Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 3, 2025

USING MACHINE LEARNING TO FORECAST COLLEGE STUDENTS' ACADEMIC PERFORMANCE: THE INFLUENCE OF EXTRACURRICULAR ACTIVITIES AND PREVIOUS ACADEMIC SCORES

Ghulam Murtaza[*1], **Dr. Amiya Bhaumik**[1]

[*1]*Department of Psychology, Lincoln University College Malaysia,*

[*1]gmurtaza@lincoln.edu.my, [1]amiya@lincoln.edu.my

DOI: https://doi.org/10.5281/zenodo.15878243

Keywords: Key Factors, Students performance, Machine Learning, Academic Performance.

Article History
Received on 27 May 2025
Accepted on 27 June 2025
Published on 03 July 2025

Corresponding Author: * **Ghulam Murtaza**

Abstract

This study explores the application of machine learning techniques to predict college students academic performance. While traditional models have primarily focused such as extracurricular activities and academic scores, this research expands the analysis by including variables like extracurricular activities and previous academic scores. The study aims to provide a more comprehensive approach to understanding student success by analyzing a features that contribute to academic outcomes. Several machine learning models are applied to predict students' academic success, and their performance is evaluated. The findings show that machine learning models are more effective in identifying students at risk of underperforming compared to traditional methods. These models offer a more understanding of student performance and can help identify at-risk students early in their academic careers. The practical implications of this research are significant for Extracurricular Activities and Previous Academic Scores educational institutions, as early identification enables timely interventions and support, improving retention rates and overall student success. The study contributes to the research on predictive analytics in education, demonstrating the potential of machine learning to enhance decision-making processes and foster better academic outcomes.

INTRODUCTION

1.1 Background

Academic performance is a crucial indicator of student success and has long been used as a measure of educational outcomes. Traditionally, academic performance has been primarily linked to abilities and previous academic achievements. However, research has increasingly shown that it is not just intellectual capacity that determines success in higher education. Many other factors contribute significantly to students' academic outcomes. These include behavioral factors like study habits, sleep patterns, and class attendance, as well as socio-economic factors such as family background, financial status, and access to resources. Together, these elements create a complex web of influences that can impact how well students perform in their courses.

In particular, extracurricular activities have gained attention for their role in influencing academic performance. Participation in activities outside the classroom, such as sports, clubs, or volunteering, can help students develop

important skills like time management, teamwork, and leadership. These skills not only contribute to personal growth but also positively affect academic achievement by enhancing a student's ability to balance responsibilities and manage time effectively. However, the relationship between extracurricular involvement and academic performance is nuanced. While moderate participation can lead to personal development and academic benefits, excessive involvement may detract from study time, leading to poorer grades and reduced academic focus.

Another key factor influencing academic success is previous academic performance. Prior academic achievement, such as grades from earlier semesters or high school GPA, has been consistently shown to be one of the strongest predictors of future success. Students who perform well in earlier academic years are more likely to continue excelling in subsequent courses due to better study habits, higher motivation, and academic resilience. The integration of past academic scores with other factors, such as extracurricular participation, can provide a more comprehensive understanding of a student's potential, helping educators develop early interventions and personalized support strategies for students at risk of underperforming. This research aims to leverage such insights to forecast student performance using machine learning, allowing for more accurate predictions and tailored educational support.

1.2 Problem Statement

Predicting academic performance is a complex task influenced by various factors beyond abilities and previous academic records. While traditional models often focus on grades and test scores, they frequently overlook other critical elements, such as participation in extracurricular activities, study habits, and social engagement. This research aims to address this gap by using machine learning techniques to develop a predictive model for college students' academic performance. By incorporating both prior academic achievements and extracurricular involvement, this study seeks to provide a more accurate prediction, ultimately helping educators identify at-risk students and enabling targeted interventions to improve outcomes.

1.3 Objectives of the Study

The primary objective of this study is to develop a machine learning-based model to predict college students' academic performance, focusing on the influence of extracurricular activities and previous academic scores. Specific objectives include:

Develop a machine learning model to predict college students' academic performance based on key factors like previous academic scores and extracurricular activities.

Identify the specific influence of extracurricular activities on academic performance, alongside traditional factors such as study habits and attendance.

Evaluate and compare the performance of different machine learning models for predicting academic success.

Assess the predictive accuracy of the models using evaluation metrics like accuracy, precision, and recall.

Provide actionable insights for educators to create targeted interventions for at-risk students.

1.4 Research Questions

This study aims to address the following research questions:

1. **What are the key factors, including extracurricular activities and previous academic scores, that most significantly influence college students' academic performance?**
   This question aims to identify and prioritize the factors that have the most significant impact on academic performance, providing a comprehensive understanding of what influences student success beyond traditional academic scores.
2. **How do different machine learning models compare in predicting academic performance based on these factors?**
   This question evaluates the effectiveness of different machine learning models in predicting student outcomes and helps

determine which model performs best based on the data provided.

3. **What is the accuracy of the developed models in predicting student outcomes, and how can they be improved for better predictions?**
   This question seeks to assess the performance and accuracy of the predictive models, exploring potential ways to improve their reliability and robustness for future applications.
4. **Can early prediction of academic performance using machine learning help identify students at risk of underperforming for timely interventions?**
   This question addresses the potential of machine learning to identify students who are at risk of poor academic performance early in the semester, enabling educators to provide proactive support and interventions.

2.5 Motivation and Scope of the Study

The motivation for this study arises from the increasing need to identify the various factors influencing college students' academic performance. Traditionally, academic performance has been primarily predicted based on abilities and previous academic scores. However, research has shown that factors such as extracurricular activities, study habits, and socio-emotional well-being also contribute significantly to student success. By incorporating these factors, this study aims to provide a more comprehensive model for predicting academic performance, ultimately helping educators better understand the complex dynamics of student achievement.

Another key motivation is the potential for early identification of students at risk of underperforming. With higher education institutions facing increasing demands to support diverse student populations, early intervention becomes crucial. By utilizing machine learning techniques, this research seeks to create a predictive model that can identify at-risk students early on, allowing educators to implement targeted interventions (Rahman et al., 2021). Such interventions may include personalized academic advising, tutoring, or access to support services, ultimately helping students stay on track for success.

The scope of this study focuses on college students, examining how their previous academic performance and participation in extracurricular activities influence their academic outcomes. The research will primarily use data from higher education institutions to develop a predictive model based on machine learning techniques. While the study will not address other external factors, such as socio-economic background, its findings will contribute valuable insights for improving prediction models and offering better support to students at risk of academic failure (Dawar et al., 2024).

1.6 Structure of the Paper

The paper is organized into several sections, each addressing key aspects of the study. Chapter 1 provides an introduction to the research, including the background, problem statement, research objectives, research questions, and the scope of the study. Chapter 2 offers a comprehensive literature review, discussing previous studies on academic performance prediction, factors influencing student success, and the application of machine learning techniques in education. Chapter 3 outlines the research methodology, detailing the design, data collection methods, data preprocessing techniques, and machine learning models used for prediction. Chapter 4 presents the results and discussions, evaluating the performance of the developed models and analyzing their effectiveness in predicting academic outcomes. Finally, Chapter 5 concludes the paper, summarizing the findings, acknowledging the study's limitations, and suggesting directions for future research. This structure ensures a logical flow from introduction to conclusion, providing insights into the research process and its outcomes.

2. Literature Review

2.1 Overview of Academic Performance Prediction

Academic performance prediction has become an increasingly important area of research in educational studies, as institutions seek to enhance student success and identify those at risk of underperforming. Traditionally, academic performance has been assessed using basic measures such as grades, test scores, and overall GPA. While these traditional methods provide valuable insights, they often overlook the broader context of student performance. The need for a more holistic approach has led to the inclusion of various additional factors, such as socio-economic background, study habits, extracurricular activities, and sleep patterns, all of which can impact a student's academic success (Al-Alawi et al., 2023).

Machine learning (ML) techniques have emerged as a powerful tool for academic performance prediction. These techniques allow for the analysis of large datasets containing a wide variety of factors that influence academic success. Unlike traditional statistical methods, machine learning models are capable of identifying complex patterns and relationships within the data, making them ideal for predicting academic outcomes. Algorithms such as decision trees, random forests, and support vector machines have been applied to student data to improve the accuracy of academic predictions. By analyzing a combination of factors—such as prior academic achievements, study habits, and involvement in extracurricular activities—machine learning models can provide a more nuanced and accurate picture of a student's academic trajectory.

The incorporation of machine learning into academic performance prediction also opens the door to more personalized interventions for students. As these models become more accurate, they can be used to identify students at risk of underperforming earlier in the academic term. Early identification allows for targeted interventions, such as personalized academic support, tutoring, or counseling, to help students improve their performance before issues become more significant. This approach not only improves the overall learning experience for students but also contributes to better retention rates and higher graduation success (Zeineddine et al., 2021). Academic performance prediction can be a powerful tool for enhancing educational outcomes.

2.2 Role of Machine Learning in Education:

Machine learning (ML) has become a transformative tool in the field of education, offering new ways to analyze and understand student data. Traditionally, educators relied on manual analysis and simple statistical methods to assess student performance. However, the vast amounts of data generated by students ranging from grades and test scores to behavioral and engagement metrics require more advanced techniques to process and interpret effectively. Machine learning enables the automated processing of large datasets, helping educators uncover patterns that may not be apparent through traditional analysis (Dabhade et al., 2021). By applying machine learning algorithms to student data, institutions can gain deeper insights into academic performance and behavior, enabling them to develop more effective strategies for improving student outcomes.

In educational settings, machine learning is used to predict a variety of student outcomes, such as academic success, dropout risk, and engagement levels. Algorithms like decision trees, random forests, and support vector machines analyze factors such as prior academic performance, attendance, study habits, and participation in extracurricular activities to predict how students will perform in future courses. Machine learning models are particularly useful because they can learn from data, continuously improving their predictions over time (Batool et al., 2023). This capability allows for the development of more accurate models that can anticipate students' needs, offering personalized support and recommendations that are tailored to individual students' needs and characteristics.

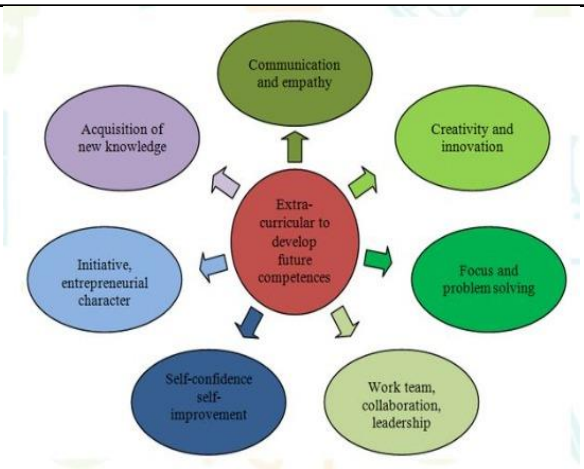

Figure 1: Impact of extracurricular activities on competencies

The role of machine learning in education also extends beyond academic performance prediction. It is increasingly being used to develop intelligent tutoring systems (ITS), which adapt to students' learning styles and needs. These systems use ML algorithms to assess a student's progress and provide customized content or feedback, improving the overall learning experience. Additionally, machine learning can help in identifying trends across large student populations, allowing for institutional improvements in teaching strategies, curriculum design, and student services. Ultimately, the integration of machine learning in education provides a data-driven approach that enhances teaching, learning, and overall student success.

2.3 Key Factors Affecting Academic Performance

Academic performance is influenced by a variety of factors that extend beyond abilities and prior achievements. While traditional models focused primarily on intellectual capacity, research now highlights the role of factors. These include behavioral, environmental, and socio-economic elements that contribute significantly to a student's success. Understanding these factors provides a more comprehensive view of academic performance.

Previous Academic Scores:

Previous academic scores are a strong indicator of future success. Students with high past performance often develop effective study habits, time management, and motivation (Suleiman et al., 2024). These factors contribute to their continued success, as previous academic achievements reflect their ability to manage coursework and understand academic concepts, fostering self-confidence and academic self-efficacy.

Extracurricular Activities

Extracurricular activities help students develop essential skills like time management, teamwork, and leadership, which contribute to academic success. Participation in sports, clubs, or volunteer work also offers emotional and social benefits. However, excessive involvement may reduce study time, leading to lower academic performance. Maintaining a balance between academics and extracurricular is key to success.

Study Habits and Time Management

Strong study habits and effective time management are crucial for academic success. Students who plan and organize their study sessions, prioritize tasks, and employ active learning techniques perform better. Poor time management and procrastination can lead to increased stress and lower grades. Students with disciplined study habits are better equipped to meet academic demands efficiently.

Sleep Patterns and Health

Adequate sleep and good health are essential for academic performance. Lack of sleep impairs functions, memory retention, and concentration. Students who maintain consistent sleep patterns and prioritize physical health tend to perform better academically. Poor sleep and health issues can lead to fatigue, stress, and decreased ability, ultimately affecting academic results.

Class Attendance and Participation

Regular class attendance and active participation directly impact academic performance. Students who attend classes consistently are more engaged

with course material, resulting in a better understanding of key concepts. Active participation, such as asking questions and discussing topics, enhances learning. Conversely, poor attendance leads to missed content, gaps in understanding, and lower academic achievement.

2.4 Machine Learning Techniques for Academic Performance Prediction

Machine learning (ML) has significantly transformed the way academic performance is predicted, as it enables the analysis of large datasets with multiple features, such as previous academic performance, study habits, extracurricular activities, and socio-economic status. By leveraging ML algorithms, educational institutions can make more accurate predictions regarding student outcomes like academic success, dropout risk, and retention (Olabanjo et al., 2022). These models analyze complex relationships between various factors, helping educators identify at-risk students early and intervene with personalized support, which can improve academic results and student retention.

Decision Trees are one of the most widely used algorithms for predicting academic performance. This model splits the dataset into subsets based on feature values, creating a tree structure where each node represents a decision based on input features. Decision trees are easy to understand and interpret, making them popular in educational data analysis. However, they can be prone to overfitting, especially when the dataset is complex or noisy. To mitigate this, more robust models like **Random Forests** are often used. Random Forests aggregate multiple decision trees, which reduces the risk of overfitting and improves the prediction accuracy by averaging the results from multiple trees.

Support Vector Machines (SVM) are another powerful algorithm used in educational data mining. SVM works by finding the optimal hyperplane that separates different classes within the dataset. It is particularly effective when dealing with high-dimensional data and complex patterns, such as predicting academic success based on numerous input factors. Another algorithm, K-Nearest Neighbors (K-NN), classifies students based on their similarity to other data points in the dataset. K-NN is intuitive and simple to implement, making it ideal for smaller datasets or when relationships between features are less clear (Johora et al., 2025). These machine learning techniques, when applied appropriately, offer valuable insights and predictive power for improving academic outcomes.

2.5 Previous Studies on Predicting Academic Performance:

Machine learning has become an increasingly important tool in predicting academic performance, offering insights into the complex factors that influence student outcomes. Early studies primarily focused on using traditional academic measures such as grades and test scores to predict success. However, with the growing availability of data, researchers have expanded their focus to include factors like extracurricular activities, socio-economic status, and mental health (So et al., 2023). These studies have shown that machine learning models can provide more accurate predictions by analyzing a wider range of factors that contribute to academic performance.

Research has demonstrated the effectiveness of various machine learning algorithms in predicting academic performance. Decision trees, random forests, and support vector machines (SVM) have been widely applied to analyze student data. These algorithms identify patterns and relationships between multiple factors, such as previous academic scores, attendance, and involvement in extracurricular activities (Issah et al., 2023). Machine learning models have been shown to outperform traditional statistical methods in terms of predictive accuracy, highlighting the advantages of these techniques in educational research.

Recent studies have further explored the use of more complex variables, such as online engagement and social media activity, to predict academic success. By integrating these modern data sources, researchers have created more sophisticated models that can provide real-time predictions of student performance (Chen & Zhai, 2023). These models enable early identification of students at risk of

underperforming, allowing institutions to offer timely support. The growing use of machine learning in predicting academic outcomes has the potential to significantly enhance the ability of educators to improve student success and retention.

3. Research Methodology

3.1 Research Design

This study adopts a quantitative research design to explore how machine learning techniques can predict college students' academic performance based on various factors. The research primarily focuses on analyzing variables such as previous academic scores and extracurricular activities. The dataset used for this study is collected from Kaggle, which includes comprehensive student data, such as academic records, demographic information, study habits, and attendance.

The research follows a structured approach, beginning with data selection and preprocessing, where missing values are handled, and categorical variables are encoded. Feature selection is then performed to identify the most relevant factors for prediction. Afterward, machine learning models are applied to the data to identify patterns and relationships between the factors and academic performance. The study then compares the performance of various models, providing a robust analysis of their predictive power. This structured methodology ensures objective and reliable results in understanding the factors affecting student success.

3.2 Data Selection:

Data selection is a critical step in predicting academic performance, as the chosen dataset must accurately represent the factors influencing student success. For this study, data is collected from Kaggle, a well-known platform that provides access to a variety of publicly available datasets. The dataset selected for this research includes detailed student information, including demographic characteristics, academic performance, and behavioral factors. Key attributes such as student GPA, class attendance, study hours, participation in extracurricular activities, and sleep patterns are included, as these are considered significant predictors of academic outcomes.

The dataset is carefully chosen to ensure a diverse sample, representing students from various academic backgrounds, socio-economic statuses, and personal habits. By utilizing this comprehensive dataset, the study aims to identify patterns and relationships between academic performance and these various factors. The data serves as the foundation for applying machine learning models to predict student success. Ethical considerations are strictly followed, with personal information anonymized to ensure privacy. The dataset's diversity and the inclusion of multiple factors provide a comprehensive view of the influences on academic performance, which allows for a more accurate and robust prediction model.

3.3 Data Preprocessing

Data preprocessing is a crucial step in preparing the dataset for machine learning models. It involves cleaning and transforming the raw data to ensure its quality and suitability for analysis. Since raw data often contains inconsistencies, missing values, and noise, preprocessing helps to address these issues and make the dataset ready for accurate model training. The following steps outline the main procedures used in preprocessing the dataset:

Handling Missing Values: Missing data in the dataset is addressed by imputing values. For numerical features, missing values are replaced with the mean or median of the respective column. For categorical variables, the mode (most frequent value) is used to fill in the missing entries, preserving the integrity of the dataset.

Encoding Categorical Variables: Categorical features such as gender, participation in extracurricular activities, or region are transformed into numerical format using one-hot encoding. This process creates binary columns for each category, enabling machine learning models to process them effectively.

Normalization: Continuous numerical variables such as study hours, GPA, and attendance are normalized using min-max scaling. This process scales the features to a range between 0 and 1,

ensuring that no single feature disproportionately influences the model due to differences in magnitude.

Outlier Detection: Outliers are identified using statistical techniques, such as Z-scores or IQR (Interquartile Range). These extreme values are either removed or adjusted to prevent them from distorting the results and negatively affecting the model's performance.

Feature Engineering: New features are created by combining existing ones, such as combining study hours with extracurricular participation to better capture the effects on academic success.

3.4 Data Splitting

Data splitting is an essential process in machine learning to ensure that models are properly trained, validated, and tested for accuracy and generalization. The dataset is typically divided into three main subsets: training, validation, and testing. The training set is the largest portion, generally comprising 70-80% of the data. It is used to train the model by allowing it to learn from the data and adjust its internal parameters. During this phase, the model identifies patterns and relationships in the data that it will use to make predictions on unseen data.

The validation set accounts for about 10-15% of the data. It is not used for training but is employed during the training process to help tune the model's hyperparameters, such as learning rate or regularization strength. This set provides valuable feedback on the model's performance, allowing adjustments to be made before finalizing the model. The **testing set** also represents 10-15% of the data and is used after the model is fully trained and validated. This set contains data that the model has never seen, providing an unbiased estimate of the model's performance on new, unseen data. It is used to evaluate the final accuracy and generalization capability of the model. By splitting the data into these three sets, the risk of overfitting is minimized, and a more accurate measure of the model's real-world performance is obtained.

3.5 Machine Learning Models Used:

This study applies a range of machine learning models to predict college students' academic performance, aiming to determine the most effective algorithm for this task. The models selected include Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), and Logistic Regression. Each model was chosen for its unique strengths in handling different types of data and prediction tasks. Decision Trees are a fundamental model that divides the dataset into smaller subsets based on feature values, creating a hierarchical structure. This model is easy to interpret and visualize, making it useful for understanding how decisions are made. However, decision trees are prone to overfitting, especially when dealing with complex datasets or noise in the data.

Random Forests build upon decision trees by constructing an ensemble of multiple trees and aggregating their predictions, making them more robust and accurate than a single decision tree. This ensemble method reduces the overfitting issue and is highly effective for datasets with many features. Support Vector Machines (SVM) are another powerful model, particularly suited for high-dimensional data. SVM works by finding the optimal hyperplane that best separates different classes in the dataset. It is particularly effective for datasets where the decision boundaries are not linearly separable, making it a strong candidate for complex prediction tasks.

Additionally, K-Nearest Neighbors (K-NN) is used, which classifies a data point based on the majority class of its closest neighbors. K-NN is simple, intuitive, and works well for smaller datasets. Logistic Regression, a statistical model, is employed for binary classification, predicting the probability of a specific outcome based on input features. These models are trained and evaluated on the dataset to determine the most effective approach for predicting academic performance and identifying at-risk students.

3.6 Evaluation and Prediction:

After training the machine learning models, evaluation and prediction are key steps in determining their effectiveness in predicting

academic performance. The evaluation process involves testing the models on a testing dataset, which has not been seen by the models during training. This ensures that the models' ability to generalize to unseen data is accurately assessed. The primary goal of evaluation is to understand how well the model predicts outcomes based on various performance metrics.

The models are evaluated using metrics such as accuracy, precision, recall, and F1 score. Accuracy measures the overall percentage of correct predictions made by the model, providing a general assessment of performance. Precision and recall are crucial for tasks like identifying at-risk students, as they evaluate the model's ability to minimize false positives and false negatives, respectively. F1 **score**, the harmonic mean of precision and recall, balances these two metrics and provides a single value for model performance. These metrics are essential for understanding the trade-offs between different types of errors and ensuring that the model performs well in practical scenarios.

After evaluation, the best-performing model is selected for prediction. This model is then used to predict the academic performance of future students based on the input features. By applying the selected machine learning model, the study aims to provide actionable insights that can help institutions identify students who may need additional academic support, ultimately improving student outcomes and retention.

4. Results and Discussion

4.1 Dataset Description

The dataset used in this study is sourced from Kaggle, and it provides a rich collection of student-related data, including demographic details, academic performance, and behavioral factors. Key attributes include students' GPA, previous academic scores, and test results, which are critical indicators of academic ability. In addition, the dataset contains demographic information such as age, gender, and socio-economic background, offering insights into how these factors may influence a student's academic success. These academic and demographic features provide a solid foundation for analyzing and predicting student performance.

Beyond academic and demographic data, the dataset also includes behavioral factors such as study hours, class attendance, and participation in extracurricular activities. These variables are important predictors of academic success, as they reflect a student's level of commitment, time management skills, and engagement in learning. The dataset also includes categorical features, which are processed using techniques like one-hot encoding to make them suitable for machine learning models. The dataset is pre-processed to handle missing values and normalize numerical features, ensuring that the data is clean and ready for analysis. This diverse factors, allows for a comprehensive exploration of the influences on academic performance and forms the basis for applying various machine learning models to predict student outcomes.

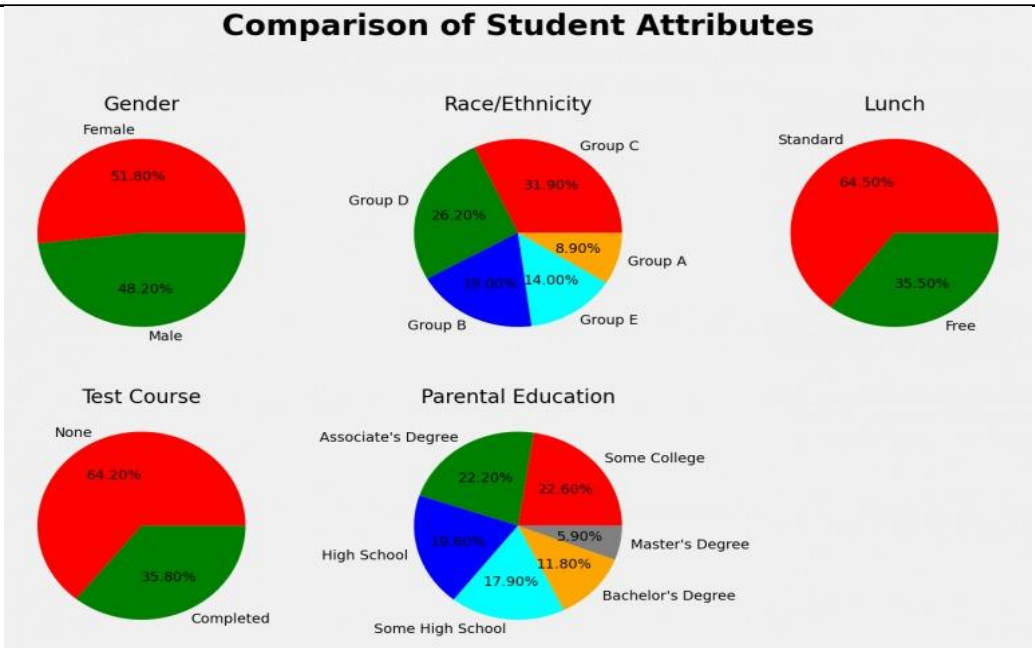

Figure 2: Comparison of Student Attributes

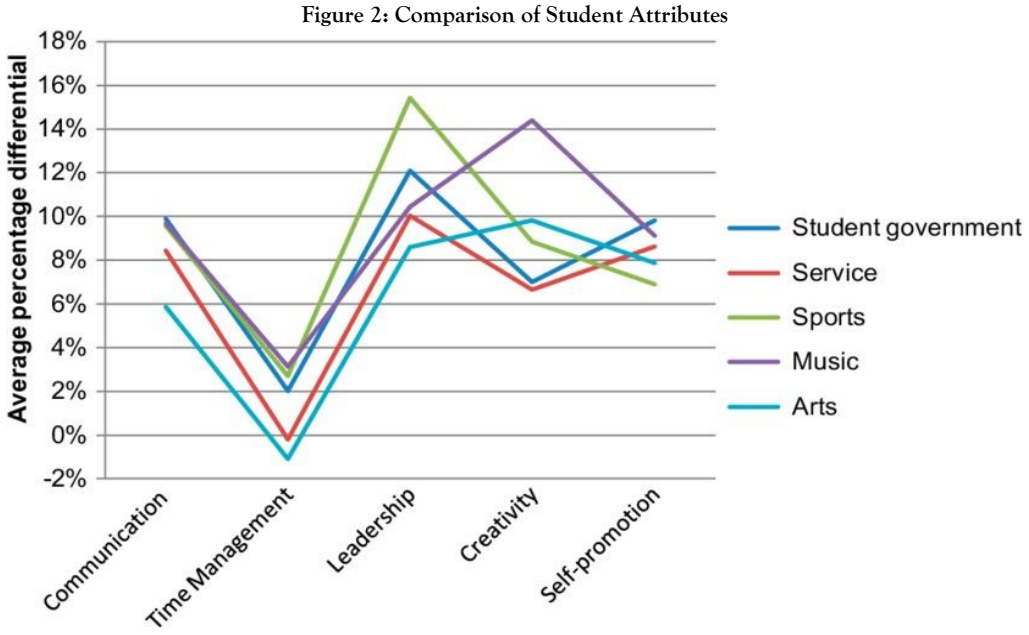

Figure 3: Comparison of skills across activities

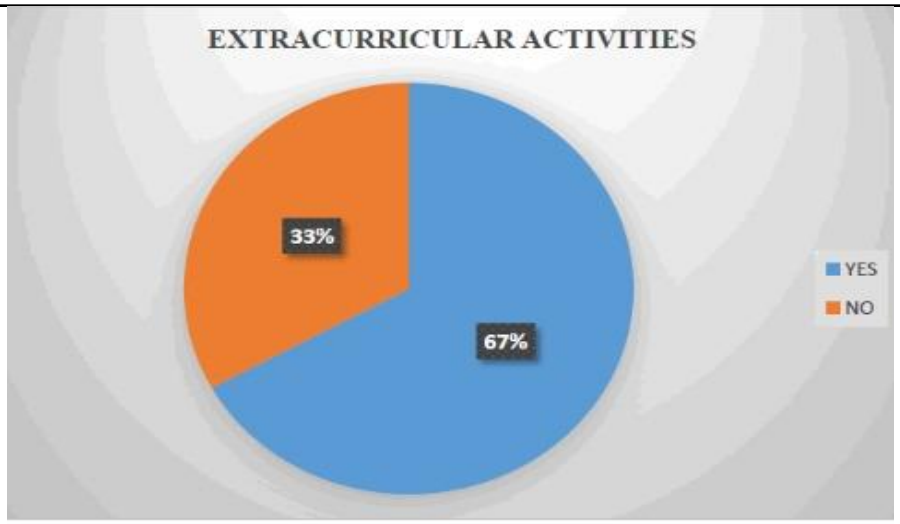

Figure 4: Distribution of extracurricular activity participation

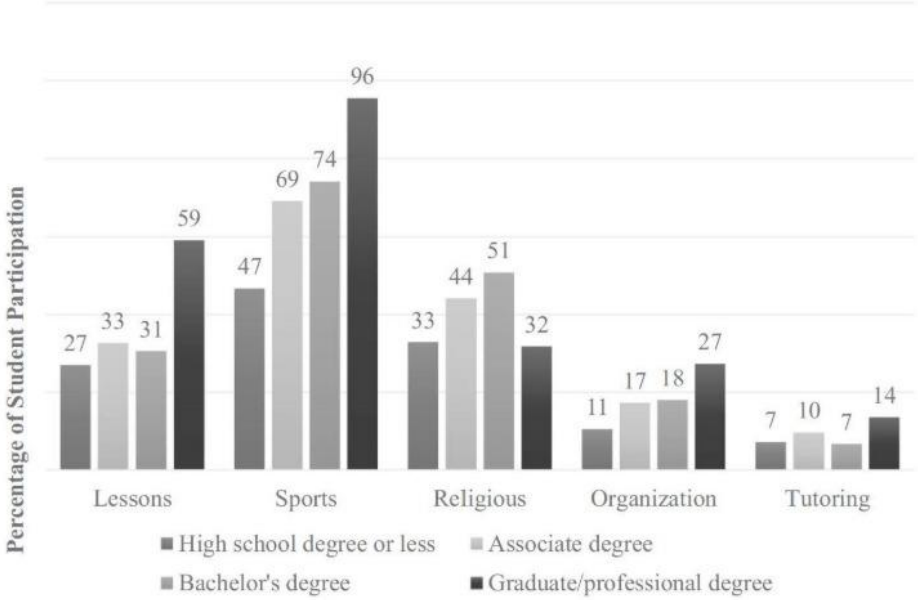

Figure 5: Student participation across educational levels

4.2 Performance Comparison of Models

Several machine learning models were evaluated to predict academic performance, and their effectiveness was compared using a variety of performance metrics, including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive assessment of each model's ability to predict student outcomes accurately. Some models performed exceptionally well at minimizing false positives, resulting in high precision, while others were more successful in identifying at-risk students, achieving higher recall. The F1 score, which balances both precision and recall, helped to further evaluate the models' overall predictive power, particularly in ensuring that predictions were accurate without sacrificing recall or precision.

The comparison of model performance revealed that ensemble models, which combine predictions from multiple simpler models, generally outperformed individual models due to their ability to reduce overfitting and improve generalization. While simpler models were faster and easier to interpret, they tended to perform

less effectively, especially with complex datasets or when faced with the challenge of balancing the trade-off between false positives and false negatives. Overall, models that offered a balance between high accuracy and reduced overfitting were found to be the most effective in predicting academic performance. This analysis provided insights into which types of models are most reliable for predicting student success and identifying those in need of intervention.

5. Contributions

5.1 Theoretical Contributions

This study provides key theoretical contributions by integrating a wide range of factors to predict academic performance. By incorporating machine learning techniques, the study enhances the understanding of how various factors influence student success. This research fills the gap in traditional academic prediction models, which mainly concentrate on factors, by offering a broader, more holistic approach. The findings are beneficial for future educational research, providing a robust framework for understanding the diverse elements that impact academic outcomes and guiding the development of more effective prediction models.

Integration of Factors

Extracurricular activities, study habits, and socio-economic background, in predicting academic success. By incorporating these factors, the study offers a more comprehensive model of academic performance, suggesting that success is influenced by a wider range of factors, not just intellectual capabilities.

Application of Machine Learning Models in Education

The study also contributes to the growing body of knowledge on applying machine learning models in education. By exploring and comparing different algorithms, this research provides insights into which models are most effective for predicting academic performance. These findings are valuable for educational institutions and researchers, helping them select the best machine learning techniques for data analysis. The study demonstrates how machine learning can be used to improve educational predictions, offering a more accurate and data-driven approach to understanding student success.

Early Identification of At-Risk Students

The study's contribution to early identification of at-risk students is another key theoretical advancement. By using machine learning to predict academic performance, the research enables the identification of students who may be at risk of underperforming. This early identification allows for timely intervention and support, such as personalized academic advising or tutoring, which can help improve student outcomes. This theoretical contribution offers a proactive approach to student success, enhancing the potential for academic institutions to provide targeted support that can lead to better retention and graduation rates.

5.2 Practical Contributions

The findings of this study provide valuable insights for educational institutions aiming to enhance student success. By applying machine learning models to predict academic performance, institutions can proactively identify at-risk students and implement timely interventions. This approach goes beyond traditional assessment methods, offering data-driven insights that enable schools to allocate resources effectively. The following sections explore the key practical applications of these findings and their potential impact on improving student outcomes and institutional effectiveness.

Early Identification and Intervention

The ability to predict student performance allows for early identification of at-risk students, enabling timely interventions such as tutoring, counseling, or academic advising. Early intervention prevents students from falling behind and provides them with the support they need to succeed academically. This proactive approach leads to improved retention rates and academic outcomes, as students are given the

resources necessary to address challenges before they become significant barriers to their success.

Integration of Machine Learning in Educational Settings

Machine learning models can be integrated into existing student information systems to provide real-time predictions about student performance. This allows educators and administrators to make informed, data-driven decisions about where to allocate resources. By incorporating these predictive models into daily operations, institutions can improve their ability to support students efficiently, ensuring that at-risk individuals receive the right help at the right time. This integration promotes a more personalized and proactive approach to student support.

Informing Curriculum and Teaching Strategies

The insights gained from predicting academic performance can inform curriculum design and teaching strategies. By identifying the factors most associated with student success, educators can adjust their approaches to better align with students' needs. This enables institutions to provide targeted support, improving the overall learning experience and ensuring that students have the tools and strategies they need to succeed academically.

Conclusion

In conclusion, this study demonstrates the effectiveness of machine learning techniques in predicting college students' academic performance. By analyzing a comprehensive dataset that includes academic scores, extracurricular activities, study habits, and socio-economic bacground, the research provides a more holistic approach to understanding the factors influencing student success. The findings show that machine learning models, particularly ensemble methods, can significantly enhance prediction accuracy, making it possible to identify at-risk students early in their academic careers. These insights offer a more comprehensive method for forecasting academic outcomes compared to traditional prediction models.

The practical implications of these findings are substantial for educational institutions. By incorporating machine learning into student information systems, schools can make data-driven decisions to better allocate resources and offer targeted interventions. Early identification of at-risk students allows for personalized support, such as tutoring or academic counseling, which can improve retention rates and academic performance. While the study's scope is limited by the data source used, future research can expand on these findings by incorporating diverse datasets and testing models in real-world academic environments. Ultimately, this research provides a strong foundation for using machine learning to improve student success and institutional effectiveness in educational settings.

Future Work

While this study provides valuable insights into predicting academic performance using machine learning, several opportunities exist for further research to expand and enhance the model's applicability. One key area for future work is the inclusion of more diverse datasets. This study relied on a single dataset, which may limit the generalizability of the results. By incorporating data from multiple educational institutions or regions, future research could provide a more representative sample, leading to more robust and widely applicable models. A diverse dataset would allow for a deeper understanding of how academic performance is influenced across different populations.

Another opportunity for future research is the exploration of more advanced machine learning techniques. Although traditional models like decision trees and random forests provided useful predictions in this study, newer approaches such as deep learning and neural networks might yield even more accurate results, particularly when dealing with large, complex datasets. These advanced techniques could help capture complex, non-linear relationships between features, improving the predictive power of the models and enabling more precise academic performance forecasts.

For example, data related to student motivation, stress levels, and mental health could provide deeper insights into their academic behavior. Longitudinal studies that track changes in students' study habits and engagement over time could also improve the model's predictive capabilities. Finally, testing these models in real-world educational environments would help validate their effectiveness and ensure they provide actionable insights for educators.

References

Al-Alawi, L., Al Shaqsi, J., Tarhini, A. and Al-Busaidi, A.S., 2023. Using machine learning to predict factors affecting academic performance: the case of college students on academic probation. *Education and Information Technologies*, *28*(10), pp.12407-12432. https://link.springer.com/article/10.1007/s10639-023-11700-0

Batool, S., Rashid, J., Nisar, M.W., Kim, J., Kwon, H.Y. and Hussain, A., 2023. Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, *28*(1), pp.905-971. https://link.springer.com/article/10.1007/s10639-022-11152-y

Chen, Y. and Zhai, L., 2023. A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, *28*(9), pp.12039-12057. https://link.springer.com/article/10.1007/s10639-023-11672-1

Dabhade, P., Agarwal, R., Alameen, K.P., Fathima, A.T., Sridharan, R. and Gopakumar, G., 2021. Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, *47*, pp.5260-5267. https://www.sciencedirect.com/science/article/pii/S2214785321042735

Dawar, I., Negi, S., Lamba, S. and Kumar, A., 2024. Enhancing Student Academic Performance Forecasting: A Comparative Analysis of Machine Learning Algorithms. *SN Computer Science*, *5*(6), p.758. https://link.springer.com/article/10.1007/s42979-024-03118-3

Falát, L. and Piscová, T., 2022. Predicting GPA of university students with supervised regression machine learning models. *Applied Sciences*, *12*(17), p.8403. https://www.mdpi.com/2076-3417/12/17/8403

Guanin-Fajardo, J.H., Guaña-Moya, J. and Casillas, J., 2024. Predicting Academic Success of College Students Using Machine Learning Techniques. *Data*, *9*(4), p.60. https://www.mdpi.com/2306-5729/9/4/60

Issah, I., Appiah, O., Appiahene, P. and Inusah, F., 2023. A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision analytics journal*, *7*, p.100204. https://www.sciencedirect.com/science/article/pii/S2772662223000449

Jin, X., 2023. Predicting academic success: machine learning analysis of student, parental, and school efforts. *Asia Pacific Education Review*, pp.1-22. https://link.springer.com/article/10.1007/s12564-023-09915-4

Johora, F.T., Hasan, M.N., Rajbongshi, A., Ashrafuzzaman, M. and Akter, F., 2025. An explainable AI-based approach for predicting undergraduate students academic performance. *Array*, *26*, p.100384. https://www.sciencedirect.com/science/article/pii/S2590005625000116

Olabanjo, O.A., Wusu, A.S. and Manuel, M., 2022. A machine learning prediction of academic performance of secondary school students using radial basis function neural network. *Trends in Neuroscience and Education*, *29*, p.100190. https://www.sciencedirect.com/science/article/pii/S2211949322000187

Rahman, S.R., Islam, M.A., Akash, P.P., Parvin, M., Moon, N.N. and Nur, F.N., 2021. Effects of co-curricular activities on student's academic performance by machine learning. *Current Research in Behavioral Sciences*, *2*, p.100057. https://www.sciencedirect.com/science/article/pii/S2666518221000449

Sharma, N. and Yadav, M., 2022, August. A Comparative Analysis of Students' Academic Performance using Prediction Algorithms Based on Their Time Spent on Extra-Curricular Activities. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)* (pp. 745-750). IEEE. https://ieeexplore.ieee.org/abstract/document/9917606/

So, J.C.H., Ho, Y.H., Wong, A.K.L., Chan, H.C., Tsang, K.H.Y., Chan, A.P.L. and Wong, S.C.W., 2023. Analytic study for predictor development on student participation in generic competence development activities based on academic performance. *IEEE Transactions on Learning Technologies*, *16*(5), pp.790-803. https://ieeexplore.ieee.org/abstract/document/10172338/

Suleiman, I.B., Okunade, O.A., Dada, E.G. and Ezeanya, U.C., 2024. Key factors influencing students' academic performance. *Journal of Electrical Systems and Information Technology*, *11*(1), p.41. https://link.springer.com/article/10.1186/s43067-024-00166-w

Yağcı, M., 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), p.11. https://link.springer.com/article/10.1186/S40561-022-00192-Z

Zeineddine, H., Braendle, U. and Farah, A., 2021. Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, *89*, p.106903. https://www.sciencedirect.com/science/article/pii/S0045790620307552