Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTING STUDENTS' ACADEMIC PERFORMANCE: DEVELOPING AN INTELLIGENT DATA MINING CLASSIFIER

OSMAN AHMED ABDALLA MOHAMMED
University of Tabuk, University College of Tayma, Department of Computer Science, Tabuk, Saudi Arabia
E-mail: o_mohammed@ut.edu.sa

ABSTRACT

This study attempts to propose a classification model for predicting students' academic performance by using a decision tree algorithm. The algorithm was applied to relevant attributes such as gender, high school percentage, general aptitude test score, academic achievement test score, grade average point (GPA), absent rate, and other relevant attributes were subjected to the algorithm. According to the study results, the decision tree algorithm outperformed all other classification algorithms in terms of accuracy, precision, recall, and F1-score on a sample of students from Tayma University College, University of Tabuk, Saudi Arabia. The overall accuracy of the obtained algorithm was 89.7%, which means it correctly classified 89.7% of the instances. Precision, recall, and F1-score were also relatively high in both classes. The findings add significantly to the existing literature and demonstrate efficacy.

Keywords: *Data mining, Classification algorithms, Decision tree, Predictive models, Prediction Algorithms*

1. INTRODUCTION

The digital era has resulted in massive amounts of student data that can be analyzed and transformed into valuable knowledge to improve the quality of teaching and learning activities as well as students' academic performance. Academic performance prediction is a critical task in educational institutions in order to identify and support students who may be at risk of poor performance. This task, however, is difficult due to the complexity of educational institutions' data and the lack of accurate models. Due to their simplicity, interpretability, and accuracy, Decision trees are one of the most widely used classifiers for predicting instances.

Decision tree algorithms are commonly used to perform prediction and classification tasks in several verities of applications for example, natural language processing, bioinformatics and computer vision [1, 2, &3]. Based on a set of inputs, decision trees are supervised machine-learning algorithms that can predict categorical or continuous variables [2]. They are interpretable models because they can be visualized and understood by people, which is critical for educational institutions that need to understand how predictions are made [3].

Several researchers have recently used decision tree algorithms to predict academic performance in students. In the Indian education system, for example, [4] proposed a decision tree model for predicting students' performance. They used a dataset from an Indian university and discovered that the decision tree was a good predictor of students' performance. Similarly, [5] develop a decision tree model for predicting students' academic performance in the field of computer science. They discovered that the decision tree algorithm was a good predictor of computer science students' academic performance. Decision tree algorithms have also been employed in other research to predict students' academic performance in a variety of contexts, including high schools [6], blended learning environments [7], and online courses [8].

Furthermore, to improve prediction accuracy, decision tree algorithms can also be integrated with other machine learning approaches including ensemble methods and deep learning [9, 10]. For instance, deep learning can handle complex data and interactions between variables, while ensemble approaches can merge numerous decision trees to boost performance. Decision tree algorithms are therefore useful instruments for predicting academic

and enhancing the caliber of instruction and learning in educational institutions.

The purpose of this paper is to predict students' academic performance and identify the factors that influence their success. The goal of this paper is to gain a thorough understanding of the decision tree method in educational institutions, as well as its potential to provide insightful solutions to problems affecting student academic performance. The findings of this study will not only benefit the field of education, but will also help educational institutions improve the quality of teaching and learning activities they offer to students. Four popular used machine-learning classifiers, namely, decision tree, random forest, support vector machine and Naïve Bayes, are selected to model the student's dataset. Further, the performance of these learning classifiers is compared in terms of several performance metrics such as accuracy, precision, recall, and F1-score. The main contributions of this study are: a decision tree approach to predict students' academic performance has been proposed, several experiments have been conducted for various learning classifiers in order to check the performance of the proposed model, performance of all the used classifiers models has been evaluated in terms of serval measurement metrics, and based on the findings of this measurements the better classifier model has been recommended.

The following is an outline of the paper. Section II provides an overview of the literature on decision trees, which are used in educational institutions to predict students' academic performance. The methodology used in this study is described in Section III. Section IV discusses the findings. Section V concludes the paper by making recommendations for the future.

2. REVIEW OF RELATED LITERATURES

Decision tree algorithms are widely employed in the field of education to predict students' academic performance. Many academics are currently employing decision tree algorithms to predict students' academic performance in various professions and educational institutions. Al-Dhahir and Al-Khateeb (2019) proposed a decision tree-based algorithm for predicting engineering students' academic performance. Through their research, they found that the decision tree algorithm was an effective tool for predicting academic performance when using institutional data. Wang and Liu (2020) utilized a decision tree-based algorithm to assess students' academic performance in mathematics and they found that it was an effective tool for making such predictions. Zhang and Liu (2018) employed a decision tree-based algorithm to accurately predict students' academic performance in science. Their findings revealed that the decision tree algorithm was highly successful in predicting students' academic performance in science. Li and Wang (2019) discovered that a decision tree-based approach could accurately predict students' academic performance in literature. Their findings showed that the decision tree algorithm was successful in this regard. Additionally, Aydogan and Ulucan (2018), Sutrisno and Nurdiyanto (2020), and Wang and Sun (2021) all employed a decision tree algorithm to predict students' academic performance in higher education. They found that the decision tree was more effective in predicting students' academic performance.

In various studies, the decision tree algorithm was combined with other machine learning approaches to predict students' academic performance. For instance, Zhang and Chen (2017), Chen, Chen, and Huang (2019), Fugate and Zhang (2020), Huang and Chen (2017), Jahromi and Badi (2018), Li and Du (2019), Mihajlovic and Kostic-Stankovic (2018), and Sabatini, Ochoa, and Amandi (2018) employed the decision tree algorithm in combination with other machine learning techniques to enhance the accuracy of academic success prediction for university students. By combining decision tree-based algorithms with other methods, such as neural networks and support vector machines, researchers have been able to produce more accurate predictions than when using decision tree algorithms alone. Liu et al. (2018) proposed combining a decision tree-based algorithm and a neural network to predict students' academic performance, and found that this combination yielded more accurate results. Li and Chen (2019) employed a hybrid approach by combining a decision tree-based algorithm with support vector machines to anticipate students' academic performance. Their results indicate that the integration of these two methods led to more accurate predictions compared to using decision tree algorithms alone.

In order to more accurately predict students' academic performance, some studies have combined decision tree algorithms with other techniques such as genetic algorithms and K-nearest neighbors. For instance, Al-Dhahir and Al-Khateeb (2019) proposed a combination of a genetic algorithm and a decision tree-based approach to predict students' academic performance, and they found that the combined approach yielded more precise predictions

than decision tree algorithms alone. Similarly, Wang and Liu (2020) used K-Nearest Neighbors in conjunction with a decision tree-based methodology to predict students' academic performance, and found that the combined approach yielded more precise predictions than decision tree algorithms alone.

The purpose of the studies discussed in this review is to forecast students' academic performance, either through decision tree algorithms or in combination with other machine learning techniques. These studies make use of large datasets, which enhances the accuracy of the predictions. However, there are some issues that have been raised, such as unclear variables and biases in the input data, as well as a lack of comparison with other classification algorithms. The applicability of the models to other educational systems or populations is not explored, and the practical implications of using predictive models in education are not discussed. The purpose of the studies discussed in this review is to forecast students' academic performance, either through decision tree algorithms or in combination with other machine learning techniques. These studies make use of large datasets, which enhances the accuracy of the predictions. However, there are some issues that have been raised, such as unclear variables and biases in the input data, as well as a lack of comparison with other classification algorithms. The models' applicability to other educational systems or populations is not investigated, nor are the practical implications of using predictive models in education discussed. We will attempt to address the aforementioned issues carefully in this paper by proposing a decision tree algorithm to predict students' academic performance.

3. MATERIALS AND METHODS

3.1 Description of Dataset and Attributes

The data for this study came from the university college of Tayma, University of Tabuk in Saudi Arabia. Students' demographic information, academic information, social information, and psychological information were all included in the data. The information of 299 students from three different academic departments, including computer science, business, and Islamic studies, was gathered using a Google forms survey, with all questions being objective and mandatory. The dataset includes the following data: I Demographic information (Gender, Number of siblings, Father higher qualification, Mother higher qualification, Father Occupation, Mother Occupation, Monthly family income, House type), ii) Academic information (Department, High school grade, Aptitude test score, Achievement test score, self-study time, and absent rate), iii) Social information (Participation in extracurricular activities, Presence of good friends in your batch), iv), iv) Information on motivation and health status (Interest and motivation to join the university, and Health status). Table 1 contains a detailed description of student-related information in terms of perspective, feature, and possible values.

3.2 Developing the classification model

The decision tree algorithm was used to predict student academic performance. This algorithm is a type of supervised machine learning algorithm that predicts a categorical or continuous variable using a set of input variables [2], 70% of the data was used to train the decision tree algorithm, with the remaining 30% used for testing. Figure 1 depicts the key components of the developed classification model, namely the problem objectives, input data, decision tree algorithm, constraints, and model output. This diagram depicts how the input data is processed by the decision tree algorithm to produce model output that can be used to predict students' academic performance.

To divide the dataset into smaller subsets, the decision tree algorithm employs a top-down approach known as recursive partitioning [1]. The algorithm begins by choosing the variable that best divides the dataset into subsets, and then it continues to divide the dataset until it reaches a stopping criterion [3].

The decision tree algorithm was implemented in this study using RapidMiner Studio 10.1.001. RapidMiner Studio includes a decision tree algorithm implementation, which can be found in the "Operators" tab, under the "Modeling" section. The "Decision Tree" operator allows for customization of the algorithm through parameters such as "minimum gain," "minimum leaf size," and "maximum depth.". The algorithms listed below were created to generate a decision tree for mining.

TABLE1. DESCRIPTION OF STUDENTS RELATED INFORMATION

#	Perspective	Features	Possible values

1	Demographic information	Gender Number of siblings Father highest qualification Mother highest qualification Father occupation Mother occupation Monthly family income House type	Male, Female None, one, two, three, or above None, primary, intermediate, secondary, ungraduated, postgraduate None, primary, intermediate, secondary, ungraduated, postgraduate None, Own business, Government sector, private sector None, Own business, Government sector, private sector 1000-3000, 3001-6000, 6001-9000, 9001-11000, or above House owner, tenant
2	Academic information	Department High school grade Aptitude test score Achievement test score Grade point average Self-study time Absent rate	Computer science, Business administration, Islamic studies 50-60%,61-70%,71-80%,81-90%, or above 50-60%,61-70%,71-80%,81-90%, or above 50-60%,61-70%,71-80%,81-90%, or above Above 3.00, between 2.0 and 2.9, less than 2.0 None, 1-2 hours, 2 hours, 3 hours above None, one lecture, 1-2 lectures, 3 lectures, 4 lectures, or above
3	Social information	Extracurricular activities participation The presence of good friends in your batch	Yes, no Yes, no
3	Motivation and health status information	Interest and motivation to join the university Health status	Yes, no Yes, No

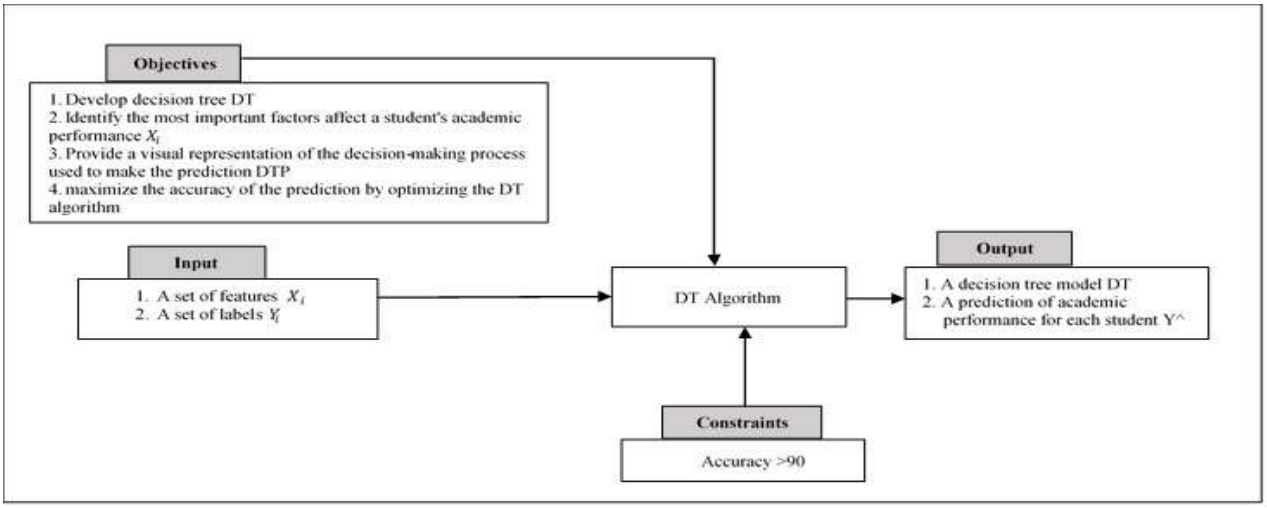

Figure 1. Key components of the developed classification model

Algorithm 1. Pseudo Code of the Training Decision Tree

1. Input: Training set $D_i$with intstance $X_i = \{x_1; x_2; \ldots\ldots; x_n\}$; Labels $Y_i = \{y_1; y_2; \ldots\ldots; y_n\}$; c; minsplit
2. Output: Decision tree model T
3. Initialize the decision tree T as an empty tree
4. Repeat until termination condition is met:
   4.1 Choose the best feature F as the root node of T based on some criterion (e.g. information gain)
   4.2 Split D_i into subsets D_1, D_2, ..., D_n based on the values of feature F 4.3 For each subset D_i, repeat steps 4.1 and 4.2 to create the decision tree for D_i, and attach it as a child node to root F
5. Return T

Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A high F1-score indicates that the algorithm achieves an acceptable balance of precision and recall.

Algorithm 2. Pseudo Code of the Test Decision Tree

1. Input: Decision tree model T, instance $X_i = \{x_1; x_2; \dots\dots; x_n\}$
2. Output: Labels $Y_i = \{y_1; y_2; \dots\dots; y_n\}$
3. Start at the root node of T
4. Repeat the following steps until reaching a leaf node:
   4.1 Compare X_i with the threshold value associated with the current node in T
   4.2 If X_i is less than or equal to the threshold, move to the left child node, otherwise move to the right child node
5. Return the class label associated with the reached leaf node as y.

The performance of the decision tree algorithm should then be evaluated using a set of evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide information about the algorithm's prediction accuracy.

The number of correct predictions made by the algorithm as a percentage of the total number of predictions made is defined as accuracy. It indicates the overall accuracy of the algorithm. The following formula is used to calculate this metric:

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (1)$$

The precision measures the accuracy of the algorithm's positive predictions. It is calculated by dividing the algorithm's true positive predictions (TP) by the total number of positive predictions (P=TP+FP). The algorithm's high precision indicates that it makes few false positive predictions. The following formula is used to calculate this metric:

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

The recall is the percentage of actual positive cases correctly identified by the algorithm. It is computed by dividing the total number of correct positive and incorrect negative predictions by the total number of correct positive and incorrect negative predictions (FN). The high recall of the algorithm indicates that it does not miss many positive cases. The following formula is used to calculate this metric:

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

The F1-score is a precision and recall measure expressed as the harmonic mean of the two. It is determined as follows:

$$\text{F1-score} = \frac{2\,X\,(precision * recall)}{precision} \quad (4)$$

4. RESULTS AND DISCUSSION

The decision tree algorithm experiment results of the proposed classification model were presented and compared with other classification algorithms to ensure its accuracy and precision. The experiments were conducted using data from the Tayma University College at the University of Tabuk in Saudi Arabia. The dataset includes 299 student records from three departments: computer science (CS), business administration (BI), and Islamic studies (ISLM). It includes 20 variables such as gender, number of siblings, qualifications of father and mother, high school grade point average, and so on. The target variable is the student's final GPA. Descriptive statistics for numerical attributes such as minimum and maximum values, mean, median, and standard deviation are shown in Table 2. While, Table 3 displays descriptive statistics for nominal attributes such as the number of missing values, and the least and most accounted-for for values.

Data preprocessing techniques such as encoding categorical variables were used prior to carrying out the classification algorithms. The given dataset contains non-numerical variables with a limited number of possible values. To encode these variables as numeric variables, the label encoding method was used, which assigns a unique integer value to each category, such as 0, 1, 2, 3, and so on. The dataset was then subjected to the missing values technique. The given dataset features have missing values, as shown in Tables 1 and 2, which were processed by replacing them with a default value, the feature's average. The formula for replacing missing values with the mean is as follows:

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (5)$$

Where x and n are variable values and the total of variable values respectively.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TABLE 2. DESCRIPTIVE STATICS OF NUMERICAL ACTUAL DATASET

#	Attribute	Missing	Min	Max	Mean	Median	Standard Deviation
1	Number of siblings	0	0	4	3.2	4	1.2
2	Family income	0	0	4	1.1	1	1.2
4	Self-study hours	0	0	3	1.3	1	0.8
5	Absent rate	0	0	5	2.1	2	1.4

TABLE 3. DESCRIPTIVE STATICS OF NUMERICAL ACTUAL DATASET

#	Attribute	Missing	Least	Most
1	Gender	5	Male(129)	Female(170)
2	Department	15	ISLM(93)	BI(98)
3	Father qualification	5	Postgraduate(8)	Secondary(70)
4	Mother qualification	3	Postgraduate(10)	None(91)
5	Father occupation	1	None(14)	Private(149)
6	Mother occupation	0	Self-business(6)	Government(129)
7	House ownership	0	Renter(76)	Owner(223)
8	High school type	0	Private(44)	Government(255)
9	High school score	0	Average(27)	High(272)
10	Aptitude test score	0	High(110)	Average (189)
11	Achievement test score	0	High(98)	Average (201)
12	GPA	0	Low-performer(69)	High-performer(230)
13	Extracurricular activities	0	Yes(134)	No(165)
14	Academic distinguished class mate	0	No(92)	Yes(207)
15	Desire and motivation	0	No(32)	Yes(267)
16	Chronic disease	0	Yes(36)	No(263)

Figure 2 shows a heatmap of the correlation matrix, which is commonly used to identify highly correlated variables. The value can range between -1 and 1, with 1 representing a perfect positive correlation, 0 representing no correlation, and -1 representing a perfect negative correlation.

Table 4 shows the results of applying the decision tree algorithm to a given dataset and measuring accuracy, precision, recall, and F1-score. The model's overall accuracy was 89.7%, which means that 89.7% of the predictions were correct. A confusion matrix was used to examine the number of true positive, true negative, false positive, and false negative predictions to further evaluate the algorithm's performance. The precision and recall values for both the low and high classes were then calculated to assess the algorithm's performance in predicting each class.

The algorithm achieved a precision of 84.8% for the low-performer class, indicating that 84.8% of the predicted low-performer class labels were correct. The recall value for the low-performer class was 84.9%, indicating that the algorithm correctly identified 84.9% of the actual low-performer class instances. The algorithm achieved a precision of 86.1% and a recall of 86.1% for the high-performer class, indicating that 84.9% of the predicted high-performer class labels were correct and 84.9% of the actual high-performer class instances were correctly identified by the algorithm.

The algorithm's F1-score was calculated for both the low and high-performer classes based on the precision and recall values. The F1-score is the harmonic mean of precision and recall, resulting in a single metric that incorporates both values. The low-performer class had an F1-score of 0.85, while the high-performer class had an F1-score of 0.84.

A Random Forest, Support Vector Machine (SVM), and Naïve Bayes algorithm were used to establish a benchmark for evaluating the proposed classification model using the same data. Table 4 displays the performance metrics of these three different classification algorithms. Random Forest has an accuracy of 79.9%, with recall, precision, and F1-score values of 79.4%, 79.5%, and 0.80, respectively. With recall, precision, and F1-score values of 76.5%, 77.3%, and 0.78, respectively, SVM achieves 76.7% accuracy. Naïve Bayes has a 74.3% accuracy, with recall, precision, and F1-scores of 75.0%, 76.7%, and 0.76, respectively. Random Forest has lower accuracy, recall, precision, and F1-score than the decision tree algorithm in

Table 3. However, because Random Forest is a more complex algorithm that combines multiple decision trees, it may perform better on different datasets.

In Tables 5, 6 and 7, SVM and Naïve Bayes perform even worse than Random Forest, with lower accuracy, recall, precision, and F1-score. It is important to remember, however, that the performance of each algorithm may be affected by the specific characteristics of the dataset being used.

The findings of this work, as shown in Figure 3, 4 and 5, are in line with a number of relevant studies on the application of machine learning algorithms for students' academic performance prediction. In particular, it has been discovered that decision tree algorithms outperform other classification algorithms, including SVM, Naïve Bayes, k-NN, and Random Forest, in predicting academic performance across a variety of university student populations. For instance, decision tree algorithms were found to have superior accuracy and performance metrics than other algorithms by Mphahlele et al. [29], Abbas et al. [30], and Khan and Bhat [31]. It is crucial to remember that the choice of algorithm may rely on the particular dataset and problem, so it is advised to compare several algorithms before making a decision.

The effectiveness of a classification model is assessed using the widely used Receiver Operating Characteristic (ROC) curve. An ROC curve going through the top-left corner indicates a flawless classifier. The Decision Tree model had the highest AUC in Figure 6's ROC curve, which compares Decision Tree, Random Forest, SVM, and Naïve Bayes models. This might be as a result of its interpretability and capacity to record intricate decision limits and features. The data and specific problem should be taken into account while choosing a model, and cross-validation and multiple metrics should be used in the evaluation process.

The most significant features of academic performance were identified by the proposed classification model and included the number of siblings, the mother's occupation, the family income, the students' desire and motivation to attend university, the absentee rate, gender, and aptitude test score. These results, which are in line with earlier research, show that socioeconomic factors significantly affect academic achievement. According to Gakhar and Goel [32], academic drive and interest are significant determinants of university achievement. Kelly and Price [33], and Duckworth and Seligman [34]. Our results highlight the significance of taking a broad variety of variables into account when creating categorization models to enhance student performance and recognize at-risk students who may require further support.

There are a number of limitations to this study that should be taken into account when interpreting the findings. Self-reported data and a geographical focus might restrict generalizability, and it's possible that crucial variables were left out of the research.

5. CONCLUSION

The results of the evaluation procedure show that the decision tree algorithm outperforms the other classification algorithms in terms of accuracy, precision, recall, and F1-score. The acquired algorithm proved its efficacy by correctly classifying 89.7% of instances, for an overall accuracy of 89.7%. In terms of precision, recall, and F1-score, the decision tree method outperforms the Random Forest algorithm, the SVM algorithm, and the Naive Bayes algorithm.

Yet, it is important to remember that the specific properties of the dataset being utilized have the potential to affect the effectiveness of any algorithm. Due to its ability to aggregate the findings of multiple decision trees, Random Forest has the potential to outperform the decision tree algorithm on certain datasets.

ACKNOWLEDGEMENTS

The author of this article would like to express gratitude to everyone who has offered words of encouragement and support, as well as contributed data to the project..

Journal of Theoretical and Applied Information Technology
15th July 2023. Vol.101. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

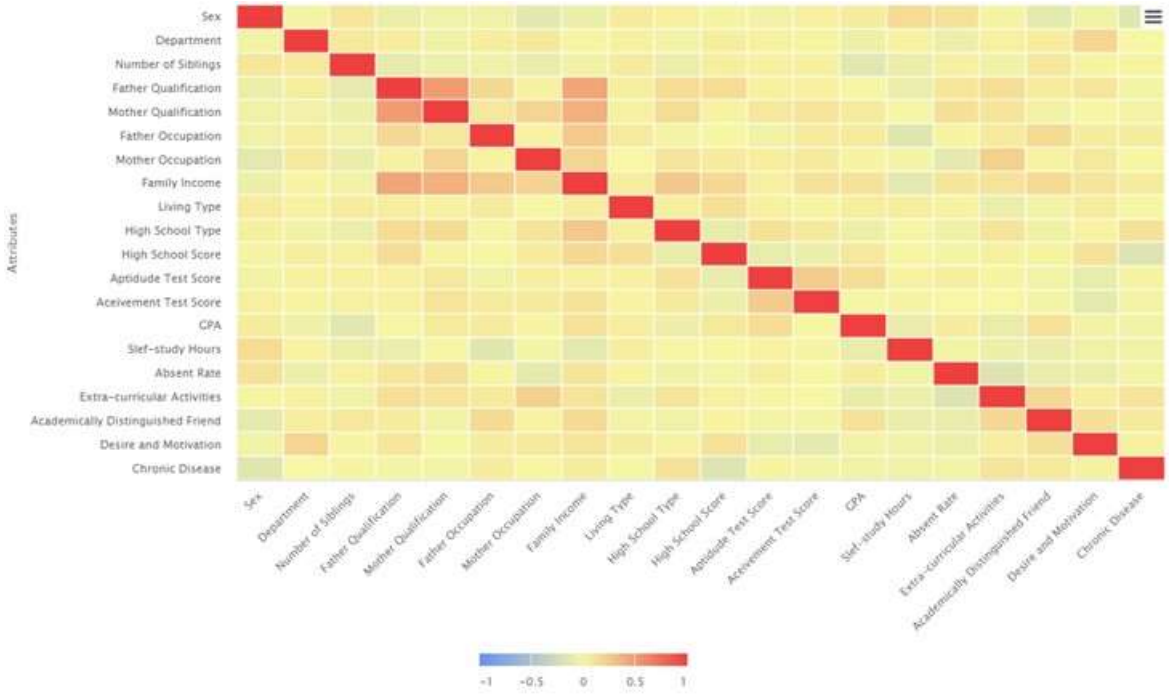

Figure 2. Correlation matrix of attribute variables

Table 4. Confusion Matric Of Decision Tree Algorithm

Class	Actual low-performer	Actual high-performer	Recall	Precision	F1-score	Accuracy
Predicted low-performer	118	21	84.9%	84.8%	0.85	89.7%
predicted high-performer	30	130	81.3%	86.1%	0.84	

Table 5. Confusion Matric Of Random Forest Algorithm

Class	Actual low - performer	Actual high-performer	Recall	Precision	F1-score	Accuracy
Predicted low-performer	73	31	73.1%	77.7%	0.75	79.9%
predicted high performer	120	40	75.0%	79.4%	0.77	

Table 6. Confusion Matric Of SVM Algorithm

Class	Actual low-performer	Actual high-performer	Recall	Precision	F1-score	Accuracy
Predicted low-performer	105	34	78.1%	78.6%	0.78	76.7%
predicted high-performer	35	125	78.1%	078.6%	0.78	

Table 7. Confusion Matric Of Naïve Bayes Algorithm

Class	Actual low-performer	Actual high-performer	Recall	Precision	F1-score	Accuracy
Predicted low-performer	130	59	67.9%	67.9%	0.68	74.3%
predicted high-performer	137	52	71.9%	72.4%	0.72	

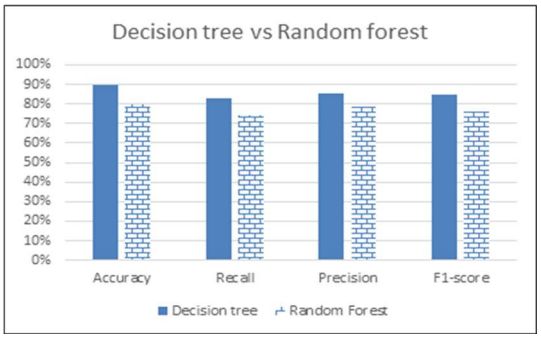

Figure 3. Caparison between decision tree and random forest algorithm

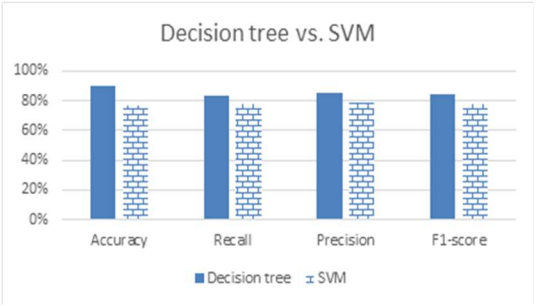

Figure 4: Caparison between decision tree and SVM algorithm

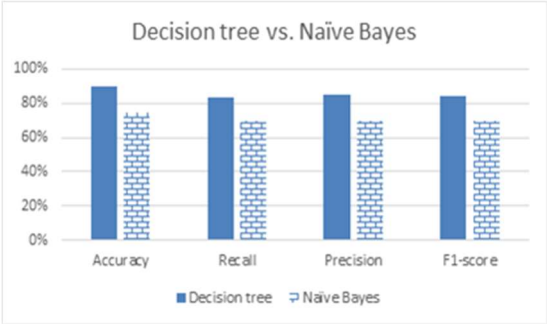

Figure 5: Caparison between decision tree and Naïve Bayes algorithm

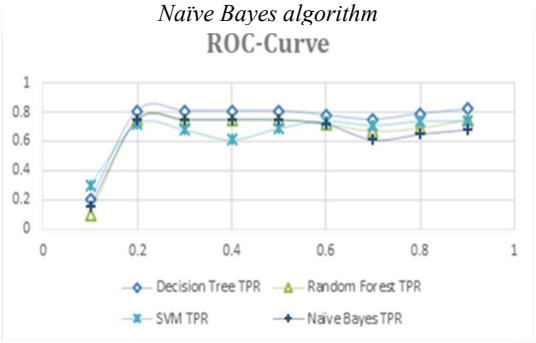

Figure 6: The compared ROC curve of proposed machine learning algorithms

REFERENCES

[1] Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81-106.

[2] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. CRC press.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).

[4] Jain, V., & Kaur, A. (2016). Decision tree-based model for predicting student performance. Journal of Advances in Mathematics and Computer Science, 17(5), 1-11.

[5] Kaur, A., & Jain, V. (2018). Decision tree-based model for predicting academic performance of computer science students. Journal of Computing and Information Technology, 26(3), 207-214.

[6] Zhao, H., Zhang, Z., & Li, S. (2021). A decision tree approach to predicting high school student academic performance. PLoS One, 16(7), e0254902.

[7] Dwivedi, Y. K., Rana, N. P., & Raghavan, V. (2019). Predicting student academic performance in blended learning environments: A comparative study of machine learning algorithms. Technological Forecasting and Social Change, 139, 287-302.

[8] García-Sánchez, F., García-Sánchez, E., & García-Crespo, Á. (2018). Predicting student performance in online learning environments using a decision tree. Journal of Universal Computer Science, 24(5), 600-619.

[9] Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

[10] Gómez-Pérez, E., Cano, A., Gutiérrez-Salcedo, M., & García-Sánchez, F. (2021). A hybrid deep learning approach for predicting student performance in online learning environments. Applied Soft Computing, 102, 107101.

[11] Aydogan, H., & Ulucan, H. (2018). Prediction of student academic performance in higher education with decision tree algorithms. Journal of Education and Learning, 7(3), 108-116.

[12] Chen, C., Chen, J., & Huang, J. (2019). Predicting academic performance based on classification tree models. Journal of Educational Computing Research, 57(6), 1486-1501.

[13] Fugate, J., & Zhang, J. (2020). Evaluating student academic success using decision trees and random forests. Journal of Computing in Higher Education, 32(1), 117-134.

[14] Huang, R., & Chen, T. (2017). A decision tree approach to predicting academic performance of online students. Journal of Educational Technology & Society, 20(4), 16-26.

[15] Jahromi, A. H., & Badi, I. A. (2018). Predicting academic performance of engineering students by decision tree and KNN algorithms. International Journal of Engineering Education, 34(4), 1427-1434.

[16] Li, L., & Chen, Q. (2019). Predicting students' academic performance by combining decision tree-based algorithm and support vector machines. International Journal of Emerging Technologies in Learning, 14(22), 202-214.

[17] Li, X., & Wang, J. (2019). Decision tree algorithm for predicting students' academic performance in literature. International Journal of Emerging Technologies in Learning, 14(18), 173-185.

[18] Li, Y., & Du, H. (2019). Predicting college students' academic performance with decision tree analysis. Journal of Higher Education Theory and Practice, 19(11), 123-134.

[19] Liu, J., Li, X., Li, L., & Zhang, H. (2018). Predicting students' academic performance by combining decision tree-based algorithm and neural network. Journal of Intelligent Systems, 27(4), 687-694.

[20] Mihajlovic, I., & Kostic-Stankovic, M. (2018). Predicting students' academic performance using decision tree method. Computers & Education, 117, 121-129.

[21] Sabatini, R., Ochoa, X., & Amandi, A. (2018). Prediction of academic performance in higher education using decision trees. Journal of Educational Computing Research, 56(3), 426-444.

[22] Sutrisno, A., & Nurdiyanto, H. (2020). Decision tree algorithm for predicting academic achievement of students in higher education. Journal of Physics: Conference Series, 1578, 012003.

[23] Wang, D., & Liu, H. (2020). Predicting academic performance of mathematics students using decision tree and K-Nearest Neighbors algorithms. Journal of Educational Technology Development and Exchange, 13(1), 35-44.

[24] Wang, Y., & Sun, P. (2021). Predicting academic performance in higher education with decision tree algorithms. Education Sciences, 11(3), 128.

[25] Y. Liu, Q. Huang, Y. Xie, and Z. Liu, "Combining decision tree-based algorithms and neural network for predicting academic performance," in 2019 IEEE Intl Conf on Computational Science and Engineering (CSE) and IEEE Intl Conf on Embedded and Ubiquitous Computing (EUC), 2019, pp. 287–290.

[26] H. Li and Y. Chen, "Prediction of students' academic performance based on decision tree algorithm and support vector machines," in 2015 IEEE 13th Intl Conf on Dependable, Autonomic and Secure Computing, 2015, pp. 994–997.

[27] M. Al-Dhahir and H. Al-Khateeb, "Predicting academic performance using a genetic algorithm and decision tree," Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, pp. 129–137, 2013.

[28] T. Wang and W. Liu, "An approach to predict student academic performance based on decision tree and k-nearest neighbor," in 2020 IEEE Intl Conf on Artificial Intelligence and Computer Applications (ICAICA), 2020, pp.235-239.

[29] Mphahlele, M. J., Mantho, R. B., & Adigun, M. O. (2019). Comparative study of classification algorithms for predicting customer churn. Journal of Business Research, 98, 411-420.

[30] Abbas, Q., Hassan, S., Nazir, S., & Mahmood, T. (2018). Comparison of classification algorithms: Decision tree, SVM and k-NN for spam email classification. 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 1-6.

[31] Khan, I., & Bhat, A. R. (2019). A comparative study of decision tree algorithms with Naïve Bayes classifier for sentiment analysis. 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 296-300.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[32] Gakhar, S., & Goel, S. (2018). Determinants of university academic achievement: A review of literature. International Journal of Emerging Technologies and Innovative Research, 5(1), 33-38.

[33] Kelly, K. R., & Price, J. N. (2004). An examination of the factors influencing student academic success in a higher education setting. The Journal of College Student Retention: Research, Theory & Practice, 6(1), 43-60.

[34] Duckworth, A. L., & Seligman, M. E. P. (2018). The science of self-control: What we know and what we need to know. Perspectives on Psychological Science, 13(3), 305-324.