Journal of Applied Research and Technology

www.jart.icat.unam.mx

ICAT Instituto de Ciencias Aplicadas y Tecnología

Journal of Applied Research and Technology 24 (2026) 368-375

Original

Predicting Academic Performance of Students Through Supervised Learning Approaches

Amjad Alqahtani[1] • Atheer Alshahrani[1] • Maram Alserhani[1]
Reem Alrasheedi[1] • Nahla Aljojo[1]* • Araek Tashkandi[1] • Areej Alshutayri[2]
Aisha Blfgeh[2] • Anas Al-Tirawi[3] • Iqbal Alsaleh[4]

[1]College of Computer Science and Engineering, Information Systems and Technology Department, University of Jeddah, Jeddah, Saudi Arabia
[2]College of Computer Science and Engineering, Department of Computer Science and Artificial Intelligence, University of Jeddah, Jeddah, Saudi Arabia
[3]College of Engineering, Computing and Design, Department of Computer Science, Dar Al-Hekma University, Saudi Arabia
[4]Faculty of Economic and Administration, Management Information System Department, King Abdulaziz University, Saudi Arabia

Received: 14 02 2025; Accepted: 17 06 2025
Available: 30 06 2026

Abstract: In the contemporary educational environment, where the success of millions of students depends on precise forecasting, it is essential to conduct in-depth research into the many factors that influence academic achievement. Beyond simply analyzing students' grades, our research aims to provide a holistic picture of student achievement by examining a wide range of student demographics, academic backgrounds, and behavioral factors. We use advanced machine learning techniques, such as regression and classification, to decipher the complex patterns embedded in the data. This enables us to gain nuanced insights into the factors that predict student performance. We hope that by using these approaches, we will not only forecast academic outcomes but also identify the underlying factors that influence overall student success. In addition, our research seeks to determine the primary factors that have the greatest impact on students' academic performance. Educators receive vital insights that enable them to personalize interventions that target both academic and non-academic aspects that affect

*Corresponding author.
E-mail address: nmaljojo@uj.edu.sa (N. Aljojo).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

student progress. After an in-depth investigation, we concluded that the Artificial Neural Network (ANN) and Decision Tree (DT) models were the most accurate predictors. These models achieved accuracy rates of 81% and 76%, respectively. The results of this study demonstrate that the use of sophisticated machine learning algorithms is an effective method for predicting student performance and guiding interventions specifically designed to support student achievement.

Keywords: **Artificial Intelligence, Machine Learning, Supervised learning, Classification, Educational Data Mining, Support Vector Machine.**

1. Introduction

In the field of education, students' academic outcomes are greatly influenced by a wide range of factors, including personal characteristics, socioeconomic backgrounds, and environmental factors. To properly manage the influence of these factors, it is essential to have a thorough understanding of the complex interactions among them and their impact on students' academic performance. The field of educational data mining, still in its early stages, has recently received considerable attention. Methods, techniques, and academic endeavors aimed at automatically extracting insights from large-scale educational databases are among the topics included in this endeavor. When it comes to effectively forecasting pupils' achievement, the wealth of data in educational databases presents a number of obstacles. Numerous studies have been conducted to explain and forecast academic achievement (Yadav & Pal, 2012). These studies have highlighted the importance of predicting student performance in educational settings.

Accurately forecasting student academic performance has significant implications for educational institutions. To allocate resources effectively and provide targeted support to students at risk of underperformance, institutions must estimate student performance well in advance. Moreover, identifying the characteristics that influence course success rates can provide valuable insights for instructional design, which in turn can inform initiatives aimed at improving course outcomes. In light of recent developments in web-based educational technology and the implementation of quality standards, researchers are uniquely positioned to investigate how students acquire knowledge and which instructional strategies are most effective in fostering success. This project has two basic goals: first, to uncover factors that influence course success rates and student achievement; second, to use these factors as early predictors of expected success rates, enabling proactive interventions to address student shortcomings. Both goals are interrelated.

In the realm of higher education, students' achievements are a fundamental component of an institution's greatness. To improve people's awareness, knowledge, and cognitive abilities, educational institutions strive to deliver education of the highest possible quality. The goal of educators is to continuously improve student accomplishment, and they do this by monitoring student performance to evaluate the effectiveness of their teaching. The performance of students, on the other hand, is a complex phenomenon influenced by a variety of factors, including the instructional environment and individual study habits. Because of the complexity of the factors influencing student performance, standard prediction models, such as regression, which rely on overly simplistic assumptions, face considerable hurdles. Utilizing data mining techniques, on the other hand, provides a more comprehensive approach to the task of capturing intricate data relationships and constantly altering outputs in reaction to new information.

The purpose of our research is to investigate large-scale datasets to identify patterns that reflect students' behavior and learning using analytics and data mining approaches. We intend to construct predictive models capable of forecasting students' academic success across a variety of educational levels and fields of study, including both high school and university education. This will be accomplished by identifying a wide range of indicators and factors.

2. Related Work

There have been many studies that have utilized machine learning approaches to forecast student performance. In this part, we shall represent some of these works. 68.7%

of students passed, while 88.8% failed, making it the most accurate method for predicting pupils' final grades. There were eleven study publications that Sandra and Lumbangaol (2021) gathered on predicting student performance. They discovered that the majority of the algorithms employed are classification-based. The most common methods used to predict students' learning effectiveness are Artificial Neural Networks (ANNs), Naïve Bayes, Logistic Regression, Support Vector Machines (SVMs), and Decision Trees. They discovered that the Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, and Decision Tree algorithms are the most frequently used in machine learning for predicting student performance.

In his research, Zafari et al. (2021) investigated the impact that features have on the overall performance of models. To determine the significance of characteristics, the Boruta algorithm was used. During training, the models are trained on all features, including both effective and ineffective ones. The ANN accuracy, which was the best in the original dataset at 0.83, and the SVM performance, which improved after ineffective features were removed, had the highest accuracy among the models at 0.78. This was based on the results. Kamal et al. (2022) used a machine-learning- and metaheuristic-based approach to determine and forecast students' performance. By using a relief algorithm, they were able to select the elements that had a strong impact on accuracy. Theresa, a wide variety of classifiers, including BPNN, RF, and NB, are utilized to classify students' academic performance data. According to their findings, BPNN achieves higher accuracy in classifying and predicting student academic achievement, with a 95 percent accuracy rate.

Alsariera et al. (2022) conducted a comparative study of various machine learning models used to forecast the performance and success of students. They read most of the articles related to this topic. Compared with previous models, they found that the ANN Algorithm performed better and achieved higher accuracy. There is 85.9% accuracy with NN. The research was conducted by Ouatik et al. (2022) using information from various sources. Because of the vast amount of data and its variety, they utilized big data techniques. In order to reduce the amount of time required for execution, they saved data on DHFS and dispersed it over numerous workers. Through the utilization of the MapReduce Processing Framework, they applied machine learning methods and property selection to HDFS files. According to the SVM algorithm, the accuracy of the prediction was 87.32% with the application of big data techniques.

During the period of July to September 2018, Cassano et al. (2019) carried out an analysis study on 43 educational systems that were part of the Eurydice Network. They came to the conclusion that there was a widespread dissemination of evaluation and self-evaluation mechanisms, and consequently, the formation of a school management culture, in order to enhance the quality of the system and the performance of the students. The findings on student performance demonstrate that the evaluation of the quality improvement of the educational system cannot be based solely on these findings. Not only can educational systems be very good, but they can also be quite bad. In order to comply with the law, there is a system that requires evaluations to be completed. Specifically, Polyzou and Karypis (2016) utilized sparse linear and low-rank matrix factorization models that are tailored to each individual course or student–course tuple in order to forecast the grade that would be assigned to the subsequent course. The application was performed on a dataset that was obtained from the University of Minnesota, and it was executed for two distinct departments that possessed distinct characteristics. According to the findings, the accuracy of grade prediction is improved when the attention is placed on data that is relevant to the course.

Determining the factors that have an impact on academic success and working to improve the areas in which these factors are lacking. In Yassein et al. (2017), an attempt was made to investigate whether or not there are any patterns that could be helpful in forecasting the performance of pupils. Small data sets consisting of 150 students were used for the research that they carried out. They came to the conclusion that there are substantial links between the inclusion of both practical work and assignments in the course and the success rate of the course. Furthermore, they discovered that the number of assignments given had a detrimental impact on the course's academic performance. The criteria that have the most impact on a student's overall performance are their academic attendance in class, as well as their grades on both the final exam and the midterm exam. In the paper by Lu and Yuan (2018), the authors suggested an integrated Optimized Ensemble Feature Selection Algorithm by Density Peaks. This algorithm picked the parameters that had an effect on the prediction of the student's grade. The results of their investigation revealed that the SVM student performance prediction model, which is based on the feature selection algorithm (DPEFS), has superior prediction performance. As a result of their findings, seven characteristics were graded

according to the importance they hold for the courses, while five characteristics were deemed neutral on the basis of their values. For the student.

A study was conducted by Wakelam et al. (2020) with a group of 23 students enrolled in the final year of a university module. The data collected from each individual student were restricted to attendance at lectures and tutorials, access to virtual learning environments, and intermediate evaluations. They discovered that there was promise in predicting individual students' interim and final assessment marks in small student cohorts with very limited features, and that these predictions could help module leaders identify students who might be considered "at risk." Over the course of various investigations, deep learning models were utilized. Baashar et al. (2022) were able to identify a pattern of the most popular artificial neural network (ANN) techniques and algorithms, with a primary emphasis on higher education. Based on the findings, it was determined that artificial neural networks (ANNs) are consistently used alongside data analysis and data mining approaches. This allows for research to evaluate the efficacy of their findings in determining academic success standards. Several studies have attempted to improve the academic performance of pupils considered weak by employing machine learning models. The C4.5, ID3, and CART decision tree algorithms were used to the data of engineering students in Yadav and Pal (2012) in order to make predictions about how well they would perform on the final exam. In the comparative examination of the findings, it is said that the prediction has assisted the students who were not as strong in their academic performance and has brought about an improvement in the outcome.

Shahiri et al. (2015) conducted research on the difficulties associated with forecasting a student's grade. Specifically, they cited two primary causes. To begin, the research that has been done on the many methods of prediction presently available is not yet adequate to determine which methods are the most appropriate for predicting the performance of students. The second reason is that there have been insufficient studies on the factors influencing the academic performance of students in certain classes in Malaysia. They came to the conclusion that the characteristics of the model play a very significant effect in its overall success. Using features, cumulative grade point average, student demographics, high school history, scholarship, and social network interactions, they discovered that Naive Bayes gave the highest performance (76%) of all the models.

3. Methodology

The term "data mining" refers to a computational approach to processing data that has been effectively implemented across a variety of fields with the objective of extracting useful knowledge from the data. Data mining techniques are used to construct a model that can identify new information and knowledge. The following are some of the most important data mining techniques that have been developed and implemented: association, classification, clustering, prediction, sequential patterns, and decision trees. This section will explain the primary methods used in the region.

3.1 Classification

Classification is a fundamental aspect of data mining and is closely tied to the principles of machine learning. Classification is the process of categorizing individual data points in a collection into predetermined classes or groupings. This foundational approach is widely used in numerous disciplines to organize and understand data for analytical purposes. The classification process employs a wide range of mathematical tools, such as decision trees, linear programming, neural networks, and statistical methodologies. Each of these methodologies possesses distinct strengths and capabilities, equipping analysts with versatile tools to efficiently address categorization tasks.

The primary goal of the classification process is to create software or algorithms that can independently learn and identify patterns in data, enabling precise categorization of items into distinct groups. The capacity to adapt is crucial for the effectiveness of classification algorithms, since it enables them to continuously improve their classification skills by incorporating fresh data inputs and adapting to changing patterns. By using classification algorithms, analysts can extract useful insights from complex datasets, enabling informed decision-making and problem-solving across various domains and applications. Classification is a strong technique used in various industries, such as marketing for consumer segmentation and healthcare for disease detection. It helps organize and comprehend data, leading to innovation and advancement.

3.2 Support Vector Machine

The "Support Vector Machine" (SVM) is a robust supervised machine learning algorithm that can handle both classification and regression tasks, but is primarily used for classification. In the SVM algorithm, each data point is represented as a point in an n-dimensional space, where n is the number of features in the dataset. The value of

each feature represents the position along a particular axis in this multidimensional space. The core premise of SVM classification is to determine a hyperplane that can efficiently separate the data points belonging to distinct classes. Support vectors, which are essential components of the SVM framework, refer to the coordinates of certain observations that are located nearest to the decision border or hyperplane.

The support vectors play a crucial role in determining the position and orientation of the hyperplane, which in turn affects the classification result. The SVM classifier serves as the hyperplane that effectively separates the two distinct classes in the dataset. Through the utilization of the SVM algorithm, analysts may proficiently navigate intricate datasets and identify patterns that facilitate precise classification of data points into separate groups. SVM's versatility and widespread application in disciplines such as image classification, text categorization, and bioinformatics make it a highly useful and popular tool.

3.3 Decision Tree

The decision tree is a fundamental machine learning technique in supervised learning (Figure 1). Conceptually, it is represented as a flowchart, providing a distinct and visual guide for making judgments or predictions. Decision trees consist of a sequence of conditional control statements that act as queries to the dataset, taking into account its attributes, and then branching into new nodes. The decision tree originates from a single decision node and then splits into several nodes, with each node representing a unique conclusion or prediction. These nodes have the ability to divide into more decision nodes or prediction nodes, creating a hierarchical structure that resembles a tree, hence the name. This approach is applicable in both classification and regression problems, providing versatility across different domains and analytical contexts. Decision trees are highly effective at identifying patterns and extracting useful insights from datasets, whether for classifying outcomes or predicting continuous variables. These insights can then be used to inform decision-making processes. Figure 1 shows the decision tree graph.

3.3 Artificial Neural Network

Neural networks are used to simulate the fundamental operations of the human brain, drawing inspiration from how the brain processes and interprets information. Due to its ability to perform computations quickly and provide rapid responses, it is used to solve a variety of real-time problems. Many components modeled after the biological

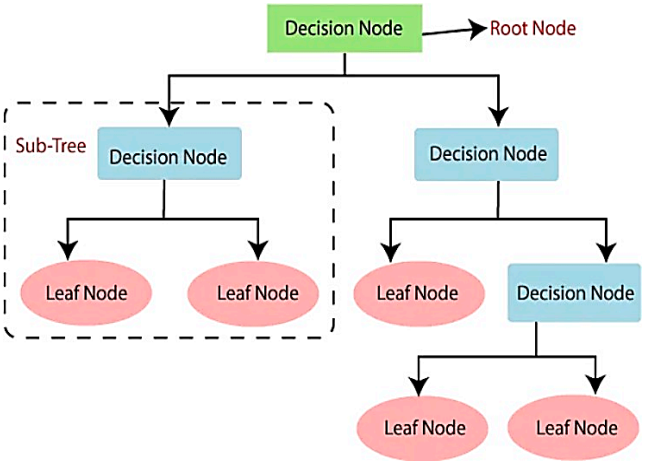


Figure 1. Decision Tree.

nervous system are included in an Artificial Neural Network (ANN). A large number of interconnected processing elements, often referred to as nodes, are required for an Artificial Neural Network to function properly.

Using a connection link, these nodes are connected to other nodes in the network. Weights are contained within the connection link, and these weights are responsible for storing information on the input signal. These weights are ultimately updated as a result of each iteration and input that is received. The Trained Neural Network is the name given to the final weights of the Neural Network together with its architecture. This occurs when all of the data instances from the training data set have been supplied. Training Neural Networks is the name given to this particular technique. By applying this trained neural network, certain problems outlined in the problem statement can be solved. In Figure 2, we see how NN operates.

3. Data Collection

The information is gathered using a learner activity tracker tool known as the Experience API (xAPI). The extensible application programming interface (xAPI) is a component of the training and learning architecture (TLA) that supports the monitoring of learning progress and learner actions, such as reading an article or viewing a training video. Learning activity providers can determine the learner, activity, and items that describe a learning experience using the experience application programming interface (API). The collection includes 480 student records from the Middle East, along with 16 attributes. The characteristics are classified into three primary categories: 1. Population characteristics, such as gender and nationality. 2. Academic background, including

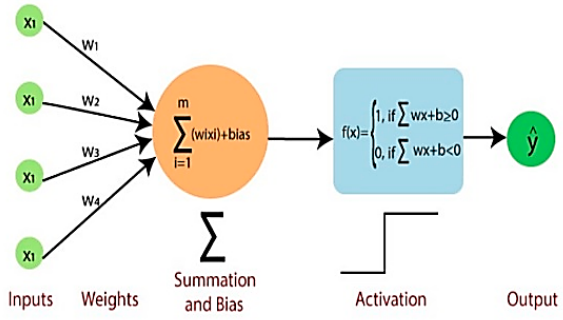


Figure 2. How NN works.

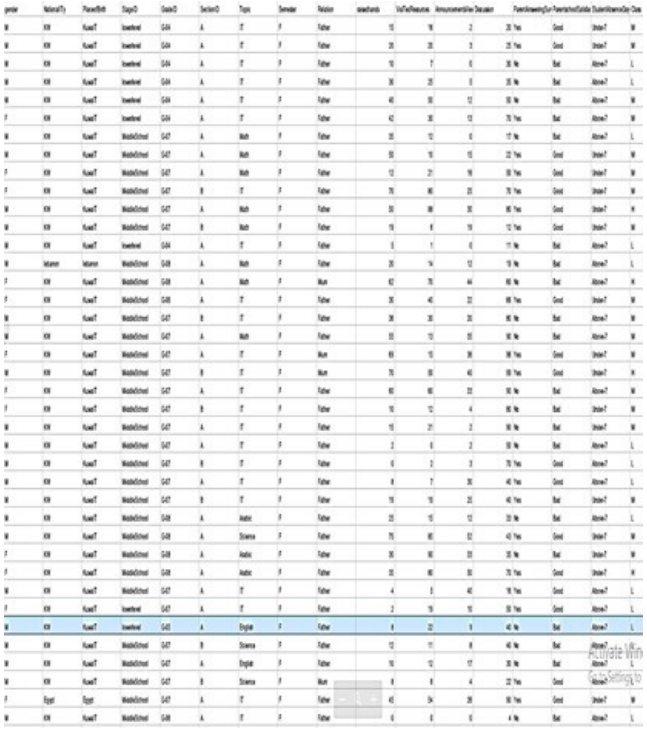


Figure 3. Students Dataset.

educational stage, grade level, and section (see Figure 3). 3. Behavioral traits, such as raising hands in class, opening resources, answering parent surveys, and school satisfaction. Information for the dataset was gathered over the course of two academic semesters: 245 student records were collected during the first semester, and 235 during the second. The dataset also includes the school attendance feature, which categorizes students into two groups based on the number of days they have missed school: 191 students have more than 7 absence days, while 289 students have fewer than 7 absence days. One of the new attribute categories included in this dataset is the concept of parent participation in the educational process. The Parent Participation feature includes two subfeatures: the Parent Answering Survey and the Parent School Satisfaction feature. 292 of the parents are content with the school, while 188 of them are not. There are 270 parents who responded to the poll, and 210 of them disagree with the results. Furthermore, we manually added thirty additional instances from Saudi Arabia, working in conjunction with the Shrooq al-Marfa school in Riyadh (Yadav & Pal, 2012).

Sources

- lElaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, The University of Jordan, Amman, Jordan, http://www.Ibrahimaljarah.com www.ju.edu.jo.
- Shrooq al-marfah School in Riyadh, https://ibnroshd.edu.sa/page/shrooq-sch/.

4. Analysis

The analysis is divided into two main sections. In the first stage, feature selection is performed to identify the most relevant features for constructing accurate predictive models. Next, the second phase involves selecting a suitable mining method for building the prediction models, with a specific emphasis on classification tasks. Feature selection is crucial in the initial stage, as it directly affects classifier performance and prediction accuracy. After careful analysis, we identified several characteristics that demonstrated compatibility with classifiers, while others were found to be less suited.

In the classification phase, we utilized knowledge from prior research, which emphasized the effectiveness of Decision Tree, Neural Network, and Support Vector Machine algorithms for classification tasks due to their high accuracy. Using these findings, we implemented the chosen procedures on our dataset and carefully documented the outcomes. Before constructing the model, the dataset was divided into training and testing subsets. The training set was accurately categorized by success rate, with three classes identified: high, medium, and low. This division into distinct levels of performance made it easier to create strong predictive models that can accurately classify outcomes.

5. Results

To accurately predict the percentage of students who will be successful, the first step is to apply a feature selection method to the records. A list of characteristics that are rated in descending order of importance is produced as a result of this process, which enables one to gain insights

into the characteristics that are the most significant. This ranked list is used to determine which characteristic is the most suitable for analysis, and then we select that feature. The characteristics that have been discovered as having the most significant impact on our selection process are depicted in Figure 4, which may be seen here.

Following that, the data, which has been augmented with the features that were chosen, is sent through a number of different classifiers. The results of these comparisons are then carefully analyzed in order to choose the classifier that is most appropriate for the data. There were four metrics that we utilized in order to evaluate the performance of the classifiers. These metrics were accuracy, precision, recall, and F-measure. Within the framework of our supervised classification approach, the data were suitably labeled in accordance with the success rates, then classified as either high, medium, or low.

After conducting an exhaustive investigation, we came to the conclusion that the Artificial Neural Network (ANN) developed into the most effective classifier, displaying an astounding accuracy rate of 81%. Table 1 presents a full summary of the findings collected from the three different classifiers.

Table 1. Measuring results of classifiers.

Classifier	Accuracy	precision	recall	F1-Score
SVM	0.63	0.65	0.36	0.46
ANN	0.81	0.78	0.73	0.75
DT	0.76	0.76	0.77	0.76

Conclusion

The primary aim of this study is to identify the key predictive variables that significantly impact student performance, with a particular focus on showcasing the immense potential of data mining applications within university management. This entails using data mining techniques to conduct a thorough analysis of historical data, thereby uncovering valuable insights. Data mining techniques serve as invaluable tools for extracting hidden patterns and elucidating relationships among features in the dataset. In this paper, we present a straightforward yet effective data mining-based prediction model. This model leverages classification techniques to pinpoint the features that influence student performance, thereby providing academic stakeholders with actionable insights to enhance academic outcomes – a central objective of our study. By employing classification techniques,

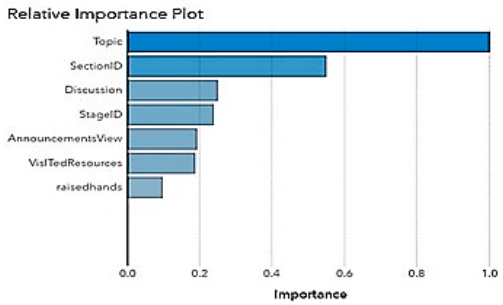


Figure 4. Feature selection.

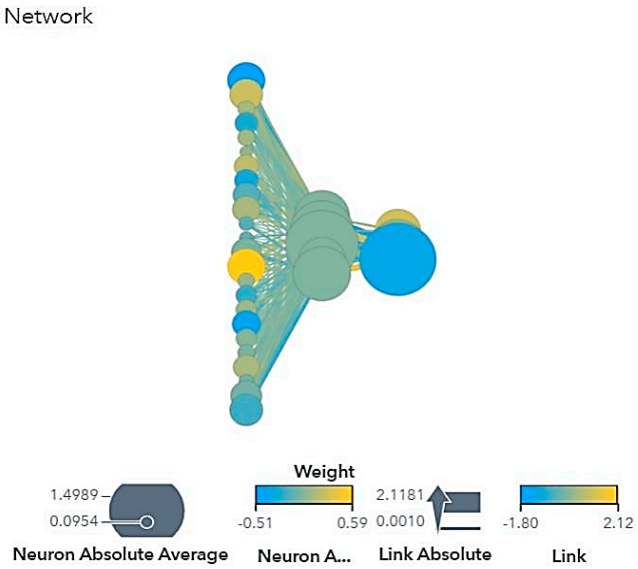


Figure 5. Architecture of ANN.

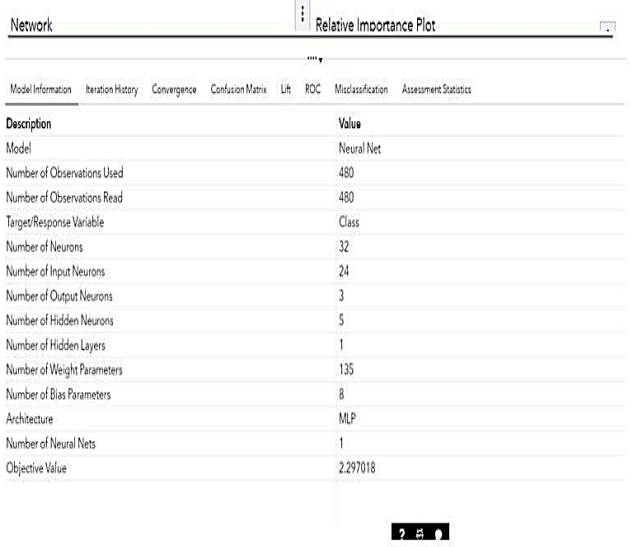


Figure 6. ANN Architecture Information.

we aim to identify and prioritize the factors that most significantly impact student performance. This knowledge empowers academic institutions to implement targeted interventions and initiatives to improve academic performance and foster student success. Ultimately, our study underscores the pivotal role of data mining in facilitating evidence-based decision-making and driving positive change within educational settings.

Conflict of interest

The authors have no conflict of interest to declare.

Funding

The authors received no specific funding for this work.

Acknowledgements

The authors acknowledge the College of Computer Science and Engineering, Information Systems and Technology Department, University of Jeddah, Jeddah, Saudi Arabia, for their support

References

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational intelligence and neuroscience*, *2022*(1), 4151487.
https://doi.org/10.1155/2022/4151487

Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., ... & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, *12*(3), 1289.
https://doi.org/10.3390/app12031289

Cassano, R., Costa, V., & Fornasari, T. (2019). An effective national evaluation system of schools for sustainable development: A comparative European analysis. *Sustainability*, *11*(1), 195.
https://doi.org/10.3390/su11010195

Kamal, M., Chakrabarti, S., Ramirez-Asis, E., Asís-López, M., Allauca-Castillo, W., Kumar, T., ... & Rahmani, A. W. (2022). Metaheuristics method for classification and prediction of student performance using machine learning predictors. *Mathematical Problems in Engineering*, *2022*(1), 2581951.
https://doi.org/10.1155/2022/2581951

Lu, H., & Yuan, J. (2018). Student performance prediction model based on discriminative feature selection. International Journal of Emerging Technologies in Learning (Online), 13(10), 55.
https://doi.org/10.3991/ijet.v13i10.9451

Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting student success using big data and machine learning algorithms. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(12), 236-251.
https://www.learntechlib.org/p/223146/

Polyzou, A., & Karypis, G. (2016). Grade prediction with models specific to students and courses. *International Journal of Data Science and Analytics*, *2*(3), 159-171.
https://doi.org/10.1007/s41060-016-0024-z

Sandra, L., & Lumbangaol, F. (2021). Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review. *TEM Journal*, *10*(4).
https://www.ceeol.com/search/article-detail?id=998905

Sandra & Lumbangaol (2021)

Shahiri, A. M., Husain, W., & Rashid, N. A. A. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, 414-422.
https://doi.org/10.1016/j.procs.2015.12.157

Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2020). The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology*, *51*(2), 347-370.
https://doi.org/10.1111/bjet.12836

Yadav, S. K., & Pal, S. (2012). Data mining: A prediction for performance improvement of engineering students using classification. *arXiv preprint arXiv:1203.3832*.
https://doi.org/10.48550/arXiv.1203.3832

Yassein, N. A., Helali, R. G. M., & Mohomad, S. B. (2017). Predicting student academic performance in KSA using data mining techniques. Journal of Information Technology & Software Engineering, 7(5), 1-5.
https://doi.org/10.4172/2165-7866.1000213

Zafari, M., Sadeghi-Niaraki, A., Choi, S. M., & Esmaeily, A. (2021). A practical model for the evaluation of high school student performance based on machine learning. *Applied Sciences*, *11*(23), 11534.
https://doi.org/10.3390/app112311534