DATA ANALYTICS AND APPLIED MATHEMATICS (DAAM)
e-ISSN: 2773-4854
VOL. 4, ISSUE 1, 40 – 48
DOI: https://doi.org/10.15282/daam.v4i1.9591

ORIGINAL ARTICLE

The development of a predictive model for students' final grades using machine learning techniques

N.H.A. Rahman[1,2], S.A. Sulaiman[2,*], and N.A. Ramli[2]

[1]Department of Business, Faculty of Management and Informatics, Universiti Islam Pahang Sultan Ahmad Shah, Jalang Gambang, 25150 Kuantan, Pahang, Malaysia

[2]Centre for Mathematical Sciences, Universiti Malaysia Pahang, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan Pahang, Malaysia

ABSTRACT – As per research, utilizing predictive analytics in education can be very beneficial. It can help educators improve students' performance by analyzing historical data through various approaches such as data mining and machine learning. However, there is a scarcity of studies on using machine learning and predictive analytics to enhance student performance in Malaysian higher education. This study used the records of 450 students enrolled in the Business Statistics course at Universiti Islam Pahang Sultan Ahmad Shah (UnIPSAS) from 2013, obtained from UnIPSAS's Learning Management System. The aim was to develop the best predictive model for forecasting students' final grades based on their performance levels, using machine learning techniques such as Decision Tree, k-Nearest Neighbor, and Naïve Bayes. The final model was developed using Python software. The results showed a strong negative correlation between the students' carry marks and their final grades, with an r-value of -0.8. Naïve Bayes was found to be the best model, with an AUC score of 0.79.

ARTICLE HISTORY
Received: 14/02/2023
Revised: 27/03/2023
Accepted: 31/03/2023

KEYWORDS
Machine learning
Predictive models
Students' performance
Education

INTRODUCTION

Predictive analytics research has grown in popularity as a result of its ability to provide useful information to educators, and potentially assisting them in enhancing students' achievement in higher education. Educators could use predictive analytics to build an effective mechanism to improve academic achievements, preventing students from dropping out and ensuring student retention [1]. The risk of failing a course, the risk of student's dropout, the grade prediction, and the graduation rate are all common prediction targets [2]. Dropouts have a negative impact on both educational institutions and stakeholders. Furthermore, with the virtual learning methods practised in today's education system, e-learning dropout rates are often greater than face-to-face education [3]. Many factors, including academic performance, health, family, and personal reasons, can lead to dropout, which varies based on the nature of the study and the higher education provider. If a huge number of students dropping out of their respective universities, the higher education provider's reputation might be dropped. Dropout would also result in a significant loss of human capital for the country, as public universities would generate fewer professionals and experts [4].

Students' grades and final marks will be disclosed after the final examination, which means that students will only be aware of their accomplishments after the faculty has announced their grade. If students fail, they will have to repeat the subject in the following semester, incurring university expenditures as well as burdening many other parties, such as parents and lecturers. As a result, students are less likely to remain motivated to learn. In addition, the financial strain on the family will also increase as the student's college loan must be paid even if they do not graduate. Therefore, one of the most effective strategies that should be considered by the education providers is to detect the tendency of failure before the students sit for final examination. Identifying students who require further assistance and taking the necessary steps to improve their performance is also critical [5]. Preliminary prediction of students' grades based on their accumulated marks and previous achievement as well as other factors, will be able to help students and lecturers to take early action to improve students' grades before the final examination. Students will be able to prepare and improve their accumulated marks so that the final grade will be better. Lecturers can also pay more attention to students with low grade predictions

Nowadays, there are a great number of researches and studies that follow the lines of forecasting students' behavior, as well as other associated educational subjects of interest. Many works on this issue have been published in journals and presented at conferences [6]. The results of the studies can be used by educators to predict and monitor the achievement of their students at an early stage and ensure to reduce student failure rates. However, in Malaysia, the machine learning (ML) technique is still underutilized in schools and higher education institutions [1]. Several colleges and universities have utilized this method to forecast a student's final grade or final score for a certain course but there is no evidence that this technique has been used in Islamic University in Malaysia, according to literature review.

The ML techniques can help higher education administrators to take preventative action by identifying students who are at risk of dropping out at an early stage. Thus, the likelihood of these students dropping out and abandoning their

*CORRESPONDING AUTHOR | S.A. Sulaiman | sahimel@ump.edu.my

studies will be reduced [7]. Therefore, the purpose of this study is to develop the predictive models using supervised ML techniques for students' final grades based for Business Statistics (BS) course at Pahang Sultan Ahmad Shah Islamic University (UnIPSAS). Besides that, this study also aimed to determine the factors that affect students' performance as well as to evaluate the performance of all developed models. This study employed the Decision Tree (DT) technique, the *k*-Nearest Neighbor (*k*-NN) and the Naïve Bayes(NB). The following sections will discuss a few of related previous studies on ML techniques used in students' final grade prediction, followed by the methodology proposed in this study. In the last two sections, the results and discussion, as well as the conclusions and some recommendations to improve this research will be elaborated.

LITERATURE REVIEW

There is no doubt that ML is entering a variety of businesses, including the education sector. In the educational area, ML techniques have been used for many topics, including enrollment, graduation forecasts, dropout rates, retention, and students' performance. ML also makes it much easier for the stakeholders, management, deans, and department heads at Higher Education Institutions to make decisions. Student achievement can be determined by various variables based on past studies. In the digital era, the amount of data being captured and consumed is exploding. Predictive analytics is increasingly being used by organizations to improve their performance and operations, minimizing risk, and detecting frauds. They also may undergo predictive analytics in order to make better judgments, which is why predictive analytics has gotten a lot of attention recently. As stated by Mishra and Silakari [8], predictive analytics consists of four steps: collecting and pre-processing raw data, transforming pre-processed data into a data that can be easily managed by the preferred ML technique, generating the training model using the transformed data, and reporting forecasts to the user using the successfully created learning model.

ML, according to Mahesh [9], is the branch of study that enables computers to learn without being pattern recognition. The surge in the amount of education data available through digitalization has resulted an increase in the usage of ML in education [10].

One of the most often used prediction techniques is the DT. Because of its simplicity and comprehensibility, most researchers have utilised this technique to explore little or huge data structures and predict the value [11]. Study by Hasma et al. [12] found that the DT is stricter than the Fuzzy Genetic Algorithm (FGA) in predicting students' academic performance among 120 and 48 students from bachelor and master degree programs respectively. To predict student grades for a research project, Abana [13] created a classification model employing DT (RepTree, Random Tree, and J48) method. Over the course of four years of study, the prediction model was tested on 133 samples with five variables. With an accuracy of 75.188 percent, the study found that Random Tree is the optimum approach.

The NB method is a straightforward probabilistic classifier based on the Bayes theorem. NB is an inductive learning technique for ML and data mining. These algorithms employ Bayesian probability theory to forecast future probabilities based on prior experience [14]. The goal of Pojon [15] research is to determine the best effective comparison-based prediction algorithm for predicting students' performance by using Grade Point Average (GPA) and final grade. Final mark and performance category were set as the target variable. According to the result, NB classification was the best algorithm for his first data set, with 98 percent accuracy, and DT was the best for his second data set, with 78 percent accuracy.

The *k*-NN method is one of the most fundamental and important classification algorithms in ML. It is non-parametric and does not make any assumptions regarding data distribution [16]. Dhilipan et al. [16] found a high accuracy value with *k*-NN in their research which is 93.71% for student's grade prediction (predicted variable) by using previous semester marks as predictor variables. The *k*-NN was one of the method employed in [17] study besides Linear Regression, LR and RF.

Finally, according to the previous studies, researchers have predicted students' final grades based on their demographic attributes such as gender, student ID, class, year intake, and religion as stated by Bujang et al. [1], Karlos et al. [18], Aman et al. [19] and others. Other attributes used to predict student's grade are continuous assessment marks such as tests, quizzes, and all assessment marks, as stated by Altabrawee et al. [5], Majeed and Junejo [20], and Khan et al. [21]. These studies proved that the combination of these variables can contribute to a high level of accuracy in the construction of predictive models for student grades.

METHODOLOGY

This section describes the methods employed in developing the predictive model to forecast students' final grade. To achieve this objective, it is necessary to experiment with some machine learning techniques. The data for this study includes demographic information (gender and course) and students' performance data (CGPA before taking statistics subjects, grade for pre-requisite subject, accumulated marks, and performance level). The process of machine learning involves data collection, data preprocessing, correlation analysis, and model development. This research proposed to employ supervised models because the aim of supervised learning is to determine the best model so that if new input or data is given, the model can forecast the output. The ML models that have been developed are the Decision Trees (DT), the *k*-Nearest Neighbor (*k-NN*), and the Naive Bayes (NB). Figure 1 shows the flowchart showing all steps in ML process employed in this study.

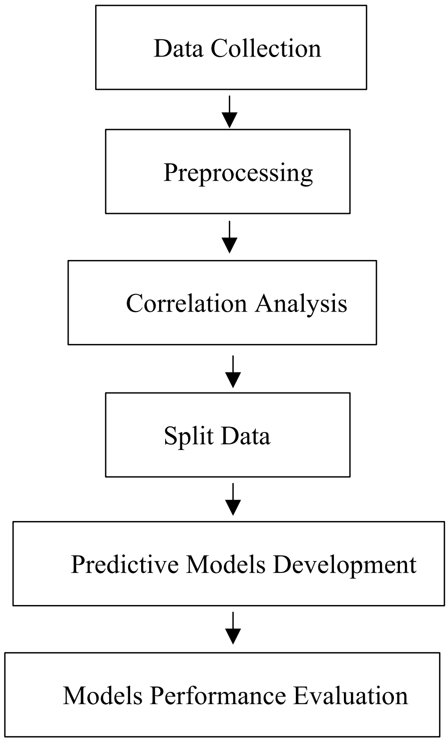


Figure 1. The flow chart of ML process

In data collection step, all the information are extracted from the Learning Management System (LMS) of UnIPSAS which provides students' personal records, course results, and grade points. The dataset was initially gathered and saved as an Excel file, and then the complete dataset will be loaded in Python as a CSV file. The dataset used in this study is the actual data of students from Diploma in Business Studies, Accounting, Marketing Management, as well as Finance and Banking students at UnIPSAS from June 2013 to June 2022. The dataset contains data of 450 students taking Business Statistics (BS) course in fourth semester. As stated in Table 1, the attributes involve are gender, course, Cumulative Grade Point Average (CGPA) for semester before taking BS subject, Business Mathematics (BM) grades, total continuous marks, and student's performance level according to grades in BS final examination.

In ML, data preprocessing refers to the process of making preparations (cleaning and organising) of raw data in order to make it appropriate for developing and training ML models. In other word, data preprocessing is a data mining approach used in ML that processes actual data into a clear and readable format. Preprocessing data is essential for reducing the attribute's high dimensional data [22]. In this study, preprosessing data is conducted in Python and involves data cleaning and encoding process. While cleaning the data, normally the analysis supposed to detect any missing values and duplicated data in the data set. It is important to deal with missing values because they can change the results of the ML models or make the models less accurate. All errors and any inconsistencies are should be cleaned. For instance, CGPA should be in reasonable range which is between 1.00 to 4.00 as well as the carry marks for BS course, the marks should lies between 0 to 50 marks.The encoding process need to be run on string variables, which are the student's gender, student's course, BS grades, and performance levels, in a numerical form as illustrated in Table 1. In Python, the encoding funtion used was LabelEncoder( ).

Meanwhile, the correlation analysis in ML is the step to identify how the independent variables correlate with the output. In order to find the factors that affected students' final grades in BS courses, the correlation analysis will be run in Python. The variables with good or high correlation value ($\pm$ 0.7 to $\pm$ 1.0) will be identified as the factor(s) that affected students' grade. Next, after the correlation analysis is done, the data were splitted into train and test. The first 80% (in cosecutive order) of the dataset will be selected for training data, and the last 20% (recent data) will be selected as testing data. The main objective of data splitting process is to avoid overfitting. Overfitting is an obstacle in ML where the selected model performs very well on training data but poorly perform on new dataset. When the data was successfully splitted, the process continued with model development.

Table 1. List of attributes

No.	Attributes	Description	Details	Encode value
1.	Gender	Student's gender	Male	1
			Female	0
2.	Program	Students' program	Diploma in Accounting	0
			Diploma in Business Studies	1
			Diploma in Finance and Banking	2
			Diploma in Marketing Management	3
3.	CGPA	CGPA for semester before taking BS subject.	1.00 – 4.00	-
4.	CM	Carry Marks for Business Statistics	0 - 50	-
5.	BM_grade	Business Mathematics grade	A	11
			A-	10
			B+	9
			B	8
			B-	7
			C+	6
			C	5
			C-	4
			D+	3
			D	2
			E	1
6.	BS_Performance level (grades)	Student's performance level according to grades in Business Statistics final examination	Excellent (A, A-)	0
			Very good (B+, B)	1
			Good (B-, C+)	2
			Pass (C, C-)	3
			Weak (D+, D)	4
			Fail (E)	5

As stated early, there were three models that have been developed. The first model was DT. The DT is a diagram similar to a flowchart that depicts the many consequences of a set of choices. A decision tree can be represented as a tree consisting of nodes from root to leaf, with inspections on characteristics set in internal nodes and class variables shown in leaf nodes [20]. To find solutions, a decision tree makes sequential, hierarchical decision about the outcomes variable based on the predictor data [5]. Changing parameters such as the quality measure, the splitting criteria, the minimum number of records per node, and the pruning procedure may enhance the accuracy of a decision tree model's predictions. In Python, the DecisionTreeClassifier will be used to develop the DT model. In Python, the DecisionTreeClassifier was used to develop the DT model and the hyperparameters that has been tuned are Gini and max_depth.

Second model was the *k*-NN model. The *k-NN* technique is a basic and straightforward supervised machine learning approach that may be used to handle both classification and regression issues. The classifier determines a data item as belonging to the training dataset class to which it is geometrically closest [23]. In addition, the *k*-NN algorithm assumes that similar things are near to each other. For classification cases, *k*-NN operates by calculating the distance between an inquiry and each example in the data, picking the *k* closest to the query, and then voting for the most frequent label. The *k* value was determined based on the error rate k value graph.

The third model was NB which is the one of the most popular supervised machine learning methods. The NB Classifier, is a straightforward probabilistic classifier based on the Bayes theorem. NB's effectiveness stems from the assumption of attribute independence, which may not hold true in many real-world data sets [24]. Several approaches have been taken to address the assumption, with attribute selection being one of the most important. On the other hand, conventional methods of attribute selection in Naive Bayes have a significant computational cost [24].

In order to determined the best model, the evaluation of model performance was very crucial. In this research, model performances were evaluated by comparing the value of accuracy, precision, recall, and f-measure, as well as the Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve. The model with higher value or score of performance is selected as the best model in predicting students' final grades for BS subject in UnIPSAS.

The number of True Positive (TP) refers to the frequency of predictions where the classifier correctly predicts the positive class to be positive. While the number of True Negative (TN) is referring to the frequency of predictions where the classifier or model successfully predicts the negative class as negative correctly. On the other hand, False Positive (FP) refers to the frequency of predictions made by the model which it incorrectly predicts the negative class as positive. While False Negative (FN) defines as the number of predictions which the model incorrectly predicts the positive class as negative.

As an assessment criterion for machine learning model, the confusion matrix is always preferable. It provides with extremely straightforward and effective performance indicators for classifiers. There are several performance measures can be use from the confusion matrix. For instance, by using the confusion matrix, the value of accuracy, precision, recall and F-measure can be obtained. The accuracy of a system may be defined as the degree of similarity between the expected and actual values of a quantity [20]. The measure of accuracy is the number of accurately detected events. The higher the accuracy value, the better the performance of a model. It is most prevalent when all classes are of equal importance. The accuracy can be measured by the following formula:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{1}$$

The precision value is the proportion of predictions as a positive class were actually positive. It is defined as:

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

The recall indicates what proportion of all positive samples the classifier accurately identified as positive. It is also referred to as the True Positive Rate (TPR), the Sensitivity, and the Probability of Detection. To calculate recall, the following formular should be considered:

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

F-Measure gives a method for combining accuracy and recall into a single metric that includes both characteristics. Traditionally, F-measure is calculated as follows:

$$\mathbf{F-measure} = \frac{(2 \times \text{Precision} \cdot \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

In order to conclude the values, precision, recall, and F-measure, there are similar descriptions, which is, the 0.5 score and less indicates poor value, more than 0.5 to 0.7 indicates moderate to good performance, more than 0.8 to 1.0 indicates the best to perfect performance value.

The One-vs-All (OvA) setting is where the AUC of ROC curve for multi-class models is most frequently applied. By classifying all classes as positive, this technique is used to create an overall AUC. This method trains a group of independent binary classifiers, with one class being positive and the others being negative. The curves' areas are averaged after calculating the AUC for each binary classifier to provide a final area under the ROC curve. In order to conclude the values of accuracy, precision, recall, F-measure, as well as the AUC of ROC curve, there are similar descriptions, which is, the 0.5 score and less indicates poor value, more than 0.5 to 0.7 indicates moderate to good performance, more than 0.8 to 1.0 indicates the best to perfect performance value.

Python is becoming one of the most prominent programming languages for scientific computing and increasingly used not just in academic contexts, but also in the business world. Scikit-learn makes use of this environment to deliver state-of-the-art implementations of plenty of well machine learning techniques, while keeping a user-friendly interface that is firmly integrated with the Python programming language. This satisfies the rising need for statistical data analysis by non-specialists in the software and online businesses, as well as in non-computer science subjects such as biology and physics [25]. Python libraries enable their users to access, analyze, and alter data for machine learning, which required regular data processing [26]. All processes involved in this study are conducted using Python software.

RESULTS AND DISCUSSION

As stated in Table1, six variables were involved which are gender, student programme, student CGPA before taking the business statistics subject, total accumulated marks, grade for the pre-requisite subject which is business mathematics, and the performance level. The first step in data preprocessing is encoding, where all the non-numeric variables will be encoded into numerical data. The LabelEncoder() function has been used to encode data for gender, course, and performance level. Before the process of modelling with machine learning, statistical analysis is necessary to comprehend the data. The correlation analysis is conducted to identify variables that affect the target variable, which is student performance level according to the final grades.

Correlation analysis was used in this study to investigate how strongly the predictors, or independent attributes, correlate to the performance level as the target variable. A heatmap is a visualisation of the correlation matrix in which the correlation values between variables are represented by colour. The study used a heatmap to visualise the findings of the correlation analysis between the variables in the dataset.

Referring Figure 2, the heatmap revealed that one factor correlated strongly negative with $\text{r} = -0.8$, which is student's carry marks, followed by the CGPA, which has a moderately positive relationship ($\text{r} = -0.49$) to the student's performance level. The carry marks have a strong negative correlation, showing that higher carry marks will give a lower code for performance level, which represents the higher level as stated previously. Apart from that, the remaining variables show a very weak relationship to the target variable. Overall, the heatmap gave a clear and straightforward visualization of the correlations in the dataset and assisted us in determining which factors were highly related to one another.

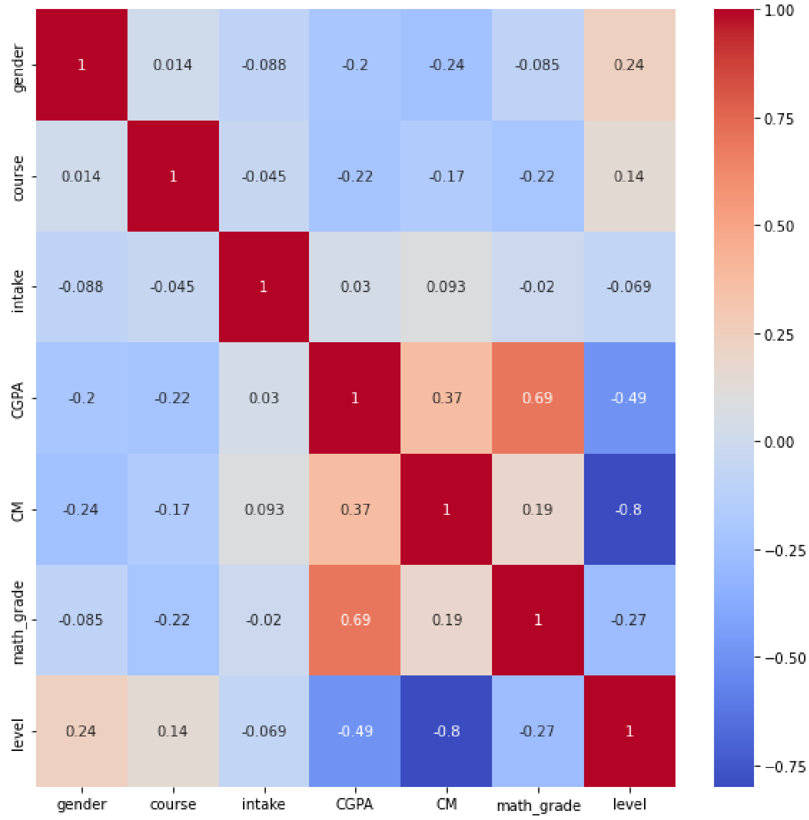
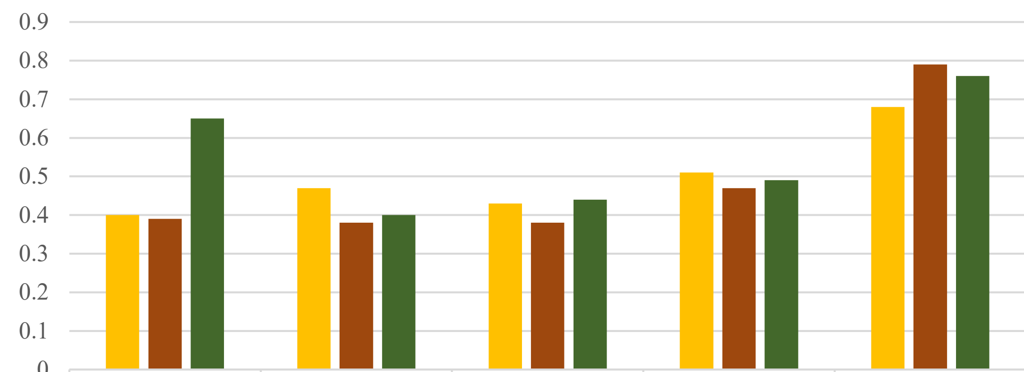


Figure 2. The correlation heatmap

Model performance based on the value of Precision, Recall, F1-score and Accuracy which were presented in the classification report, as well as the Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve for each model are visualize in Figure 3. Since these models were trained to predict six classes (six performance levels 0 to 5), the AUC performance was determined by calculating the average AUC value of all classes.



	Precision	Recall	f1-score	Accuracy	Average AUC
DT	0.4	0.47	0.43	0.51	0.68
NB	0.39	0.38	0.38	0.47	0.79
kNN	0.65	0.4	0.44	0.49	0.76

Figure 3. Comparison of all classification models' performances

The above figure also shows that the NB performed the highest average AUC of ROC curve value which is 0.79. The AUC of ROC curve for the NB model is presented in Figure 4. By comparing the AUC scores for each class, insights into the relative predictive performance of the model for different classes could be gained. According to Figure 4 , class 5 has an AUC rating of 0.98. This shows that the model does even better when distinguishing instances of student performance that result in failure from those that do not. In addition, this model also shows a very good performance in differentiating the situation of students who get an excellent level (class 0) from other levels of performance, but slightly less than the performance of class 5. In fact, the results of the overall analysis have found that the AUC score for each ML model has performed very well in distinguishing class 5 from other classes. As shown in Figure 5 and Figure 6, the *k*-NN and DT gave the AUC score for the ROC curve 1.00 and 0.99 respectively for class 5.

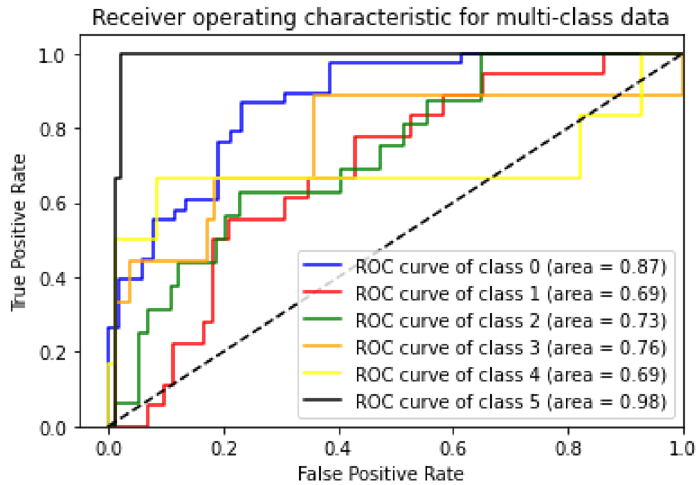


Figure 4. Receiver operating characteristic for NB model

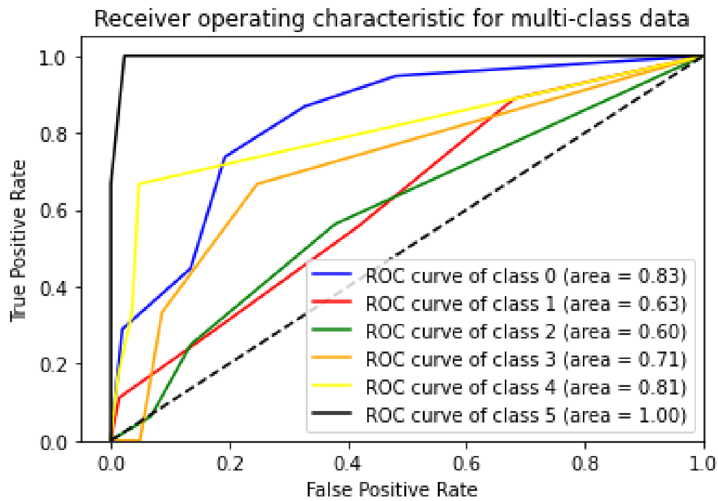


Figure 5. Receiver operating characteristic for *k*-NN model

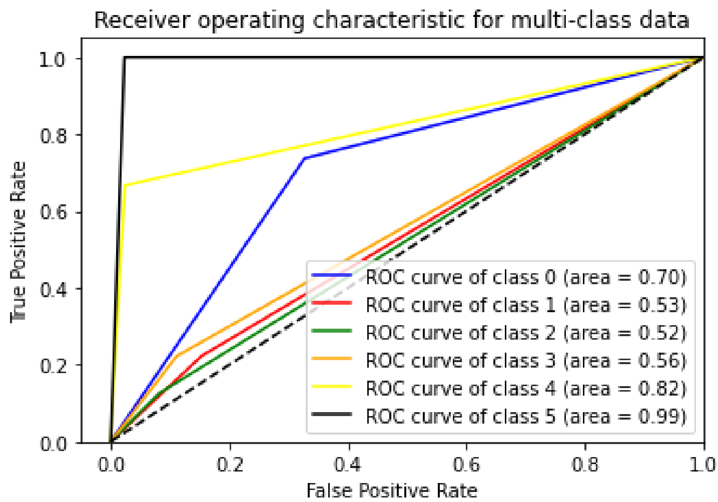


Figure 6. Receiver operating characteristic for DT Model

CONCLUSION

In conclusion, the objectives of the study are successfully achieved. The NB model using the AUC score of ROC Curve 0.79 is the best in predicting student final grades for the business statistics course in UnIPSAS. The AUC score is the best evaluation value for the performance of each developed model. While the accuracy value could not give a good value when it only recorded a value of 0.49 for *k*-NN, 0.47 for NB, and 0.51 for DT. It is needed to state that the final exam marks are worth 50 percent of the total grade, which is quite large weightage. Therefore, achieving perfect accuracy on this data set is quite impossible [20]. In certain situations, students' performance on the final examination is different from their performance for assessment marks. Some students were unable to study for their final examinations due to illness or an emergency and ended up with lower grades. While some other students concentrated more and excelled at their final exams, but they did not perform well in their assessments before the final exam. Since the model cannot be done based on these types of factors, predicting the final grade with 100 percent accuracy is impractical [20].

The student whose achievement is most accurately predicted is the student who gets an E. Hence, this is very useful in identifying students who may fail, and early action can be taken by lecturers to improve student performance. If the lecturers can predict the students who get an E or fail before the final exam, these students can improve their carry marks so that they will not fail on the actual results. Carry marks have proven to be the most strongly correlated factor and greatly influence students' final grades.

In the future, it is suggested that the data of students taking this subject should be added to help the increament of model accuracy. In addition, other ML models that can also be developed to predict student grades are Support Vector Machines, Multiclass Logistic Regression, and ensemble method such as Random Forest.

REFERENCES

[1] S.D.A Bujang, A. Selamat and O. Krejcar, "A predictive analytics model for students grade prediction by supervised machine learning," IOP Conference Series: Materials Science and Engineering, vol. 1051, p. 012005, 2021.

[2] A.E. Tatar and D. Düştegör, "Prediction of academic performance at undergraduate graduation: course grades or grade point average?," Applied Sciences, vol. 10, no. 14, p. 4697, 2020.

[3] J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses using different machine learning techniques," Applied Sciences, vol. 11, no. 7, p. 3130, 2021.

[4] N.S. Sani, A.F.M. Nafuri, Z.A. Othman, M.Z.A. Nazri and K.N. Mohamad, "Drop-out prediction in higher education among B40 students," International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, pp. 550-559, 2020.

[5] H. Altabrawee, O.A.J. Ali and S.Q. Ajmi, "Predicting students' performance using machine learning techniques," Journal of University of Babylon for Pure and Applied Sciences, vol. 27, no. 1, pp. 194-205, 2019.

[6] J.L. Rastrollo-Guerrero, J.A. Gómez-Pulido and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: a review," Applied Sciences, vol. 10, no. 3, p. 1042, 2020.

[7] N. Roslan, J.M. Jamil and I.N.M. Shaharanee, "Prediction of student dropout in Malaysian's private higher education institute using data mining application," Turkish Journal of Computer and Mathematics Education, vol. 12, no. 3, pp. 2326-2334, 2021.

[8] N. Mishra and D.S. Silakari, "Predictive analytics: a survey, trends, applications, oppurtunities & challenges," International Journal of Computer Science and Information Technologies., vol. 3, no. 3, pp. 4434-4438, 2012.

[9] B. Mahesh, "Machine learning algorithms-a review," International Journal of Science and Research, vol. 9, no. 1, pp. 381-386, 2020.

[10] N. Mduma, K. Kalegele and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," Data Science Journal, vol. 18, no. 1, pp. 1-10, 2019.

[11] A.M. Shahiri, W. Husain and N.A Rashid, "A review on predicting student's performance using data mining techniques," Procedia Computer Science, vol. 72, pp. 414-422, 2015.

[12] H. Hamsa, S. Indiradevi and J.J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm," Procedia Technology, vol. 25, pp. 326-332, 2016.

[13] E.C. Abana, "A decision tree approach for predicting student grades in research project using Weka," International Journal of Advanced Computer Science and Applications., vol. 10, no. 7, pp. 285-289, 2019.

[14] F. Ofori, E. Maina and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: a literature based review," Journal of Information and Technology, vol. 4, no 1, pp. 23-45, 2020.

[15] M. Pojon, "Using machine learning to predict student performance," M.Sc. thesis, Faculty of Natural Sciences Software Development, University of Tampere, Tampere, Finland, 2017.

[16] J. Dhilipan, N. Vijayalakshmi, S. Suriya and A. Christopher, "Prediction of students performance using machine learning," IOP Conference Series: Materials Science and Engineering, 1055, p. 012122, 2021.

[17] J. Xu, K.H. Moon and M. van der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742-753, 2017.

[18] S. Karlos, G. Kostopoulos and S. Kotsiantis, "Predicting and interpreting students' grades in distance higher education through a semi-regression method," Applied Sciences, vol. 10, no. 23, p. 8413, 2020.

[19] F. Aman, A. Rauf, R. Ali, F. Iqbal and A.M Khattak, "A predictive model for predicting students academic performance," in Proceeding of 2019 10th International Conference on Information, Intelligence, Systems and Applications, Greece, 2019, pp. 1-4.

[20] E.A. Majeed and K.N. Junejo, "Grade prediction using supervised machine learning techniques," in Proceeding of the 4th Global Summit on Education Conference, Malaysia, 2016, pp. 222-234.

[21] B. Khan, M.S.H. Khiyal and M.D. Khattak, "Final grade prediction of secondary school student using decision tree," International Journal of Computer Applications., vol. 115, no. 21, pp. 32-36, 2015.

[22] M.Y.I. Basheer, S. Mutalib, N.H.A. Hamid, S. Abdul-Rahman, A.M.A. Malik, "Predictive analytics of university student intake using supervised methods," IAES International Journal of Artificial Intelligence, vol. 8, no. 4, pp. 367-374, 2019.

[23] T. Anderson and R. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," Global Journal of Business Pedagogy, vol. 1, no. 3, pp. 13-22, 2017.

[24] A.C.Y. Hong, K.W. Khaw, X.Y. Chew and W.C. Yeong, "Prediction of US airline passenger satisfaction using machine learning algorithms," Data Analytics and Applied Mathematics, vol. 4, no. 1, pp. 8–24, 2023.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[26] S. Kumar, (2021, June 23). Why Python is Best for AI, ML, and Deep Learning. Available: https://www.rtinsights.com/why-python-is-best-for-ai-ml-and-deep-learning/