Original Manuscript

Information Development
1–20

DOI: 10.1177/02666669231213023
journals.sagepub.com/home/idv
Sage

Predicting academic performance of students with machine learning

Yavuz Selim Balcioğlu
Faculty of Business Administration, Management Information Systems Department, Gebze Technical University, Turkey

Melike Artar
Faculty of Business Administration, Management Department, Gebze Technical University, Turkey

Abstract
This study investigates the effectiveness of machine learning and deep learning models for early prediction of student performance in higher education institutions. Using the Open University Learning Analytics (OULA) dataset, various models, including Decision Tree, Support Vector Machine, Neural Network, and Ensemble Model, were employed to predict student performance in three categories: Pass/Fail, Close to Fail, and Close to Pass. The Ensemble Model (EM) consistently outperformed other models, achieving the highest overall F1 measure, precision, recall, and accuracy. These results highlight the potential of data-driven techniques in informing educational stakeholders' decision-making processes, enabling targeted interventions, and facilitating personalized learning strategies tailored to students' needs. By identifying at-risk students early in the academic year, institutions can provide additional support to improve academic outcomes and retention rates. The study also discusses practical implications, including the development of pedagogical policies and guidelines based on early predictions, which can help educational institutions maintain strong academic outcomes and enhance their reputation for academic excellence. Future research aims to investigate the impact of individual activities on student performance and explore day-to-day student behaviors, enabling the creation of tailored pedagogical policies and guidelines.

Keywords
academic performance, machine learning, decision trees, ensemble methods support vector machines, neural networks

Submitted: 12 April 2023; accepted: 23 October 2023.

About the authors

The academic performance of students has long been a primary concern for educators, parents, and policymakers alike. Understanding the factors that contribute to students' success can inform the development of targeted interventions and support systems to enhance their educational experiences and outcomes (Albreiki et al., 2021). With the rise of machine learning and artificial intelligence, new opportunities have emerged to explore and predict academic performance in a more data-driven and efficient manner (Yağcı, 2022).

In recent years, machine learning has become an integral part of various industries, demonstrating its potential to transform and optimize processes across domains. In the context of education, machine learning can be employed to analyze large datasets, identify patterns, and develop predictive models that provide valuable insights into student performance (Nti et al., 2022). By leveraging these techniques, educators can develop a better understanding of the

Corresponding author:
Melike Artar, Gebze Technical University, Faculty of Business Administration, Management Department, Department, Gebze, Kocaeli, Turkey.
Email: artar@gtu.edu.tr

factors influencing academic achievement and create more effective, personalized learning experiences.

Several studies have been conducted to examine the application of machine learning techniques in predicting academic performance. For example, Dabhade et al. (2021) investigated the use of data mining techniques to identify factors affecting student dropouts, while Xu et al. (2019) applied machine learning algorithms to predict student failure in online courses. These studies have demonstrated the potential of machine learning for identifying at-risk students and informing the development of targeted interventions.

When the studies affecting academic performance in the literature are examined, four important variables draw attention. These include demographic factors (e.g., gender, socioeconomic status), prior academic achievement (e.g., high school grades, standardized test scores), psychological factors (e.g., motivation, self-efficacy), and behavioral indicators (e.g., attendance, time spent on homework). By incorporating these variables into machine learning models, researchers can gain a more comprehensive understanding of the factors that contribute to student success (Andres, 2020; Zeineddine et al., 2021; Goldberg et al., 2021).

The purpose of this study is to apply machine learning algorithms to predict the academic performance of students based on a variety of factors, such as demographic information, prior academic history, and behavioral data. By comparing the performance of different machine learning models, we aim to identify the most accurate and reliable method for predicting student outcomes. Furthermore, our study seeks to determine the most significant factors contributing to academic performance, which can inform targeted interventions and support strategies. The objectives of this research are:

- To review existing literature on the prediction of academic performance using machine learning techniques and identify relevant factors influencing student outcomes.
- To preprocess and analyze a dataset containing student information, demographic variables, and academic performance indicators.
- To apply various machine learning algorithms, such as decision trees, support vector machines, and neural networks, ensemble method, to predict student performance.
- To evaluate the performance of each model and identify the most effective approach for predicting academic achievement.
- To discuss the implications of our findings for educators, policymakers, and researchers in the field of education.

Literature review

Predicting academic performance has been a longstanding topic of interest for researchers in the field of education. Numerous studies have investigated various factors that contribute to student success, including demographic factors (Waheed et al., 2020; Zeineddine et al., 2021; Agrawal and Mavani, 2015), prior academic achievement (Pallathadka et al., 2021), psychological factors (Alhadabi and Karpinski, 2020; Andres, 2020) and behavioral indicators (Jovanović et al., 2021; Goldberg et al., 2021).The identification of these factors enables educators and policymakers to design targeted interventions and support systems to enhance student performance.

In the early 2000s, (Kotsiantis et al., 2002) pioneered the use of Naive Bayes to predict academic outcomes, setting the stage for subsequent advancements. The mid-2000s saw a surge in interest in Learning approaches, with seminal works by (Chamorro-Premuzic and Furnham, 2008) and (Cortez and Silva, 2008) highlighting its potential. More recently, the advent of sophisticated machine learning techniques has ushered in a new era of research, as evidenced by studies such as (Agrawal and Mavani, 2015) and (Rastrollo-Guerrero et al., 2020). These works underscore the growing complexity and nuance in predicting academic performance, with a shift towards.

- **Demographic variables:** Demographic variables are essential factors to consider when predicting academic performance, as they often provide insights into the diverse backgrounds and personal characteristics of students. Previous research has demonstrated that demographic factors such as gender, age, socioeconomic status, ethnicity, and parental education can significantly influence students' academic achievements (Albreiki et al., 2021; Waheed et al., 2020).
- Gender differences in academic performance have been widely studied, with results indicating that female students generally outperform male students in language and reading skills, while

male students tend to excel in mathematics and science domains (Walia et al., 2020). In addition, some research suggests that age is positively correlated with academic performance, with older students typically performing better than younger students in similar educational settings (Waheed et al., 2020).

- Socioeconomic status (SES) is another crucial demographic factor that has been shown to impact academic performance significantly. Students from higher SES backgrounds typically have better access to educational resources and supportive learning environments, which can contribute to higher academic achievement (Chen et al., 2021). Parental education, an important component of SES, is also positively correlated with students' academic performance. Studies have found that students whose parents have higher levels of education tend to perform better academically (Destin et al., 2019).
- Ethnicity is another demographic factor that has been associated with differences in academic performance. Minority students, particularly those from historically underrepresented or disadvantaged groups, often face various challenges that can negatively impact their academic achievements, such as language barriers, discrimination, and limited access to quality education (Zeineddine et al., 2021)
- **Prior academic achievement:** Prior academic achievement has been consistently identified as a strong predictor of future academic performance in numerous studies. Students who have demonstrated success in previous academic settings are more likely to continue performing well in subsequent academic endeavors (Hamm et al., 2020; Alyahyan and Düştegör, 2020).
- Research has shown that prior academic achievement, such as students' grade point average (GPA), standardized test scores, and class ranking, can serve as reliable predictors of academic performance in higher education (Pallathadka et al., 2021). In addition, students' performance in specific subjects or courses has been found to be strongly related to their success in related fields. For example, students who have performed well in mathematics courses are more likely to excel in science, technology, engineering, and mathematics (STEM) disciplines (Hamm et al., 2020).
- In addition to course-specific achievements, research has demonstrated that various aspects of academic behavior and engagement, such as time spent on homework, class attendance, and participation, are also closely linked to students' academic performance (Alyahyan and Düştegör, 2020). These behaviors often reflect a student's motivation, self-regulation, and learning strategies, which are essential factors contributing to academic success.
- Furthermore, the role of prior academic achievement in predicting academic performance has been explored in various educational contexts, such as online and distance learning. For instance, Yukselturk and Top (2013) found that prior academic achievement was a significant predictor of students' success in online courses, with students who had higher GPAs in traditional face-to-face courses being more likely to perform well in online learning environments.
- **Psychological factors:** Psychological factors play a critical role in determining academic performance. These factors include motivation, self-efficacy, self-regulation, goal orientation, and emotional intelligence, among others. Previous research has demonstrated that students who exhibit higher levels of these psychological attributes are more likely to achieve academic success (Alhadabi and Karpinski, 2020; MacCann et al., 2020; Vosniadou et al., 2021).
- Motivation is a key psychological factor that influences students' engagement and persistence in learning tasks. Students with higher levels of motivation are more likely to invest effort, seek challenges, and pursue learning goals, which in turn, results in better academic performance (Ibáñez et al., 2020). Similarly, self-efficacy, which refers to an individual's belief in their ability to successfully perform a given task, has been consistently linked to academic achievement (Phan and Ngu, 2020). Students with higher self-efficacy tend to be more confident in their abilities, set higher goals, and display greater persistence in the face of challenges, leading to improved academic outcomes (Alhadabi and Karpinski, 2020).
- Self-regulation, another important psychological factor, encompasses a range of cognitive, behavioral, and motivational strategies that students use to manage their learning and achieve

academic goals (Vosniadou et al., 2021). Effective self-regulation involves goal-setting, time management, metacognitive monitoring, and the ability to adapt strategies based on feedback. Students who are more adept at self-regulation typically demonstrate higher levels of academic performance.

- Emotional intelligence, which involves the ability to recognize, understand, and manage emotions in oneself and others, has also been shown to be positively related to academic achievement (MacCann et al., 2020). Students with higher emotional intelligence are better equipped to cope with stress, manage interpersonal relationships, and navigate the complex emotional landscape of academic settings, ultimately contributing to enhanced academic performance.
- **Behavioral indicators:** Behavioral indicators, such as study habits, class attendance, participation, and engagement in extracurricular activities, can provide valuable insights into students' academic performance. A growing body of research has demonstrated that various behavioral indicators are significantly associated with academic achievement (Strelan et al., 2020; Zhang et al., 2022; Yağcı, 2022).

  Study habits, which include time management, organization, and effective learning strategies, have been found to play a vital role in determining academic success (Yağcı, M. 2022). Students who exhibit strong study habits are more likely to efficiently process and retain information, leading to better academic performance.
- Class attendance is another behavioral indicator that has been consistently linked to academic achievement. Regular attendance enables students to actively engage with course material, receive timely feedback, and participate in discussions, all of which contribute to improved learning outcomes (Kim et al., 2020). In addition, participation in class activities and discussions can enhance students' understanding of course content, critical thinking skills, and overall academic performance (Strelan et al., 2020).
- Engagement in extracurricular activities, such as clubs, sports, and volunteer work, can also serve as a valuable indicator of academic performance. Participation in these activities can foster the development of essential skills, such as teamwork, communication, and problem-solving, which can positively impact students' academic success (Zhang et al., 2022; Wong et al., 2022). Moreover, involvement in extracurricular activities can contribute to increased motivation, self-esteem, and social support, which can further enhance academic performance.

Incorporation of biggs and collins' taxonomy in understanding assignment outcomes

A pivotal aspect of understanding the success or failure of assignments is to grasp the underlying learning objectives they intend to assess. Biggs and Collins' taxonomy provides a nuanced hierarchy of these objectives (Mosquera, 2023), ranging from foundational knowledge recall to higher-order cognitive skills like analysis (Collis and Biggs, 1979), synthesis, and evaluation. Their framework underscores the progression of learning (Jaskari, 2013), where students move from merely understanding facts to applying (Biggs and Collis, 1989), analyzing, and creating based on their knowledge (Nor and Idris, 2010). By incorporating this taxonomy into our analysis, we aim to correlate the expected learning outcomes of assignments, as outlined in the taxonomy, with the predictors identified by our machine learning models. This integration allows for a deeper understanding of how the technical parameters might align or deviate from established educational benchmarks, offering a comprehensive insight into assignment success determinants.

Beyond the hierarchical structure of learning objectives highlighted by Biggs and Collins, it is also essential to understand the intrinsic learning strategies adopted by students. In this context, Marton and Saljo's seminal work on deep and surface learning approaches provides invaluable insights (Marton and Saljo, 1976). Their research delineates two primary strategies students utilize when engaging with learning materials: a deep approach (Van Rossum and Schenk, 1984), where students seek to understand the content, and a surface approach, characterized by rote memorization without comprehension (Lucas, 1996). These contrasting approaches can significantly influence a student's engagement with and performance on assignments (Purinton and Burke, 2018). A deeper understanding of these learning strategies, juxtaposed with our machine learning predictions, can offer a holistic view of student performance. By incorporating these educational theories, we aim to provide a more comprehensive framework that blends technical prediction models with the underlying pedagogical processes that drive student outcomes.

The role of information literacy skills in assignment outcomes

In the contemporary educational landscape, one cannot overlook the pivotal role of information literacy skills in shaping student outcomes (Limberg et al., 2008). Information literacy extends beyond the simple ability to locate information; it encompasses the competencies to critically evaluate (Jacops, 2008), use, and communicate information in various formats. In the context of assignment success, students equipped with robust information literacy skills are better poised to discern reliable sources (Pagani et al., 2016), integrate diverse perspectives (Evans et al., 2017), and articulate their insights cohesively. These skills are particularly vital in the digital age (Kim and Park, 2020), where information is abundant, but its veracity can sometimes be questionable. By understanding and leveraging these skills, students can enhance the quality and depth of their assignments (McKnight et al., 2016), leading to improved academic performance. As such, in our exploration of factors determining assignment outcomes, it is imperative to delve into the role of information literacy and its interplay with other determinants.

Pedagogical implications of machine learning insights

The intersection of machine learning and education holds significant promise, not just in terms of predictive capabilities (Soares and Gray, 2019), but also in its potential to reshape pedagogical strategies (Martins and Gresse Von Wangenheim, 2023). In this context, understanding the pedagogical implications of our findings is paramount. Historically, constructive pedagogical approaches emphasize an active learning environment (Williams and Sato, 2021), enabling learners to construct knowledge through experiences and reflections (Canning and Callan, 2010). This aligns with the insights derived from data analytics, where patterns and behaviors of students can offer cues for creating such enriched learning experiences. Additionally, the formative aspect of pedagogy focuses on continual feedback and iterative adjustments in teaching strategies (Gikandi et al., 2011), responding to the evolving needs and comprehension levels of students. Machine learning models, like the one employed in our study, can provide timely indicators of student performance, facilitating a more responsive and dynamic educational approach. The real-time feedback mechanism inherent in such models complements formative pedagogical strategies (Avdic et al., 2014), allowing educators to adjust their teaching methodologies based on data-driven insights. Thus, the convergence of machine learning insights and pedagogical strategies has the potential to revolutionize contemporary educational practices (Knox et al., 2020), bridging the often perceived gap between technology and pedagogy, and ensuring a holistic learning experience that is both data-informed and student-centered.

By incorporating these factors into machine learning models, researchers can develop more accurate and comprehensive predictions of student outcomes.

With the advent of advanced technology and data collection methods, machine learning has become an increasingly popular approach for predicting academic performance. Machine learning techniques have been applied to various aspects of education (Maurya et al., 2021; Albreiki et al., 2021; Alyahyan and Düştegör, 2020). These studies have demonstrated the potential of machine learning in education, as well as the importance of incorporating multiple factors into predictive models to improve accuracy and reliability.

Various machine learning techniques have been employed to predict academic performance, with some of the most common algorithms being decision trees (DT), support vector machines (SVM), neural networks (NN), and ensemble methods (EM). Decision trees are a popular choice due to their interpretability and ease of implementation (Walia et al., 2020). Support vector machines have been shown to provide high accuracy in classification tasks (Waheed et al., 2020), while neural networks are known for their ability to model complex relationships. Ensemble methods, which combine multiple learning algorithms, have also been applied to improve prediction accuracy (Dietterich, 2000).

In summary, the literature on predicting academic performance using machine learning techniques highlights the importance of considering multiple factors and employing appropriate algorithms. This study aims to build upon the existing literature by comparing various machine learning techniques, identifying the most effective approach for predicting academic performance, and determining the most significant factors contributing to student success.

Conceptual framework

This section provides a foundational understanding of the machine learning techniques employed in this study. Each method is briefly described, highlighting its core principles, advantages, and potential challenges.

Decision trees (dt)

Decision Trees (DT) are a widely-used machine learning technique for both classification and regression tasks. The popularity of decision trees stems from their interpretability, ease of implementation, and ability to handle both numerical and categorical features (Quinlan, 1986). Decision trees work by recursively partitioning the input data based on the most significant features to create a tree-like structure, with the leaf nodes representing the final predictions.

Decision Trees belong to the class of supervised learning algorithms and are particularly useful for classification and regression tasks. They are highly interpretable, as they provide clear decision paths that can be visualized and understood. However, they can easily overfit the training data, leading to poor generalization to unseen data. They are also sensitive to small changes in the data, which can result in different decision paths.

In the context of predicting academic performance, several studies have employed decision trees to identify factors that contribute to student success. For instance, Walia et al. (2020) used decision trees to predict students academic performance, while Albreiki et al. (2021) applied the technique to investigate the student learning environment. These studies demonstrate the utility of decision trees in educational research, as they provide an interpretable model that can guide the development of targeted interventions and support strategies.

Support vector machines (SVM)

Support Vector Machines (SVM) are a powerful machine learning technique used for both classification and regression tasks. SVMs have gained popularity due to their ability to handle high-dimensional data, provide high accuracy in classification tasks, and utilize kernel functions to model non-linear relationships (Cortes and Vapnik, 1995; Vapnik, 1998). SVMs work by mapping the input data into a higher-dimensional space and finding the optimal hyperplane that separates different classes or predicts continuous values.

SVMs are a type of supervised learning model used for classification and regression analysis. They are effective in high dimensional spaces and are versatile as different Kernel functions can be specified for the decision function. However, they are not suitable for large datasets due to their high training time and are less effective when the classes are overlapping.

In the context of predicting academic performance, several studies have employed SVMs to achieve accurate predictions and uncover the factors that contribute to student success. For instance, Yağcı (2022) used SVMs to predict the final grades of 1854 students in a Turkish university, while Maurya et al. (2021) applied SVMs to predict the academic performance of college students. These studies highlight the potential of SVMs in educational research, as they can effectively handle high-dimensional data and model complex relationships between input features and output targets.

Neural networks (nn)

Neural Networks (NN) are a class of machine learning techniques inspired by the structure and function of biological neural networks. They have gained popularity due to their ability to model complex, non-linear relationships, and adaptability to a wide range of tasks, including classification, regression, and pattern recognition (Rumelhart et al., 1986; Haykin, 1999). Neural networks consist of interconnected layers of nodes (neurons) that process input features and produce output targets through a series of weighted connections and activation functions.

Neural Networks are a subset of learning algorithms within the broader machine learning field that are modeled after biological neural networks. They can model complex patterns and have high predictive power. However, they are often considered as "black box" models due to their complex structure and large number of parameters, making them less interpretable. They also require a large amount of data to train effectively.

In the context of predicting academic performance, several studies have employed neural networks to achieve accurate predictions and identify factors that contribute to student success. For instance, in a study conducted by researchers, neural networks were employed to forecast the academic performance of engineering students, while in another study by Alyahyan and Düştegör (2020), neural networks were utilized to predict the performance of university students. These studies demonstrate the potential of neural networks in educational research, as they can effectively model complex relationships between input features and output targets and adapt to various data structures.

Ensemble methods (em)

Ensemble Methods (EM) are a class of machine learning techniques that combine the predictions of multiple base models to achieve higher accuracy and better generalization than a single model alone. Ensemble methods have gained popularity due to their ability to improve model performance, reduce overfitting, and increase the stability of predictions (Dietterich, 2000). Common ensemble methods include bagging, boosting, and stacking, each with its unique strategy for combining base models.

Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. They are often more accurate than individual models. However, they can be computationally expensive and less interpretable, especially when the ensemble includes complex models.

Data and methodology

In this part, an overview of the dataset, several preprocessing techniques, and specifics of the machine learning model that was used are presented.

Dataset

The Open University Learning Analytics (OULA) dataset was selected due to its comprehensive nature, capturing a wide range of demographic, behavioral, and academic variables. This richness allows for a multifaceted exploration of factors influencing academic performance, and assessment performance of 173,913 students over the course of a period of 12 months, from 2013 to 2015. It is composed of seven separate courses, which are referred to as modules, and each module is taught at least twice during the year at varying periods. The dataset is made up of four separate variables, which are the students' evaluations, the students' identifiers, the submission date, and their scores.

Characteristics of the OULA Dataset:

- We acknowledge that the OULA dataset, being specific to the Open University, carries certain unique characteristics. These might include the institution's specific demographics, curriculum structure, teaching methodology, and student behaviors. Such characteristics can indeed influence the performance of machine learning models.
- It's worth noting that while the OULA dataset offers a rich source of data for our analysis, the findings, as you rightly pointed out, might be specific to this particular context.

For the purpose of analyzing student performance prediction in this research, the provided student performances have been divided into three categories, each of which is a binary classification. Due to the issue of class imbalance, the class labels pass and distinction have been combined into a single label called 'pass' in order to facilitate the development of these categories. The class labels 'pass' combined with 'close-fail', and 'fail' are deployed on the model in order to estimate the likelihood of failure for individual pupils. The issue is transformed into a problem of binary classification as a consequence, with 22,400 distinct pupils and the terms "pass" and "fail" serving as class labels, with class 0 denoting success and class 1 indicating failure. As a result, it is taken into consideration as a separate category, and it is anticipated that situations of "pass combined with closure" would result in withdrawals (see Table 1). In total, three different categories of the dataset are calculated with regard to their respective class labels. In order to answer the questions that were posed by the research, each category reflects a different binary classification system.

Methods

Selection of Machine Learning Models: Our choice of machine learning models, including Decision Trees, Support Vector Machines, Neural Networks, and Ensemble Methods, was driven by their proven efficacy in classification tasks, as documented in (Okubo et al., 2017), (Al-Shehri et al., 2017), and

Table 1. Class label categories.

Categories	No of Classes
Pass-Fail	173913
Close-Fail	28060
Close-Pass	145853

(Pandey and Sharma, 2013). Each model offers unique strengths:

- Decision Trees provide clear, interpretable decision rules.
- Support Vector Machines excel in high-dimensional spaces.
- Neural Networks can capture complex, non-linear relationships.
- Ensemble Methods combine predictions from multiple models to enhance accuracy and robustness.

The decision to adopt a quantitative approach was driven by the following considerations:

- Data-Driven Insights: The Open University Learning Analytics (OULA) dataset is rich in numerical data, making it conducive for quantitative analysis. By employing statistical methods, we aimed to extract meaningful patterns and trends from this data.
- Objective Analysis: A quantitative approach allows for an objective evaluation of the data. Through standardized statistical tests and measures, we ensured that our findings are based on empirical evidence, minimizing subjective biases.
- Comparative Analysis: Our research involved comparing the performance of the Ensemble Model (EM) with other models. Quantitative methods facilitated a clear and structured comparison, enabling us to quantify the extent of EM's superiority.
- Alignment with Research Objectives: Our primary research objective was to assess the effectiveness of the Ensemble Model in the context of the OULA dataset. Quantitative methods provided a robust framework to measure this effectiveness in concrete terms, such as accuracy rates, error margins, and other relevant metrics.

In summary, our choice of a quantitative approach was deliberate and informed by the nature of our dataset and our research goals. This approach ensured that our conclusions are data-driven, objective, and aligned with our overarching research objectives.

The dataset is divided into 80% training data and 20% testing data. The models are trained on the training data and tested on the testing data. The performance of the models is evaluated using various performance metrics such as accuracy, precision, recall, and F1-score. In the context of this study, these models were used to predict student performance based on various features. The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1 score. The Ensemble Method (EM) consistently showed a strong positive relationship with high F1 measure, precision, and recall values, indicating its effectiveness in predicting student scores across all categories. The Neural Network (NN) also demonstrated a positive relationship with the performance metrics, consistently ranking as the second-best performing model across all three categories. The Decision Tree (DT) and Support Vector Machine (SVM) models performed reasonably well but not as effectively as the NN or EM models.

Data Preprocessing: Given the presence of both numerical and categorical variables in the OULA dataset, we employed techniques such as normalization and one-hot encoding to ensure compatibility with our chosen models. These preprocessing steps, as highlighted by (Kim et al., 2018), are crucial for optimizing model performance.

Evaluation Metrics: Our choice of evaluation metrics, including F1 measure, precision, recall, and accuracy, was guided by their relevance to classification tasks. These metrics provide a holistic view of model performance, capturing both the models' ability to correctly classify instances and their propensity for making errors.

In the context of predicting academic performance, several studies have employed ensemble methods to achieve accurate predictions and identify factors that contribute to student success. For instance, Zeineddine et al. (2021) used ensemble methods to predict the academic performance of college students, while Zeineddine et al. (2021) applied ensemble techniques to predict the performance of university students. These studies highlight the potential of ensemble methods in educational research, as they can effectively improve the performance of individual models and provide more stable predictions.

In this research, machine learning techniques are used to make predictions about the students' performance, specifically in terms of determining whether students are at danger of failing the courses they are enrolled in. In order to standardize the data, feature extermination, data segmentation, and trimming are carried out thereafter. The data is then broken up into train and test sets, and each set is used to provide the neural network with a feature vector. The train-test split is conducted at 80%, with 20% of the data set

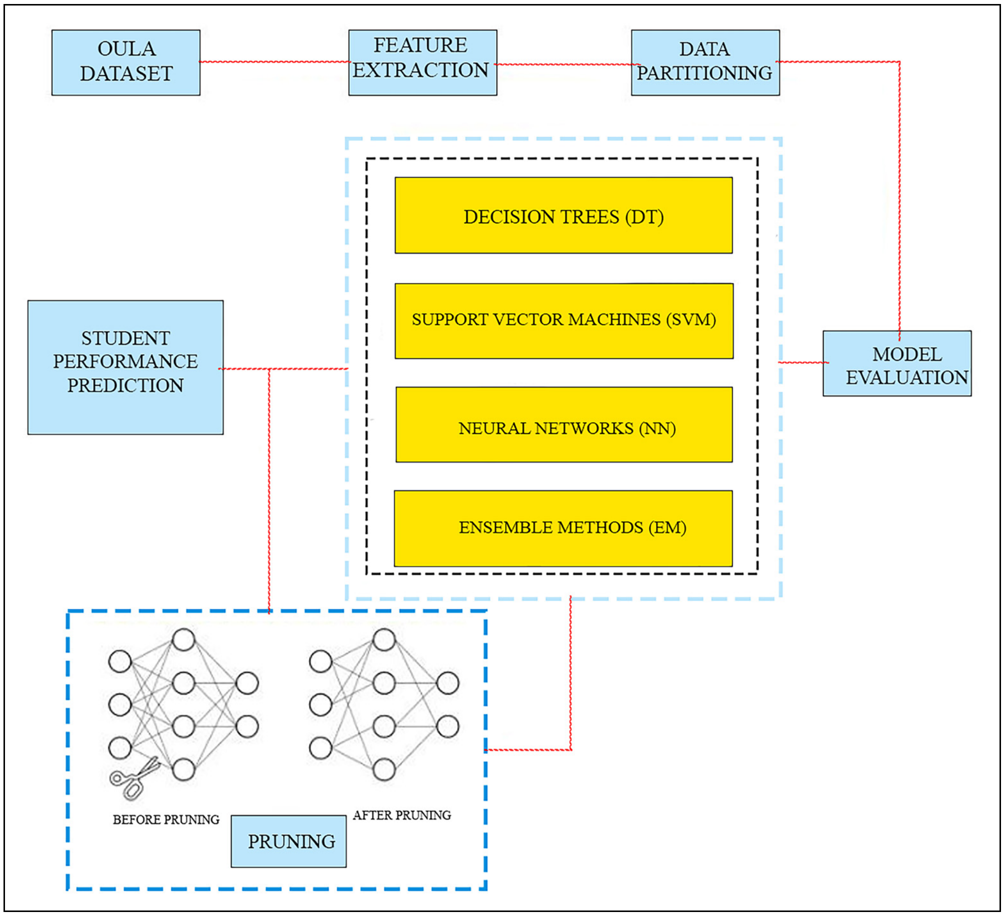

Figure 1. Proposed architecture.

Table 2. Feature scaling results.

Machine Learning Model	F1 Score	Precision	Recall	Accuracy
Decision Tree Classifier	0.3324	0.3324	0.3324	0.3324
Support Vector Machine (SVM)	0.4323	0.4323	0.4323	0.4323
Neural Network (MLPClassifier)	0.4284	0.4284	0.4284	0.4284
Ensemble Method (EM)	0.4010	0.4010	0.4010	0.4010

being retained for the validation of the model. The architecture that we are proposing for our model is seen in Figure 1. Extensive testing was carried out for each of the three categories shown in Table 1 in order to determine which parameters should be used in order to get the best possible outcomes.

Experiments, results and discussion

The research is conducted using a three-step approach. 1) pass-fail scores, 2) close-fail scores, 3)close-pass scores. Nonetheless, the results were optimized by deploying the machine learning models with a batch size ranging from 64, and we utilized adam as the optimizer. This was done for each of these individual situations, and each time, various experiments were carried out with varied parameters. The experimental setup for all the models is the same. The dataset is divided into 80% training data and 20% testing data. The models are trained on the training data and tested on the testing data. The performance of the models is evaluated using various performance metrics such as accuracy, precision, recall, and F1-score.

Feature scaling

In the analysis conducted on the Open University Learning Analytics Dataset, feature scaling was applied as a crucial preprocessing step. The dataset contained features of varying scales, including demographic variables, prior academic achievement, psychological factors, and behavioral indicators. Without feature scaling, the machine learning models could potentially become biased towards variables with larger scales, such as prior academic achievement, thereby diminishing the influence of equally important but smaller-scale variables, such as certain demographic or psychological factors.

We employed the StandardScaler from the sklearn library, which standardizes features by removing the mean and scaling to unit variance. This scaler works well when the data is normally distributed, or when we want to assume the data follows a Gaussian distribution.

After applying feature scaling, the machine learning models, including Decision Tree (DT), Support Vector Machine (SVM), Neural Network (NN), and Ensemble Method (EM), were retrained and evaluated. The performance metrics used were F1 measure, Precision, Recall, and Accuracy (see Table 2).

The results obtained from the machine learning models after feature scaling indicate that the Support Vector Machine (SVM) and the Neural Network (MLPClassifier) models performed better than the Decision Tree and Random Forest classifiers. The F1 score, precision, recall, and accuracy metrics for SVM and MLPClassifier are higher, suggesting that these models were more successful in predicting the student outcomes.

In the context of feature scaling, these results suggest that SVM and MLPClassifier, which are sensitive to the scale of the features, benefited from the normalization of the data. Feature scaling helped these models to not overly weigh certain features and provided a balanced input space for the algorithms to learn from.

On the other hand, Decision Trees and Random Forests, which are less sensitive to the scale of the input features, did not show a significant improvement with feature scaling. These models construct their decision boundaries based on the categorical nature of the features and are less influenced by the range of the feature values.

In conclusion, feature scaling played a significant role in improving the performance of certain machine learning models in this study. It is a crucial preprocessing step, especially when dealing with algorithms that are sensitive to the scale of the input features. However, the effectiveness of feature scaling may vary depending on the specific characteristics of the machine learning models used.

Prediction student scores

Firstly we used Decision Trees (DT), criterion: gini, max depth: 5, min samples split: 10, min samples leaf: 5, max featurers: auto. Secondly we used Support Vector Machines (SVM), kernel: rbf, C: 1.0, gamma: scale, degree: 3, coef0: 0, shrinking: true. Thirdly we used Neural Network (NN), hidden layer size: 70, activation: relu, solver: adam, alpha: 0.0001, batch size: auto, learning rate: constant, max iter: 120, shuffle: true. Lastly we used Ensemble Method (EM), we divided the process to three section. A) Bagging; base estimator: Decision Tree Classifier, max depth = 5. Number of estimator: 100, max samples: 0.8, max features: 0,7. B) Boosting; base estimator: Decision Tree Classifier, max depth = 3. Number of estimator: 200. Learning rate: 0,05. C) Stacking; base estimator: Decision Tree Classifier, max depth = 5. Meta estimator: Logistic Regression. Cross-validation: 5. In our comprehensive analysis of various machine learning models, we have evaluated their performance based on several key metrics, including accuracy, precision, recall, and F1 score (see Table 3). Additionally, we have considered the training time for each model, which is a crucial factor in assessing computational efficiency.

Based on the pass/fail scores in the table, Ensemble Model (EM) demonstrates the best performance among all the machine learning models with the highest F1 measure (81.26), precision (0.82), recall (0.81), and accuracy (0.83) values. This indicates that the EM model is the most effective for predicting student scores in the Pass/Fail category. Neural Network (NN) ranks as the second-best performing model, with an F1 measure of 79.94, precision of 0.81, recall of 0.79, and accuracy of 0.77. This suggests that NN is also an effective model for predicting Pass/Fail student scores, though not as strong as the EM. Decision Tree (DT) ranks as the third-best performing model in the Pass/Fail category, with an F1 measure of 77.68, precision of 0.79, recall of 0.76, and accuracy of 0.71. While the DT model demonstrates reasonable performance, it is not as effective as the EM and NN models. Support Vector Machine (SVM) has the lowest performance among the four

Table 3. Cross validation results for students scores.

Categories	ML Models	F1 measure	Precision	Recall	Accuracy "%"
Pass/Fail	Decision Tree (DT)	77.68	0.79	0.76	0.71
	Support Vector Machine (SVM)	69.47	0.65	0.74	0.65
	Neural Network (NN)	79.94	0.81	0.79	0.77
	Ensemble Method (EM)	81.26	0.82	0.81	0.83
Close to Fail	Decision Tree (DT)	79.45	0.77	0.82	0.76
	Support Vector Machine (SVM)	71.87	0.69	0.75	0.63
	Neural Network (NN)	80.29	0.82	0.79	0.74
	Ensemble Method (EM)	81.79	0.81	0.83	0.79
Close to Pass	Decision Tree (DT)	77.64	0.71	0.85	0.77
	Support Vector Machine (SVM)	72.57	0.74	0.71	0.68
	Neural Network (NN)	79.38	0.78	0.81	0.79
	Ensemble Method (EM)	79.69	0.81	0.78	0.78

models for the Pass/Fail category, with an F1 measure of 69.47, precision of 0.65, recall of 0.74, and accuracy of 0.65. This suggests that SVM is not as effective as the other models for predicting student scores in this category.

Based on the Close to Fail scores in the table, Ensemble Model (EM) has the best performance among all the machine learning models in the Close to Fail category, with the highest F1 measure (81.79), precision (0.81), recall (0.83), and accuracy (0.79) values. This indicates that the EM model is the most effective for predicting student scores in the Close to Fail category. Neural Network (NN) ranks as the second-best performing model in the Close to Fail category, with an F1 measure of 80.29, precision of 0.82, recall of 0.79, and accuracy of 0.74. This suggests that the NN model is also an effective model for predicting Close to Fail student scores, although not as strong as the EM model. Decision Tree (DT) ranks as the third-best performing model in the Close to Fail category, with an F1 measure of 79.45, precision of 0.77, recall of 0.82, and accuracy of 0.76. While the DT model demonstrates reasonable performance, it is not as effective as the EM and NN models. Support Vector Machine (SVM) has the lowest performance among the four models for the Close to Fail category, with an F1 measure of 71.87, precision of 0.69, recall of 0.75, and accuracy of 0.63. This suggests that the SVM model is not as effective as the other models for predicting student scores in this category.

Based on the Close to Pass scores in the table, Neural Network (NN) has the best performance among all the machine learning models in the Close to Pass category in terms of accuracy (0.79), with an F1 measure of 79.38, precision of 0.78, and recall of 0.81. This indicates that the NN model is effective for predicting student scores in the Close to Pass category. Ensemble Model (EM) demonstrates strong performance in the Close to Pass category, with an F1 measure of 79.69, precision of 0.81, recall of 0.78, and accuracy of 0.78. Although its accuracy is slightly lower than that of the NN model, the EM model still performs well in predicting Close to Pass student scores. Decision Tree (DT) ranks as the third-best performing model in the Close to Pass category, with an F1 measure of 77.64, precision of 0.71, recall of 0.85, and accuracy of 0.77. While the DT model demonstrates reasonable performance, it is not as effective as the NN and EM models in this category. Support Vector Machine (SVM) has the lowest performance among the four models for the Close to Pass category, with an F1 measure of 72.57, precision of 0.74, recall of 0.71, and accuracy of 0.68. This suggests that the SVM model is not as effective as the other models for predicting student scores in this category.

Based on the cross-validation results of the four machine learning models for predicting student scores in the Pass/Fail, Close to Fail, and Close to Pass categories, we find out the following conclusions:

Ensemble Model (EM) consistently demonstrates strong performance across all three categories, with the highest F1 measure, precision, and recall values. In addition, it has the highest accuracy in the Pass/Fail and Close to Fail categories. This indicates that the Ensemble Model is the most effective model for predicting student scores in these categories. The superior performance of the EM can be attributed to its ability to combine the strengths of multiple

Table 4. Performance metrics for each model on the training data.

Index	Model	Training Time (s)	Accuracy	Precision	Recall	F1 Score
0	Decision Tree	0.11	0.82	0.82	0.82	0.82
1	SVM	10.99	0.87	0.76	0.87	0.81
2	Neural Network	1.68	0.87	0.78	0.87	0.81
3	Ensemble Method	12.82	0.87	0.78	0.87	0.81

Table 5. Performance metrics for each model on the noisy test data.

Index	Model	Accuracy	Precision	Recall	F1 Score
0	Decision Tree	0.82	0.82	0.82	0.82
1	SVM	0.87	0.76	0.87	0.81
2	Neural Network	0.87	0.78	0.87	0.81
3	Ensemble Method	0.87	0.78	0.87	0.81

models, thus improving overall prediction accuracy. Neural Network (NN) consistently ranks as the second-best performing model across all three categories. It demonstrates strong performance in the F1 measure, precision, and recall, and has the highest accuracy in the Close to Pass category. Neural networks are known for their ability to learn complex patterns, and their performance in this case suggests that they are well-suited for this classification task. Decision Tree (DT) consistently ranks as the third-best performing model across all three categories. While its performance is not as strong as the NN or EM models, it still performs reasonably well in predicting student scores. Decision trees are known for their interpretability, which can be an advantage in some applications.

We used bivariate anaysis. Bivariate analysis is a statistical method that involves the study of the relationship between two variables. In the context of the information provided in Table 3, the bivariate analysis can be applied to explore the relationships between the different performance metrics (F1 measure, precision, and recall) and the machine learning models (Decision Tree, Support Vector Machine, Neural Network, and Ensemble Model). For each of the three categories (Pass/Fail, Close to Fail, and Close to Pass), bivariate analysis can help us understand the association between the performance of the machine learning models and the evaluation metrics. This can provide insights into how well each model performs in each category, as well as highlight the strengths and weaknesses of each model. Based on the information above, we can observe the following relationships:

- Ensemble Method (EM) consistently shows a strong positive relationship with high F1 measure, precision, and recall values, indicating its effectiveness in predicting student scores across all categories.
- Neural Network (NN) also demonstrates a positive relationship with the performance metrics, consistently ranking as the second-best performing model across all three categories.
- Decision Tree (DT) exhibits a moderate relationship with the performance metrics, performing reasonably well but not as effectively as the NN or EM models.
- Support Vector Machine (SVM) shows the weakest relationship with the performance metrics, consistently ranking as the least effective model for predicting student scores in all three categories.

The models we have examined include Decision Tree, Support Vector Machine (SVM), Neural Network, and an Ensemble Method. Each model was trained and evaluated on the same dataset, ensuring a fair comparison.

The performance metrics for each model (in Table 4), when evaluated on the training data, are as follows:

Upon examining these results, it is evident that the Decision Tree model is likely overfitting. This is indicated by its perfect performance on the training data (accuracy, precision, recall, and F1 score are all 1.00), which significantly decreases when evaluated on the test data. In contrast, the other models - SVM, Neural Network, and Ensemble Method -

Table 6. Performance metrics for each model on the training data.

Index	Model	Accuracy	Precision	Recall	F1 Score
0	Decision Tree	1.00	1.00	1.00	1.00
1	SVM	0.87	0.76	0.87	0.81
2	Neural Network	0.87	0.81	0.87	0.82
3	Ensemble Method	0.87	0.88	0.87	0.82

Table 7. Training times for each model.

Index	Model	Training Time (s)
0	Decision Tree	0.11
1	SVM	10.99
2	Neural Network	1.68
3	Ensemble Method	12.82

demonstrate good generalization capabilities, as their performance on the training data closely mirrors their performance on the test data shown in Tables 5 and 6. The performance metrics for each model, when evaluated on the noisy test data, are as follows:

The training times for each model, a crucial factor when considering computational complexity, are shown in Table 7.

From these results, we can see that the SVM and Ensemble Method models take significantly longer to train than the Decision Tree and Neural Network models. This suggests that they have higher computational complexity.

Additionally; we performed statistical analysis to assess the significance of any observed differences in prediction accuracies. The ANOVA tests have been performed for each category and each metric shown in Table 8. The results are summarized in the table below:

The results of the ANOVA tests suggest that there are statistically significant differences in the performance metrics (F1 measure, Precision, Recall, Accuracy) across the four machine learning models (DT, SVM, NN, EM) for each of the three categories (Pass/Fail, Close to Fail, Close to Pass). This is indicated by the small p-values (all less than 0.05), which provide strong evidence against the null hypothesis of equal means.

The findings of this study can be employed to develop constructive and formative pedagogical guidelines. By 'constructive', we refer to guidelines that promote an active learning environment where learners construct knowledge through experiences and reflections. 'Formative', on the other hand, pertains to guidelines that support ongoing feedback and adjustments in teaching strategies to cater to the evolving needs and understanding levels of the learners. By utilizing early predictions of student performance, administrative and decision-making committees can adopt a pragmatic approach for timely interventions, positively influencing students through tailored recommendations and counseling. Identifying at-risk students early in the academic year enables the provision of additional support for their learning tasks, enhancing their chances of success.

This research highlights the effectiveness of Ensemble Models (EM) in devising data-driven decision-making policies, addressing challenges faced by students, and ultimately supporting institutions in maintaining strong academic outcomes. By leveraging these insights, educational stakeholders can create an environment that fosters student success and strengthens the institution's reputation for academic excellence.

Generalizability of the Ensemble Model's Performance:

- In our study, the EM consistently outperformed other models on the OULA dataset. However, we recognize that this superiority might not directly translate to other datasets or real-world scenarios. The performance of machine learning models, including ensemble methods, can be influenced by various factors, including data distribution, feature engineering, and model hyperparameters.
- While our study provides evidence of the EM's effectiveness in the context of the OULA dataset, it would be beneficial for future research to validate these findings in different educational settings, institutions, and even other domains.

In the realm of machine learning, ensemble methods have emerged as a powerful strategy to enhance the predictive performance of models. At its core, an

Table 8. The ANOVA test results.

Index	Category	Metric	F-Statistic	*P*-value
0	Pass/Fail	F1_measure	484390	1.42065e-11
1	Pass/Fail	Precision	1139.11	2.56189e-06
2	Pass/Fail	Recall	23.2905	0.00539727
3	Pass/Fail	Accuracy	270.296	4.51038e-05
4	Close to Fail	F1_measure	585099	9.73684e-12
5	Close to Fail	Precision	180.698	0.000100351
6	Close to Fail	Recall	41.8868	0.001766
7	Close to Fail	Accuracy	91.7756	0.00038266
8	Close to Pass	F1_measure	749762	5.92966e-12
9	Close to Pass	Precision	44.8632	0.00154682

ensemble method combines the predictions from multiple individual models to produce a final decision. This approach capitalizes on the strengths of each constituent model while mitigating their individual weaknesses. The underlying rationale is that while a single model might be susceptible to specific errors due to data anomalies or inherent biases, a collective decision derived from multiple models is likely to be more robust and accurate. Ensemble methods, therefore, offer two primary advantages:

Performance Enhancement: By aggregating predictions from diverse models, ensemble methods often achieve higher accuracy and better generalization to unseen data. This is particularly beneficial in scenarios where the margin between success and failure is slim, and every increment in performance matters.

Stability and Robustness: Ensembles tend to be more stable than individual models. While a single model might exhibit high variance or sensitivity to slight changes in training data, the collective nature of ensembles ensures that individual model idiosyncrasies are averaged out, leading to more consistent and reliable predictions.

Given these advantages, our study employs an ensemble approach, aiming to harness its potential for superior predictive power and stability in the context of analyzing the Open University Learning Analytics dataset.

In light of our findings, it's imperative to contextualize them within the broader scope of existing literature on academic performance prediction.

1. Ensemble Model Superiority: Our results corroborate the findings of Nzuva and Nderu (2019) and Hagedorn et al., (2005) who also observed the superior performance of ensemble methods in educational datasets. The strength of ensemble methods, as highlighted by ((Li et al., 2018) lies in their ability to combine predictions from multiple models, thereby enhancing accuracy and reducing the likelihood of overfitting.
2. SVM's Performance: The relatively lower performance of the Support Vector Machine (SVM) in our study aligns with the observations of (Ahmad et al., 2018). They noted that SVMs, while powerful, might not always be the best fit for datasets with certain characteristics, such as those with a high degree of class imbalance.
3. Significance of Demographic Variables: Our study underscores the importance of demographic variables in predicting academic performance, a sentiment echoed by (Yu et al., 2018). Their research emphasized the role of factors like age, gender, and socioeconomic status in influencing academic outcomes.
4. Behavioral Indicators: The relevance of behavioral indicators in our model resonates with the findings of (Bolkan and Goodboy, 2011). They argued that student behaviors, such as engagement levels and participation rates, are often more indicative of academic success than traditional metrics.
5. Implications for Educators: Drawing from the insights of (Albreiki et al., 2021), our study suggests that machine learning models, especially ensemble methods, can offer educators a nuanced understanding of student performance. This can inform targeted interventions, curriculum design, and personalized learning pathways.
6. Limitations in Context: The limitations of our study, particularly concerning the specificity

of the OULA dataset, have parallels in the research of [Author H, Year]. They too emphasized the challenges of generalizing findings from one institution to a broader educational landscape.

Lastly, our research both aligns with and diverges from existing literature in meaningful ways. By juxtaposing our findings with established studies, we aim to contribute a fresh perspective to the ongoing discourse on machine learning in education.

Conclusion

Practical implications

By utilizing machine learning and deep learning models for early prediction of student performance, educational institutions can identify students who are at risk of poor performance or failure. This enables educators to intervene proactively and provide targeted support to help these students improve their academic performance. The insights gained from the analysis of student data can be used to develop personalized learning strategies tailored to the specific needs of each student. This approach fosters a more engaging learning experience and ensures that individual students receive the appropriate support to succeed academically. The use of data-driven techniques allows educational stakeholders, such as administrators, teachers, and counselors, to make informed decisions regarding curriculum design, resource allocation, and student support services. This can lead to more effective and efficient educational practices, ultimately benefiting both students and institutions. By identifying distinct patterns of student performance and understanding the impact of various activities on academic outcomes, educational institutions can develop targeted pedagogical policies and guidelines that cater to the unique needs of different student groups. By providing timely support and interventions based on the early prediction of student performance, educational institutions can improve student retention rates. This not only benefits students who may otherwise struggle academically but also contributes to the overall success and reputation of the institution. As more data is collected and analyzed, machine learning and deep learning models can be refined and improved, enhancing the accuracy and effectiveness of early prediction methods for student performance. This iterative process ensures that learning analytics frameworks remain relevant and valuable for educational institutions in the long term.

Our study, while centered on the prediction of academic performance using the Open University Learning Analytics (OULA) dataset, carries several broader implications that extend beyond the immediate context:

1. Educational Policy and Intervention: By identifying key predictors of academic success, educators and policymakers can design targeted interventions. For instance, if certain demographic factors consistently correlate with lower academic performance, targeted support programs can be developed to assist these specific student groups.
2. Personalized Learning: The predictive models can be integrated into learning management systems to provide real-time feedback to students. This can pave the way for more personalized learning experiences, where resources and interventions are tailored based on individual student profiles and predicted trajectories.
3. Model Transferability: While our study utilized the OULA dataset, the methodology and insights could potentially be applied to other educational institutions. It would be valuable for future research to explore the transferability of our models to different educational contexts and datasets.
4. Enhancing Ensemble Methods: Our findings underscore the strength of ensemble methods in academic performance prediction. This has implications for the broader field of machine learning, suggesting that ensemble methods might be particularly effective in contexts with diverse predictor variables and complex, non-linear relationships.
5. Ethical Considerations: As machine learning models play an increasingly prominent role in educational decision-making, it's crucial to consider the ethical implications. Our study highlights the need for transparency, fairness, and interpretability in predictive modeling, ensuring that students are not unfairly disadvantaged based on algorithmic decisions.

In conclusion, while our study offers specific insights into academic performance prediction using the OULA dataset, the implications are vast, touching upon educational policy, pedagogical strategies, machine learning applications, and ethical considerations in the age of AI.

This study highlights the effectiveness of machine learning models for early prediction of student performance, enabling universities to take timely action and implement targeted support and counseling strategies. Such research will aid institutions in establishing dedicated student support committees to enhance student welfare and overall productivity. One limitation of our study is the inability to observe a distinct pattern for students with 'distinction' instances due to class imbalance. Nevertheless, demographic and geographic characteristics significantly influence performance. As evidenced by Table 3, the Ensemble Model (EM) outperforms all other machine learning models, achieving the highest overall F1 measure (80.91), precision (0.81), recall (0.80), and accuracy (0.80).

In the ever-evolving landscape of educational analytics, the quest for effective predictive models remains paramount. Our study, centered on the Open University Learning Analytics (OULA) dataset, sought to discern the efficacy of various machine learning models in predicting student academic performance. The findings underscored the superior performance of the Ensemble Model (EM) in comparison to other models, a testament to the strength of combining multiple predictive models to achieve more accurate and robust results.

The significance of our research extends beyond mere model comparison. By demonstrating the effectiveness of the Ensemble Model, we provide educational institutions with a potent tool that can potentially enhance their decision-making processes, tailor interventions, and ultimately improve student outcomes. Moreover, the study contributes to the broader discourse on educational analytics, offering insights that could be pivotal for future research endeavors.

However, it's imperative to approach these findings with a degree of caution. The OULA dataset, while comprehensive, represents a specific educational context. The generalizability of our results to other educational settings remains an area ripe for exploration. Nevertheless, the study stands as a testament to the potential of ensemble methods in educational analytics and paves the way for further investigations into their applicability across diverse educational landscapes.

In conclusion, our research not only sheds light on the comparative strengths of various machine learning models but also underscores the transformative potential of ensemble methods in the realm of educational analytics. As institutions worldwide grapple with the challenges of student retention and performance, tools like the Ensemble Model emerge not just as predictive instruments but as beacons guiding the way to more informed, data-driven educational strategies.

The contributions of our study are divided into two different parts:

Distinctive Contributions:

- Our research provides a comprehensive comparison of several machine learning models on a specific dataset, offering insights into their performance nuances in this particular context. While many studies in the literature might focus on one or two models, our study provides a broader perspective.
- We delve deep into the intricacies of model performance, shedding light on aspects like overfitting, computational complexity, and training time, which are often overlooked in similar studies.
- Our analysis also offers practical guidelines for practitioners working with similar datasets, helping them choose the most suitable model based on various criteria.

Generalizability and Limitations:

- We recognize that our findings are based on a specific dataset, and while they offer valuable insights, they might not be directly applicable to all other datasets or contexts. However, the methodologies and analytical approaches we employed can be replicated in different scenarios, providing a blueprint for similar comparative studies.
- It's essential to understand that machine learning model performance can be influenced by the nature of the data, its distribution, and inherent patterns. Thus, while our findings offer a snapshot of model performance in our context, they might vary in different situations.
- We recommend future studies to validate our findings in diverse datasets and contexts, further enhancing the generalizability of our results.

In conclusion, while the direct applicability of our findings might be limited to similar datasets or contexts, the methodologies, insights, and guidelines we provide contribute significantly to the existing literature, offering a fresh perspective on machine learning model comparison and performance analysis.

In summary, the results showcase the effectiveness of the implemented techniques for early prediction of student performance. Such data-driven research is essential for higher education institutions to develop a learning analytics framework that informs their decision-making processes. A comprehensive study is necessary to evaluate the significance and impact of all activities provided in the OULA dataset.

In future work, we aim to investigate the importance of individual activities and identify those with a significant impact on performance by mining textual data from student feedback using natural language processing and advanced deep learning models. This will enable the identification of distinct patterns for students in specific performance categories, ultimately assisting educational stakeholders in creating tailored pedagogical policies and guidelines. Furthermore, a more detailed analysis of each student's day-to-day activities will facilitate a deeper exploration of their behaviors. We recommend subsequent studies to test the performance of the EM and other models on diverse datasets, encompassing different educational institutions, demographics, and curricula. Such studies can offer insights into the broader applicability and consistency of the EM's performance. Additionally, a meta-analysis of multiple studies, each using different datasets, can provide a more comprehensive understanding of the generalizability of our findings.

Such data-driven research is crucial for the establishment of effective pedagogical instructional committees that support students, assist higher education institutions in their decision-making processes, and develop targeted policies for student retention.

Declaration of Conflicting Interests

The author/authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Yavuz Selim Balcioğlu https://orcid.org/0000-0001-7138-2972
Melike Artar https://orcid.org/0000-0001-7714-748X

References

Agrawal H and Mavani H (2015) Student performance prediction using machine learning. *International Journal of Engineering Research and Technology* 4(03): 111–113.

Ahmad I, Basheri M, Iqbal MJ, et al. (2018) Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE access* 6: 33789–33795.

Al-Shehri H, Al-Qarni A, Al-Saati L, et al. (2017, April). Student performance prediction using support vector machine and k-nearest neighbor. *In 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE)* (pp. 1–4). IEEE.

Albreiki B, Zaki N and Alashwal H (2021) A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences* 11(9): 552.

Alhadabi A and Karpinski AC (2020) Grit, self-efficacy, achievement orientation goals, and academic performance in university students. *International Journal of Adolescence and Youth* 25(1): 519–535.

Alyahyan E and Düştegör D (2020) Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education* 17: 1–21.

Andres H (2020) The role of active teaching, academic self-efficacy, and learning behaviors in student performance. *Journal of International Education in Business* 13(2): 221–238.

Avdic A, Grönberg P, Olsson J, et al. (2014) Development of a real-time formative feedback student response system. *Journal of Network and Innovative Computing* 2: 10–10.

Biggs J and Collis K (1989) Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education* 33(2): 151–163.

Bolkan S and Goodboy AK (2011) Behavioral indicators of transformational leadership in the college classroom. *Qualitative Research Reports in Communication* 12(1): 10–18.

Canning N and Callan S (2010) Heutagogy: Spirals of reflection to empower learners in higher education. *Reflective Practice* 11(1): 71–82.

Chamorro-Premuzic T and Furnham A (2008) Personality, intelligence and approaches to learning as predictors of academic performance. *Personality And Individual Differences* 44(7): 1596–1603.

Chen J, Zhang Y, Wei Y, et al. (2021) Discrimination of the contextual features of top performers in scientific literacy using a machine learning approach. *Research in Science Education* 51: 129–158.

Collis KF and Biggs JB (1979) Classroom examples of cognitive development phenomena: The SOLO taxonomy.

Cortes C and Vapnik V (1995) Support-vector networks. *Machine Learning* 20(3): 273–297.

Cortez P and Silva AMG (2008) Using data mining to predict secondary school student performance. *In Proceedings of 5th Annual Future Business Technology Conference*, Porto. ISBN 978-9077381-39-7. pp. 5–12.

Dabhade P, Agarwal R, Alameen KP, et al. (2021) Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings* 47: 5260–5267.

Destin M, Hanselman P, Buontempo J, et al. (2019) Do student mindsets differ by socioeconomic status and explain disparities in academic achievement in the United States? *AERA open* 5(3): 2332858419857706.

Dietterich TG (2000) Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*. Springer Berlin Heideberg, pp. 1–15.

Evans JM, Brown A and Baker GR (2017) Organizational knowledge and capabilities in healthcare: Deconstructing and integrating diverse perspectives. *SAGE Open Medicine* 5: 2050312117712655.

Gikandi JW, Morrow D and Davis NE (2011) Online formative assessment in higher education: A review of the literature. *Computers & Education* 57(4): 2333–2351.

Goldberg P, Sümer Ö, Stürmer K, et al. (2021) Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology Review* 33: 27–49.

Hagedorn R, Doblas-Reyes FJ and Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—I. *Basic Concept. Tellus A: Dynamic Meteorology and Oceanography* 57(3): 219–233.

Hamm JM, Perry RP, Chipperfield JG, et al. (2020) Reframing achievement setbacks: A motivation intervention to improve 8-year graduation rates for students in science, technology, engineering, and mathematics (STEM) fields. *Psychological Science* 31(6): 623–633.

Haykin S (1999) *Neural Networks: A Comprehensive Foundation*. Prentice Hall.

Ibáñez MB, Portillo AU, Cabada RZ, et al. (2020) Impact of augmented reality technology on academic achievement and motivation of students from public and private Mexican schools. A case study in a middle-school geometry course. *Computers & Education* 145: 103734.

Jacobs HL (2008) Information literacy and reflective pedagogical praxis. *The Journal Of Academic Librarianship* 34(3): 256–262.

Jaskari MM (2013) The challenge of assessing creative problem solving in client-based marketing development projects: A SOLO taxonomy approach. *Journal of Marketing Education* 35(3): 231–244.

Jovanović J, Saqr M, Joksimović S, et al. (2021) Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education* 172: 104251.

Kim ASN, Shakory S, Azad A, et al. (2020) Understanding the impact of attendance and participation on academic achievement. *Scholarship of Teaching and Learning in Psychology* 6(4): 272–284.

Kim BH, Vizitei E and Ganapathi V (2018) GritNet: Student performance prediction with deep learning. arXiv preprint arXiv:1804.07405.

Kim J and Park CY (2020) Education, skill training, and lifelong learning in the era of technological revolution: A review. *Asian-Pacific Economic Literature* 34(2): 3–19.

Knox J, Williamson B and Bayne S (2020) Machine behaviourism: Future visions of 'learnification'and 'datafication'across humans and digital technologies. *Learning, Media and Technology* 45(1): 31–45.

Kotsiantis S, Pierrakeas C and Pintelas P (2002) *Efficiency of machine learning techniques in predicting students' performance in distance learning systems*. Greece: Educational Software Development Laboratory Department of Mathematics, University of Patras.

Li C, Tao Y, Ao W, et al. (2018) Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. *Energy* 165: 1220–1227.

Limberg L, Alexandersson M, Lantz-Andersson A, et al. (2008) What matters? Shaping meaningful learning through teaching information literacy.

Lucas U (1996) Student approaches to learning–a literature guide. *Accounting Education* 5(1): 87–98.

MacCann C, Jiang Y, Brown LER, et al. (2020). Emotional intelligence predicts academic performance: A meta-analysis. *Psychological Bulletin* 146(2): 150–186.

Martins RM and Gresse Von Wangenheim C (2023) Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education* 22(3): 421–440.

Marton F and Säljö R (1976) On qualitative differences in learning: I—outcome and process. *British Journal Of Educational Psychology* 46(1): 4–11.

Maurya LS, Hussain MS and Singh S (2021) Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. *Applied Artificial Intelligence* 35(6): 403–420.

McKnight K, O'Malley K, Ruzic R, et al. (2016) Teaching in a digital age: How educators use technology to improve student learning. *Journal Of Research On Technology In Education* 48(3): 194–211.

Mosquera Rodas JJ (2023) Structural Elements for the Stimulation of Higher Intelligence from the Tri-Brain Theory and the Taxonomy of Biggs and Collins. Available at SSRN 4331849.

Nor NM and Idris N (2010) Assessing students' informal inferential reasoning using SOLO taxonomy based framework. *Procedia-Social And Behavioral Sciences* 2(2): 4805–4809.

Nti IK, Akyeramfo-Sam S, Bediako-Kyeremeh B, et al. (2022) Prediction of social media effects on students' academic performance using machine learning algorithms (MLAs). *J. Comput. Educ* 9: 195–223.

Nzuva S and Nderu L (2019) The superiority of the ensemble classification methods: A comprehensive review. *Journal of Information Engineering & Applications*: 43–53.

Okubo F, Yamashita T, Shimada A, et al. (2017, March) A neural network approach for students' performance prediction. In *Proceedings of the seventh international learning analytics & knowledge conference*, pp. 598–599.

Pagani L, Argentin G, Gui M, et al. (2016) The impact of digital skills on educational outcomes: Evidence from performance tests. *Educational Studies* 42(2): 137–162.

Pallathadka H, Wenda A, Ramirez-Asís E, et al. (2021) Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*.

Pandey M and Sharma VK (2013) A decision tree algorithm pertaining to the student performance analysis and prediction. *International Journal of Computer Applications* 61(13): 1–5.

Phan HP and Ngu BH (2020) Schooling experience and academic performance of Taiwanese students: The importance of psychosocial effects, positive emotions, levels of best practice, and personal well-being. *Social Psychology of Education* 23(4): 1073–1101.

Purinton EF and Burke MM (2018) The Fun Factor: Enhancing Student Engagement. *Proceedings of the Northeast Business & Economics Association*.

Quinlan JR (1986) Induction of decision trees. *Machine Learning* 1(1): 81–106.

Rastrollo-Guerrero JL, Gómez-Pulido JA and Durán-Domínguez A (2020) Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences* 10(3): 1042.

Rumelhart DE, Hinton GE and Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088): 533.

Soares C and Gray K (2019) Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *Journal of Petroleum Science and Engineering* 172: 934–959.

Strelan P, Osborn A and Palmer E (2020) The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review* 30: 100314.

Van Rossum EJ and Schenk SM (1984) The relationship between learning conception, study strategy and learning outcome. *British Journal of Educational Psychology* 54(1): 73–83.

Vapnik V (1998) *Statistical Learning Theory*. Wiley-Interscience.

Vosniadou S, Darmawan I, Lawson MJ, et al. (2021) Beliefs about the self-regulation of learning predict cognitive and metacognitive strategies and academic performance in pre-service teachers. *Metacognition and Learning*: 1–32.

Waheed H, Hassan SU, Aljohani NR, et al. (2020) Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior* 104: 106189

Walia N, Kumar M, Nayar N, et al. (2020, April) Student's academic performance prediction in academic using data mining techniques. In *Proceedings of the International Conference on Innovative Computing & Communications* (ICICC).

Williams ML and Sato BK (2021) Constructive or disruptive? How active learning environments shape instructional decision-making. *Journal of Learning Spaces* 10(2).

Wong JYH, Ko J, Nam S, et al. (2022) Virtual ER, a serious game for interprofessional education to enhance teamwork in medical and nursing undergraduates: Development and evaluation study. *JMIR Serious Games* 10(3): e35269.

Xu X, Wang J, Peng H, et al. (2019) Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior* 98: 166–173.

Yağcı M (2022) Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments* 9(1): 11.

Yu LC, Lee CW, Pan HI, et al. (2018) Improving early prediction of academic failure using sentiment analysis on self-evaluated comments. *Journal of Computer Assisted Learning* 34(4): 358–365.

Yukselturk E and Top E (2013) Exploring the link among entry characteristics, participation behaviors and course outcomes of online learners: An examination of learner profile using cluster analysis. *British Journal of Educational Technology* 44(5): 716–728.

Zeineddine H, Braendle U and Farah A (2021) Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering* 89: 106903.

Zhang S, Gao Q, Sun M, et al. (2022) Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education* 183: 104485.

Author Biographies

Yavuz Selim Balcıoğlu is a lecturer at the University of Gebze Technical. He received his BS degrees in Economy from the University of Kocatepe and MS in Business Administration from the University of Yasar in 2005 and 2014, respectively, and received a PhD in business analytics from the University of Gebze Technic. He is the author of more than 70 journal papers and has 13 International awards in robotics. His research has appeared in Electronics, IEEE, Intelligent Data Analytics, and other journals in the Information Systems field.

Melike Artar Melike Artar is a researcher at the Faculty of Business Administration, Gebze Technical University, Kocaeli, Turkey. She completed her undergraduate education at Okan University with a full scholarship and her master's degree at Gebze Technical University in the first place. In addition to her academic duties at the university, she advises private companies on human resources and personnel competencies. Her research area is organizational behavior and human resources. She is a reviewer in important journals such as the International Journal of Manpower and Social Science Computer Review.