Journal of Applied Data Sciences
Vol. 5, No. 4, December 2024, pp. 2025-2038

ISSN 2723-6471
2025

Leveraging Data Analytics for Student Grade Prediction: A Comparative Study of Data Features

Misinem[1,*], Tri Basuki Kurniawan[2], Deshinta Arrova Dewi[3], Mohd Zaki Zakaria[4], Che Mohd Alif Nazmi[5]

[1]*Faculty of Vocational, Universitas Bina Darma, Palembang, Indonesia*

[2]*Postgraduate Program, Universitas Bina Darma, Palembang, Indonesia*

[3]*Faculty of Data Science and Information Technology, INTI International University, Malaysia*

[4,5]*Faculty of Computer & Mathematics Sciences, University Technology Mara, Malaysia*

(Received: June 23, 2024; Revised: August 21, 2024; Accepted: October 29, 2024; Available online: November 7, 2024)

Abstract

In educational settings, a persistent challenge lies in accurately identifying and supporting students at risk of underperformance or grade retention. Traditional approaches often fall short by applying generalized interventions that fail to address specific academic needs, leading to ineffective outcomes and increased grade repetition. This study advocates for integrating machine learning algorithms into educational assessment practices to address these limitations. By leveraging historical and current performance data, machine learning models can help identify students needing additional support early in their academic journey, allowing for precise and timely interventions. This research examines the effectiveness of three machine learning algorithms: Naive Bayes, Deep Learning, and Decision Trees. Naive Bayes, known for its simplicity and efficiency, is well-suited for initial data screening. Deep Learning excels at uncovering complex patterns in large datasets, making it ideal for nuanced predictions. Decision Trees, with their interpretable and actionable outputs, provide clear decision paths, making them particularly advantageous for educational applications. Among the models tested, the Decision Tree algorithm demonstrated the highest performance, achieving an accuracy rate of 86.68%. This high precision underscores its suitability for educational contexts where decisions need to be based on reliable, interpretable data. The results strongly support the broader application of Decision Tree analysis in educational practices. By implementing this model, educational administrators can better identify at-risk students, tailor interventions to meet individual needs, and ultimately improve student success rates. This study suggests that Decision Trees could become a vital tool in data-driven strategies to enhance student retention and optimize academic outcomes.

Keywords: Student Grade Prediction, Data Analytic Model, Machine Learning, Education Quality

1. Introduction

Grading in education refers to the process of applying standardized measurements to assess varying levels of achievement within an educational course. Grades are often represented by letters, such as A to F, or numerically, where values range from 1 to 6, with 1 typically being the highest and 6 the lowest. This measurement reflects the percentage of questions answered correctly or the number of points achieved out of a total possible score, such as 20 or 100 points.

In some countries, grading systems are based on the GPA (Grade Point Average) format, which averages grades from all courses a student takes in a semester to create an overall score used for academic evaluation [1]. In Malaysia, for instance, GPA is used exclusively for university-level students, while primary and secondary school students receive grades based on individual test scores, rated from 0 to 100 [2]. Typically, the highest marks correspond to an A, while the lowest are assigned an F. However, in primary and secondary schools, grades from early, middle, and final terms are not cumulative; each term's grades stand independently. In contrast, university students use both GPA and CGPA

***Corresponding author: Misinem (misinem@binadarma.ac.id)**
DOI: https://doi.org/10.47738/jads.v5i4.442

 ISSN 2723-6471
2026

(Cumulative Grade Point Average) formats, where CGPA reflects an average of all semester grades, giving a comprehensive measure of academic performance over time [3].

A common issue in educational institutions is grade retention, where students who do not meet the required standards must repeat a grade. This often results from a failure to diagnose and support at-risk students before their final examinations. When early intervention is lacking, teachers are unable to provide the necessary support to students who may require additional help, affecting their final academic performance and resulting in lower grades. Consequently, teachers may not have a clear understanding of each student's unique needs, leading to a uniform approach that may not be effective for all learners [4]. To improve the support provided to students, advanced analytical tools, including machine learning algorithms such as Naive Bayes, Deep Learning, and Decision Trees, are increasingly used. These tools analyze student data to identify patterns and predict outcomes like grades or the likelihood of retention.

This study compares the effectiveness of Naive Bayes, Deep Learning, and Decision Trees in educational contexts. Naive Bayes offers simplicity and efficiency, particularly with large datasets, making it suitable for initial screenings. Deep Learning models excel in capturing complex relationships within data, offering robust predictions that are adaptable to various educational environments. Decision Trees, on the other hand, produce interpretable models that can guide educators in implementing targeted interventions based on clear decision paths.

By comparing these algorithms, the research aims to identify the most suitable model for educational purposes, balancing accuracy, interpretability, and practical implementation [5]. Implementing the most effective algorithm can help institutions support students more effectively through personalized learning strategies, potentially enhancing academic success and reducing grade retention.

2. Factors Affecting Student's Grade

The excellence of the students' performances is still a main priority for teachers and educators. It is created by making differences locally, regionally, and nationally, and it can be global. Educators, teachers, trainers, and researchers have been aware of and interested in researching all the variables that can help them effectively improve the quality of their student's performances. These variables are found inside and outside the school and can affect students' academic performance. Some of these variables, also called factors, can be described as student, school, family, financial, and peer factors [6].

Some of the student's performances may depend on the differences in their socioeconomic, environmental factors, and psychology [7]. Some students have financial problems studying, which can affect their performance badly. As college education costs have rapidly increased in the past few years, extended graduation time has become an important factor in the overgrowing student loan debt [8]. Some other factors compromise a student's learning ability, which can affect the student's performance. These factors are caused by course management, which provides courses that require much time, which makes the students have a hard time following what they are studying and learning. These will cause the students to feel frustrated in their studies.

In a study concerning factors that can affect the students' performances, researchers found that anxiety and stress were major factors that negatively affect the students' performance. Based on research in a medical school, the researchers found that 92% of the preclinical and clinical students confessed that most of them were suffering from anxiety and stress [9]. Factors such as brain processing, creative thinking, and culture can influence a student's learning style [10]. Some students prefer learning by hearing, while others prefer a visual study style. Some students make reading and writing their first choice, alternatives, or preferences to comprehend and accommodate information [11]. Teachers must understand their students' learning preferences and make plans and strategies to help them learn.

3. Methodology

Research methodology defines the activities of a particular research, ways to proceed, and ways to establish success in an organized way [12]. It can also be defined as how knowledge and data are gained. This research methodology consists of four phases: preliminary study, theoretical and empirical study, architecture design, and system development and evaluation.

 ISSN 2723-6471

3.1. Research Framework

The Research Framework illustrated in figure 1 provides a systematic approach to developing technological solutions, ensuring that projects are effective and efficient in addressing the identified needs [13].

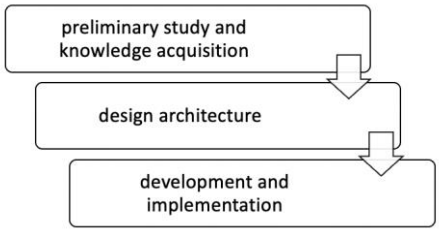

Figure 1. Research Frameworks

The framework comprises three primary stages, starting with the preliminary study and knowledge acquisition phase. In this stage, researchers concentrate on developing a thorough understanding of the problem domain and collecting necessary data through literature reviews, feasibility studies, and consultations with experts. This foundational work ensures the project is grounded in solid knowledge and provides direction for the subsequent phases.

With a strong base established, the process moves into the design architecture phase, where a detailed blueprint of the solution is developed. This blueprint outlines the core components, their interactions, and how they will collectively address the identified issue. The final stage, development and implementation, transitions from theory to application, where the actual construction, coding, and deployment of the solution take place. During this phase, rigorous testing is conducted to verify that the solution meets the initial quality standards and functions as intended. Additionally, the methodology involves a theoretical study to understand the factors influencing student grades, empirical investigations, architectural design, and evaluation. This initial study focuses on identifying the main predictors of student performance and suitable techniques to guide future development.

3.2. Theoretical Study

Theoretical study was the first thing done in doing this research. To achieve the objective of identifying the factors that affect the student's grades. The first approach that was taken was doing a preliminary study. It was being done to study scope, problem statements, objectives, domain, and the factors that affect the student's grades, and also what the normal grading system the government sets, what the grade needed for the students to pass or make them fail. Through this approach, an introduction, problem statement, research question, objective, scope, significance, and Literature Review is formulated.

3.3. Data Understanding

The approach used in this phase was data acquisition. Data acquisition is acquiring information and data from external sources, such as experts, journals, articles, websites, and others. The activities in this phase include finding suitable datasets for the study, understanding the data, and acquiring all the useful information to design and develop the system. In this phase, finding the data was the most important thing to be used in developing the system. In this phase, there will be a study about the factors that affect the student's grades, how the government is setting the grading system, the marks given that decide whether they pass or fail, and how helpful it is when grades can be foreseen and predicted earlier than the actual test.

3.4. System Architecture Design

The process begins with data pre-processing, followed by data analysis, and finally, the design of the prediction model. The accuracies of the Naïve Bayes, Random Forest, and Decision Tree models were calculated using the Auto Model function in RapidMiner Studio [14], [15]. This tool facilitates faster model building and validation by addressing three major types of tasks: prediction, clustering, and outlier detection. For prediction tasks, Auto Model can handle both classification and regression challenges. When used for clustering, it organizes data into groups based on similarity, identifying clusters within the dataset. In the clustering task, Auto Model classifies data points that are closely related, as illustrated in figure 2.

 ISSN 2723-6471
2028

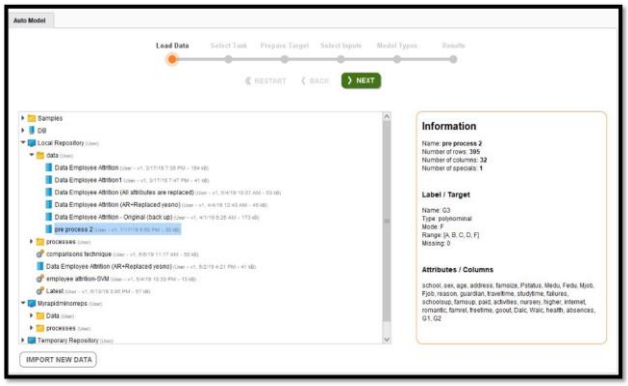

Figure 2. Data Selection for Auto Model

Data was loaded from the local repository after having been preprocessed earlier. The chosen data was Data Preprocessing. After the data is selected, the Auto Model gives three options for selecting a suitable task. For this research, Prediction was chosen to be used in the Auto Model.

3.5. System Development

After the design is complete, the system can then be developed. This research used a dashboard to develop the system and visualize the data analysis. To succeed, this project needs to visualize the data analysis and do testing and evaluation. The approach turns all the research and methodologies into real-world solving problems in system development. System development is an important part of this study for a successful project. Besides that, system development is important to achieve the third objective, which is to develop a system that can predict the students' grades [16], [17].

3.6. System Testing and Implementation

The system is then tested using the testing dataset, and the result will determine whether it is a success or needs more fixing and modifying. When the results are satisfying, the system can be implemented and used by other users.

4. Results and Discussion

After the system's development is completed, the following phase is to analyze its results or findings. The project's result is described as an outcome derived from the data analytical technique used to generate data, and it is concluded with an output after going through the required process.

4.1. Correlation Between Attributes

Following data pre-processing and attribute analysis, the correlation between variables was assessed to determine which attributes most significantly influenced others. A correlation heatmap, shown in figure 3, visually represents the relationships among the 16 attributes, providing insights into the factors affecting academic performance, particularly the final_score. This analysis reveals that parental education has a positive correlation with academic outcomes, with mother_education and father_education each correlating at 0.22 with final_score, suggesting that higher parental education levels are linked to better student performance. Study_time also shows a positive correlation of 0.25 with final_score, indicating that increased study hours are associated with higher grades. Conversely, failures display a strong negative correlation of -0.36 with final_score, implying that students with prior course failures are more likely to achieve lower final grades. This finding underscores the value of early intervention in addressing academic difficulties to prevent a cycle of underperformance [18].

Interestingly, absences display a slight positive correlation of 0.09 with the final_score, suggesting that occasional absences may not significantly impact academic performance. This also indicates that students with higher absences might still perform well academically, though they may benefit from improvement in attendance. Alcohol consumption shows a minor negative correlation with grades, with weekday_alcohol_usage and weekend_alcohol_usage correlating at -0.054 and -0.052, respectively. While these correlations are weak, they point to a subtle adverse effect of alcohol on academic performance.

 ISSN 2723-6471
2029

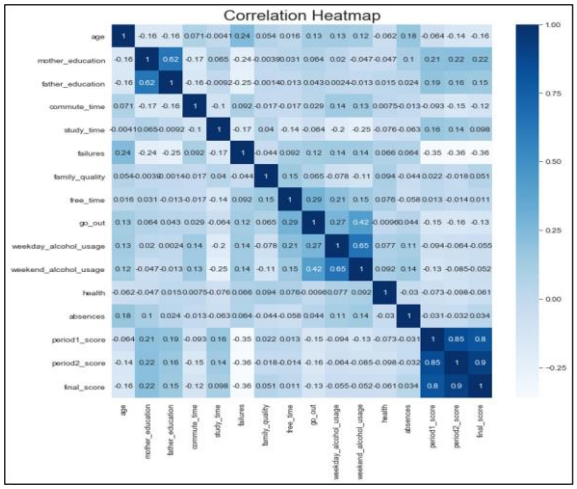

Figure 3. Correlation Heatmap Between Attributes

A critical insight from the heatmap is the strong positive correlation between final_score and scores from earlier periods (with period1_score and period2_score correlating at 0.85 and 0.91, respectively). This underscores that performance in earlier grading periods is a significant predictor of final academic outcomes, highlighting the importance of consistent academic engagement throughout the year. Additionally, the heatmap reveals that factors like health and commute time have minimal correlation with the final score, suggesting that while these factors may impact general well-being, they do not directly influence academic results in this dataset. This analysis sheds light on the key predictors of academic success, such as parental education, study habits, and early academic performance. Recognizing these relationships can assist in developing targeted strategies to support students, potentially enhancing their overall educational outcomes.

4.2. Diagnostic Analysis on Factors that Affecting the Students' Grade

Using diagnostic analytics as the data analytical method, some factors that correlate with final grade attributes were analyzed to see what effect the attributes have and which aspect will affect the student's final grade. Diagnostic Analytics is one type of advanced analytics that examines data to answer the question of "Why did it happen?". It is characterized by techniques such as drill-down, data mining, data discovery, and correlations.

4.3. Desire for Higher Education vs Final Grade

The relationship between students' aspirations for higher education and their academic performance is a valuable area of research, offering insights into predictors and outcomes related to final grades. This connection is essential for educators, policymakers, and students, as it impacts educational strategies, resource allocation, and students' future career paths. Understanding how aspirations shape academic outcomes can inform support strategies to help students achieve their goals [18]. This study examines whether students with strong aspirations for higher education achieve higher final grades, proposing that these students may be more motivated by their goals and expectations. It also considers the influence of factors like socioeconomic status, access to resources, and personal motivation in shaping this relationship. By exploring these dynamics, the study aims to inform more effective educational practices and policies that support students in achieving both short-term academic targets and broader educational aspirations, as depicted in figure 4.

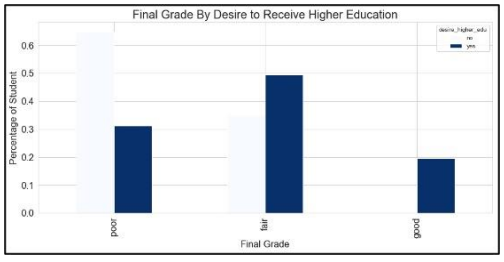

Figure 4. Graph Analysis on Desire vs Final Grade

 ISSN 2723-6471

Figure 4 illustrates the relationship between students' aspirations for higher education and their academic performance, categorizing outcomes as "poor" or "fair." The data shows that students with aspirations for higher education (marked as "yes" in blue) consistently achieve better grades than those without such aspirations (marked as "no"). A greater proportion of students with higher education goals attained "fair" grades, while fewer fell into the "poor" category, suggesting that motivation for future academic pursuits may lead to greater commitment and improved academic outcomes. However, the chart lacks data for higher performance categories, such as "good" or "excellent," which could provide a more comprehensive view of how aspirations affect academic achievement. Additionally, the figure does not consider other influential factors, such as socioeconomic background, quality of resources, and instructional support, which also impact academic success. Overall, this trend underscores the positive role of higher education aspirations in enhancing academic performance, indicating that educational strategies should prioritize fostering motivation and providing resources to sustain students' aspirations, thereby supporting overall academic achievement.

4.4. Study Time vs Desire for Higher Education

The relationship between the time students dedicate to studying and their aspirations for higher education is a significant area of educational research, providing insights into how motivation shapes academic behavior and success. Examining this interaction reveals how study habits align with students' long-term educational goals. This study investigates whether students who allocate more time to academic activities generally have higher educational aspirations, driven by their commitment to achieving specific goals and recognizing the role of further education in their career prospects. Additionally, it explores if students with strong aspirations for higher education are more likely to invest substantial time in their studies as a deliberate strategy to ensure that their academic performance supports their ambitions [19], [20].

The violin plot in figure 5 visually represents how study time varies among students of different ages, with a clear distinction between those aspiring to higher education and those without such goals. The data shows a consistent trend: students with higher education aspirations (indicated in pink) dedicate more study time across all age groups compared to their peers without these aspirations (in beige), suggesting a strong link between the desire for further education and academic commitment. Age also influences study habits, with notable fluctuations as students mature. Younger students (ages 15-17) display a broad range of study times, with increased variability and higher medians among those aiming for higher education. At ages 18-20, study time peaks, likely due to preparations for major exams, while students aged 21-22 show more consistent study routines, especially among those with academic goals. These findings underscore the importance of fostering educational aspirations early, as motivation significantly impacts study engagement. Educational programs focusing on study skills and time management could be especially valuable for students pursuing higher education, enabling them to optimize their study habits and achieve long-term academic and career objectives.

Figure 5. Graph Analysis on Study Time vs Desire on Higher Education

4.5. Romantic Status vs Final Grade

The relationship between students' romantic status and their academic performance is a complex area in educational research, examining how personal relationships may influence academic outcomes. This intersection offers insights that could inform educational strategies and student support services, addressing whether romantic involvement has a positive or negative impact on academic performance. Understanding this dynamic provides valuable perspectives on the broader social factors affecting academic success, helping educators and policymakers consider how personal relationships might contribute to students' educational experiences and outcomes.

 ISSN 2723-6471

This study investigates the effects of romantic relationships on students' academic performance, aiming to determine if those in relationships perform differently than their single peers. The hypothesis suggests that romantic involvement could offer emotional and mental support, potentially enhancing academic outcomes, or it could serve as a distraction, leading to reduced focus and lower grades. Figure 6 illustrates this relationship by showing the distribution of final grades among students based on romantic status. Students with poor grades appear slightly more common among those who are single, suggesting that romantic involvement does not significantly harm academic performance at lower grade levels. However, a notable difference emerges in good grades, where students without romantic commitments outperform those who are involved, hinting that they may experience fewer distractions and therefore better academic focus. This trend suggests an academic advantage for single students at higher performance levels. Yet, the minimal impact on poor and fair grades implies that other factors, such as age, academic year, extracurricular activities, and personal stressors, likely play significant roles. The analysis highlights the need for a balanced approach in supporting students' academic and emotional well-being, as romantic status is only one of many factors affecting success. Further research could deepen our understanding of how to design educational strategies that holistically support students in managing both academic and personal responsibilities.

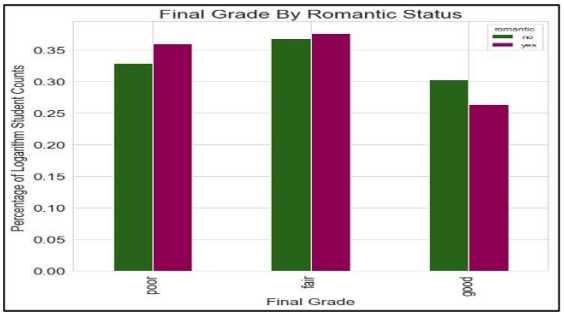

Figure 6. Graph Analysis on Romantic Status vs Final Grade

4.6. Frequency of Going Out vs Final Grade

The relationship between students' social activities—specifically, the frequency of going out—and their academic performance presents an intriguing subject for investigation. This research aims to analyze how often students participate in social outings and how this correlates with their final academic grades, providing valuable insights into the balance between social life and academic responsibilities. The study hypothesizes that frequent social activities could have dual effects on academic performance. On one hand, increased social outings might reduce study time, potentially leading to lower grades. On the other hand, social engagement could alleviate stress and improve mental well-being, which may positively influence academic outcomes. By examining this relationship, the research seeks to offer insights that could help educators and students find an optimal balance between academic and social life, ultimately promoting both well-being and academic success, as depicted in figure 7.

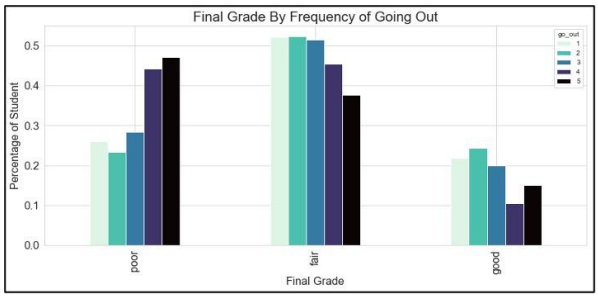

Figure 7. Graph Analysis on Frequency of Going Out vs Final Grade

The bar chart explores the relationship between students' final grades and their frequency of social outings, categorized into five levels ranging from minimal to very frequent. Analyzing this data reveals notable trends in how social activity levels correlate with academic outcomes. For students with poor grades, those who rarely go out (level 1) exhibit a significant proportion of low marks, but this proportion declines as social activity increases, with the lowest percentage of poor grades observed among the most socially active students (level 5). For fair grades, the distribution remains

 ISSN 2723-6471

relatively stable across all levels of social frequency, with a slight increase among students who participate in moderate social outings (level 3).

Regarding good grades, the highest academic performance is seen in students who engage in moderate social activities (levels 2 and 3), suggesting that a balanced approach to socializing supports academic success. Conversely, students with the most frequent outings (level 5) show the fewest good grades, indicating that excessive socializing may detract from academic performance. These findings emphasize the value of moderation, as students who balance academic responsibilities with moderate social engagement seem to benefit from improved mental well-being and stress relief, which can positively impact grades. This insight is useful for students and educators seeking to optimize academic success through balanced lifestyle choices, promoting social routines that support both well-being and academic performance.

4.7. Living Area vs. Final Grade

The relationship between a student's living area and their academic performance offers valuable insight into how environmental factors impact educational outcomes. This study examines the correlation between living environments—urban, suburban, or rural—and students' final grades, hypothesizing that socio-economic and cultural factors unique to each setting can significantly influence academic achievement. By understanding these dynamics, educators and policymakers can develop targeted strategies and allocate resources more effectively, helping to ensure that students receive equitable support for academic success regardless of their geographic location, as illustrated in figure 8.

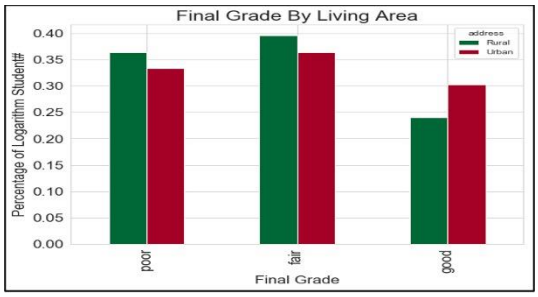

Figure 8. Graph Analysis on Living Area vs Final Grade

The bar chart compares final grades of students based on whether they reside in rural or urban areas, providing a clear view of how the living environment may influence academic performance. For poor and fair grades, the percentages are almost identical between rural (green bars) and urban (red bars) students, suggesting that the type of living area does not significantly impact students who achieve lower or average grades. However, a notable difference appears with good grades, where urban students have a higher percentage than their rural peers. This disparity implies that urban settings might provide certain advantages, such as better access to educational resources or more supportive academic environments, which could contribute to higher academic performance. The analysis highlights that while rural or urban residence does not markedly affect students with lower grades, it does seem to influence the likelihood of achieving good grades, with urban students outperforming those from rural areas. These findings could be crucial for policymakers and educators aiming to improve educational outcomes, particularly in rural areas where additional support might help bridge the performance gap.

4.8. Weekend Alcohol Consumption vs Final Grade

The relationship between weekend alcohol consumption and academic performance offers valuable insights into how lifestyle choices impact student outcomes. This research examines the correlation between students' weekend alcohol consumption and their final grades, hypothesizing that higher alcohol intake may be linked to lower academic performance due to its negative effects on cognitive function, time management, and overall health. Conversely, minimal or no alcohol consumption is expected to correlate with higher academic performance. Understanding these patterns can guide the development of targeted interventions that encourage healthier lifestyle choices, ultimately supporting improved educational outcomes, as illustrated in figure 9.

 ISSN 2723-6471

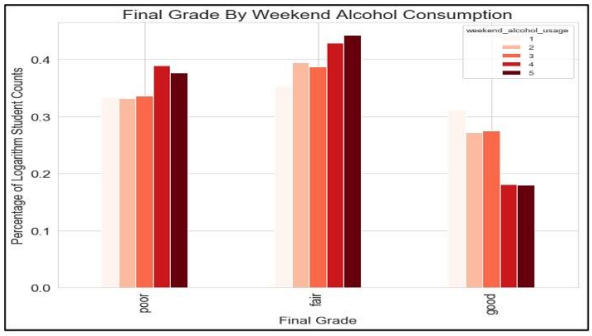

Figure 9. Graph Analysis on Alcohol Consumption vs Final Grade

The bar chart in figure 9 illustrates the relationship between students' weekend alcohol consumption and their final grades, categorized as poor, fair, and good across varying levels of alcohol use from 1 (lowest) to 5 (highest). For poor grades, the data reveals a slight increase in the percentage of students with poor performance as alcohol consumption rises, suggesting that higher weekend drinking may correlate with lower academic outcomes. In the fair grades category, percentages remain relatively stable across different alcohol levels, indicating that moderate drinking does not significantly impact students in the average performance range. The most pronounced trend appears in the good grades category, where a clear decline is observed; students with the lowest alcohol consumption (level 1) have the highest percentage of good grades, with this percentage consistently decreasing as alcohol use increases. This trend points to a negative correlation between alcohol consumption on weekends and the likelihood of achieving higher grades, highlighting the potential adverse effects of alcohol on cognitive function and time management, which are crucial for academic success. These findings underscore the importance of promoting responsible alcohol use among students, offering valuable insights for educational institutions and health professionals to develop programs that support healthier lifestyle choices and improved academic outcomes within the student population.

4.9. Result of Classification using Auto Model

This research used a prediction function to test the accuracy of each of the classifiers. The classifiers used are Naïve Bayes, Deep Learning, and Decision Trees. Auto Model can accelerate the process faster and make the user understand results better, especially for Deep Learning classifiers, as the inner logic might be hard to understand. A classifier with the highest accuracy was chosen to be the engine for predicting the employee attrition system. The results of each accuracy were explained further in the next subtopic.

4.10. Classification using Naive Bayes

$$P(A|B) = \frac{\left(P(B|A)P(A)\right)}{\left(P(B)\right)} \quad (1)$$

Using the Bayes theorem, A represents the hypothesis, and B is the evidence. According to the assumptions made, the features or predictors are independent, where any one feature does not affect others. This research coded the Naive Bayes classifier using Rapid Miner Studio, and the classification results displayed a confusion matrix where the classifier's recall, precision, and accuracy were calculated.

Table 1. Naïve Bayes Confusion Matrix

	A	**B**	**C**	**D**	**F**	**Precision**
A	6	7	0	0	0	93.88%
B	0	6	6	0	0	50.00%
C	1	0	6	11	1	60.00%
D	0	1	3	12	4	46.15%
F	0	0	0	3	46	31.58%
Recall	**90.20%**	**42.86%**	**46.15%**	**85.71%**	**40.00%**	

From the confusion matrix, the recall and the precision of the classification, as shown in table 1, were calculated using the formula below:

 ISSN 2723-6471

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \quad (3)$$

The accuracy of Naïve Bayes achieved 67.31%. The accuracy of the classification using Naive Bayes classifier was calculated using the formula:

$$\text{Accuracy} = \frac{\text{TP}}{(\text{TP+FP})} \quad (4)$$

The study utilizes Naïve Bayes under the assumption of feature independence, implemented via Rapid Miner Studio. A confusion matrix is provided, which calculates precision, recall, and accuracy for the classifier:

Precision and Recall: These metrics are derived from the confusion matrix for each class labeled from A to F. Precision is defined as the ratio of true positive predictions to total positive predictions, and recall is the ratio of true positives to the actual positives in the data.

Accuracy: The Naïve Bayes classifier achieved an overall accuracy of 67.31%, indicating its efficacy in correctly predicting outcomes based on the given data.

4.11. Decision Tree

Decision Tree is a commonly used classification algorithm in machine learning, employing a set of if-else rules to enable machines to make decisions. Its straightforward, rule-based structure makes it suitable for classification tasks. Table 2 provides insights into the optimal depth for the Decision Tree model. The tree's performance peaks at a depth of 2, achieving an accuracy of 84.8%, while accuracy stabilizes around 78.9% at greater depths (4 and beyond), indicating that increasing depth does not necessarily improve performance.

Table 2. Optimal Depth for Decision Tree

Maximal Depth	Performance
2	0.848
4	0.806
7	0.789
10	0.789
15	0.789
25	0.789

Table 3 shows the Decision Tree's confusion matrix along with precision and recall metrics. Precision across the classes is generally high, with notable performance in Class A (96.23%) and Class D (100.00%). Recall scores also remain robust, particularly in Class B (93.33%) and Class F (81.25%). The model achieves an overall accuracy of 86.68%, underscoring its effectiveness for practical classification applications.

Table 3. Confusion Matrix Decision Tree

	A	B	C	D	F	Precision
A	5	0	0	0	0	96.23%
B	1	14	0	0	0	93.33%
C	0	1	13	2	0	60.87%
D	0	0	3	14	6	100.00%
F	0	0	0	2	51	81.25%
Recall	**89.47%**	**93.33%**	**77.78%**	**83.33%**	**81.25%**	

4.12. Deep Learning Result

The exploration of deep learning results in predictive modeling provides valuable insights into the strengths of advanced machine learning techniques. This study specifically examines the performance of deep learning models

compared to traditional approaches, such as Naïve Bayes and Decision Trees, with an emphasis on accurate classification and prediction outcomes. Here, we analyze results from a deep learning model applied to predict employee attrition, utilizing complex neural networks that can process large, intricate datasets and often outperform simpler models when dealing with high-dimensional data. Our evaluation highlights key performance metrics—precision, recall, and overall accuracy—as depicted in the confusion matrix shown in table 4. These metrics are essential for gauging the model's effectiveness in practical applications, where accurate predictions are crucial for informed decision-making.

Table 4. Confusion Matrix for Deep Learning

	A	B	C	D	F	Precision
A	30	2	1	1	1	79.00%
B	3	25	2	1	1	80.60%
C	2	1	20	1	1	66.70%
D	1	2	2	15	1	17.40%
F	1	0	1	0	4	25.00%
Recall	**90.50%**	**89.20%**	**86.90%**	**85.70%**	**80.00%**	

The confusion matrix for the Deep Learning model provides an in-depth assessment of its performance across five classes (A, B, C, D, F), allowing us to evaluate precision and recall—two essential metrics for gauging classification accuracy. For high-performing classes, Classes A and B demonstrate strong results, with Class A achieving 79.00% precision and 90.50% recall, and Class B showing 80.60% precision and 89.20% recall, indicating the model's effectiveness in accurately identifying these categories with minimal misclassification. Class C shows moderate performance, with 66.70% precision and 86.90% recall, suggesting that while the model frequently recognizes Class C, it also misclassifies other classes into this category. In contrast, Classes D and F have lower precision, at 17.40% and 25.00% respectively, although their recall rates (85.70% for D and 80.00% for F) are somewhat better; this low precision indicates a high rate of false positives, where other classes are mistakenly labeled as D or F. The model's performance varies significantly across classes, excelling in some while struggling with precision in others, especially D and F. This inconsistency may stem from imbalanced training data or insufficient training for less common classes. Overall, these findings highlight the need for further refinement, such as rebalancing the dataset or adjusting the model's architecture, to enhance precision for underperforming classes. These improvements could lead to more uniform accuracy across all classes, ultimately increasing the model's reliability in practical applications.

4.13. Comparison Result Machine Learning with Deep Learning

In machine learning, comparing traditional algorithms with deep learning techniques provides crucial insights into these approaches' strengths and limitations. This section of the document, "Comparison Result Machine Learning with Deep Learning," aims to critically assess the performance differences between conventional machine learning models and more complex deep learning frameworks. Table 5 reveals that the Decision Tree classifier achieves the highest accuracy at 86.7%, outperforming both the Deep Learning and Naïve Bayes classifiers, which have accuracies of 74.4% and 67.3%, respectively. Despite its high accuracy, Decision Tree has a modest error rate of 13.3% and a short runtime of 2 seconds, making it both effective and efficient. In comparison, the Deep Learning classifier, with an error rate of 25.6% and a runtime of 3 seconds, ranks second in accuracy but demands more computational resources. Naïve Bayes, though fast with a runtime of 0.89 seconds, has the lowest accuracy and the highest error rate at 32.7%.

Table 5. Classifiers Result Comparison

Classifier	Accuracy	Error Rate	Runtime (s)
Naïve Bayes	67.3%	32.7%	0.89
Decision Tree	86.7%	13.3%	2
Deep Learning	74.4%	25.6%	3

This analysis focuses on comparing the efficacy of traditional machine learning models, such as Decision Trees and Naïve Bayes, against deep learning for predictive applications, particularly in employee attrition prediction. Deep

learning is celebrated for its capacity to manage large datasets and recognize intricate patterns but requires considerable computational power and may lack the interpretability of traditional models. Traditional algorithms like Decision Trees and Naïve Bayes provide simplicity and speed, which can be advantageous in settings with less complex data needs. This initial analysis prepares for a detailed discussion of these models' empirical results, offering insights to inform model selection for predictive tasks in organizational contexts, especially regarding interpretability, efficiency, and accuracy.

Table 6 compares precision and recall across three machine learning models: Naïve Bayes, Deep Learning, and Decision Tree, each showing unique strengths in these metrics essential for evaluating practical effectiveness. Naïve Bayes achieves high precision at 93.88% and recall at 90.20%, striking a reliable balance by minimizing false positives while identifying most positive cases, making it suitable where both precision and recall are critical but some error margin is acceptable. Deep Learning, with a precision of 87.27% and the highest recall at 94.12%, excels in capturing the most positive cases, though its slightly lower precision suggests a higher rate of false positives. This model is ideal where missing a positive is more problematic than having false positives, as in fields like medical diagnostics or fraud detection. The Decision Tree model stands out for perfect precision at 100%, meaning it makes no false positives, and achieves a high recall of 93.33%, though slightly below Deep Learning. This model is best for scenarios where avoiding false positives is paramount, ensuring only true positives are flagged. Each model offers distinct advantages—Decision Tree for flawless precision, Deep Learning for highest recall, and Naïve Bayes for balanced performance—allowing for strategic model selection based on whether the priority is reducing false positives or capturing all positive cases. This tailored approach supports informed decision-making aligned with specific application goals and constraints.

Table 6. Precision and Recall Comparation

Model	Precision	Recall
Naïve Bayes	93.88%	90.20%
Deep Learning	87.27%	94.12%
Decision Tree	100.00%	93.33%

5. Conclusion and Recommendation

This research aimed to explore the factors influencing students' grades and academic performance, applying data analysis techniques to assess these factors and develop a predictive model for forecasting student outcomes. The study provided a comprehensive framework, progressing from introductory concepts and literature reviews to research methodology and results. A grade prediction system benefits students, educators, and institutions by identifying at-risk students and supporting academic achievement. This research developed a data analysis model that highlights factors contributing to academic success and constructs a predictive model for grade forecasting. Analysis indicates several influential factors, including the motivation to pursue further studies, which correlates with effective time management and academic dedication. In contrast, a lack of aspirations is associated with reduced academic effort. Romantic relationships may have a negative effect on academic performance, as time spent socially could otherwise be allocated to studying. Additionally, excessive social outings were found to be detrimental, whereas residence location (urban vs. rural) showed no significant impact on performance. Alcohol consumption displayed a negative correlation with grades, as higher intake was linked to poorer outcomes. Among the classifiers tested, the Decision Tree model achieved the highest accuracy (87.6%) and perfect precision (100%). Deep Learning followed with an accuracy of 74.4% and precision of 94.12%, while Naïve Bayes recorded the lowest accuracy at 63.33%, establishing Decision Tree as the most effective model in this study.

This project successfully identifies factors affecting student performance, potentially helping students avoid certain pitfalls and improve academically. The system also offers a user-friendly interface, making it accessible for all users. However, the project faced a limitation due to a small dataset, which may restrict the accuracy of findings compared to analyses conducted on larger datasets. While a predictive model was developed, it has not yet evolved into a fully integrated system allowing real-time grade predictions.

Future research could investigate additional classifiers, such as Support Vector Machine and Random Forest, to potentially improve accuracy. Expanding the model into a comprehensive system would enhance its practical application, enabling direct grade predictions. Researchers are also encouraged to use larger datasets to obtain more robust results and compare them to smaller datasets to examine accuracy variations. Further refinement of the Decision Tree model with deeper levels and minimal leaf adjustments could enhance predictive accuracy, advancing the model into a fully operational system for real-time grade prediction in educational settings.

6. Declarations

6.1. Author Contributions

Conceptualization: M., T.B.K., D.A.D., M.Z.Z., C.M.A.N.; Methodology: D.A.D.; Software: M.; Validation: M., D.A.D., dan C.M.A.N.; Formal Analysis: M., D.A.D., dan C.M.A.N.; Investigation: M.; Resources: D.A.D.; Data Curation: D.A.D.; Writing Original Draft Preparation: M., D.A.D., dan C.M.A.N.; Writing Review and Editing: D.A.D., M., dan C.M.A.N.; Visualization: M.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Elbadrawy, R. S. Studham, and G. Karypis, “Collaborative multi-regression models for predicting students’ performance in course activities,” *ACM Int. Conf. Proc. Ser.*, vol. 2015, no. 3, pp. 103–107, Mar. 2015.

[2] K. al Hazaa, A.-S. G. Abdel-Salam, R. Ismail, C. Johnson, R. A. N. Al-Tameemi, M. H. Romanowski, A. BenSaid, M. B. H. Rhouma, and A. Elatawneh, “The effects of attendance and high school GPA on student performance in first-year undergraduate courses,” *Cogent Educ.*, vol. 8, no. 1, p. 1956857, 2021.

[3] A. E. Tatar and D. Düştegör, “Prediction of Academic Performance at Undergraduate Graduation: Course Grades or Grade Point Average?” *Appl. Sci.*, vol. 10, no. 14, pp. 1–14, 2020.

[4] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis and H. Rangwala, "Predicting Student Performance Using Personalized Analytics," *in Computer*, vol. 49, no. 4, pp. 61-69, Apr. 2016, doi: 10.1109/MC.2016.119

[5] V. Sheth, U. Tripathi, and A. Sharma, “A Comparative Analysis of Machine Learning Algorithms for Classification Purpose,” *Procedia Comput. Sci.*, vol. 215, no. 1, pp. 422–431, 2022.

[6] M. S. Farooq and G. Berhanu, “Factors affecting students’ quality of academic performance: A case of secondary school level,” *Journal of Quality and Technology Management*, vol. 7, no. 2, pp. 1-14, 2011.

[7] M. A. Qureshi, A. Khaskheli, J. A. Qureshi, S. A. Raza and S. Q. Yousufi “Factors affecting students’ performance through collaborative learning and engagement,” *Interactive Learning Environments*, vol. 3, no. 1, pp. 1–10, 2006.

[8] J. Xu, K. H. Moon, and M. Van Der Schaar, “A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs,” *IEEE J. Sel. Topics Signal Process.*, vol. 4553, no. c, pp. 1–12, 2017.

[9] S. Mirhosseini, M. Bazghaleh, M. H. Basirinezhad, A. Abbasi, and H. Ebrahimi, “The relationship between depression and academic satisfaction in medical science students,” *J. Ment. Health Train. Educ. Pract*., vol. 16, no. 2, pp. 99–111, 2021.

[10] S. Y. Chen, C. F. Lai, Y. H. Lai, and Y. S. Su, “Effect of project-based learning on development of students’ creative thinking,” *Int. J. Eng. Educ*., vol. 59, no. 3, pp. 232–250, 2019.

[11] D. L. Baker, L. Santoro, G. Biancarosa, S. K. Baker, H. Fien, and J. Otterstedt, “Effects of a read aloud intervention on first grade student vocabulary, listening comprehension, and language proficiency,” *Read. Writ.*, vol. 33, no. 10, pp. 2697–2724, 2020.

[12] J. K. Lê and T. Schmid, “The Practice of Innovating Research Methods,” *Organ. Res. Methods*, vol. 25, no. 2, pp. 308–336, 2022.

[13] C. Cruz Villazón, L. Sastoque Pinilla, J. R. Otegi Olaso, N. Toledo Gandarias, and N. de Lacalle, “Identification of Key Performance Indicators in Project-Based Organisations through the Lean Approach,” *Sustainability*, vol. 12, no. 15, pp. 1–15, 2020.

[14] S. S. Maidin, L. Fan, L. B. Y. Simon, D. S. Dinda, A. Zahra, Z. Oughannou, and M. Frikha, “From Theory to Practice: Understanding the Factors Affecting the Development of Digital Community Education in China,” *J. Theor. Appl. Inf. Technol*., vol. 102, no. 2, pp. 1–12, 2024.

[15] Hery and A. E. Widjaja, “Analysis of Aprioriand FP-Growth Algorithms for Market Basket Insights: A Case Study of The Bread Basket Bakery Sales,” *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, pp. 63-83, 2024.

[16] Henderi and Q. Siddique, “Anomaly Detection in Blockchain Transactions within the Metaverse Using Anomaly Detection Techniques”, *J. Curr. Res. Blockchain.*, vol. 1, no. 2, pp. 155–165, Sep. 2024.

[17] D. Sugianto and A. R. Hananto, “Geospatial Analysis of Virtual Property Prices Distributions and Clustering,” *Int. J. Res. Metav.*, vol. 1, no. 2, pp. 127-141, 2024.

[18] A. D. Terfassa, "The Relationship Between Parental Education and Children’s Academic Performance: The Case of Genda Tesfa Primary School, Dire Dawa," *Research on Humanities and Social Sciences*, vol. 8, no .1, pp. 10-16, 2018.

[19] C. Masui, J. Broeckmans, S. Doumen, A. Groenen, and G. Molenberghs, "Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education," *Studies in Higher Education*, vol. 39, no. 1, pp. 621–643, 2014

[20] S. Nonis and G. Hudson, "Performance of College Students: Impact of Study Time and Study Habits," *Journal of Education for Business*, vol. 85, no. 1, pp. 229-238, 2010.