Issues in Information Systems
Volume 26, Issue 3, pp. 291-301, 2025

DOI: https://doi.org/10.48009/3_iis_2025_2025_123

Using machine learning to analyze factors influencing grades in upper-level information technology courses

Jenq-Foung Yao, *Georgia College and State University, jf.yao@gcsu.edu*
Daniel Wu, *Georgia College and State Universit, daniel.wu@gcsu.edu*
Yu-Hsiang (John) Huang, *Georgia College and State University,john.huang@gcsu.edu*
Troy Strader, *Drake Universit, troy.strader@drake.edu*
Tsu-Ming Chiang, *Georgia College and State University,tm.chiang@gcsu.edu*

Abstract

Understanding the determinants of academic success in upper-level Information Technology (IT) courses is critical for improving student outcomes and informing pedagogical strategies. This study explores the predictive relationships among attendance patterns, exam performance, homework scores, and final grades using machine learning techniques. Drawing on data from 162 students enrolled in four advanced IT courses, we employed three classification models—Decision Tree (DT), Artificial Neural Network (ANN), and Naïve Bayes (NB)—to assess their effectiveness in forecasting student performance. Results revealed that ANN outperformed the other models with a prediction accuracy of 79.01%, indicating that complex, non-linear interactions among academic factors are influential. DT and NB also surpassed the baseline model, highlighting the relevance of key predictors, especially exam scores and first-day attendance. Our findings underscore a strong link between early attendance and academic achievement, suggesting that absence on the first day of class is a significant early-warning indicator of poor performance. The study concludes with practical implications for early intervention strategies and proposes future research directions involving expanded datasets and cross-disciplinary analysis.

Keywords: course grades, class attendance, information technology, undergraduate education, machine learning

Introduction

Understanding the factors that contribute to academic success in information technology (IT)-related courses in higher education remains a central concern for educators and researchers. Identifying these key determinants enables the development of targeted interventions and pedagogical strategies aimed at enhancing student learning outcomes. In the dynamic field of IT, where a strong theoretical foundation coupled with practical skills is crucial, the ability to predict and influence student performance is particularly valuable. This study aims to address this need by employing machine learning techniques to analyze academic data from advanced undergraduate IT courses.

Although a negative relationship between absences and grades might seem self-evident, scholarly inquiry necessitates robust investigation to confirm and elaborate upon such presumptions (Kuh, 2009). Our study utilizes machine learning techniques to rigorously examine the interplay of class attendance patterns, homework grades, exam performance, and overall grades in upper-level information technology curricula. A particular focus is placed on the impact of attendance on the first day of classes, a session that is often overlooked by some students. Machine learning analysis not only substantiates the expected trends but also provides a more granular and potentially less biased understanding of these relationships than previous studies relying on traditional statistical methods (Domingos, 2012; Goodfellow et al., 2016). The adoption of machine learning offers a contemporary, powerful, and potentially more insightful approach to analyzing these educational dynamics. Therefore, our key findings underscore the critical role of class attendance in academic success and support the use of attendance policies as an effective method for identifying struggling students early and improving their learning outcomes.

The remainder of this paper is structured as follows: In Related Work section, we provide a review of related literature on factors influencing academic performance and the application of machine learning in educational settings. This is followed by details of the methodology employed, including the data collection process, the selected machine learning algorithms, and the evaluation metrics used. We then present and discuss the correlation analysis and the predictive performance of the models. In conclusion, we elaborate on the recommendations for educational practice and outline directions for future work.

Related Work

The increasing availability of educational data has opened significant opportunities for applying data-driven approaches to understand and improve student learning processes (Romero & Ventura, 2010). Machine learning algorithms, with their capacity to identify complex patterns and generate predictions from large datasets, provide a powerful toolkit for this endeavor (Jain, 2014). Consequently, researchers have explored various predictors of academic performance, encompassing factors such as prior academic achievement, study habits, engagement levels, and demographic characteristics (Habley et al., 2004). Within the specific context of IT education, investigations have also considered the influence of programming experience, problem-solving skills, and engagement with technology (Lee et al., 2013). Urban-Lurain and Weinshank (2000) specifically investigated the relationship between attendance and outcomes for non-computer science students in large introductory computer science labs without lectures, concluding that overall class attendance was the strongest predictor of student performance, even when considering factors like year-in-school and prior computing experiences.

A key aspect of student engagement, and a potential indicator of their commitment and involvement in a course, is class attendance. The relationship between attendance and academic performance has been a subject of considerable research, yielding varied findings. Some studies have reported a strong positive correlation between class attendance and grades in disciplines such as business (e.g., Broucek & Bass, 2011; Kuh, 2009) and business law (Davenport, 1990). Millis et al. (2009) found a significant link between classroom participation and examination performance in a first-year medical school class. In contrast, Aldosary (1995) observed a stronger correlation between final scores and homework performance compared to attendance among Environmental Design students. Clump et al. (2003) supported the influence of class attendance on student grades in a general psychology course, while a study in Physiology Education classes (Hammen & Kelland, 1994) suggested that regular attendance was helpful but not the sole determinant of learning human physiology. In particular, Yao and Chiang (2011) statistically examined the correlation between class attendance and grades, also exploring the impact of attending the first day of class on overall grades. Yadav (2022) employed linear regression analysis to investigate the relationship between absenteeism and scores.

Furthermore, the ability to accurately predict student performance holds significant value for enabling early intervention strategies. Identifying students at risk of underperforming enables educators to provide timely support and resources, ultimately aiming to improve learning outcomes. Yu et al. (2010) explored the use of machine learning techniques to analyze student data and identify patterns and correlations that can inform strategies for improving educational outcomes. Their study highlighted the potential of data-driven approaches in educational settings to provide actionable insights for educators and administrators, ultimately aiming to support student success and optimize teaching practices. Hence, to identify the key factors contributing to academic success in higher education, our study utilizes machine learning techniques to rigorously examine the interplay among class attendance patterns, homework grades, exam performance, and overall grades in upper-level IT courses.

Methodology

Participants

This study collected data from 162 students' attendance records and corresponding grades in four different upper-level computer science courses: Databases, Discrete Structures, Operating Systems, and Data Mining. Among the 162 students, 18 are female and 144 are male. We have excluded gender as a predictor in this study due to the significant discrepancy between the two groups, which could skew the results. In addition, 138 students attended the first class while 24 were absent. The summary statistics are provided in Table 1 and Table 2. The attendance data represents attendance rate, while the exam data are the grades after curve.

Table 1. Data Statistics

Attributes	minimum	Maximum	Mean	Std
Homework	0	100	85.49	16.38
Attend I	0	100	80.34	25.72
Exam I	0	108.8	76.39	19.2
Attend II	0	100	71.97	27.97
Exam II	0	105	69.91	20.89

Table 2. Grade Distribution

Grade	A (88, 100]	B (78, 88]	C (68, 78]	D (58, 68]	F [0, 58]
# of Students	27	49	56	19	11

Data Collection Process

At the beginning of each class, students signed in with a notification that there would be no penalties for missed classes. We calculated the number of attendances and its percentage after each semester. The dataset comprises seven attributes: First Class, Attend1, Attend2, Homework, Exam1, Exam2, and Grade. To identify students who missed the first day of class, the variable "First Class" was assigned a value of 0; otherwise, 1. The attribute Attend1 represents the attendance record before the first exam, while the attribute Attend2 corresponds to the attendance record before the second exam. The remaining attributes are self-explanatory. Notably, the attributes First Class, Homework, Attend1, Exam1, Attend2, and Exam2 are predictors, with the attribute "Grade" as the target (class). The final exam was excluded from the analysis because its comprehensive nature made it less meaningful for predicting the overall grade. It is important to note that the grade distribution is as follows: Exam I, Exam II, homework, and the final exam each constitute 25% of the total grade. Additionally, students who dropped the class were excluded from the dataset to ensure the integrity and accuracy of the analysis.

Machine Learning Algorithms
We developed a decision tree (DT) classifier (Quinlan, 1996), an Artificial Neural Network (ANN) (Goodfellow et al., 2016), and a Naïve Bayes (NB) classifier (Schütze et al., 2008) to predict students' overall grades based on the aforementioned attributes. The implementation of these models, namely the decision tree, the artificial neural network, and the Naïve Bayes classifier, was facilitated by the open-source Weka software package (Witten et al., 2016).

Decision Tree (DT)
The decision tree construction involves evaluating entropies and information gain based on predictors derived from historical data. The entropy function is as follows:

$$Entropy(P) = -\sum_{i=1}^{n} P_i log_2 P_i$$

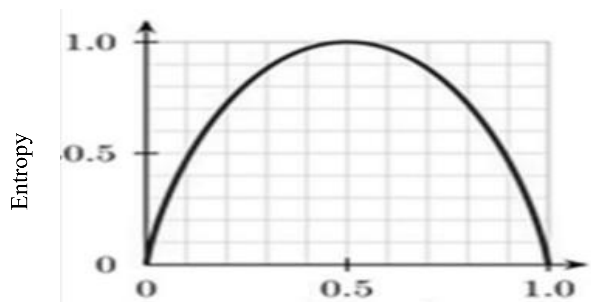

Figure 1. Entropy of Decision Tree

Lower entropy corresponds to higher certainty and predictability, making it a superior predictor. Following the entropy for all predictors, the predictor with the smallest entropy emerges as the optimal choice, occupying the top position in the decision tree (See Figure 1). This procedure is applied iteratively to remaining predictors in distinct branches. In essence, this method establishes the significance of predictors, assigning the most influential ones by placing them higher in the tree structure.

Artificial Neural Network (ANN)
An ANN model utilizes predictors as input data and propagates them through a network with weighted connections. The combination of inputs and weights passes through a function that generates temporary outputs. These outputs subsequently become new inputs and undergo a similar process, ultimately producing the final outputs. The results are then compared with actual data. In cases of discrepancies or errors, the errors are retroactively fed back to adjust the inputs, thereby modifying the weights and forwarding these changes to generate new outputs. This iterative process continues until the model stabilizes. Our ANN model consists of three layers: an input layer, a hidden layer, and an output layer (Figure 2).

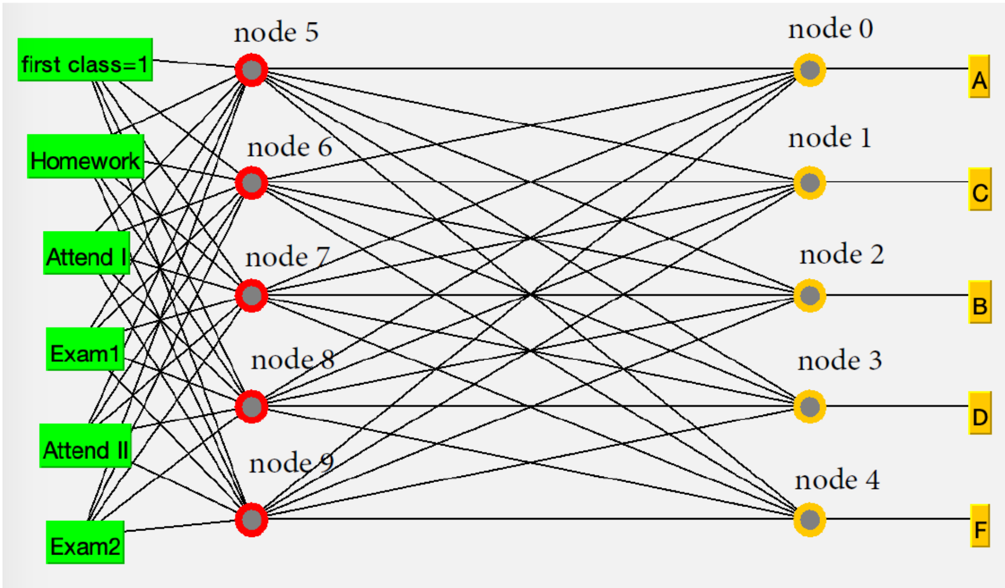

Figure 2. Artificial Neural Network

The sigmoid function, a widely used activation function in artificial neural networks, is defined as:

$$f(I) = \frac{1}{1 + e^{-I}}$$

where $0 < f(I) < 1$ and $I = \sum_1^n w_u x_j$ with w_u represents the weights and x_j the inputs. The output of each node is determined by its *f(I)*. In our model, nodes 5, 6, 7, 8, 9 use the predictors from the input layer along with their weights to generate an output *f(I)*. These outputs then serve as inputs to nodes 0, 1, 2, 3, 4, producing another set of *f(I)* values that lead to the final outputs *A, B, C, D, F*.

Naïve Bayes (NB) Classifier

The Naïve Bayes classifier employs Bayes' Theorem to integrate all predictor variables and the target variable within its probabilistic framework. This approach relies on two key assumptions: the equal importance of all attributes and their conditional independence given the class label, meaning that the value of one attribute provides no information about others within the same class. This algorithm is often surprisingly effective, easy to implement, and computationally efficient.

The equation is as follows:

$$P(C|X) = P(X_1|C) * P(X_2|C) * \dots * P(X_n|C) * P(C),$$

where C represents the target/class, and $X_1, X_2, \dots, X_n$ are the predictors. $P(C|X)$ denotes the probability of class C given predictor X, while $P(X_i|C)$ denotes the probability of predictor X_i given class C and $i = 1,2,\dots,n$. *P(C)* is the overall probability of class *C*. The Naïve Bayes Classifier uses this known data to predict *P(C|X)*.

In our model, the equation appears as follows:

$$P(\text{grade} \mid \text{all-predictors}) = P(\text{first-class} \mid \text{grade}) * P(\text{Homework} \mid \text{grade}) * P(\text{Attend1} \mid \text{grade}) * P(\text{Exam1} \mid \text{grade}) * P(\text{Attend2} \mid \text{grade}) * P(\text{Exam2} \mid \text{grade}) * P(\text{grade})$$

We utilize the known data *P*(first-class | grade), *P*(Homework | grade), *P*(Attend1 | grade), *P*(Exam1 | grade), *P*(Attend2 | grade), *P*(Exam2 | grade), and *P*(grade) to calculate *P*(grade | all_predictors), which represents the probability of the grade given all predictors.

The three classification models employed a 10-fold cross-validation approach, where the dataset is divided into 10 segments. In each iteration, one segment is used for testing while the other nine are combined for training. Each segment takes its turn as the testing set, allowing the dataset to be trained and evaluated in ten different ways. Initially, each of the three algorithms was trained on 90% of the dataset to create a model. The remaining 10% served as the testing dataset to assess the model's accuracy. This process was repeated ten times with different segment divisions.

Results and Discussion

Attending the First Class

The attendance of students on the first day of class continues to significantly influence their overall grades, as observed in comparison to a previous study (Yao & Chiang, 2011). The disparity between attending and not attending the initial class session is noteworthy, with a 23.54% (or 14.78 difference) advantage favoring those who attend. Additionally, the group that attended the first day of class exhibited a smaller standard deviation, signifying not only superior grades but also greater consistency within that particular group (see Table 3).

Table 3. Grade Comparison Between Attending or Not-Attending First Class

Attend First Class?	Mean	Std	Max	Min
No	62.79	25.12	87.53	1.25
Yes	77.57	11.02	93.59	66.65

Overall Model Performance

We use the ZeroR model as a baseline for prediction performance comparison. Table 4 summarizes the performance of the four models in terms of their ability to correctly classify instances.

Table 4. Correctness of Model Prediction

Model	Correct	Wrong	Correct (%)
ZeroR	56	106	34.57
Artificial Neural Network (ANN)	128	34	79.01
Decision Tree (DT)	112	50	69.14
Naïve Bayes (NB)	106	56	65.43

The baseline model achieved an accuracy rate of 34.57%. In contrast, the Decision Tree model reached a prediction rate of 69.14%, while the Artificial Neural Network model excelled with an impressive rate of 79.01%. The Naïve Bayes model also performed well with a prediction rate of 65.43%. These results highlight the effectiveness of all three models, showing their ability to outperform the baseline and reinforcing their reliability in predicting factors that influence a student's grade. Table 5 presents a comparative analysis of three machine learning models in predicting each of the grades (*A, B, C, D, F*) and several key performance metrics: Accuracy, Precision, Recall, F1-Score, and Specificity.

In particular, ANN demonstrates the strongest overall performance. It consistently achieves the highest average scores across all metrics. This suggests that the ANN model is effectively capturing the underlying patterns in the data and generalizing well across different classes. Decision Tree shows intermediate performance. While its average accuracy and specificity are reasonably high, its precision, recall, and F1-Score are noticeably lower than the ANN. This indicates a potential trade-off, possibly with more false positives or false negatives compared to the ANN. Naïve Bayes exhibits the weakest average performance. Its average scores across all metrics are generally lower than the other two models. This suggests that the assumptions of independence between features made by Naïve Bayes might not hold strongly for this dataset, leading to suboptimal performance.

Table 5. Model Performance Metrics for Each Class

Classifier	Grade	Accuracy	Precision	Recall	F1-Score	Specificity
ANN	A	93.8%	79.3%	85.2%	82.1%	95.6%
	B	81.5%	69.4%	69.4%	69.4%	86.7%
	C	87.0%	81.8%	80.4%	81.1%	90.6%
	D	97.5%	94.1%	84.2%	88.9%	99.3%
	F	98.1%	83.3%	90.9%	87.0%	98.7%
	Average	91.6%	81.6%	82.0%	81.7%	94.2%
Decision Tree	A	93.2%	83.3%	74.1%	78.4%	97.0%
	B	77.8%	61.0%	73.5%	66.7%	79.6%
	C	80.2%	74.0%	66.1%	69.8%	87.7%
	D	92.0%	66.7%	63.2%	64.9%	95.8%
	F	95.1%	63.6%	63.6%	63.6%	97.4%
	Average	87.7%	69.7%	68.1%	68.7%	91.5%
Naïve Bayes	A	93.2%	78.6%	81.5%	80.0%	95.6%
	B	75.9%	66.7%	60.7%	63.6%	84.0%
	C	76.5%	61.2%	61.2%	61.2%	83.2%
	D	91.4%	63.2%	63.2%	63.2%	95.1%
	F	93.8%	53.3%	72.7%	61.5%	95.4%
	Average	86.2%	64.6%	67.9%	65.9%	90.6%

Predictor Influence

Based on the observed performance of the different models presented in Table 5, we may infer some potential strengths and weaknesses of the underlying factors. The superior performance of the ANN suggests that the predictors likely have complex, non-linear relationships with the final grades. ANNs are well suited to model such intricate patterns and interactions between features that might be missed by simpler linear models like Naïve Bayes or even tree-based models to some extent. This could indicate that factors like a combination of attendance and performance in different types of assignments (exams and homework) might interact in a non-additive way to determine the final grade. Compared to ANN, the weaker performance of Naïve Bayes strongly hints that the predictors are likely not statistically independent. For example, a student who performs well in one exam is also likely to perform well in another exam, violating the independence assumption. Similarly, attendance might be correlated with assignment scores. The violation of Naïve Bayes' core assumption would potentially lead to its lower predictive accuracy.

The Decision Tree's intermediate performance suggests that some predictors are likely more influential than others in determining the final grade. The tree-based structure inherently tries to identify the most discriminatory features at each split. Its lower precision and recall compared to the ANN might indicate that while it captures some key decision boundaries, it might be more susceptible to overfitting on less important features or struggling with complex combinations of features that the ANN can handle more effectively. The varying performance across classes (e.g., lower for *B* and *F grades*) could also point to certain predictors being less effective in distinguishing these grade levels. Specifically, a set of rules can be extracted from the Decision Tree model (Figure 3). These rules implicitly suggest the relative importance of the predictors. Exam 1 and Exam 2 scores appear to be the most frequently used and have the most significant thresholds in determining final grades, notably for the higher grades. Homework plays a more compensatory role, helping students achieve higher grades even with slightly lower exam scores. Attendance, primarily in the "first class" and "Attend I/II," appears to have an influence, especially in borderline cases or when combined with other factors.

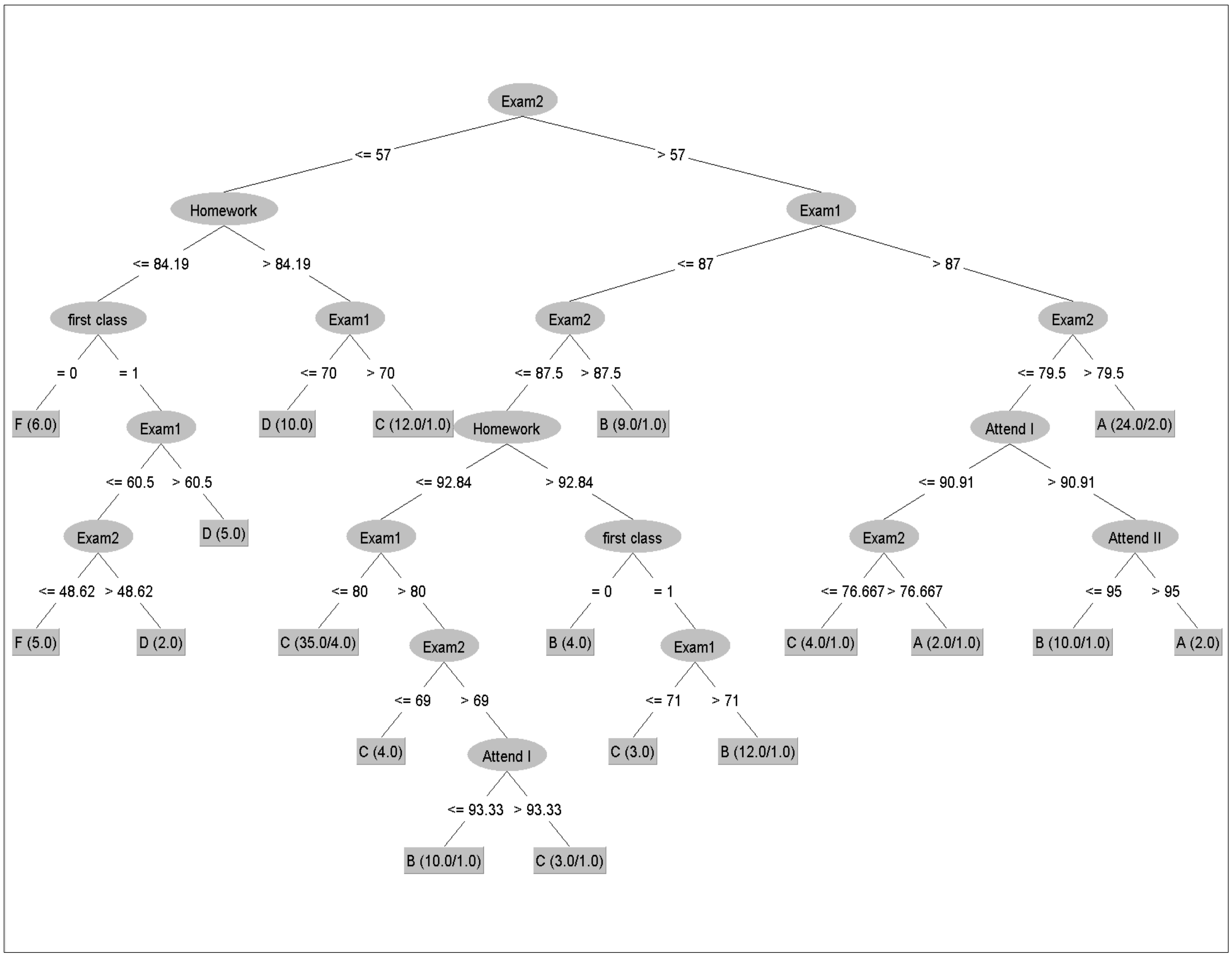

Figure 3. Decision Tree

Consequently, comparing these extracted explicit rules with the overall performance metrics of the ANN and Naïve Bayes provides a richer understanding. The ANN's superior performance suggests it might be capturing more subtle and complex interactions that are not explicitly represented by these individual rules. The weaker performance of Naïve Bayes reinforces the idea that the features are likely not independent, as the Decision Tree finds meaningful splits based on combinations of feature values. Furthermore, the varying performance of all models across different grade levels (*A*, *B*, *C*, *D*, *F*) implies that the strength and relevance of the predictors might not be uniform across all grade outcomes. The relatively high accuracy for *A*, *D*, and *F* grades across all models suggests that the predictors might be quite effective in identifying high-achieving students and those who might fail. The ANN's strong performance here indicates that it leverages the predictors most effectively for these outcomes. The lower precision and recall for grades *B* and *C*, particularly for the Decision Tree and Naïve Bayes, could suggest that the predictors are less effective at cleanly separating students who will achieve these middle grades. Based on our analysis, this might be due to two possibilities: (1) overlapping characteristics — students in the *B* and *C* range might exhibit more varied combinations of predictor values, making them harder to distinguish based on simple rules or independent probabilities; and (2) less discriminatory features for these boundaries — the features used might be more strongly correlated with the extremes of performance (*A* and *F*) than with the nuances of achieving a *B* or *C*.

Acknowledgements

This research was supported by The J. Whitney Bunting College of Business & Technology at Georgia College & State University.

Conclusion and Future Work

Numerous studies have emphasized that identifying the factors contributing to academic success in IT-related courses remains a significant concern for educators and researchers in higher education. This study employs machine learning algorithms to analyze various factors influencing students' academic performance. Notably, the algorithms used include the Naïve Bayes model, Decision Tree model, and Artificial Neural Network model. The results demonstrate the effectiveness of all three models, confirming their reliability in predicting the factors that influence students' grades.

A more detailed evaluation of the models' performance in predicting individual grade categories offers valuable insights. These machine learning classifiers also serve as effective tools for forecasting students' future academic outcomes. The superior performance of the Artificial Neural Network (ANN) suggests that student grades are influenced by complex, non-linear relationships among predictors—patterns that simpler models, such as Naïve Bayes, fail to adequately capture. The comparatively weaker performance of Naïve Bayes indicates that key predictors, such as exam scores and attendance, are likely not statistically independent. Meanwhile, the Decision Tree model demonstrates intermediate performance, highlighting the varying importance of predictors, with exam scores emerging as the most influential factor. Notably, the variation in model accuracy across different grade levels suggests that predictors differ in relevance; the models are more effective at identifying high (*A*) and low (*F*) achievers, but less accurate when predicting middle-range grades (*B* and *C*).

Our key findings reveal a strong association between poor attendance and lower academic performance, with a particularly robust correlation observed between overall grades and absenteeism on the first day of class. Specifically, students who were absent on the first day earned average grades that were 23.54% lower than those of their peers who attended. This underscores the critical importance of attendance in shaping academic outcomes. Notably, first-day absence emerges as a powerful early indicator of increased risk of academic failure. As such, implementing policies that promote attendance may significantly benefit student learning. Moreover, identifying students absent on the first day as an "at-risk" group could enable instructors to better communicate course expectations and use early attendance patterns as a practical tool for early intervention, thereby supporting improved academic performance.

In future research, we intend to broaden the scope of our study by incorporating additional academic disciplines and data from a wider range of universities. We also aim to compare lower-level and upper-level courses, as well as analyze differences between male and female students. The methodology will remain consistent with the current study; however, a key challenge will be maintaining data consistency, given that grading practices may vary across instructors.

References

Aldosary, A. S. (1995). The correlation between final grade score, attendance, and homework in the performance of CED students. *European Journal of Engineering Education, 20*(4), 481-486.

Clump, M. A., Bauer, H., & Whiteleather, A. (2003). To attend or not to attend: Is that a good question? *Journal of Instructional Psychology, 30*(3), 220.

Credé, M., Roch, S. G., & Kieszczynka, U. (2010). Class attendance in college: A meta-analytic review of the relationship between class attendance and grades. *Review of Educational Research*, *80*(2), 272-295.

Davenport, W. S. (1990). *A study of the relationship between attendance and grades of three business law classes at Broome Community College*. Unpublished doctoral dissertation, Nova University, Davie, FL.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78-87.

Friedman, P., Rodriguez, F., & McComb, J. (1995). Why does missing class matter? Evidence from introductory microeconomics. *The Journal of Economic Education*, *26*(4), 346-359.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: The MIT Press.

Habley, W. R., Bloom, J. L., & Robbins, S. (2004). *Increasing persistence: Research-based strategies for college student success*. Jossey-Bass.

Hammen, C. S., & Kelland, J. L. (1994). Attendance and grades in a human physiology course. *American Journal of Physiology, 267*(6 Pt 3), S105-S108.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666.

Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, *141*, 5–20.

Lee, S. H., Lim, J. H., & Grabowski, B. L. (2013). Predictors of success in online learning for undergraduate students. *British Journal of Educational Technology*, *44*(6), 1060-1079.

Millis, R. M., Dyson, S., & Cannon, D. (2009). Association of classroom participation and examination performance in a first-year medical school course. *Advances in Physiology Education, 33*(3), 139-143.

Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Computing Surveys, 28*(1), 71-72.

Romero, C., & Ventura, S. (2010). Educational data mining: A survey from 1995 to 2009. *Expert Systems with Applications*, *37*(10), 6019-6029.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval*. Cambridge University Press.

Urban-Lurain, M., & Weinshank, D. J. (2000). Attendance and outcomes in a large, collaborative learning, performance assessment course. In *Proceedings of the Annual Meeting of the American Educational Research Association*. New Orleans, LA.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.

Yadav, R. S. (2022). A study of relationship to absentees and score using machine learning method: A case study of linear regression analysis. *IARS'International Research Journal*, *12*(1), 33-39.

Yao, J. F., & Chiang, T. M. (2011). Correlation between class attendance and grade. *Journal of Computing Sciences in Colleges, 27*(2), 142-147.

Yu, Z., Li, N., Wang, P., & Zhao, W. (2010). Mining student data to improve academic performance. *International Journal of Information and Education Technology*, *1*(1), 59-62.