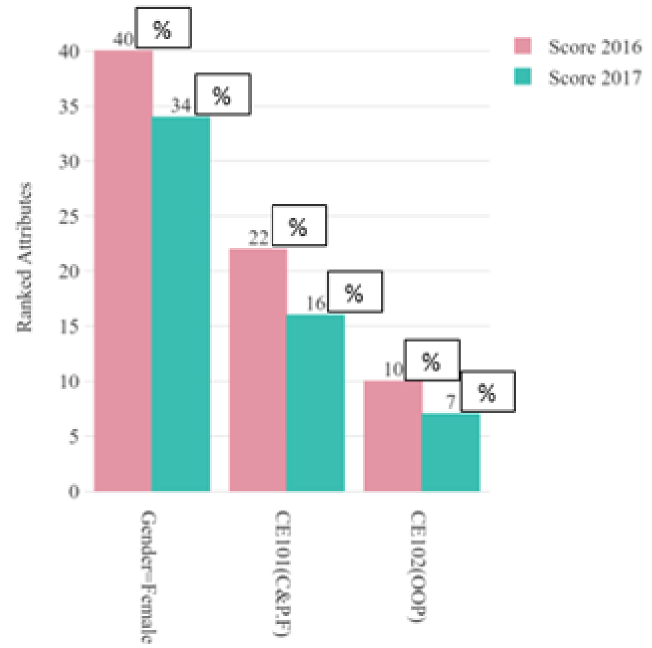

Figure 8. Features ranking using PCA

may be used to rank input variables based on their importance for predicting the output variable. Figures 9 and 10 depict the correlation matrix plots of batch 2016 and 2017, respectively. The estimated Pearson correlation values are marked in the upper triangle, which is symmetrical around the main diagonal [27]. Each of the programming courses' gender is displayed on the main diagonal. The color intensity of points represents the marks obtained in each course scatter plot (the darker the point, the more significant the correlation between the variables). We can observe from Figure 9 (which is the correlation matrix of batch 2016) and Figure 10 (which is the correlation matrix of batch 2017) that the courses CE-205 (DS.&A) and CE-206 (ADV.OOP) are highly correlated. According to the above analyses, the most common grades predicted for all courses are 'B,' 'C+,' and 'F.' Many students, particularly those at risk, struggle to pick their subsequent courses after completing their first year and throughout their studies, particularly with electives and general studies courses. Students on probation may face academic expulsion if their Grades fall below a certain threshold [26]. They have a higher

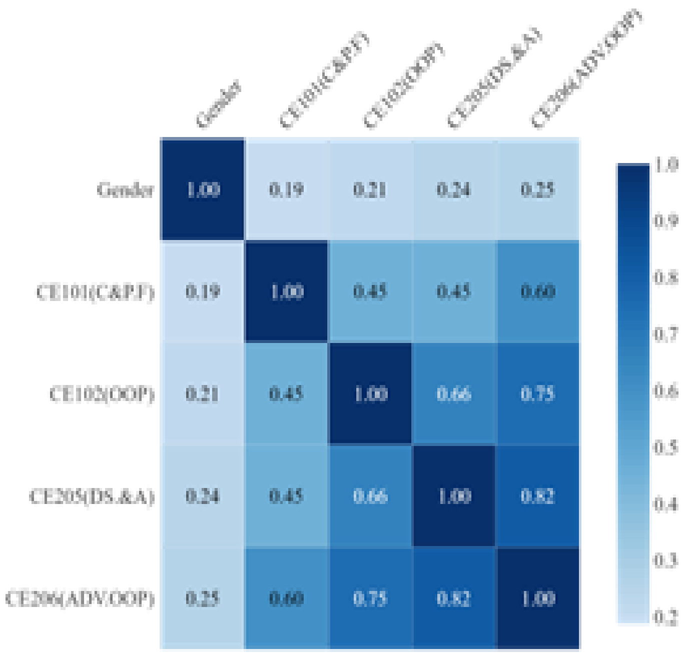

Figure 9. Correlation matrix representation of batch 2016

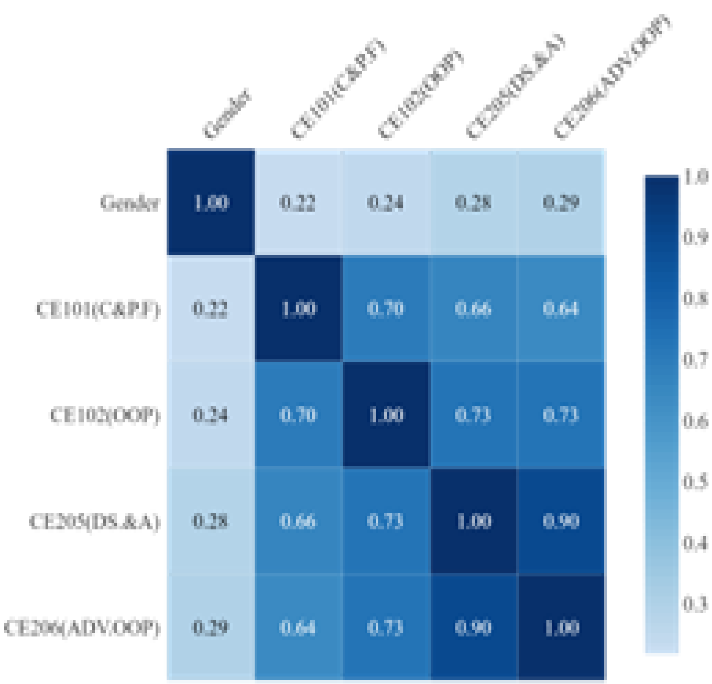

Figure 10. Correlation matrix representation of batch 2017

chance of getting out of probation if they choose a course with the highest possibility of success. In addition, the institution encourages entrepreneurship and is transitioning to flexible degrees that allow students to tailor their studies and earn multidisciplinary credentials. As a result, predicting a student's success in a course can substantially assist students in making course selection decisions for the reasons stated above [12].

6 Conclusion

The contribution to student achievement is significant from various stakeholders, including students themselves, their relationships, higher education institutions (HEIs), society, and nations. Predicting students' future performance holds potential benefits for their educational journey. Educators and advisors can make well-informed decisions to support students effectively. In this study, an ensemble approach is employed to present a strategy for predicting student performance, specifically their grades, in four foundational programming courses within an engineering degree program during early semesters. The choice of an ensemble approach over a single model is motivated by two interconnected factors. Firstly, ensemble methods tend to yield superior predictions and performance compared to individual models. Secondly, they enhance the robustness of forecasts and mitigate model variability.

Utilizing ensemble techniques can yield insights that enable the creation of informed decisions to assist students in improving their academic performance. A 10-fold cross-validation procedure is applied to ensure the reliability of the results. The study demonstrates an intelligent evaluation of student course grade data, employing readily available machine learning tools like Weka. Among the techniques evaluated, stacking achieved the highest accuracy, registering 84.11% on the 2017 batch data, while Random Forest achieved the highest accuracy at 84.70% for the 2016 batch. Notably, students from the 2016 batch earned the highest marks in the Advanced Object-Oriented Programming course, whereas the 2017 batch excelled in Data Structures & Algorithms.

Across both batches, the top three influential features, as determined by Principal Component Analysis (PCA), were Gender (with a predominance of Female), CE-101 (C&P.F), and CE-102 (OOP). It's worth highlighting that even though female students constituted only a quarter of the male student population, 92% of them passed all programming courses, compared to 64% of male students. This gender comparison revealed a significant disparity in grade performance, indicating the potential for more accurate grade predictions within a subset of these courses. These findings underscore the importance of focusing on low-performing students to enhance overall educational quality.

Future research endeavors may involve replicating the same analysis with a new batch of engineering students to ascertain if the identified courses maintain their impact. Additionally, expanding the dataset with more features could lead to more precise findings. Consequently, future research may delve into combining ensemble approaches with various base classifiers and exploring additional hyperparameter tuning to further enhance predictive capabilities.

Author Contributions

Khalid Mahboob: Writing – original draft, Investigation. **Priha Bhatti / Abeer Javed Syed**: Visualization, Investigation, Writing – original draft. **Sarfaraz Abdul Sattar Natha / Samra Mehmood**: Comparative Analysis, writing – review and editing. **Syed Saood Zia:** Proofreading, Editing and Supervision.

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

Funding Information

NO funding associated with this research.

References

[1] J. Xu, K. H. Moon, and M. Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017. DOI: 10.1109/JSTSP.2017.2692560.

[2] A. Tekin, "Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach," *Eurasian Journal of Educational Research*, vol. 14, no. 54, pp. 207–226, 2014. DOI: 10.14689/ejer.2014.54.12.

[3] M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," *International Journal of Information and Education Technology*, vol. 6, no. 7, pp. 528–533, 2016. DOI: 10.7763/ijiet.2016.v6.745.

[4] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. Hernandez-Torrano, "Predicting Students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors," *IEEE Access*, vol. 5, pp. 23792–802, 2017. DOI: 10.1109/ACCESS.2017.2740980.

[5] K. Mahboob, S. A. Ali, D. U. R. Khan, and F. Ali, "A comparative study of engineering students pedagogical progress," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 323–31, 2018. DOI: 10.14569/IJACSA.2018.090646.

[6] F. F. Patacsil, "Survival analysis approach for early prediction of student dropout using enrollment student data and ensemble models," *Universal Journal of Educational Research*, vol. 8, no. 9, pp. 4036–47, 2020. DOI: 10.13189/ujer.2020.080929.

[7] B. Al Breiki, N. Zaki, and E. A. Mohamed, "Using Educational Data Mining Techniques to Predict Student Performance," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2019. DOI: 10.1109/ICECTA48151.2019.8959676.

[8] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, pp. 1–21, 2019. DOI: 10.1016/j.heliyon.2019.e01250.

[9] A. E. Tatar and D. Düştegör, "Prediction of academic performance at undergraduate graduation: Course grades or grade point average?" *Applied Sciences*, vol. 10, no. 14, pp. 1–15, 2020. DOI: 10.3390/app10144967.

[10] S. Malla, J. Wang, W. Hendrix, and K. Christensen, "Predicting Success for Computer Science Students in CS2 using Grades in Previous Courses," in *Proceedings - Frontiers in Education Conference (FIE)*, 2019, pp. 1–5. DOI: 10.1109/FIE43999.2019.9028577.

[11] S. Randhir and P. Saurabh, "Application of Machine Learning Algorithms to Predict Students Performance," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 7249–61, 2020.

[12] A. G. Fateh Allah, "Using machine learning to support students academic decisions," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 18, pp. 3778–96, 2020.

[13] M. I. Hoque, A. K. Azad, T. M. Abu Hurayra Tuhin, and Z. U. Salehin, "University students result analysis and prediction system by decision tree algorithm," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 3, pp. 115–22, 2020. DOI: 10.25046/aj050315.

[14] A. Behr, M. Giese, K. D. Teguim, and K. Theune, "Early prediction of university dropouts - A random forest approach," *Jahrb Natl Okon Stat*, vol. 240, no. 6, pp. 743–89, 2020. DOI: 10.1515/jbnst-2019-0006.

[15] H. Hassan, N. B. Ahmad, and S. Anuar, "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining," *Journal of Physics: Conference Series*, vol. 1529, no. 5, pp. 1–8, 2020. DOI: 10.1088/1742-6596/1529/5/052041.

[16] N. Alangari and R. Alturki, "Predicting students final GPA using 15 classification algorithms," *Romanian Journal of Information Science and Technology*, vol. 23, no. 3, pp. 238–49, 2020.

[17] Y. Lijuan and L. Yanshen, "An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning," *Symmetry*, vol. 12, no. 728, pp. 1–17, 2020.

[18] F. F. Patacsil, "Predicting University Students' Academic Success Using Different Tree Classifiers And Ensemble Approaches To Suggest Suitable Program," *International Journal of Scientific Technology Research*, vol. 9, no. 02, pp. 6001–9, 2020.

[19] K. P. Wahome, W. Bongo, and R. R. Maina, "Towards Effective Data Preprocessing for Classification Using WEKA," *International Journal of Science and Research (IJSR)*, vol. 5, no. 8, pp. 1210–1214, 2016. DOI: 10.21275/ART2016999.

[20] P. Dhakate, S. Patil, K. Rajeswari, and D. Abin, "Preprocessing and Classification in WEKA Using Different Classifiers," *Journal of Engineering Research and Applications*, vol. 4, no. 8, pp. 91–93, 2014. Available from: www.ijera.com.

[21] Y. Zhu and D. Zhou, "Online and Offline Teaching Effect of C Language in Colleges and Universities Based on Principal Component Analysis and Neural Network," *2021 2nd International Conference on Computers, Information Processing and Advanced Education*, pp. 733–737, 2021. DOI: 10.1145/3456887.3457054.

[22] R. O. Odegua, "An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking)," *Deep Learn IndabaX*, vol. 12, no. 10, pp. 1–10, 2020.

[23] F. Leon, S. A. Floria, and C. Badica, "Evaluating the effect of voting methods on ensemble-based classification," *Proceedings - 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications, INISTA 2017*, pp. 1–6, 2017. DOI: 10.1109/INISTA.2017.8001122.

[24] Z. Iqbal, A. Qayyum, S. Latif, and J. Qadir, "Early Student Grade Prediction: An Empirical Study," *2019 2nd International Conference on Advancements in Computational Sciences, ICACS 2019*, 2019. DOI: 10.23919/ICACS.2019.8689136.

[25] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, 2021. DOI: 10.3390/a14110318.

[26] M. Ali, "Novel technique to analyze the effects of cognitive and non-cognitive predictors on students course withdrawal in college," *Technology Interface International Journal*, vol. 20, no. 2, pp. 36-41, 2020.

[27] A. Almasri, E. Celebi, and R. S. Alkhawaldeh, "EMT: Ensemble meta-based tree model for predicting student performance," *Scientific Programming*, vol. 2019, pp. 1–13, 2019. DOI: 10.1155/2019/3610248.

[28] W. Nuankaew and J. Thongkam, "Improving Student Academic Performance Prediction Models using Feature Selection," *17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2020*, pp. 392–395, 2020. DOI: 10.1109/ECTICON49241.2020.9158286.

[29] H. Gunawan, Hanes, and Catherine, "Information Systems Students' Study Performance Prediction Using Data Mining Approach," *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, 2019. DOI: 10.1109/ICIC47613.2019.8985718.

[30] R. Singh and S. Paul, "Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 3970–3976, 2020. DOI: 10.30534/ijatcse/2020/221932020.

[31] A. Vihavainen, "Predicting students' performance in an introductory programming course using data from students' own programming process," *Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013*, pp. 498–499, 2013. DOI: 10.1109/ICALT.2013.161.

[32] 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," *2019 10th International Conference on Information and Communication Systems, ICICS 2019*, pp. 104–109, 2019. DOI: 10.1109/IACS.2019.8809106.