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

PREDICTING ACADEMIC OUTCOMES IN COLLEGE STUDENTS WITH MACHINE LEARNING: THE ROLE OF STUDY AND SLEEP HOURS

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.15878011

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

Article History
Received on 23 May 2025
Accepted on 23 June 2025
Published on 03 July 2025

Corresponding Author: * Ghulam Murtaza

Abstract

This study examines the application of machine learning models to predict academic performance in college students, with a focus on the impact of study hours and sleep hours as key factors influencing academic success.Tradtional method uses annual test base methods. This research aims to provide a more comprehensive approach by incorporating study habits and sleep patterns into the evaluation. Different machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Decision Tree, used to predict students' academic outcomes based on these factors. The findings demonstrate that study and sleep hours significantly affect academic performance, with different models offering varying levels of accuracy in predictions. The study highlights the importance of factors in academic prediction and suggests that machine learning can be a valuable tool for identifying at-risk students early. By providing insights into the factors influencing academic success, the research also offers practical implications for educators and administrators, enabling them to design more targeted interventions and support strategies. The results of this study contribute to the developing field of educational data science, offering a data-driven approach to improving student success and fostering a proactive learning environment.

INTRODUCTION

1.1 Background

In recent years, predicting academic outcomes in college students has become a significant area of research in education. Traditional assessment methods, such as grades and attendance, offer limited insights into the multifaceted factors that influence student performance. As more data on student behaviors, such as study habits, sleep patterns, and extracurricular activities, becomes available, there is an increasing opportunity to explore how these factors can contribute to academic success. Machine learning, with its ability to analyze large and complex datasets, offers a promising approach for predicting student performance by uncovering hidden patterns and relationships that are often difficult to identify using traditional methods.

Study habits and time management are widely recognized as key determinants of academic success. Students who allocate more time to their studies and develop effective study strategies tend to perform better in their courses. However, the mere quantity of study hours does not necessarily

guarantee success. The quality of study time, the strategies employed, and the ability to manage competing demands also play significant roles. Additionally, sleep patterns have been shown to have a substantial impact on performance. Adequate sleep is essential for processes such as memory consolidation, focus, and problem-solving abilities, all of which are critical for academic achievement. Inadequate sleep, on the other hand, can hinder a student's ability to concentrate and retain information, leading to lower academic performance.

Machine learning offers a powerful tool for predicting academic outcomes by analyzing various factors simultaneously, including study and sleep hours. By examining how these factors interact and contribute to student success, machine learning models can provide valuable insights into which elements have the most significant impact. This predictive capability can help educators and administrators identify at-risk students early, allowing them to implement targeted interventions to improve academic outcomes. This study aims to explore how study and sleep hours influence academic performance, using machine learning to create more accurate predictions and ultimately enhance student success.

1.2 Problem Statement

Predicting academic outcomes in college students is inherently complex due to the multiple factors influencing student success. While traditional measures like grades and attendance offer some insight, they fail to consider important factors such as study habits and sleep patterns. This research aims to fill this gap by using machine learning to predict academic performance, specifically analyzing how study hours and sleep hours contribute to student success. By understanding the relationship between these factors and academic outcomes, the study hopes to identify key predictors, enabling more targeted interventions to improve student performance and overall academic achievement.

1.3 Objectives of the Study

The primary objective of this study is to develop a machine learning model to predict academic outcomes in college students, focusing on the impact of study and sleep hours. Specifically, the study aims to:

Identify and analyze the influence of study hours on academic performance.

Investigate the relationship between sleep hours and academic success.

Develop a machine learning model that incorporates both study and sleep hours to predict academic outcomes.

Evaluate the accuracy and effectiveness of the model in predicting student performance.

Provide insights for educators and administrators to design targeted interventions to improve student success.

1.4 Research Questions

This study aims to answer the following research questions:

1. **How do study hours influence the academic performance of college students?**
   This question aims to explore the correlation between the amount of time students dedicate to studying and their overall academic performance. It will examine whether longer or more structured study hours lead to improved outcomes.
2. **What is the relationship between sleep hours and academic success in college students?**
   This question seeks to determine how sleep patterns, particularly the number of hours of sleep students get, impact their abilities and academic performance.
3. **How can machine learning models predict academic outcomes based on study and sleep hours?**
   This question focuses on the ability of machine learning models to process data

related to study and sleep hours and accurately predict academic success. It will explore the predictive power of these models.

4. **Which of the two factors, study hours or sleep hours, has a stronger impact on academic performance?**
   This question aims to assess whether study hours or sleep hours play a more significant role in determining a student's academic performance. It will analyze which factor has a stronger influence.
5. **Can early identification of students at risk of poor academic outcomes be achieved using the proposed machine learning model?**
   This question examines whether the machine learning model can identify students who are at risk of underperforming, allowing for timely interventions to improve their academic success.

2.5 Motivation and Scope of the Study

The motivation for this study arises from the need to improve the prediction of academic outcomes in college students, as traditional metrics like grades and attendance often fail to capture the broader factors influencing academic success. While these measures offer useful insights, they overlook critical aspects, such as study habits and sleep patterns. These factors significantly impact a student's ability to retain information, focus during lectures, and perform well in exams, yet they remain largely unaccounted for in many academic performance models (Alghwiri et al., 2021). This study aims to address that gap.

Machine learning offers a promising approach to better understand and predict academic performance. By analyzing the relationship between study hours, sleep hours, and academic outcomes, machine learning can reveal complex patterns that traditional methods miss. This research aims to build a predictive model that incorporates these factors, offering more accurate and actionable insights into student success. The ability to identify at-risk students based on study and sleep patterns can enable timely interventions and personalized support, ultimately improving academic achievement.

The scope of this study is specifically focused on the role of study and sleep hours in predicting academic success. While other factors like socio-economic background and previous academic performance may influence outcomes, this study narrows its focus to study habits and sleep patterns. By leveraging machine learning, the study seeks to create a model that predicts student performance based on these two key factors, providing valuable insights for educational institutions to improve student outcomes through targeted interventions.

1.6 Structure of the Paper

This paper is organized into several sections that provide a comprehensive examination of the research topic. The Introduction introduces the background, outlines the problem statement, and defines the objectives of the study. The Literature Review presents an in-depth discussion of existing research on predicting academic performance, focusing on the roles of study and sleep hours, and the application of machine learning models in educational settings. In the Research Methodology, the paper details the data collection process, the machine learning models employed, and the evaluation metrics used to assess model performance. The Results and Discussion section presents the findings from the analysis, compares the performance of the machine learning models, and discusses the implications of the results. Finally, the Conclusion summarizes the key findings of the study, acknowledges its limitations, and suggests potential directions for future research, along with practical recommendations for improving academic performance prediction.

2. Literature Review

2.1 Overview of Academic Performance Prediction

Predicting academic performance has become a crucial area of research, as educational institutions increasingly aim to enhance student success. Traditional methods of academic

assessment, such as grades, standardized test scores, and class attendance, provide valuable insights but often fail to account for the multifaceted nature of academic achievement. Academic performance is influenced not only by elements, such as study habits, sleep patterns, time management, and socio-economic background (Rajendran et al., 2022). As these factors become more recognized, there is a growing need for more comprehensive models that can predict student success by considering a wider range of influencing variables.

Machine learning has become a powerful tool in predicting academic performance due to its ability to process large datasets and uncover complex, non-linear relationships among various factors. Unlike traditional statistical methods, machine learning models can analyze and learn from vast amounts of student data, identifying patterns that may not be obvious at first glance. For example, study hours and sleep patterns directly affect functions like memory, concentration, and problem-solving abilities, all of which play a significant role in academic success. By integrating data on study habits, sleep, and other relevant factors, machine learning models can make more accurate predictions about student performance than conventional models.

In addition to improving prediction accuracy, machine learning offers the potential for early identification of students at risk of poor academic performance. By detecting patterns that suggest a student may be struggling—such as inadequate study time or sleep deprivation—educators and administrators can implement timely interventions. These interventions could include tutoring, counseling, or changes to study plans. Ultimately, machine learning models allow for a more proactive approach to supporting student success, shifting the focus from reactive measures to early, personalized interventions that can lead to improved academic outcomes, better retention rates, and greater overall student satisfaction.

2.2 Key Factors Affecting Academic Performance

Academic performance is influenced by a combination of behavioral, and external factors. Understanding these factors is essential to improving student outcomes and predicting academic success more accurately. The key factors affecting academic performance include:

Study Hours: One of the most significant predictors of academic success is the amount of time students dedicate to studying. Research shows that students who allocate more hours to studying generally perform better academically. However, it is not just the quantity of study hours that matters, but also the quality. Effective study habits, such as active learning, proper time management, and minimizing distractions, can significantly improve the benefits of study time, helping students retain information and perform well on exams and assignments.

Sleep Quality and Quantity: Adequate sleep plays a critical role in functions such as memory consolidation, focus, and problem-solving abilities, all of which are necessary for academic success. Students who get sufficient sleep tend to have better concentration, greater retention of information, and more effective problem-solving skills (Yao & Wang, 2023). In contrast, sleep deprivation can impair performance, leading to reduced focus and memory recall, and increasing the likelihood of poor academic outcomes. Research shows that students who maintain regular sleep patterns generally perform better academically.

Prior Academic Performance: A student's previous academic performance is often a strong indicator of future success. Students with a history of high academic achievement are likely to continue excelling due to the development of good study habits, higher motivation, and increased confidence in their abilities. Students with lower prior academic performance may struggle more, but with targeted support and interventions, they can improve.

Socio-Economic Background: A student's socio-economic status plays a significant role in their academic success. Students from higher socio-economic backgrounds often have greater access to educational resources, such as tutoring, technology, extracurricular activities, and a supportive home environment. These advantages contribute to better academic performance.

Emotional and Mental Well-Being: Emotional well-being and mental health are crucial to academic success. Students facing high levels of stress, anxiety, or depression may find it difficult to concentrate, retain information, and stay engaged with their studies. Mental health challenges can detract from academic performance, making emotional and psychological support vital for student success.

2.3 Importance of Study Hours in Academic Performance

Study hours play a fundamental role in academic success, as the amount of time a student dedicates to studying is directly related to their understanding of course material and overall performance. Consistent and focused study time allows students to grasp key concepts, prepare for exams, and complete assignments effectively. Research has consistently shown that students who spend more time studying are better equipped to perform well academically, as they have more opportunities to review and reinforce what they have learned. In essence, the time invested in study sessions increases the likelihood of academic achievement.

However, the relationship between study hours and academic performance is not purely about quantity. It is equally important how students spend their study time. Simply increasing the number of study hours does not guarantee improved academic results if the study time is spent inefficiently. Effective study habits, such as active learning techniques, structured study schedules, and focused sessions, can significantly enhance the quality of study hours (Dabhade et al., 2021). Students who incorporate techniques like spaced repetition, self-testing, and active problem-solving are likely to retain information more effectively and perform better academically compared to those who study for longer periods without employing these strategies.

Moreover, study hours are intertwined with other factors such as sleep, attendance, and participation. It is essential for students to maintain a balanced approach, ensuring they dedicate adequate time to study while also getting sufficient rest to support function. Sleep deprivation can hinder concentration and memory, reducing the benefits of study hours. Therefore, successful students not only prioritize study time but also manage their overall well-being, striking a balance between academic demands and self-care. This holistic approach to time management ultimately leads to better academic performance and long-term success.

2.4 Impact of Sleep Hours on Academic Performance

Sleep plays a crucial role in academic success, significantly influencing function, memory retention, and overall performance in academic tasks. Research has shown that adequate sleep is essential for students to process and consolidate information learned throughout the day (Medikonda, 2023). During sleep, the brain organizes and stores memories, which is vital for recalling information during exams and assignments. Students who consistently get enough sleep are better able to concentrate, think critically, and perform well in class. In contrast, sleep deprivation can impair abilities, leading to decreased attention, memory lapses, and slower processing speeds, all of which can negatively affect academic performance.

The impact of sleep on academic performance is particularly evident when comparing students who have irregular sleep patterns to those who maintain a consistent sleep schedule. Irregular sleep can disrupt the body's natural circadian rhythm, affecting alertness and overall function.

Students who experience poor sleep quality or insufficient sleep are more likely to struggle with focus and comprehension during class, leading to lower participation and engagement (Atlam et al., 2022). Moreover, lack of sleep increases the likelihood of academic procrastination, as students feel fatigued and unable to efficiently complete assignments and study effectively. As a result, they are less prepared for exams and often perform poorly.

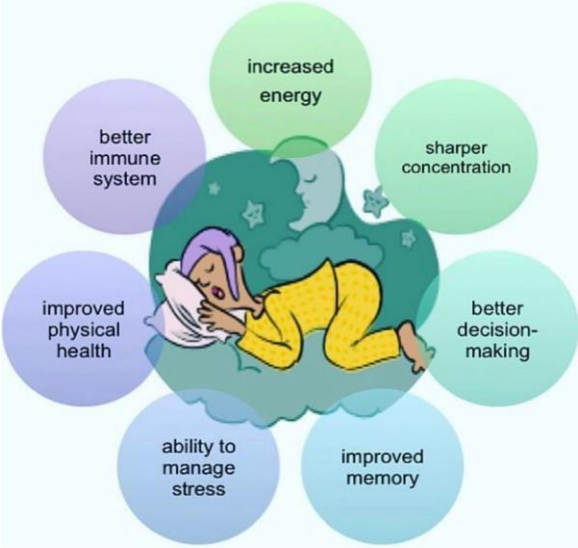

Figure 1: Impact of Sleep on Academic Performance

On the other hand, students who prioritize sleep and maintain a healthy sleep routine tend to have better academic outcomes. Studies indicate that students who sleep 7 to 9 hours per night are more likely to have better grades and higher function than those who consistently sleep less. A good night's sleep supports not only academic performance but also emotional well-being, allowing students to manage stress and maintain a positive attitude toward learning. Therefore, adequate sleep is an indispensable factor in achieving academic success, highlighting its critical role alongside study hours and other academic habits.

2.5 Role of Machine Learning in Education

Machine learning (ML) has rapidly become a powerful tool in education, revolutionizing how institutions assess and improve student performance. Unlike traditional educational methods that rely on grades, test scores, and attendance, machine learning models can process a wide range of data, including study habits, sleep patterns, socio-economic factors, and more. By analyzing these factors, machine learning algorithms can uncover complex relationships and patterns in student behavior, allowing for more accurate predictions of academic success. This capability can help educators make data-driven decisions to identify students who may need additional support or interventions.

One of the key benefits of machine learning in education is its predictive power. ML algorithms can be trained to analyze historical student data and forecast future academic outcomes. This can help in identifying students at risk of underperforming or dropping out, enabling early intervention strategies such as targeted tutoring, mentoring, or counseling (Bitkina et al., 2022). Furthermore, as new data is continuously fed into the system, machine learning models can adapt and improve over time, ensuring that predictions are constantly refined. This dynamic approach to student success not only helps prevent academic failure but also helps create a more proactive educational environment.

In addition to prediction, machine learning plays a significant role in personalizing the learning experience for students. Adaptive learning systems powered by machine learning are capable of tailoring lessons to suit each student's learning pace and style. These systems analyze a student's performance on assignments, quizzes, and tests to identify areas of strength and weakness. Based on this information, they adjust the difficulty level and type of content provided, ensuring that students receive the support they need to progress at their own pace. This personalized approach helps engage students more effectively, improving overall learning outcomes.

2.6 Machine Learning Models for Academic Prediction:

Machine learning models have become widely adopted tools for predicting academic performance, thanks to their ability to handle complex datasets and identify patterns that are often hidden in traditional analyses. These models can incorporate various factors, such as study habits, sleep patterns, attendance, and socio-economic background, to provide more accurate predictions of student outcomes. By using machine learning for academic prediction, institutions can gain a deeper understanding of the variables that influence student success and design more targeted interventions to support struggling students.

One of the most commonly used algorithms in academic performance prediction is **Random Forest**. This ensemble learning method constructs multiple decision trees and aggregates their predictions to enhance accuracy. Random Forest is highly effective at handling both numerical and categorical data, making it ideal for analyzing various factors that influence academic success. It is particularly robust when working with large datasets that contain complex, non-linear relationships. Additionally, Random Forest can assess the importance of different features, helping to identify key predictors of student performance. Another powerful model is **Support Vector Machine (SVM)**, a supervised learning algorithm used for classification and regression tasks. SVM excels in situations with high-dimensional data, finding the optimal hyperplane that separates classes. This model is especially useful for identifying students at risk of underperforming by classifying them into distinct performance categories based on their study habits and sleep patterns.

The **Decision Tree** algorithm also plays a key role in academic prediction. It recursively splits data based on feature values, creating a tree-like structure that is simple to interpret. This transparency makes Decision Trees valuable in educational settings, where understanding the reasoning behind predictions is crucial. However, they are prone to overfitting, so techniques such as pruning are used to improve generalization. **Logistic Regression** is another widely used model, especially in binary classification tasks, such as predicting whether a student will pass or fail based on study and sleep hours (Anbarasi et al., 2022). Despite its simplicity, Logistic Regression is effective when the relationship between input features and the output is linear. Lastly, **Naive Bayes**, a probabilistic classifier based on the assumption of feature independence, is highly effective for large datasets with categorical data. It estimates the likelihood of different academic outcomes, making it a useful tool for predicting student performance.

2.7 Previous Research on Study and Sleep Hours in Academic Prediction

Numerous studies have explored the relationship between study hours and academic performance, consistently showing that the more time students dedicate to studying, the better their academic outcomes. However, it is not just the quantity of study hours that matters, but also how effectively that time is used. Students who employ effective time-management strategies and engage in active learning techniques during their study sessions are more likely to achieve higher academic success. The quality of study time plays a significant role in determining its impact on academic performance.

In addition to study time, sleep has been identified as a critical factor in influencing academic performance. Adequate sleep supports functions such as memory consolidation, attention, and problem-solving, which are essential for academic achievement. On the other hand, sleep deprivation can lead to difficulties with concentration, memory retention, and information recall, all of which impair academic performance. Students who maintain regular sleep patterns tend to perform significantly better academically than those who experience sleep deprivation.

Numerous machine learning studies have incorporated both study hours and sleep patterns to predict academic performance more accurately. It has been demonstrated that machine learning models incorporating both study and sleep data

are more effective in predicting student success than those relying solely on academic history or grades. Additionally, including these behavioral factors in predictive models improves their accuracy in forecasting student grades, showing that a holistic approach to academic prediction, which includes study and sleep patterns, can lead to better results.

3. Research Methodology

3.1 Research Design

This study adopts a quantitative research design to investigate the use of machine learning models for predicting the academic performance of college students, with a particular focus on the role of study hours and sleep hours. Quantitative research is suitable for this study as it allows for the application of statistical, computational, and machine learning techniques to analyze numerical data and derive insights. The primary objective is to build predictive models that accurately estimate academic performance based on the identified key factors—study hours and sleep hours—while exploring their impact on student success.

The research follows a structured approach, beginning with data collection, followed by data preprocessing to clean and prepare the dataset. Feature selection will be carried out to identify the most significant predictors, and machine learning algorithms, such as Random Forest, Decision Tree, and Support Vector Machine, will be applied for prediction. The final step involves evaluating the models' performance using various metrics, including accuracy, precision, recall, and F1 score, to ensure reliable predictions of student outcomes.

3.2 Data Collection

Data collection for this study focuses on obtaining quantitative data regarding study hours, sleep hours, and academic performance from college students. The primary aim is to gather relevant data to create predictive models for academic success. Key variables to be collected include the number of hours students dedicate to studying each week, the average number of sleep hours they get nightly, and their academic performance, which is assessed using GPA or final grades. In addition to these main variables, the survey will inquire about other factors that may influence academic outcomes, such as class attendance, extracurricular involvement, and participation in assignments.

A structured survey will be administered to collect the data, asking students to self-report their study and sleep habits. Ethical considerations are crucial during data collection. Participation in the study is voluntary, and confidentiality will be maintained to protect students' personal information. Informed consent is obtained from all participants, ensuring they understand the purpose of the study and how their data will be used. The collected data will serve as the foundation for the machine learning models used to predict academic performance. These models will identify patterns and correlations between study hours, sleep hours, and academic outcomes, helping to reveal the most influential factors impacting student success.

3.3 Data Preprocessing

Data preprocessing is crucial for preparing raw data for machine learning models. It ensures that the data is clean, consistent, and in the right format to extract meaningful insights. Since raw data often contains errors, missing values, outliers, or irrelevant information, preprocessing steps are necessary to enhance the quality of the dataset. Proper preprocessing helps improve the accuracy and effectiveness of the predictive models by transforming the data into a form suitable for analysis. The following steps outline the key processes involved in data preprocessing:

Handling Missing Values: Missing values in numerical data are replaced with the mean or median, depending on the variable's distribution. For categorical variables, missing values are filled with the most frequent value (mode).

Normalization of Continuous Variables: Continuous variables such as study hours and sleep hours are normalized using min-max scaling or z-score normalization to ensure consistency

and avoid disproportionate influence from features with larger scales.

Outlier Detection and Removal: Outliers are identified using statistical methods, such as Z-scores, and removed if they significantly deviate from the mean, as they may distort the results and impact the predictive power of the model.

Categorical Data Encoding: Categorical variables, such as extracurricular activity participation and class attendance, are encoded using one-hot encoding, transforming them into numerical values suitable for machine learning models.

Data Transformation: Skewed continuous variables undergo transformations, such as log transformations, to make them more normally distributed, improving the model's ability to capture relationships accurately.

Feature Selection: Irrelevant or highly redundant features are removed through feature selection methods, such as correlation matrices, to ensure only the most impactful variables are used in the predictive models.

3.4 Data Splitting

Data splitting is a crucial step in machine learning, as it ensures that the model can be properly trained, fine-tuned, and evaluated for its generalization ability. In this study, the dataset is divided into three main subsets: the training set, validation set, and testing set. The training set typically comprises 60-70% of the total data. This subset is used to train the machine learning models, allowing the algorithms to learn the relationships between the input features, such as study hours and sleep hours, and the dependent variable, academic performance. The model learns from this data and adjusts its parameters accordingly to identify patterns and improve predictive accuracy.

The remaining 30-40% of the data is split into the validation and testing sets. The validation set, representing 15-20% of the data, is used during the training process to fine-tune the model's hyperparameters. It helps assess the model's performance on unseen data and prevent overfitting by providing regular feedback during training. Once the model is trained and tuned, the testing set, also making up 15-20% of the total data, is used to evaluate the final model's performance. The testing set provides an unbiased evaluation of the model's predictive capabilities, using metrics such as accuracy, precision, recall, and F1 score. This three-way data splitting ensures that the model is trained effectively, fine-tuned for optimal performance, and validated for real-world applicability, reducing the risk of overfitting and ensuring more reliable predictions.

3.5 Machine Learning Models Used (Random Forest, Decision Tree, SVM)

Various machine learning algorithms are employed in this study to predict academic performance based on key factors, such as study hours and sleep hours. Each algorithm offers distinct advantages, depending on the complexity of the data and the type of predictions needed. The choice of models is critical to ensuring reliable and accurate predictions. The following machine learning algorithms are used to evaluate their effectiveness in predicting student outcomes: Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and Naive Bayes. Each model is selected for its ability to handle different aspects of the data, from simplicity and interpretability to handling high-dimensional relationships and non-linearity in the data.

Logistic Regression: Logistic Regression is a widely used statistical model for binary classification tasks. It predicts the probability of an outcome by modeling a linear relationship between the input features and the log-odds of the target variable. For this study, logistic regression can be applied to predict binary academic outcomes, such as whether a student is likely to pass or fail based on factors like study hours and sleep hours. Despite its simplicity, logistic regression is effective for understanding

relationships in data and offers ease of interpretation with its output probabilities.

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy. Each tree is trained on a random subset of the data, and their individual predictions are aggregated to produce a final outcome. This approach helps reduce overfitting and increases model robustness. Random Forest is capable of handling both numerical and categorical data and is particularly useful for identifying the most important features influencing academic performance.

Decision Tree: Decision Tree is a simple, yet powerful, algorithm that splits data into subsets based on feature values, creating a tree structure. It is highly interpretable, making it useful for understanding how decisions are made based on the input features. However, Decision Trees are prone to overfitting, especially with complex datasets. Pruning techniques are often applied to reduce overfitting and enhance generalization.

Support Vector Machine (SVM): Support Vector Machine (SVM) is a robust classifier that finds an optimal hyperplane to separate different classes in the feature space. It is particularly effective for high-dimensional and complex data, making it ideal for predicting academic performance based on study and sleep hours. SVM handles non-linear relationships by applying kernel functions, which transform the data into higher dimensions, allowing better classification accuracy.

K-Nearest Neighbors (KNN): Naive Bayes is a probabilistic classifier that assumes the features used to predict the target variable are independent. Despite this assumption, it performs well in tasks like text classification and predicting academic outcomes. It is especially effective with large datasets and categorical data. In this study, Naive Bayes helps predict academic performance by estimating the probability distribution of features such as study hours and sleep hours, offering a reliable model for student success prediction.

3.6 Evaluation and Prediction:

Evaluation and prediction are critical steps in assessing the performance of machine learning models. In this study, the effectiveness of the predictive models is determined using several evaluation metrics: accuracy, precision, recall, and F1 score. Accuracy is the most straightforward metric, representing the overall proportion of correct predictions made by the model. However, in cases of class imbalance, accuracy alone may not provide a complete picture. Precision measures the proportion of true positive predictions among all positive predictions made by the model, ensuring that the predictions made are relevant and minimizing false positives. Recall focuses on the proportion of actual positive instances correctly identified by the model, highlighting the model's ability to detect all relevant cases.

F1 score is the harmonic mean of precision and recall, offering a balanced evaluation metric when both false positives and false negatives carry significant weight. This is especially useful when the dataset has an unequal distribution of classes. The performance of each machine learning model, such as Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and Naive Bayes, is assessed using these metrics to determine which model performs best in predicting academic outcomes based on study and sleep hours. By using multiple metrics, the study ensures that the selected model not only makes accurate predictions but also balances between precision and recall, providing a more comprehensive assessment of model performance.

4. Results and Discussion

4.1 Dataset Description

The dataset used in this study is sourced from Kaggle, a platform for data science and machine learning projects. This dataset includes a variety of attributes related to college students' academic performance, study habits, and lifestyle factors. The primary attributes in the dataset are **study hours** (the number of hours spent studying each week) and **sleep hours** (the average number of hours of sleep per night). These factors are

expected to significantly impact academic performance, measured by GPA or final course grades.

In addition to study and sleep hours, the dataset also includes other attributes such as class attendance, participation in extracurricular activities, socio-economic background, previous academic performance, and study techniques. These attributes provide a more comprehensive understanding of the various factors that contribute to student success. The data is pre-processed to handle any missing values, outliers, and normalization of continuous variables, ensuring its suitability for machine learning analysis. This dataset enables the exploration of the relationships between study and sleep hours and academic outcomes, making it ideal for developing predictive models that aim to improve student success based on these key variables.

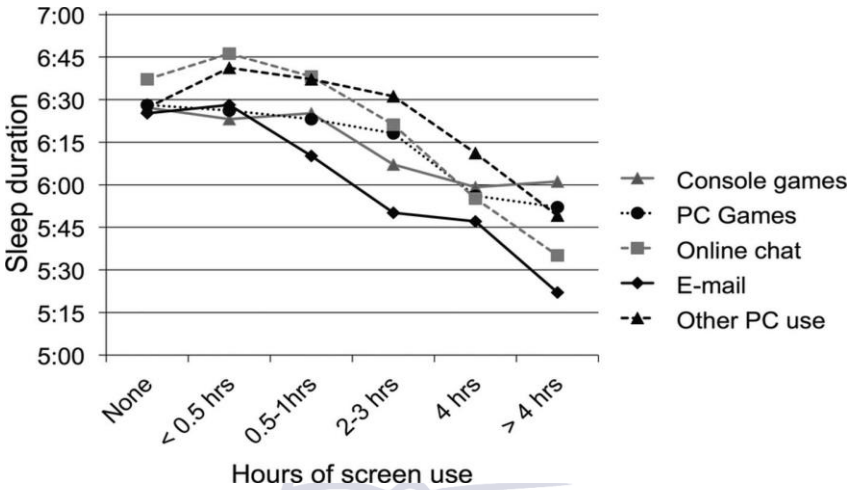

Figure 2: Relationship between Screen Time and Sleep Duration

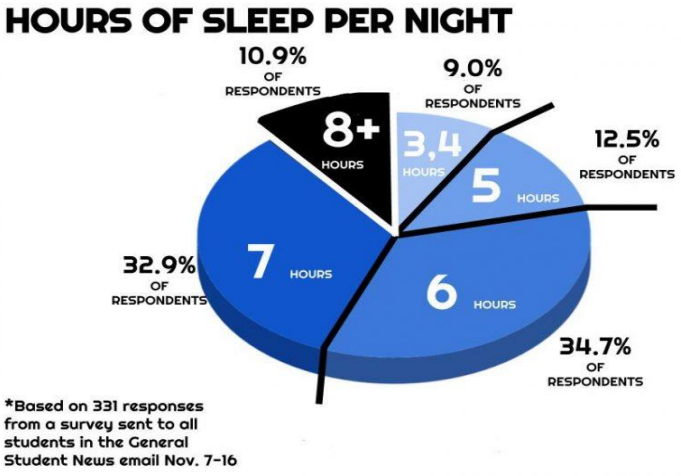

Figure 3: Distribution of Sleep Hours among Respondents

Figure 4: Recommended Hours of Sleep by Age Group

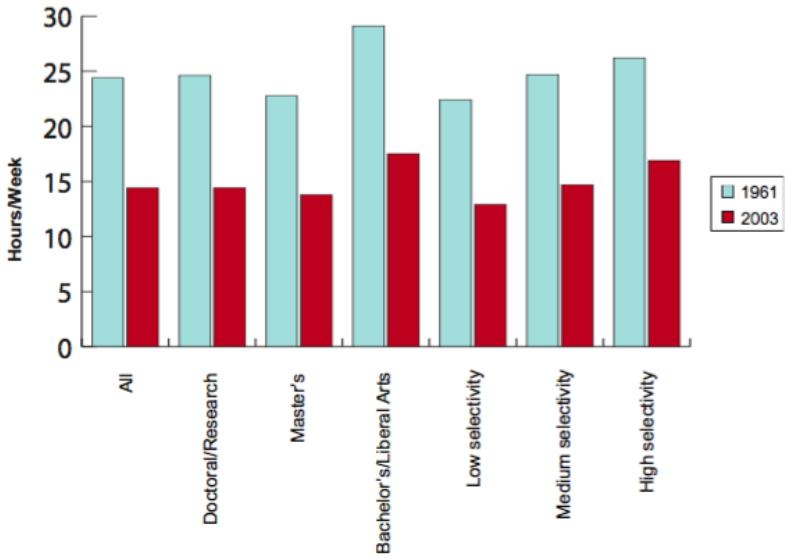

Figure 5: Correlation Study hours

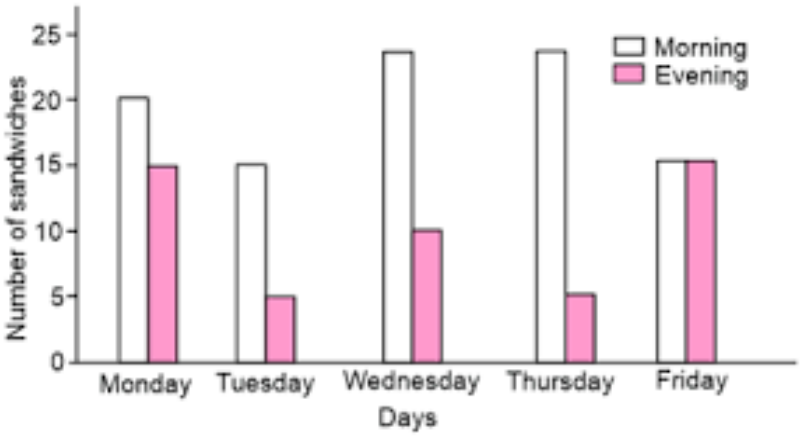

Figure 6: Study hours during week days

4.2 Performance Comparison of Models

In this study, several machine learning models were trained and evaluated to predict academic performance based on factors such as study hours and sleep hours. The models used in this analysis include Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and Naive Bayes. To assess their predictive accuracy and reliability, the models are evaluated using various metrics, including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive view of model performance, helping to identify which model is most effective for predicting student academic outcomes. Accuracy reflects the overall correctness of predictions, while precision and recall balance false positives and false negatives, and the F1 score offers a combined measure of both.

The performance of each model is compared by analyzing these evaluation metrics on the testing dataset, which was kept separate from the training data to ensure unbiased evaluation. Random Forest and Support Vector Machine (SVM) tend to perform well in handling complex and non-linear relationships within the data, making them strong contenders for the best predictive models. Logistic Regression and Decision Trees, while simpler, provide insights into the relationships between the key features and academic performance. Naive Bayes, with its simplicity and effectiveness in handling categorical data, also offers competitive results. The comparison of these models helps determine which algorithm provides the best balance of accuracy, precision, and generalizability for predicting academic performance based on study habits and sleep patterns.

5. Contributions

5.1 Theoretical Contributions

This study offers several significant theoretical contributions to the field of academic performance prediction using machine learning. By incorporating factors, such as study hours, sleep hours, and socio-economic background, alongside traditional academic measures, the research challenges conventional academic success models that focus primarily on abilities. It expands existing theories by providing a more comprehensive understanding of how various behavioral and environmental factors contribute to student outcomes. This broader perspective enriches academic performance frameworks, offering valuable insights into the predictors of academic success and suggesting areas for future research and model development.

Expanding Understanding of Academic Success

This study significantly enhances the theoretical understanding of academic success by integrating factors such as study habits and sleep patterns with traditional academic metrics like grades and attendance. This approach moves beyond focused models and emphasizes the influence of behavioral factors on academic performance. By providing a holistic view, the research broadens the scope of what contributes to student achievement, offering new insights into the diverse factors that affect success. The findings encourage a reevaluation of existing theories and suggest that integrating factors is essential for more accurate academic performance predictions.

Contributions to Machine Learning Applications in Education

This study contributes to the theoretical application of machine learning in educational contexts by evaluating and comparing various models. The research identifies the most effective models for predicting academic performance based on study and sleep hours, highlighting which algorithms work best for educational data analysis. These findings offer valuable insights for selecting appropriate machine learning methods in academic prediction tasks. Moreover, the study suggests future directions, including integrating factors into machine learning models to improve accuracy and make them more comprehensive.

5.2 Practical Contributions

The findings of this study provide valuable insights that can be applied in educational settings to improve student performance. By understanding the impact of study hours and sleep habits on academic outcomes, educators

and administrators can develop more effective strategies and interventions tailored to students' needs. The ability to predict academic success based on these factors allows for better-targeted support that can enhance overall student achievement.

Targeted Support Strategies

Understanding how study habits and sleep patterns influence academic success enables educators to design interventions that address. Workshops on effective study techniques, time management, and promoting healthy sleep habits can help students improve their academic performance. These strategies help institutions provide a more comprehensive approach to supporting students, fostering both academic and personal growth.

Early Identification of At-Risk Students

Machine learning models used in this study can predict students who may be at risk of underperforming by analyzing their study and sleep hours in relation to their academic outcomes. By identifying students at risk early, administrators can implement timely interventions, such as personalized counseling, tutoring, or academic workshops. This proactive approach helps prevent students from falling behind, increasing the likelihood of academic success and improving retention rates.

Optimizing Resource Allocation

Data-driven insights from this study enable institutions to allocate resources more effectively. By identifying students who need additional support, educators and administrators can prioritize resources like tutoring services, counseling, and academic workshops to those who would benefit most. This approach ensures that resources are used efficiently, providing targeted help where it is needed most.

Promoting Holistic Student Well-Being

In addition to academic interventions, understanding the role of sleep and study habits can lead to initiatives that promote overall student well-being. By encouraging healthy sleep routines and balanced study schedules, institutions can help students not only improve their academic outcomes but also support their mental and physical health. This holistic approach to student success contributes to a healthier, more engaged student body.

Conclusion

In conclusion, this study demonstrates the potential of machine learning in predicting academic performance. such as study hours and sleep hours. The findings highlight the significant impact that study hours and sleep time patterns have on student success, providing a more holistic approach to understanding academic achievement. By integrating these factors into academic performance models, the study offers a deeper insight into the variables that influence student outcomes and challenges traditional approaches that focus solely on abilities.

The results of this research have valuable practical implications for educators and administrators. By using machine learning to predict academic outcomes, institutions can identify at-risk students early and implement targeted interventions tailored to their needs. The insights gained from this study enable institutions to optimize resource allocation, design personalized support strategies, and improve overall student success. This data-driven approach not only contributes to better academic performance but also fosters more supportive and proactive learning environments. Ultimately, the study demonstrates the importance of considering both factors in academic prediction and provides actionable recommendations for enhancing student achievement and well-being in educational settings.

Future Work

While this study has successfully demonstrated the applicability of machine learning models in predicting academic performance, future research can build upon these findings by incorporating additional factors that influence student success. While study hours and sleep patterns were the focus of this study, other factors such as stress, mental health, motivation, time management, and social support can provide valuable insights.

Including these variables could significantly improve the accuracy of predictive models and offer a more comprehensive understanding of the factors contributing to academic success. Expanding the dataset to include a larger and more diverse sample from various educational institutions could further enhance the generalizability of the findings.

Moreover, future studies can explore the use of more advanced machine learning techniques, such as deep learning and neural networks, to handle more complex, non-linear relationships within the data. Longitudinal studies that track students' academic progress over multiple semesters or years could provide a deeper understanding of how study habits and sleep patterns evolve over time. Additionally, integrating real-time data from wearable devices or learning management systems would allow for continuous monitoring of student behavior. This would enable more dynamic, timely interventions tailored to individual student needs. Such advancements would enhance the predictive power of machine learning models, creating adaptive, data-driven support systems that optimize student outcomes and improve academic success in educational environments.

References

Alghwiri, A.A., Almomani, F., Alghwiri, A.A. and Whitney, S.L., 2021. Predictors of sleep quality among university students: the use of advanced machine learning techniques. *Sleep and Breathing*, *25*, pp.1119-1126. https://link.springer.com/article/10.1007/s11325-020-02150-w

Anbarasi, L.J., Jawahar, M., Ravi, V., Cherian, S.M., Shreenidhi, S. and Sharen, H., 2022. Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. *Health and Technology*, *12*(4), pp.825-838. https://link.springer.com/article/10.1007/s12553-022-00674-7

Atlam, E.S., Ewis, A., Abd El-Raouf, M.M., Ghoneim, O. and Gad, I., 2022. A new approach in identifying the psychological impact of COVID-19 on university student's academic performance. *Alexandria Engineering Journal*, *61*(7), pp.5223-5233. https://www.sciencedirect.com/science/article/pii/S1110016821007171

Bitkina, O.V., Park, J. and Kim, J., 2022. Modeling sleep quality depending on objective actigraphic indicators based on machine learning methods. *International Journal of Environmental Research and Public Health*, *19*(16), p.9890. https://www.mdpi.com/1660-4601/19/16/9890

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

Holicza, B. and Kiss, A., 2023. Predicting and comparing students' online and offline academic performance using machine learning algorithms. *Behavioral Sciences*, *13*(4), p.289. https://www.mdpi.com/2076-328X/13/4/289

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

Medikonda, J., 2023. A clinical and technical methodological review on stress detection and sleep quality prediction in an academic environment. *Computer methods and programs in biomedicine*, *235*, p.107521.

https://www.sciencedirect.com/science/article/pii/S0169260723001864

Nan, J., Herbert, M.S., Purpura, S., Henneken, A.N., Ramanathan, D. and Mishra, J., 2024. Personalized machine learning-based prediction of wellbeing and empathy in healthcare professionals. *Sensors*, *24*(8), p.2640. https://www.mdpi.com/1424-8220/24/8/2640

Rahman, A.A., Siraji, M.I., Khalid, L.I., Faisal, F., Nishat, M.M., Ahmed, A. and Al Mamun, M.A., 2022, June. Perceived stress analysis of undergraduate students during covid-19: a machine learning approach. In *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)* (pp. 1129-1134). IEEE. https://ieeexplore.ieee.org/abstract/document/9843081/

Rajendran, S., Chamundeswari, S. and Sinha, A.A., 2022. Predicting the academic performance of middle-and high-school students using machine learning algorithms. *Social Sciences & Humanities Open*, *6*(1), p.100357. https://www.sciencedirect.com/science/article/pii/S2590291122001115

Schmickler, J.M., Blaschke, S., Robbins, R. and Mess, F., 2023. Determinants of sleep quality: a cross-sectional study in university students. *International journal of environmental research and public health*, *20*(3), p.2019. https://www.mdpi.com/1660-4601/20/3/2019

Yao, N. and Wang, Q., 2023. Technostress from smartphone use and its impact on university students' sleep quality and academic performance. *The Asia-Pacific Education Researcher*, *32*(3), pp.317-326. https://link.springer.com/article/10.1007/s40299-022-00654-5

Zhang, W., Wang, Y. and Wang, S., 2022. Predicting academic performance using tree-based machine learning models: A case study of bachelor students in an engineering department in China. *Education and Information Technologies*, *27*(9), pp.13051-13066. https://link.springer.com/article/10.1007/s10639-022-11170-w