JURNAL PENDIDIKAN SAINS DAN MATEMATIK MALAYSIA
Jilid 15 Isu 1 2025 / ISSN 2232-0393 / eISSN 2600-930

Predicting Students' Performance in Mathematics Subjects at Kolej MARA Banting using Machine Learning Methods

Ramalan Prestasi Pelajar dalam Mata Pelajaran Matematik di Kolej MARA Banting Menggunakan Kaedah Pembelajaran Mesin

Ahmad Akif Ibrahim[1], Nor Azuana Ramli[2*] & Sahimel Azwal Sulaiman[2]

[1]Mathematics Department, Kolej MARA Banting, Jalan Labohan Dagang, Bukit Changgang, 42700 Banting, Selangor, Malaysia.
[2]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia.

**Corresponding author: azuana@umpsa.edu.my*

Diterima: 28 Ogos 2024; **Disemak semula**: 25 Januari 2024 **Diterima:** 06 April 2025; **Diterbitkan:** 28 April 2025

To cite this article (APA): Ibrahim, A. A. ., Ramli, N. A. ., & Sulaiman, S. A. . (2025). Predicting Students' Performance in Mathematics Subjects at Kolej MARA Banting using Machine Learning Methods. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, *15*(1), 19-31. https://doi.org/10.37134/jpsmm.vol15.1.2.2025

To link to this article: https://doi.org/10.37134/jpsmm.vol15.1.2.2025

ABSTRACT

Predicting students' performance is crucial for personalised and educational success for individuals. However, no standard procedure or method considers external factors to predict students' performance in mathematics at Kolej MARA Banting (KMB). This research aims to address this problem by exploring the potential of machine learning methods for predicting students' performance in mathematics at KMB. The study follows a machine learning process: data collection, attribute selection, pre-processing, model training, and evaluation. A sample of 703 data points on students' demographics, academic records, and mathematics performance were collected and pre-processed. Machine learning models such as support vector machine, decision tree, k-nearest neighbours, Naïve Bayes, Random Forest, AdaBoost, and stacking model were applied in this study. The accuracy and performance of these models were assessed to determine which model outperformed the others and its effectiveness in predicting students' mathematics performance. The study findings demonstrate that the stacking model exhibited superior performance in accuracy (71.43%), precision (68.73%), recall (71.43%), and F1-score (69.80%) compared to the other models. Nevertheless, it is essential to note that the stacking model achieved moderate accuracy. This could be attributed to the inherent difficulties in constructing a precise predictive model for student performance, such as the models failing to sufficiently reflect the complexities within the dataset, resulting in underfitting. Additionally, the target attribute, International Baccalaureate (IB) grade, is imbalanced, with more high performers than low performers, causing the models to be biased towards the majority class and impacting overall accuracy. The performance of the models in this study could be improved by adding more features related to students' performance, such as anxiety, depression, well-being, and others, to capture enough complexity in the data. It is also suggested that samples from other colleges with a balanced grade distribution be obtained compared to students at KMB.

Keywords: Machine Learning, Students' Performance, Mathematics Subjects, International Baccalaureate, Predictive Modelling

ABSTRAK

Meramal prestasi pelajar adalah penting bagi kejayaan peribadi dan pendidikan seseorang. Walaubagaimanapun, sehingga kini tiada prosedur standard dan kaedah yang mempertimbangkan faktor luaran untuk meramalkan prestasi pelajar dalam subjek matematik di Kolej MARA Banting (KMB). Kajian ini bertujuan untuk membangunkan model ramalan dengan menggunakan kaedah pembelajaran mesin bagi meramal prestasi para pelajar dalam subjek matematik di KMB. Kajian ini dijalankan mengikut kerangka kajian sistematik yang terdiri daripada pengumpulan data, pemilihan data, pra-pemprosesan, latihan bagi model, dan penilaian. Sebanyak 703 sampel data yang mengandungi demografi pelajar, rekod akademik, dan prestasi matematik telah dikumpulkan dan dipreproses. Teknik mesin pembelajaran seperti mesin sokongan vektor, pokok keputusan, k-nearest neighbour, Naïve Bayes, hutan rawak, AdaBoost, model gabungan dan tumpukan telah diaplikasikan dalam kajian ini. Ketepatan dan prestasi model-model ini telah dinilai untuk menentukan model yang mana mempunyai prestasi paling tinggi berbanding model-model lain dan keberkesanannya dalam meramal prestasi para pelajar dalam subjek matematik. Hasil kajian menunjukkan bahawa model tumpukan menunjukkan prestasi yang lebih unggul dalam ketepatan (71.43%), kejituan (68.73%), ulangan (71.43%), dan skor F1 (69.80%) berbanding dengan model-model lain. Walau bagaimanapun, model tumpukan ini mencapai ketepatan sederhana. Ini mungkin kerana beberapa batasan dalam membangunkan model ramalan untuk meramal prestasi pelajar dengan tepat, contohnya, model mungkin tidak menangkap kerumitan yang mencukupi dalam data atau dengan kata lain, ia tidak sesuai. Selain itu, ciri sasaran, gred IB adalah tidak seimbang, dengan lebih banyak prestasi tinggi daripada prestasi rendah dan ini membawa model ke arah kelas majoriti, memberi kesan kepada ketepatan keseluruhan. Prestasi model dalam kajian ini boleh dipertingkatkan dengan menambah lebih banyak ciri yang berkaitan dengan prestasi pelajar seperti kebimbangan, kemurungan, kesejahteraan, dan lain-lain untuk menangkap kerumitan yang mencukupi dalam data dan mendapatkan sampel daripada kolej lain yang mempunyai gred keseimbangan berbanding pelajar di Kolej MARA Banting.

Kata kunci*: Kaedah Mesin Pembelajaran, Prestasi Pelajar, Subjek Matematik, Model Ramalan, International Baccalaureate*

INTRODUCTION

The International Baccalaureate (IB), previously known as The International Baccalaureate Organization, was established in 1968 in Geneva, Switzerland. The IB collaborates with schools, ministries, and international organisations to offer rigorous international education programmes and examinations, including the Primary Years Programme, Middle Years Programme, Diploma Programme, and Career-related Programme. The IB Diploma Programme, with its emphasis on critical thinking, research skills, and understanding the interconnectedness of different subject areas, is designed to provide students with a well-rounded education. This unique focus underscores the importance of these skills in student development. Students must take six subjects from different subject groups, including a second language, and complete three components: The Theory of Knowledge, the Extended Essay, and the Creativity, Activity, and Service programme. Additionally, the IB Diploma Programme has been found to promote intercultural understanding and international-mindedness among students.

The IB Diploma Programme is a rigorous and comprehensive pre-university education programme that provides students with numerous benefits, including higher academic achievement, critical thinking skills, and intercultural understanding. It focuses on developing well-rounded individuals capable of contributing to a better and more peaceful world, which is known to be one of its many strengths (IB, 2022). Since higher academic achievement is essential for the IB Diploma Programme, predicting students' performance is critical for fostering early interventions and guiding continuous improvement for students who are predicted to have low grades. Before predicting students' performance, it is necessary to comprehensively analyse past students' performance and choose the proper method for developing predictive modelling. In this context, machine learning, a subfield of artificial intelligence (AI) and computer science that simulates human learning using data and algorithms while continually enhancing accuracy, plays a crucial role. The use of machine learning in this research study represents an innovative and advanced approach in the field of education. Algorithms are trained to make classifications or predictions using statistical methods, demonstrating significant findings within data mining projects. Furthermore, many organisations increasingly use predictive analytics to improve performance and operations, minimise risk, and detect fraud. They may also undergo predictive analytics to make better judgments, which has recently garnered much attention (Alloghani, 2020).

Student performance prediction has become a significant focus for educational institutions worldwide. The goal is to enhance learning outcomes and ensure student success. Accurate predictions can enable early intervention and personalized support for struggling students, ultimately affecting graduation rates and workforce readiness. Predicting student performance can help educators identify at-risk students and prepare them for university entrance. This study aims to develop a machine learning model that can accurately predict student performance in mathematics at Kolej MARA Banting (KMB). Such a model has the potential to significantly impact educational practices and student success by minimizing reliance on subjective judgment and potential biases, thus allowing for early intervention and targeted support for at-risk students. The findings of this research can benefit KMB by identifying attributes that significantly impact student achievement and demonstrating the reliability of KMB's predicted grade to the IB organization. Overall, this study aims to explore the potential of machine learning in improving the prediction of students' performance in mathematics at KMB, with the ultimate goal of enhancing educational practices and student success.

RESEARCH PROBLEM

Following the International Baccalaureate (IB) curriculum, students at Kolej MARA Banting (KMB) must select six subjects from various combinations, including at least one mathematics subject. Academic performance in mathematics at KMB is crucial for determining students' future success. The mathematics curriculum at KMB offers three choices: Mathematics Analysis and Approaches at Higher and Standard Levels (MAA HL/SL) and Mathematics Application and Interpretation at a Higher Level (MAI HL). This curriculum stands out due to its unique combination of syllabus content and teaching methods. These courses aim to develop students' mathematical reasoning skills and problem-solving abilities, providing a solid foundation in mathematical concepts and techniques while preparing them for real-world applications. Prestigious educational institutions highly value the IB program's mathematics curriculum, as it demonstrates its ability to nurture intellectually bright and curious learners well-prepared for advanced academic pursuits. Traditionally, predicting student performance has relied on previous grades and standardised test scores, but it often overlooks external factors. However, predicting student performance in mathematics subjects at KMB has been challenging due to the lack of reliable prediction tools. This obstacle hampers KMB's ability to identify students at risk of underperforming and intervene early to provide additional support, affecting students' overall educational experience and future success. Lack of prediction also impacts educators' ability to predict students' grades for university placement accurately. To address this issue, an effective machine learning model should be developed to predict students' performance in mathematics subjects at KMB. This model should consider factors such as students' previous academic performance and demographic information. Predicting students' performance is vital for educators as it enables them to identify students who may require assistance. This study explores alternative methods to improve the current approach to predicting students' performance and their upcoming IB exam performance based on their data and past results.

RESEARCH OBJECTIVES

1. To identify attributes used in predicting students' performance.
2. To develop the best predictive model for students' performance in mathematics in KMB using machine learning methods.
3. To evaluate the predictive models' performance using classification metrics

RESEARCH QUESTIONS

1. To identify attributes used in predicting students' performance.
2. To develop the best predictive model for students' performance in mathematics in KMB using machine learning methods.
3. To evaluate the predictive models' performance using classification metrics.

RESEARCH METHODOLOGY

Research Framework

The research framework for this study was structured to systematically address the objectives of predicting student performance in mathematics subjects at Kolej MARA Banting (KMB) using machine learning (ML) methods. The study follows a detailed, step-by-step approach encompassing data collection, pre-processing, model development, and evaluation, as shown in Figure 1. This design ensures the robustness and reliability of the predictive models developed. The process started with data collection and preparation to gather relevant information on students' demographics, academic records, and performance in mathematics subject. Then, attribute selection and data partition were applied to enhance data quality. Subsequently, various ML algorithms were utilized during the model development. The study mainly emphasises supervised machine learning methods to predict student performance. In other words, the study used the classification approach to train and validate the predictive model. Finally, the model's accuracy and performance were meticulously evaluated, providing a comprehensive understanding of the model's capabilities. Based on the performance, the best model is chosen, and it is tested with new data to see if there is a massive gap between the actual data and predicted values provided by the best model.

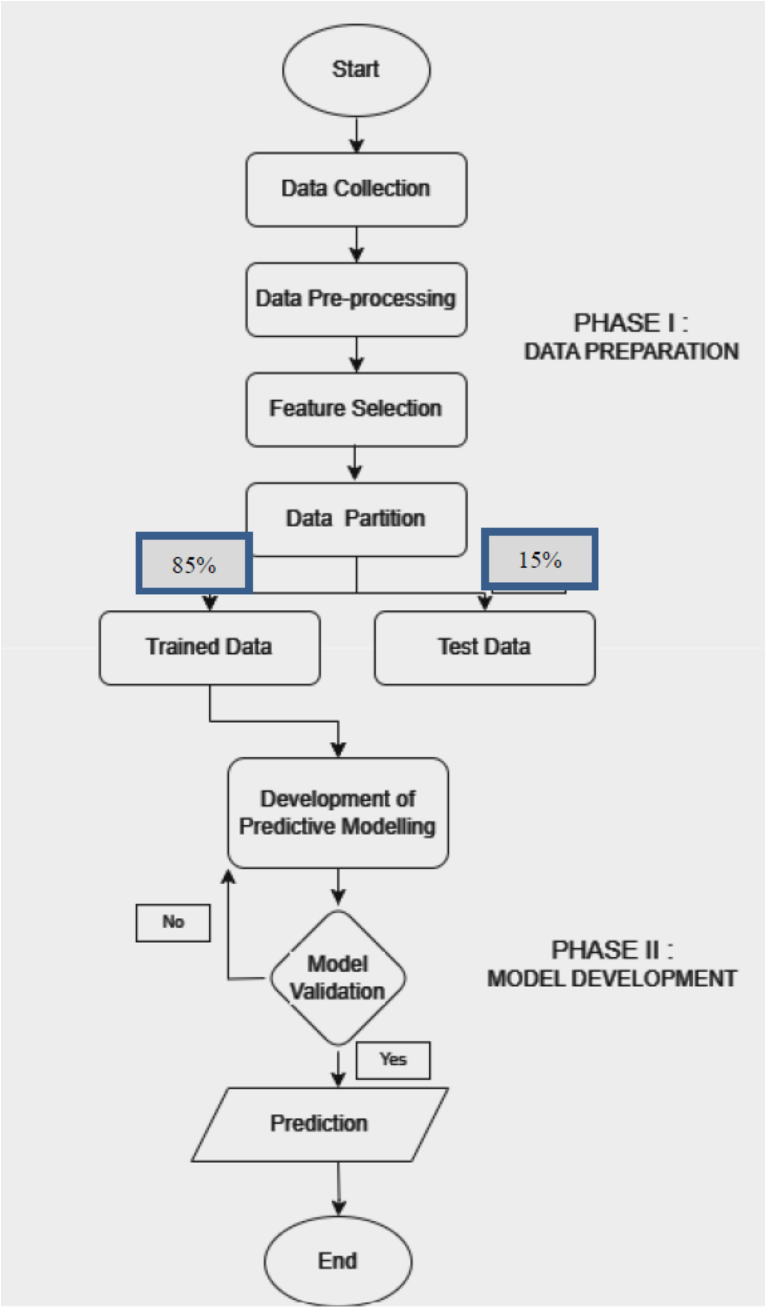

Figure 1: Research framework using machine learning methods

The ML process requires a structured approach in which a computer system is trained to recognize patterns and make predictions or decisions without explicit programming. This study is divided into two-phase methodology: data preparation and model development, as shown in Figure 1. To analyze the datasets, the initial step involved assessing the attributes' significance using statistical methods such as the Chi-square test since the attributes data are categorical. For the model development phase, this research employed all supervised models, including single classifier and ensemble classifiers, such as support vector machine (SVM), Random Forest (RF), decision tree (DT), k-nearest neighbor (k-NN), Naïve Bayes, Random Forest, AdaBoost, and stacking. Supervised learning aims to identify the most effective model to accurately predict the output when provided with new input or data.

A. Phase I: Pre-processing

In machine learning (ML), data pre-processing is crucial for preparing raw data for developing and training ML models. It involves cleaning and organizing the data to ensure its suitability for analysis. Data pre-processing is a data mining technique within machine learning, as it transforms actual data into a clear and structured format. It is essential for reducing the dimensionality of attributes (Basheer, 2019). Raw data quality may vary in terms of reliability, noise, and completeness. Utilizing such data in modeling processes may result in potentially misleading outcomes. The data for this study was extracted from the KMB database, which consists of students' personal information, semester results, and grade points. The data was initially saved as an access file and then loaded into Python as a CSV file. The dataset used in this study is the actual data of KMB students from June 2019 to June 2022, containing information from 703 students.

Subsequently, the data pre-processing in this study was conducted using the Python programming language. It encompassed various steps, including data cleaning, encoding, and data balancing. Addressing errors and inconsistencies is essential once the data has been successfully loaded. Then, encoding converts string attributes, such as intake, course, and math subjects, into numerical representations. This encoding process was accomplished by using the Label Encoder () function. Another critical step in machine learning is balancing data, as imbalanced datasets pose a significant challenge, with classes like the minority class significantly outnumbering the others. This imbalance can lead to models biased towards the majority class, resulting in poor performance of the minority class, which is often the class of greater interest. Therefore, a data balancing technique was employed using random under-sampling (RUS) to address this, underscoring its importance in machine learning. This technique was chosen since it can address imbalanced datasets by reducing the size of the majority class.

B. Phase II: Model Development

The main focus of this study revolves around implementing supervised machine-learning techniques to predict students' final grades. In supervised learning, algorithms are trained using labeled examples, such as a collection of inputs for students' assessment marks for which the desired output is known. The learning algorithm receives a set of attributes and their corresponding correct outputs, and it learns by comparing its predictions with the proper outputs to assess its performance. The hyperparameters are tuned iteratively until optimal performance is achieved. Predictive statistical analysis is a sort of statistical analysis that studies data to determine historical patterns and forecast future outcomes. It is essential to evaluate the accuracy of multiple machine learning models to determine which best predicts students' performance and is most effective at improving it (Abdullah, 2018). Some machine learning methods used by most researchers (Alija et al., 2023; Mduma, 2023; Yağcı, 2022; Ashfaq et al., 2020) who make predictions for student achievement include the Naïve Bayes, logistic regression, decision tree, Random Forest, support vector machine, and the k-nearest neighbor model. In this study, more methods are proposed to obtain the best method. The methods proposed are single classifiers, including decision tree, support vector machine, Naïve Bayes, and k-nearest neighbor, and ensemble methods such as Random Forest, AdaBoost, and stacking.

FINDINGS

A. Dataset and Chi-square test

This study utilised the Kolej MARA Banting (KMB) database, covering 2019 to 2022. The analysis was limited to 703 KMB students enrolled in mathematics subjects. Several statistical analyses were conducted to understand the dataset better. The considered attributes included gender, household income, type of secondary school, and grades in three semesters. The dataset consists of 703 rows and 20 columns. The columns in the dataset include the following:

Table 1: List of attributes for the dataset

Attribute	Description	Data Type
Student ID	Student ID range 001-703	Ordinal
Gender	Student gender (Male, Female)	Nominal
Year Intake	Student year intake (2019, 2020, 2021)	Nominal
Course	Student Course (Social Science, Life Science & Science)	Nominal
Family Income	Family household income (B40, M40, T20)	Ordinal
School Type	Previous School Type in Malaysia (SBP/MRSM/SMKA/SMA/PRIVATE/OTHER)	Nominal
Mathematics Subjects	Math subject taken (MAA HL / MAA SL / MAI HL)	Nominal
Sem 1 Exam Grade	Semester 1 exam grade (range 1-7)	Ordinal
Sem 2 Exam Grade	Semester 2 exam grade (range 1-7)	Ordinal
Sem 3 Exam Grade	Semester 3 exam grade (range 1-7)	Ordinal
Total Point	IB Total Point from a combination of six subjects (range from 24-45)	Ordinal
IB Grade	IB examination grade (range 1-7)	Ordinal

First, data was prepared by checking missing values and duplicate rows from the dataset. The dataset was cleaned and pre-processed before it could be used for further analysis and to train the machine learning model. From Python, it can be seen that all values are present, and no observations are missing. This study employed the IB grade attribute as the target variable. A student's lowest grade on the IB mathematics examination is grade 1. Similarly, an IB grade of 7 indicates the highest score in the mathematics subject. Then, statistical analysis was performed to achieve the first objective of this study. A Chi-squared test was performed to obtain the significant attributes of the categorical attributes. Based on the Chi-square test, the findings reported that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, course $\chi2$ (15, N = 703) = 95.17, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, that intake $\chi2$ (10, N = 703) = 70.99, p = 0.000, and math subject $\chi2$ (10, N = 703) = 177.58, p = 0.000 are significant when $p<0.05$. Based on socio-demographics, intake, course, and mathematics subjects are related to IB results. Meanwhile, gender $\chi2$ (5, N = 703) =10.15, p = 0.07, family income $\chi2$ (10, N = 703) = 0.219, p = 0.511 and school type $\chi2$ (45, N = 703) =55.28, p = 0.648 are not significant since $p>0.05$. It can be concluded that family income and school type are not related to IB grades. The attributes analyzed include student intake year, course of study, gender, previous school type, mathematics subjects taken, semester mathematics grades (for semesters 1, 2, and 3), IB total point score, and the target variable, the IB grade. While a study by Salleh & Othman (2019) found no significant gender differences in the impact of emotional intelligence on mathematics performance, the present study identifies a relationship between gender and students' performance.

B. Model Development

The development of machine learning (ML) models using seven algorithms, including support vector machine (SVM), k-nearest neighbour (k-NN), Naïve Bayes (NB), Random Forest (RF), decision tree (DT), AdaBoost, and stacking, is a crucial stage. The results of this stage are pivotal for achieving the second and third objectives of the study: developing the predictive model and evaluating its performance. Attribute and target variables must be considered before splitting the dataset into training and testing sets. The target variable in this dataset likely represents different categories of student performance, which might not be equally distributed. Imbalanced datasets pose a significant challenge in ML, as models can become biased towards the majority class, leading to poor performance in predicting the minority class, which is often of greater interest. To address this, a data balancing technique was employed. Random Under Sampling (RUS) is one such technique that reduces the size of the majority class by removing instances until the number of cases in each class is approximately equal. This ensures the model receives a balanced representation of all performance levels, leading to more accurate and unbiased predictions. After balancing, the dataset was divided into training and testing sets. The training set, constituting around 85% of the data, serves as the model's learning ground. In comparison, the remaining 15% forms the testing set as a benchmark for evaluating the model's performance on unseen data. The ML models were developed using Python. The process involved several single classifiers—k-NN, DT, NB, and SVM—alongside powerful ensemble techniques such as RF, AdaBoost, and stacking. Hyperparameter tuning was conducted to optimize the models. The best parameters for DT were the Gini criterion, a minimum sample leaf of four, and a minimum sample split of five. For k-NN, the best parameters were *k* equal to 5, and the Manhattan distance. For SVM, the optimal settings were a *C* value of 0.1 and a linear kernel.

DISCUSSIONS

A. Model Comparison

This study explores the strengths and weaknesses of various machine learning (ML) models in predicting student performance, specifically in International Baccalaureate (IB) mathematics subjects. Each ML model, including support vector machine (SVM), Naïve Bayes (NB), decision tree (DT), Random Forest (RF), stacking, and AdaBoost, has distinct characteristics that influence its effectiveness. The study reveals significant discrepancies in their potential efficiency, which hinges on the specific attributes of the dataset.

	Accuracy	Precision	Recall	F1-score
Stacking	0.7143	0.6873	0.7143	0.6980
SVM	0.6762	0.6839	0.6762	0.6730
AdaBoost	0.6381	0.6562	0.6381	0.6457
RF	0.6381	0.6404	0.6381	0.6311
DT	0.5905	0.6213	0.5905	0.6034
NB	0.6000	0.6468	0.6000	0.6018
KNN	0.5905	0.6048	0.5905	0.5934

Figure 2: Model comparison of several ML models with the evaluation metrics values

The evaluation metrics—accuracy, precision, recall, and F1-score—provided valuable insights into the models' strengths and weaknesses, enabling further refinement and optimisation. As illustrated in Figure 2, the accuracy values of the models ranged from 59.05% (DT and k-NN) to 71.43% (stacking). Stacking emerged as the most accurate model, followed by SVM (67.62%), AdaBoost, and

RF (63.81%), with NB achieving 60.00% accuracy. These results suggest that while the models accurately predict student performance, inherent limitations exist. Several factors could explain these limitations. Firstly, the data size might be insufficient to establish meaningful relationships and generalise effectively. The dataset also contains numerous categorical variables, introducing complexity that could hinder the models' predictive accuracy. Additionally, the target attribute (IB grade) is imbalanced, leading to a bias towards the majority class and affecting overall accuracy. This underfitting issue indicates that the models might not capture the complex variations necessary for precise predictions.

Stacking, a robust ensemble learning technique, proved to be the most suitable model for managing imbalanced datasets. By leveraging diverse models to capture distinct data characteristics, stacking mitigates the bias inherent in individual models. This consolidation approach reduces overfitting tendencies and enhances the generalisation ability, making it particularly valuable for imbalanced datasets. Consequently, stacking emerged as the optimal predictive model for assessing student performance in this study. To further evaluate the efficacy of the stacking model, this study analysed its performance in classifying data into distinct groups using the Area Under the Curve (AUC) metric for the Receiver Operating Characteristics (ROC) curve. As shown in Figure 3, the AUC scores for each class within the stacking model ranged from 0.54 (Class 6) to 0.95 (Class 7), indicating its ability to discriminate between various performance levels. Despite the lowest AUC value for Class 6, the stacking model demonstrated satisfactory performance with an average AUC value of 0.81. While specific models showed considerable achievements, their accuracy remained inadequate, prompting a deeper exploration of the underlying factors hindering accurate predictions. The primary issue identified was the inadequate data size, which may have been insufficient for establishing robust relationships and practical generalisation of unseen data. The complexity of categorical data also posed a significant hurdle, as failing to address this complexity can hinder the models' predictive accuracy. Moreover, the skewed distribution of student performance grades created an imbalance, favouring the majority class and leading to inaccurate predictions for the underperforming minority. This imbalance added another layer of complexity, necessitating the need for models that can effectively handle such distributions. While the stacking model emerged as this study's most promising predictive model, the limitations identified highlight the need for further research. Increasing the data size, addressing the complexity of categorical data, and managing the imbalance in the target attribute are crucial steps toward enhancing the predictive accuracy of ML models in educational settings.

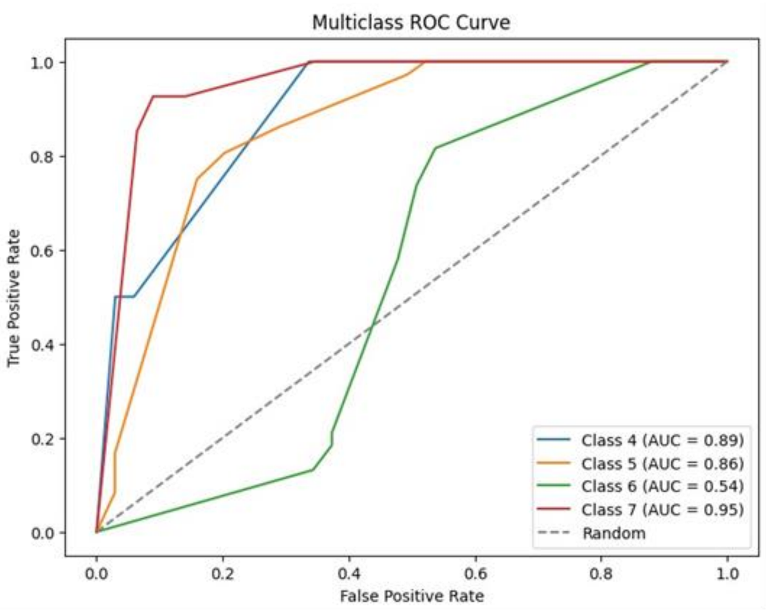

Figure 3: Area under the ROC curve for stacking model testing

B. Model Testing

Conducting data validation is a crucial step in assessing the performance and generalisation capabilities of the predictive model. The initial stage involves verifying that the dataset was prepared correctly and pre-processed using the same methods as the training datasets. For the data validation, the testing dataset of 20 students was carefully examined and analysed for duplicate entries and missing values, but none were found. Consequently, the predictions were generated using the best predictive model, stacking. Table 3 lists the grades of 20 students enrolled in mathematics subjects, both actual and predicted. Additionally, Figure 3 visually compares the actual and predicted grades. The IB grade from the dataset determines the students' actual grade, whereas the expected grade refers to the grade predicted using the best predictive model, stacking. These predictions were based on the students' performance levels in the IB examination results.

Table 3: The list of actual and predicted grades of 20 students using stacking model

ID Number	Actual Value	Predicted Value	Difference
1	7	7	0
2	5	5	0
3	5	5	0
4	5	5	0
5	5	5	0
6	5	5	0
7	6	5	1
8	5	5	0
9	5	5	0
10	4	5	-1
11	6	5	1
12	6	6	0
13	6	7	-1
14	7	7	0
15	5	6	-1
16	5	5	0
17	6	6	0
18	6	5	1
19	5	5	0
20	6	5	1

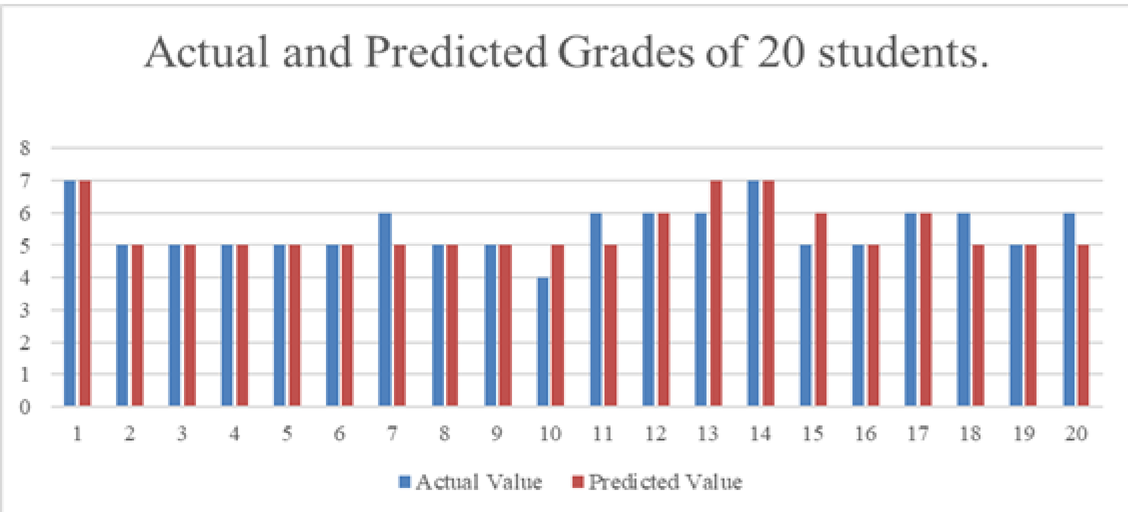

Figure 4: Comparison bar chart of actual and predicted grades of 20 students

Figure 4 presents the grades predicted to be the same as the actual grades using the stacking model. However, a few data points deviate from the actual grades. For example, ID numbers 7, 11, 18, and 20 were predicted to be lower than the actual grades, while ID numbers 10 and 15 were predicted to be one point higher than the actual grade. These discrepancies indicate instances where the predictions do not align with the actual grades. Therefore, it is important to consider this limitation when interpreting the results and the potential implications for any decisions or actions based on these predictions. Every predictive model needs to have advantages for the relevant parties.

RESEARCH SIGNIFICANCE

Predicting student performance in International Baccalaureate (IB) mathematics subjects is a complex task, heavily reliant on data characteristics, sample size, and attribute selection. This study enhances our understanding of the critical factors influencing academic success and offers practical tools for educators. By leveraging the most effective predictive models, educators can better identify students needing additional support, optimise educational strategies, and make informed decisions for university placements. The insights gained from this research are instrumental in shaping an education management system that proactively addresses students' needs, fostering academic improvement in mathematics. Ultimately, these findings empower educators to tailor interventions, personalise learning approaches, and ensure student success throughout their academic journey. The ability to predict student performance accurately can significantly impact how educators support and guide students, making this research valuable for enhancing educational outcomes in IB programs.

CONCLUSION

In conclusion, this study provides significant insights into predicting academic achievement, particularly in mathematics subjects. By identifying critical attributes through the Chi-square test, such as student intake, course, gender, specific semesters' grades, and IB total points, the research aligns with prior studies that emphasise the influence of demographics and previous academic performance on student outcomes. Notably, it reveals that family income and the last type of school do not significantly impact IB grades in mathematics at KMB. The research further explores the efficacy of various machine learning (ML) methods in developing the most effective predictive model. Among the algorithms tested, the stacking classifier stands out with superior performance across evaluation metrics, achieving a notable accuracy of 71.43% and a precision of 68.73%. These findings have practical implications for educators and policymakers, as they can use the predictive model to identify students who may need additional support. Despite its strengths, the study acknowledges the limitations of the stacking classifier, particularly with imbalanced datasets, which might explain its moderately high accuracy. The model's performance was tested with new data from the latest KMB database to ensure robust validation. The analysis confirmed the stacking model's effectiveness, with an average AUC score of 0.81 on the ROC curve. The bar chart comparison between actual and predicted data highlights the model's accuracy in forecasting student performance in mathematics. Future research could explore how the shift to online distance learning (ODL) has affected students' academic performance, particularly in mathematics and IB assessments, and compare the findings based on Muhammat Pazil et al. (2022).

ACKNOWLEDGEMENT

This research was funded by Universiti Malaysia Pahang Al-Sultan Abdullah under Fundamental Research Grant RDU230378.

REFERENCES

Abdullah, N. A. H., Mohamad, M. S., Abdullah, S. S., & Ahmad, N. (2018). Predicting students' academic performance in mathematics using data mining techniques. *Journal of Telecommunication, Electronic and Computer Engineering*, 10(1-8), 123-127.

Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K. (2023). Predicting students performance using supervised machine learning based on imbalanced dataset and wrapper feature selection. Informatica. https://doi.org/10.31449/inf.v47i1.4519

Alloghani, M., Aljaaf, A., Hussain, A. et al. (2020). Correction to: implementation of machine learning algorithms to create diabetic patient re-admission profiles. BMC Med Inform Decis Mak 20, 93. https://doi.org/10.1186/s12911-020-1102-7

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/4151487

Altabrawee, H., Ali, O. A. J., & Ajmi, S. Q. (2019). Predicting students' performance using machine learning techniques. *Journal of University of Babylon for Pure and Applied Sciences*, 27(1), 194–205. https://doi.org/10.29196/jubpas.v27i1.2108

Ashfaq, U., P. M., D. B., & Mafas, R. (2020). managing student performance: a predictive analytics using imbalanced data. International Journal of Recent Technology and Engineering (IJRTE). 8(6), 2277–2283. https://doi.org/10.35940/ijrte.e7008.038620

Basheer, M. Y. I., Mutalib, S., Hamid, N. H. A., Abdul-Rahman, S., & Malik, A. M. A. (2019). Predictive analytics of university student intake using supervised methods. *IAES International Journal of Artificial Intelligence*, *8*(4), 367–374. https://doi.org/10.11591/ijai.v8.i4.pp367-374

Bujang,A., S. D., Selamat, A., & Krejcar, O. (2021). A predictive analytics model for students' grade prediction by supervised machine learning. *IOP Conference Series: Materials Science and Engineering*, 1051(1), 012005. https://doi.org/10.1088/1757-899x/1051/1/012005

Draelos, V. A. P. B. R. (2020). Measuring performance: auc (auroc). *Glass Box*. https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/

Dhilipan, J., Vijayalakshmi, N., Suriya, S., & Christopher, A. (2021). Prediction of students performance using machine learning. *IOP Conference Series: Materials Science and Engineering*, *1055*(1), 012122. https://doi.org/10.1088/1757-899x/1055/1/012122

El-Hajj, M., Yassine, A., Mourad, A., & Hamadé, R. (2020). Predicting academic performance using artificial neural network: The case of Lebanese secondary schools. *International Journal of Emerging Technologies in Learning*, 15(18), 113-128.

Fatima, R., Saleem, M., & Hamid, M. (2019). Predicting academic performance of secondary school students using discriminant analysis. *International Journal of Educational Research and Development*, 8(1), 27-37.

Halim, N. D. A., Ismail, N. A. A., Kamarudin, S., & Mohamad, A. S. (2020). Predicting students' academic performance in english using demographic data, self-efficacy and attitude. *International Journal of Advanced Science and Technology*, 29(6), 3896-3903.

IB, Find countries and universities that recognise the IB. (n.d.). Retrieved (2022) from https://www.ibo.org/university-admission/find-countries-and-universities-that-recognize-the-ib/

Khan, S. A., & Ahmad, R. (2019). Predicting academic performance using decision tree and support vector machine algorithms. *Journal of Education and Learning*, 8(5), 382-391.

Lenin, T., & Chandrasekaran, N. (2021). Learning from imbalanced educational data using ensemble machine learning algorithms. *Webology,* 18(Special Issue 01), 183–195. https://doi.org/10.14704/web/v18si01/web18053

M., R., F., N., & A., A. (2018). Predicting and analysis of students' academic performance using data mining techniques. *International Journal of Computer Applications*, 182(32), 1–6. https://doi.org/10.5120/ijca2018918250

Md Salleh, N., & Othman, I. (2019). Integrating Emotional Intelligence In Curriculum: Do Female Students Perform Better Than Male Students? (70 - 80). *Jurnal Pendidikan Sains Dan Matematik Malaysia*, 4(1), 70-80.

Mduma, N. (2023). Data balancing techniques for predicting student dropout using machine learning. *Data*, 8(3), 49. https://doi.org/10.3390/data8030049

Muhammat Pazil, N. S., Mahmud, N., & Azman, N. A. N. (2022). The Impact of COVID-19 on Academic Performance of Bachelor's Degree Students. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, 12(1), 93-100.

Norhuda, M. S., Zulkifli, N. N., Rosli, N. M., & Jali, M. F. (2021). Predicting academic performance in bahasa melayu subject among secondary school students. *International Journal of Advanced Science and Technology*, 30(5), 1545-1554.

Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. *Indonesian Journal of Electrical Engineering and Computer Science*, 16(3), 1584. https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1). https://doi.org/10.1186/s40561-022-00192-z

Zohair, L. M. A. (2019). Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*. https://doi.org/10.1186/s41239-019-0160-3