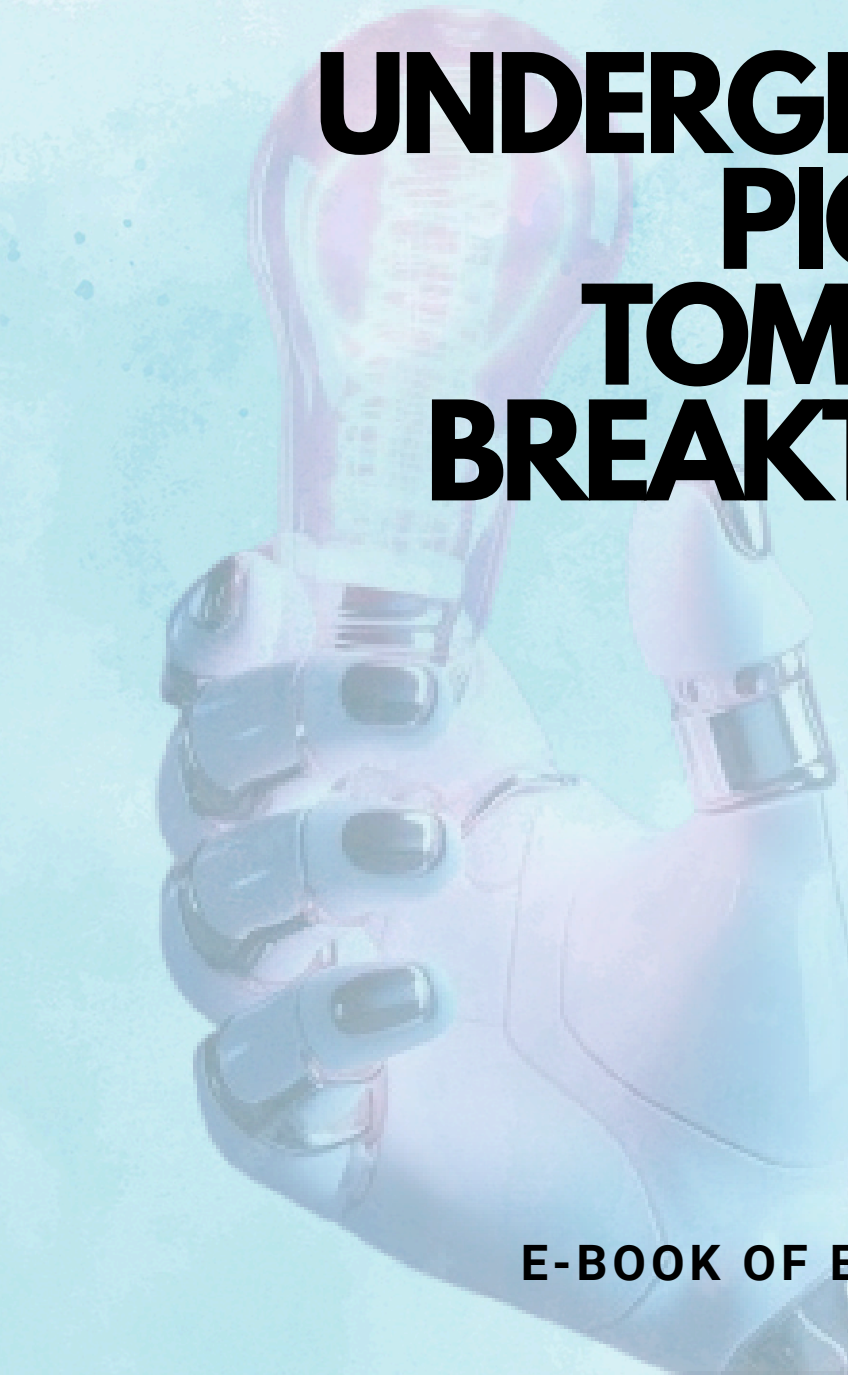
UNIVERSITI TEKNOLOGI MARA

I-URID 2025

INTERNATIONAL UNDERGRADUATE RESEARCH INNOVATION, INVENTION AND DESIGN

UNDERGRADUATES PIONEERING TOMORROW'S BREAKTHROUGH



E-BOOK OF EXTENDED ABSTRACTS

STUDENTS' ACADEMIC PERFORMANCE PREDICTION USING DECISION TREE

Daniel Zameer Shah Mohd Sabri[1], Rozianiwati Yusof[2*], Nooradilla Abu Hasan[3], and Norasma Mohd Zin[4]

[1]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Terengganu, Kampus Kuala Terengganu, Terengganu
[2,3] Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban
[4] Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Kelantan, Kampus Machang.

*rozian696@uitm.edu.my

ABSTRACT

Academic performance is a key indicator of both institutional quality and student development. High dropout rates negatively affect the university's reputation, making early identification of at-risk students crucial. While many researchers have developed prediction models for academic performance, but this remains a challenge due to the large number of influencing factors. This makes it harder to identify students at-risk and give them effective support. This study develops an accurate predictive model for students' academic performance using a Decision Tree Classification algorithm. This study begins with the preliminary studies and then continues with the data acquisition, data preprocessing, prediction model development and the web-based prototype development. The best model achieved an accuracy of 90.14% using 10-fold cross-validation with a precision of 90.73%, recall of 90.14%, and specificity of 97.74%. These results show the model's potential to assist educators and administrators in identifying at-risk students and taking timely action to support their academic success.

Keywords: Prediction, Academic Performance, Decision Tree, Classification

INTRODUCTION

Student academic performance is a crucial indicator of both institutional quality and individual development. The students who perform well will graduate, while those who do not, are at risk of dropping out. High dropout rates are major concern for university administration as they can affect the reputation and resource allocation (Roslan et al., 2024). Depending on the study's nature and the higher education provider, dropout rates might be influenced by a variety of factors, including academic performance, health, family, and personal reasons.

Early academic performance prediction is crucial for students, educators and institution. Students who have been denoted to have a bad grade will receive extra session and additional material from the educator (Ouyang, 2024). Educators can identify students who are at risk to do poorly and implement the required interventions to improve the learning process. Furthermore, understanding the factors affecting performance helps in designing strategies to enhance academic outcomes (Karmagatri et al., 2023). The factors which affect the performance of the students could also be discovered through the use of an accurate prediction model. In this manner, the educational institutions will identify the underperforming students early enough to provide support to them (Sagala et al., 2022).

However, predicting students' performance is complex due to the wide range of influencing variables, including psychological traits, educational background, personal characteristics, and demographics. Although various predictive models have been proposed, many fail to account for all relevant factors or lack interpretability. Therefore, this study aims to develop an accurate and interpretable predictive model for student academic performance using the Decision Tree algorithm. Decision Tree is an interpretable and

transparent machine learning approach for classification problems which representing the decision process as a sequence of simple, intuitive rules. Specifically, each node in a decision tree corresponds to a feature and a decision threshold, and the path from the root to a leaf node represents a series of decisions based on the feature values (Mienye and Jere, 2024). It is well known that decision trees are simple to understand. The intuitive structure of decision tree algorithm is based on the tree-like model of decisions, gives clear and active insights and thus is appropriate for prediction and analysing pattern in educational data.

This study intents to produce a prediction of student academic performance model which involves using both historical and current data to project outcomes in the future. Data mining and machine learning techniques have been employed to identify at-risk students and support timely interventions that enhance overall academic success. The prediction model will also help identify the most influential features contributing to student success or underperformance.

RESULTS AND DISCUSSION

The predictive model for students' academic performance was developed using data from the UC Irvine Machine Learning Repository, containing 4,424 student records with 36 features related to demographics, socio-economic status, academic history, and performance in the first two semesters. Using a Decision Tree Algorithm, the model was evaluated using a few metric measurements before the deployment phase. The following presents and discusses the model's performance and key findings derived from the dataset.

Experiments were performed with the goal of obtaining the best model performance using Decision Tree while comparing a model produced using a full attribute and a subset of attributes using feature reduction. In addition, 10-Fold Cross Validation and 5-Fold Cross Validation were tested to get the best model. The performance of the model was evaluated using accuracy, precision, recall and specificity. The table 1 belows show the result of the experiment.

Table 1: Experiment Results Using Different Approaches

Method	K-Fold Cross Validation	Accuracy	Precision	Recall	Specificity
Full Attributes	5 Folds	0. 8873	0.8898	0.8873	0.9532
	10 Folds	0.8895	0.8926	0.8895	0.9576
After Reduct	5 Folds	0.9008	0.9064	0.9008	0.9755
	10 Folds	0.9014	0.9073	0.9014	0.9774

For the full set attributes, the model performed consistently well across all folds and all metrics. 10-fold cross validation shows slightly better performance as compared with 5 folds cross validation with an accuracy of 0.8895 than 0.8873 for 5 folds which indicates the model more reliable due to better generalization. Meanwhile, the precision and recall were balanced, meaning that the model can accurately identify both dropout and graduate performance cases. The specificity metric consistently above 0.95 indicates strong ability to identify students who are not at risk.

Meanwhile, the experiment after reduction process shows the performance improve across all metrics. The accuracy increases over 90% for both 5 folds and 10 folds showing that the model become more efficient and accurate with fewer attributes. The values for precision and recall also increase which indicate consistency in predicting at-risk students. Specificity values increase significantly for both 5 folds and 10 folds which around 0.9774, showing the stronger capability in identifying not at risk students correctly.

From this observation, it shows that feature reduction approaches improve all the result in all metrics which indicated that removing irrelevant features help the model to focus on the most impactful attributes thus enhancing accuracy and model interpretability. Besides that, the models produce from all k-fold techniques are stable because they are not highly sensitive to the splitting technique. The balance between precision and recall is important because it can avoid from both false positive and false negative cases.

After developing the best model using the Decision Tree algorithm, a web-based prototype was created for end users. The website integrates the prediction model, allowing users such as students, educators, and

administrators to make academic performance predictions. Figure 1 below shows the web prototype interface.

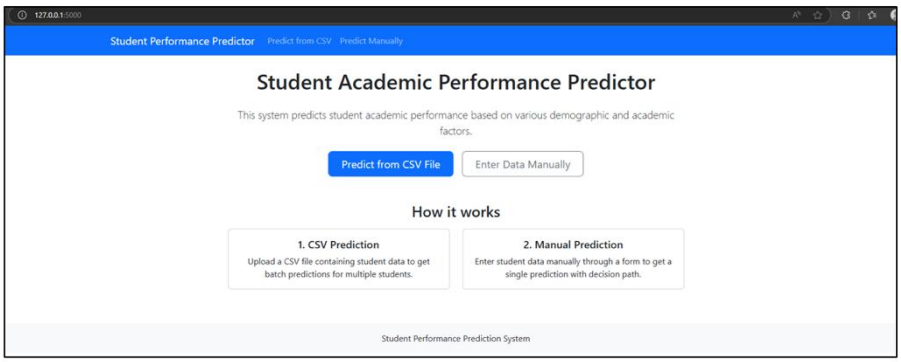


Figure 1: Web Prototype Interface

Users can interact with the Student Academic Performance Predictor system through two simple pathways depending on their specific needs as shown in Figure 1. For educators or administrators who need to analyse multiple students in a cohort, they would click the "Predict from CSV File" button and upload a CSV file with the student demographic and academic information for their class, cohort, or students. For the academic advisors, counsellors, or educators who are gaining insights for a specific student directly in a meeting or advisement session, they can click the "Enter Data Manually" button, which opens a form interface where they enter information sourced from the individual student including their grades, attendance, demographic factors of relevance, and other academic indicators. The single-student result provides not just a single prediction, but a prediction that includes the decision path, helping the user understand the most contributory factors influencing the prediction as shown in Figure 2 belows.

Prediction Result

Predicted Performance: Graduate

Decision Path:

1. "Curricular units 2nd sem (approved)" is greater than 3.00
2. "Tuition fees up to date" is greater than 0.00
3. "Curricular units 2nd sem (approved)" is greater than 4.00
4. "Curricular units 2nd sem (evaluations)" is less than or equal to 9.00
5. "Curricular units 1st sem (credited)" is less than or equal to 8.00

Top Important Features:

Feature	Importance Score
Curricular units 2nd sem (approved)	0.7817
Tuition fees up to date	0.1007
Curricular units 1st sem (enrolled)	0.0491
Curricular units 2nd sem (evaluations)	0.0205
Curricular units 1st sem (approved)	0.0100

Figure 2: Sample Output

CONCLUSION

This study successfully developed a predictive model for students' academic performance using the Decision Tree algorithm. They demonstrated strong predictive capability for academic performance, especially after feature reduction. With an accuracy of 90.14% and high precision, recall, and specificity, it offers a practical and interpretable solution to support educators in monitoring and improving student outcomes in real academic environments and enabling timely interventions.To enhance practical application, a web-based prototype was also developed, integrating the prediction model for use by

students, educators, and administratrors. This system allows users to input relevant student data and receive predictive insights, enabling early intervention and informed decision-making. Overall, the proposed solution not only supports academic monitoring but also contributes to reducing dropout rates and improving student outcomes.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all individuals and Universiti Teknologi MARA who contributed to this study. Special thanks go to the academic staff and students who participated in the data collection process, as well as to the university for providing the necessary resources and support throughout the research. Their cooperation and insights were invaluable to the success of this work.

REFERENCES

Karmagatri, M., Kurnianingrum, D., Suciana, M. R., & Utami, S. A. (2023). Predicting factors related to student performance using decision tree algorithm. *In 2023 5th International Conference on Cybernetics and Intelligent System (ICORIS), Pangkalpinang, Indonesia*, pp. 1-6. https://doi.org/1010.1109/ICORIS60118.2023.10352269

Mienye, I. D., & Jere, N. (2024). A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access*, 12, 86716–86727. https://doi.org/10.1109/ACCESS.2024.3416838

Ouyang, M. (2024). Evaluation and prediction of student's learning status in e-commerce smart classroom based on improved support vector machine algorithm. *In 2024 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, pp. 1-4. https://doi.org/10.1109/ICDSNS62112.2024.10691105

Roslan, N., Mohd Jamil, J., Mohd Shaharanee, I. N., & Sultan Juma Sultan Alawi. (2024). Prediction of student dropout in Malaysian's private higher education institute using data mining application. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 45(2), 168–176. https://doi.org/10.37934/araset.45.2.168176

Sagala, T. M. N., Permai, S. D., Gunawan, A. A. S., Barus, R. O., & Meriko, C. (2022). Predicting computer science student's performance using logistic regression. *In 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia,* pp. 817-821. https://doi.org/10.1109/ISRITI56927.2022.10052968