Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Prediction of Student Decision-making Behaviour based on Machine Learning Algorithms

Hao Luo[1], Nor Azura Husin[2], Teh Noranis Mohd Aris[3], Maslina Zolkepli[4], Mohd Yunus Sharum[5]

[1, 2, 3, 4, 5]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia*

[1]*Faculty of Big Data, University of BaoShan, BaoShan, China*

Abstract: - The study of student decision-making behaviour holds immense importance in the realm of education as it aids educators in comprehending students' learning patterns, preferences, and possible scholastic achievements. Nevertheless, the majority of existing research on behavioural decision-making primarily examines the influence of decision-making on students using questionnaires, disregarding the significance of technological methods in forecasting decisions. This study presents a technique for forecasting student decision-making patterns through the utilisation of machine learning techniques. This work utilises many machine learning techniques, including k-nearest neighbour, decision trees, support vector machines, and linear regression, to develop prediction models. These models are created by analysing students' historical data and attributes. The experimental findings substantiate that the machine learning algorithms put forth have the capability to accurately forecast student decision-making behaviour with a significant level of precision. The prediction outcomes can serve as significant guidance for educators to design customised teaching approaches, thus enhancing students' academic achievements and contentment.

Keywords*: Machine learning algorithms, decision-making behaviour, predictive models.*

1. Introduction

With the continuous changes in society and the development of the educational environment, research on student behaviour has gradually become an important issue of concern. Student decision-making behaviour encompasses a wide range of factors, including learning habits, interests, and potential academic performance (Breşfelean, 2007). Analysis of these factors has a positive impact on understanding students' learning habits and improving students' learning efficiency. Specifically, studying student behaviour can help us better understand their motivations and preferences (Könings et al., 2011). By delving into the reasons why students choose certain behaviours over others, we can gain a better understanding of their needs and expectations, thus better meeting their learning and developmental needs. Secondly, research on student behaviour can reveal various factors that influence their decision-making (Moogan et al., 1999, Halabi & Hands, 2018). Understanding these factors, such as personal values, social norms, peer influence, and family background, can assist educators and policymakers in developing appropriate intervention measures to promote positive behaviour and provide support. Additionally, studying student behaviour can help us identify the challenges and obstacles that students may face. By understanding how students perceive and evaluate the risks associated with different behaviours, we can provide them with the necessary support and guidance to make wise decisions (Wise & Jung, 2019). However, the majority of studies simply takes into account how various behaviors affect students' academic achievement, without considering the influence of students' different behavioural habits on their decision-making.

Previously, researchers have mostly focused on predicting student performance by considering factors such as their grades, personal attributes, and other assessments (Olivier et al., 2019). Various advanced technological methods, such as recommendation algorithms, have been used in different learning platforms to anticipate students' interest in courses and suggest relevant courses for them to learn (Luo et al., 2023). Nevertheless, there is a lack of comprehensive analysis of student behavioural data, and it is equally imperative to understand the fundamental elements associated with students' decision-making behaviour.

In this study, we propose a machine learning-based student behavior prediction method to address these issues. To be more specific, (1) to addresses the limitations of traditional questionnaire-based methods by leveraging the power of machine learning algorithms. By utilizing historical data and student features, various algorithms like k-nearest neighbors, decision trees, support vector machines, and linear regression can be used to build predictive models. These models can then anticipate student choices and inform educators about their learning habits, interests, and potential academic performance. (2) To predict the impact of certain behavioral attributes of students on learning outcomes in advance, we shift the research focus from the impact of behavioral attributes to outcome prediction. Specifically, we go beyond simply studying the impact of student attributes on students, and instead focus on developing tools to predict these decisions in advance. This proactive approach allows for targeted interventions and personalized learning experiences to improve student learning outcomes. The purpose of this study is to investigate how different machine learning techniques, including k-nearest neighbor, decision trees, support vector machines, and linear regression, may be applied to predict students' decision-making behavior. Through the use of these algorithms, teachers can create individualized lesson plans by learning more about the preferences, habits, and possible difficulties of their pupils. Algorithms using machine learning have drawn a lot of attention in the realm of education by evaluating pertinent aspects and historical data. Teachers can improve student results and their teaching tactics by having a deeper understanding of and ability to predict student behavior.

In conclusion, the purpose of this work is to contribute to the growing body of information concerning the application of machine learning techniques to the prediction of the decision-making behaviour of students. By utilising these algorithms, teachers will be able to improve their understanding of their students and their ability to support them, which will ultimately lead to improved learning outcomes and experiences. Our main contributions are summarized as follows:

- We first introduce different machine learning methods into the research of student behavior habit prediction, and predict the results of student behavior habits through technical means.
- To verify the impact of different attributes on student behavioral habits, we can utilize various machine learning algorithms for prediction and compare the predictive accuracy of different algorithms.

The following section of the paper is organised as follows: The second section of the study focuses on the analysis of previous research on behaviour prediction. Section 3 will propose the methodology. Section 4 provides an analysis of the conducted experiments and presents the corresponding findings. The conclusion is located in the final section which is Section 5.

2. Relative Work

2.1 Student Behavior Analysis

Traditionally, understanding student behavior relied on questionnaires, observations, and performance data. While valuable, these methods often lacked depth or relied on subjective interpretations. The rise of educational data mining (EDM) (Dutt et al., 2017) and learning analytics has revolutionized the field by leveraging data from learning platforms and student interactions (Romero & Ventura, 2020). Techniques like clustering and association rule mining uncover hidden patterns, allowing educators to personalize learning, predict student struggles, and provide timely support (Dol & Jawandhiya, 2023). Affective computing is further pushing boundaries by analyzing emotions like boredom or engagement, promising even more targeted interventions and personalized learning pathways (Wu et al., 2016). However, challenges like data privacy and algorithmic bias remain, calling for ethical considerations and careful model development (Alani et al., 2018).

2.2. Machine Learning Approach

Currently, a prominent field of study focuses on forecasting student achievement and involvement in digital educational settings. Researchers (Viberg et al., 2020), (Xu et al., 2019), and (Sekeroglu et al., 2019) have started investigating the application of machine learning algorithms, such as decision trees, support vector machines, and Naïve Bayes classifiers, to analyse patterns in student interaction data. Their aim is to identify the crucial factors that influence academic success and offer timely interventions for students who are at risk (Aissaoui et al., 2019). An instance of a supervised learning algorithm is the decision trees method described by Charbuty (Charbuty & Abdulazeez, 2021). Decision tree classifiers are user-friendly and comprehensible. Decision trees are commonly employed in data mining to examine and predict data. This model is beneficial for pupils' categorization. Song & Liu (2015) conducted a study demonstrating the extensive utilisation of decision tree approaches in classification and prediction. Utilising decision tree models offers numerous advantages in explaining research findings and evaluating various decision tree methods, including CART, C4.5, CHAID, and QUEST.

Furthermore, Osisanwo presented a machine learning model based on learning strategies, motivation, views of social support, socioeconomic variables, health issues, and academic performance aspects (F.Y et al., 2017). Using this method, the study project predicted academic attainment and dropout rates. According to the study, background information was most important in detecting dropouts, but learning tactics were the predictive factor that had the most influence on predicting GPA.

These datasets will be used to build decision trees that predict students' learning behaviors. In 2009, a model proposed by Petre that predicts students' academic achievement using machine learning classification methods like k-nearest neighbour, logistic regression, logistic forests, nearest neighbor, support vector machines, and naïve bayes (Petre, 2009). It obtained a 70–75% accuracy rate. In order to predict students' overall performance, Ahmad (Ahmad et al., 2015) used machine learning algorithms such as decision trees, logistic regression, support vector machines (SVMs), artificial neural networks (ANNs), and Naïve Bayes classifiers. In order to forecast the challenges that each student will encounter while studying, the writers conducted experiments.

3. Research Methodlogy

There are four primary stages in the process of building a model to anticipate how students will make decisions: pre-processing, attribute extraction, the suggested approach, and evaluation. The overview of the research methodology framework is represented as in Figure 1.

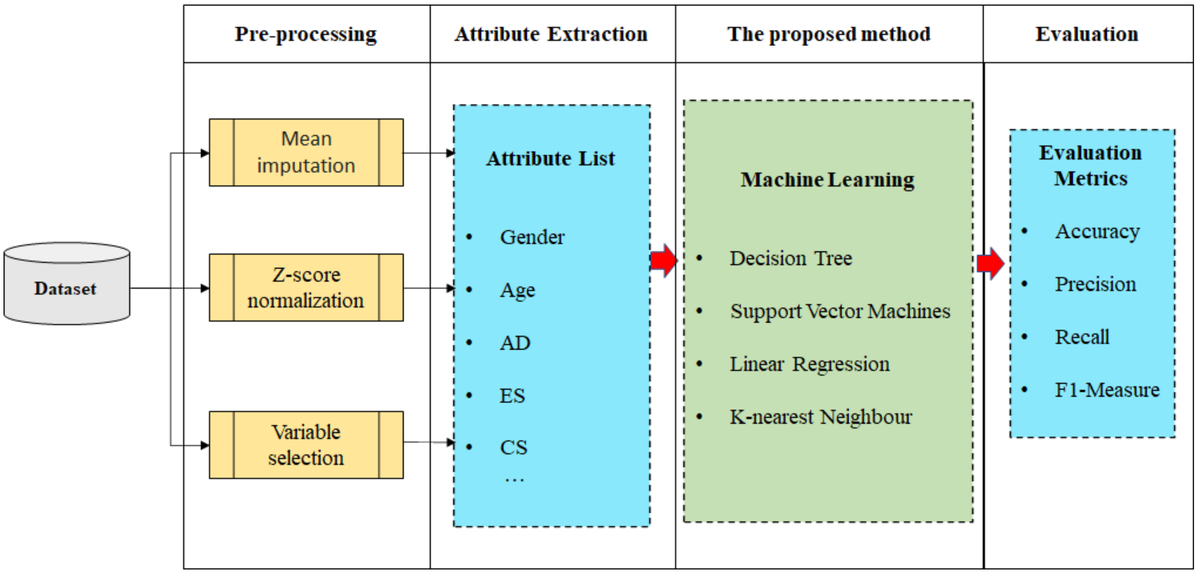

Figure. 1. The research methodology framework.

3.1 Pre-Processing and Attribute Extraction

The procedure of gathering data is the most crucial part of creating the model. Accurate predictions are ensured by a full and thorough data collection. We used a Google form to administer a survey to University Putra Malaysia undergraduate, graduate, and PhD students as part of our data collection phase. Ten factors and 421 samples from kids with various educational backgrounds make up the data. The appropriate student behaviors were carefully selected and constructed in Table 1 in order to provide reliable forecasts.

Table 1: Attribute Description and Possible Values

Variables	Description	Possible values
Gender	Student Gender	Male, Female
Age	Students age	From 17 to 41
AD	Academic degree	Bachelor, Master, PhD
ES	Economic status	Good, Medium
CS	Courses status	Online courses, Course cancellation
SMA	Social media activities	High, Medium
SL	Stress Level	Good, average, bad
IE	Internship Experience	Yes, no
SB	Students' behaviour	Optimistic, Pessimistic, Trustworthy, Untrustworthy
DML (Target Variable)	Decision-Making Level	Easy, average, hard

After that, the data pre-processing was done as it contained null and missing values which were then filled by the most repetitive values in a CSV file dataset. In the pre-processing phase, Mean imputation, Z-score normalization, and variable selection play crucial roles in optimizing the quality and relevance of the dataset for machine learning applications (Kang & Tian, 2018). To address missing values, mean imputation tackles missing data by simply plugging the average value of the existing data into the empty slots. Subsequently, Z-score normalization standardizes numerical features, promoting uniformity and robustness to varying scales (Patro & sahu, 2015). The incorporation of variable selection involves identifying and retaining the most influential variables through techniques such as correlation analysis, feature importance assessment, and dimensionality reduction (Chandrashekar & Sahin, 2014). This sequential approach enhances the dataset's predictive power by addressing missing data, normalizing numerical features, and emphasizing the significance of relevant variables, culminating in a refined dataset conducive to more accurate and efficient machine learning model training.

The resulting pre-processed student dataset is divided into a training and a test dataset using a 70:30 ratio, and saved in CSV format afterwards. Of the total data, 70% is the training dataset and 30% is the test dataset.

3.2. The Proposed Method

Selecting algorithms that are accurate and efficient is necessary in order to forecast the best outcomes. These are the algorithms that we have selected for this use:

A. Decision Tree (DT)

A decision tree is organized in a tree structure, with the uppermost level referred to as the root node. Each internal node signifies a test on a specific attribute, while each branch denotes the output associated with that attribute. Finally, each leaf node conveys information about the class or distribution of classes (K et al., 2010).

Decision trees can deal with classification and regression problems (Loh, 2011). Classification problems mainly refer to the target marked as category data, while regression problems mainly refer to the target marked as continuity value. They use the recursive construction of prediction models to classify datasets and fit simple models together.

Various well-known decision tree algorithms exist, including ID3, C4.5, LMT, and others. The ID3 algorithm, pioneered by J. R. Quinlan in 1986, selects attributes based on the smallest entropy or the largest information gain (Quinlan, 1986). To determine the root node, the algorithm iterates through all unused attributes, dividing the chosen attribute into data subsets. This recursive process persists, with the algorithm continuing to recurse on each subset until it reaches a point where further recursion is not possible.

The information entropy of the current sample set D is defined as the proportion PK of class K samples in D ：

$$\mathrm{Ent}(\mathrm{D}) = -\sum_{k=1}^{|y|} pk\, log_2\, pk \cdot \quad (1)$$

The discrete attribute "a" has V possible values $\{a^1, a^2, \ldots, a^v\}$, and the sample set with a value of a^von attribute "a" is denoted as D^v.

The "information gain" resulting from dividing the sample set D by attribute "a" is expressed as follows (2):

$$Gain(\mathrm{D}, \mathrm{a}) = Ent(D) - \sum_{v=1}^{v} \frac{|D^v|}{|D|} Ent(D^v) \cdot \quad (2)$$

The C4.5 algorithm, developed by J. R. Quinlan in 1992, enhances the accuracy of the ID3 algorithm by introducing the information gain ratio (J.Ross Quinlan, 1993).

Information Gain Ratio is calculated from information gain and entropy. “C4.5, the most popular algorithm, is used in the machine learning methods,” said by Hormann (Hormann, 1964). In the WEKA tool, the C4.5 algorithm is implemented for constructing decision trees under the classifier name J48.

The C4.5 algorithm utilizes the information gain rate, as expressed in (3):

$$Gain\ ratio\ (D, a) = \frac{Gain(D,a)}{IV(a)} \quad (3)$$

The intrinsic value of attribute "a," denoted as IV(a), is represented as shown in (4):

$$IV(\mathrm{a}) = -\sum_{v=1}^{v} \frac{|D^v|}{|D|} log_2 \frac{|D^v|}{|D|} \cdot \quad (4)$$

LMT, which is a combination of logistic regression and decision tree learning, is a classification model. LMT(Landwehr et al., 2003) is an associated supervised training algorithm. Because the leaves include linear regression functions, it uses logistic regression instead of linear regression to build logistic regression models. Reptree (Mohamed et al., 2012) is a rapid decision tree learner that constructs decision trees using information gain as a partition criterion and applies error reduction pruning algorithms for tree pruning.

B. Support Vector Machines (SVM)

A supervised machine learning technique used for regression and classification problems is called Support Vector Machines (SVM) (Kecman, 2005). It involves analysing data to construct a predictive model, enabling the classification of new data points into distinct categories. Within SVM, the algorithm identifies an optimal hyperplane to effectively separate data points belonging to different classes, striving to maximize the margin between them (Brereton & Lloyd, 2010). This hyperplane acts as a decision boundary, partitioning the feature space into regions corresponding to various classes. The objective is to locate the hyperplane that maximizes the distance between itself and the nearest data points of each class, commonly referred to as support vectors.

SVM possesses the capability to address both linearly separable and non-linearly separable data through the application of the kernel trick (Cossalter et al., 2011). By employing a kernel function, the input data is transformed into a higher-dimensional feature space, simplifying the identification of a linear separation.

Commonly used kernel functions encompass linear, polynomial, radial basis function (RBF), and sigmoid, as outlined in the work by Girosi (Girosi et al., 1995).

SVM finds extensive application across diverse domains, including image recognition, text classification, and behavior prediction (Salman & Kecman, 2012). However, it is crucial to emphasize that the efficacy of employing SVM to predict student decision-making behavior hinges on the quality and relevance of the collected data, coupled with the meticulous selection and optimization of the SVM model.

C. Linear Regression (LR)

Linear regression (Hsu, n.d.) is a statistical modeling technique employed to examine the connection between a dependent variable and one or more independent variables. It presupposes a linear relationship among the variables, positing that the dependent variable can be predicted based on the values of the independent variables (Maulud & Abdulazeez, 2020).

In the context of linear regression, the objective is to identify the optimal-fitting line that minimizes the disparity between the anticipated and actual values of the dependent variable (Uyanık & Güler, 2013). This line is expressed through the linear equation $y = mx + b$, where y denotes the dependent variable, x is the independent variable, m represents the slope of the line, and b is the y-intercept (Twomey & Kroll, 2008).

In the process of linear regression, the estimation of slope and y-intercept values is carried out based on the provided data (Green et al., 1994). Commonly employed in this procedure is a technique known as least squares, which aims to minimize the sum of squared differences between the predicted and actual values.

Linear regression finds utility in a multitude of applications, including predicting forthcoming values of the dependent variable, unravelling the relationships between variables, and discerning the influence of independent variables on the dependent variable (Baroni, 2014). Its extensive applicability spans across diverse fields, encompassing economics, finance, social sciences, and machine learning.

D. K-nearest Neighbour (KNN)

The K-nearest Neighbour (KNN) algorithm stands as a fundamental approach in both classification and regression tasks (Cunningham & Delany, 2021). It is based on a simple idea: if the majority of a sample's k nearest neighbours belong to a certain category, then it is likely that the sample belongs to that category (Kataria & Singh, 2013).

In the K-nearest Neighbour algorithm, the first step is to choose a suitable value for k, which determines how many nearest neighbours to consider (Batista & Monard, 2002). Subsequently, to classify a sample, the distances between it and all samples in the training set are computed. Common distance metrics, such as Euclidean distance and Manhattan distance, are often utilized for this purpose (Hidayati & Hermawan, 2021). Following the distance calculations, the k nearest samples to the sample being classified are selected based on their proximity. Finally, by counting the number of samples in each category among the k nearest neighbours, the category of the sample being classified is determined (Lubis et al., 2020).

The advantages of the K-nearest Neighbour algorithm are its simplicity, ease of implementation, and good adaptability to nonlinear data (Sutton, 2012). However, it also has some drawbacks, such as lower efficiency in handling high-dimensional data and large-scale datasets. The K-nearest Neighbour algorithm finds broad applications in pattern recognition, data mining, machine learning, and especially in solving classification problems.

4. Experiments and Results

For conducting experiments, the evaluation of decision-making predictions for student behavior relies on widely used metrics, including Accuracy, Precision, Recall, and F1-Measure (Powers, 2020), which are based on equations (5), (6), (7), and (8). These metrics help assess the effectiveness of the decision-making predictions. The experiments also employ K-fold and percentage split methods (Rodríguez et al., 2010). Accuracy represents the proportion of correctly calculated predictions out of the total predictions. Precision is the ratio of correctly classified cases to the total number of misclassified and correctly classified cases. Recall is the ratio of correctly classified samples to the total number of unclassified instances and correctly classified cases. Additionally, the F1-measure is utilized, combining both recall and precision.

$$\textbf{Accuracy}=\frac{TP+TN}{TP+FN+FP+TN} \tag{5}$$

$$\textbf{Precision}=\frac{TP}{TP+FP} \tag{6}$$

$$\textbf{Recall}=\frac{TP}{TP+FN} \tag{7}$$

$$\textbf{F1-Measure}=2*\frac{Precison*Recall}{Precision+Recall} \tag{8}$$

The accuracy, precision, recall and F1 measure of all the algorithms are summarized in Table 2.

Table 2: Performance Metrics

Model	Accuracy	Precision	Recall	F1-Measure
DT	0.836	0.835	0.843	0.841
SVM	0.808	0.783	0.833	0.831
LR	0.734	0.769	0.817	0.811
KNN	0.817	0.832	0.749	0.825

As shown in Table 2, it is clearly shown that Decision Tree (DT) gave the highest result, 83.6%, and Linear Regression (LR) has the lowest accuracy, 73.4%. Compare the performance metrics with Precision, Recall and F1-Measure, we also find the Decision Tree has the best results.

In this study, C4.5 algorithm has been used to build a decision tree model. C4.5 algorithm improves the disadvantages of the ID3 algorithm. This algorithm selects and splits nodes based on measuring the maximum information gain rate of attributes. During the measurement of the maximum information gain rate, we should know the information entropy and information gain. Gain (D, Gender) =0.007 as the maximum information gain can be obtained by calculation, so we can use gender as the root node. Figure 2 shows the decision tree construction and also shows that the attributer of academic degree is the main impact effect on their decision-making behaviours.

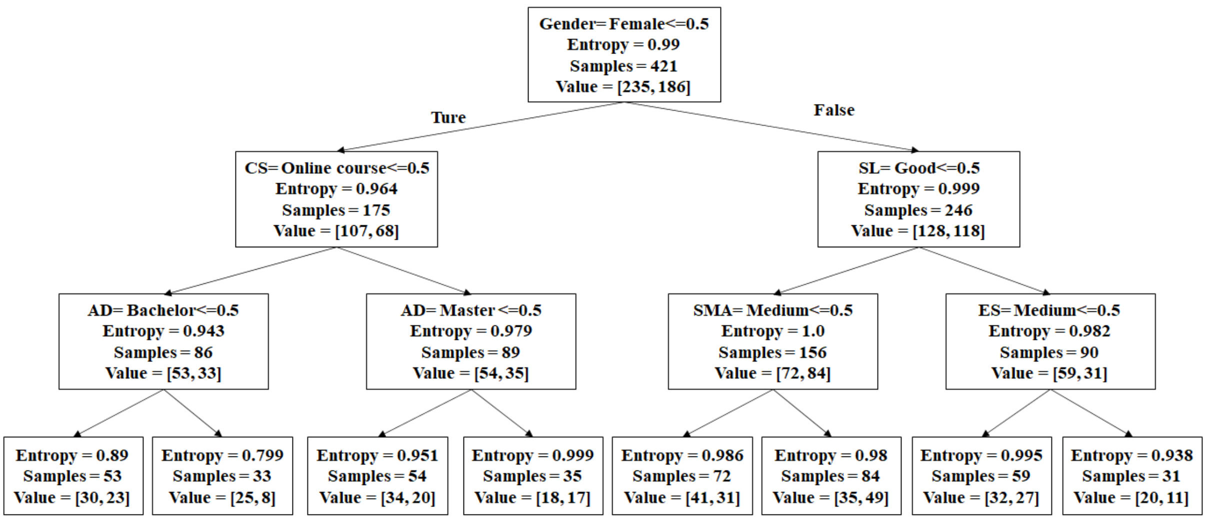

Figure. 2. Decision tree construction.

Because decision trees have the best performance, typical decision tree algorithms such as J48, Reptree, LMT were chosen to find the best DT algorithm to predict the student decision-making behaviour. The tool WEKA(Neville, 1999) was used for data analysis and building of a predictive model. This is because WEKA is a free software, it has high portability by using the Java programming language and running on most of the platforms. The data was created and loaded into WEKA explorer. In this study, we chose 10-fold cross-validation and percentage split methods to evaluate the classification model. The algorithms of J48, Reptree and LMT are used for classification. We have compared these three algorithms and the results are shown in Table 3 by using the cross-validation method and Table 4. by using the percentage split method. The J48 algorithm has the best results in these three different decision tree algorithms by using the 10-fold cross-validation method and the percentage split method. Reptree and LMT have less accuracy in predicting the students' behaviours.

Table 3: Different decision tree algorithms results using 10-fold cross validation

Decision Tree	Accuracy (%)	Time (Sec)	Correctly instances	Incorrectly instances
J48	83.61	0.05	352	69
Reptree	79.81	0.02	336	85
LMT	82.18	0.35	346	75

Table 4: Different decision tree algorithms results using percentage split method

Decision Tree	Accuracy (%)	Time (Sec)	Correctly instances	Incorrectly instances
J48	86.01	0.05	123	20
Reptree	84.63	0.02	121	22
LMT	82.52	0.40	118	25

5. Conclusions

The research delves into students' decision-making behavior and analyzes the impact of extracted behaviors on their decision-making processes. Four classification algorithms DT, SVM, LR, and KNN were comparatively analyzed using the WEKA tool. The study aimed to gain insights into student behavior, categorizing decision-making levels into Easy, Average, and Hard. Extensive experiments on the collected datasets affirmed the superiority of the Decision Tree method in prediction, outperforming various machine learning methods. Among the algorithms tested, J48 demonstrated the highest classification accuracy compared to Reptree and LMT. The academic degree attribute was identified as a significant factor influencing decision-making behaviors. Future work intends to extend the study by comparing deep learning methods with machine learning methods to predict students' decision-making behaviors.

Acknowledgement

This study was supported by Yunnan Province Local Universities Joint Special Youth Project(202101BA070001-270) and Teaching Quality and Teaching Reform Project of Baoshan University in 2022-2023(ZHP202344)

References

[1] Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, *9*(129), 6415–6426. https://doi.org/10.12988/ams.2015.53289

[2] Aissaoui, O. El, El Madani, Y. E. A., Oughdir, L., & Allioui, Y. El. (2019). Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. *Procedia Computer Science*, *148*, 87–96. https://doi.org/10.1016/j.procs.2019.01.012

[3] Alani, M. M., Tawfik, H., Saeed, M., & Anya, O. (2018). Applications of Big Data analytics: Trends, issues, and challenges. *Applications of Big Data Analytics: Trends, Issues, and Challenges*, *July*, 1–214. https://doi.org/10.1007/978-3-319-76472-6

[4] Baroni, M. (2014). *Regression 1 : Linear Regression*. 1–6.

[5] Batista, G. E. A. P. A., & Monard, M. C. (2002). A study of k-nearest neighbour as an imputation method. *Frontiers in Artificial Intelligence and Applications*, *87*(May), 251–260.

[6] Brereton, R. G., & Lloyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

[7] Breşfelean, V. P. (2007). Analysis and predictions on students' behavior using decision trees in weka environment. *Proceedings of the International Conference on Information Technology Interfaces, ITI*, *July 2007*, 51–56. https://doi.org/10.1109/ITI.2007.4283743

[8] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, *40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[9] Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

[10] Cossalter, M., Yan, R., & Zheng, L. (2011). Adaptive Kernel approximation for large-scale non-linear SVM prediction. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, 409–416.

[11] Cunningham, P., & Delany, S. J. (2021). K-Nearest Neighbour Classifiers-A Tutorial. *ACM Computing Surveys*, *54*(6). https://doi.org/10.1145/3459665

[12] Dol, S. M., & Jawandhiya, P. M. (2023). Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey. *Engineering Applications of Artificial Intelligence*, *122*(February), 106071. https://doi.org/10.1016/j.engappai.2023.106071

[13] Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, *5*, 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247

[14] F.Y, O., J.E.T, A., O, A., J. O, H., O, O., & J, A. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128–138. https://doi.org/10.14445/22312803/ijctt-v48p126

[15] Girosi, F., Jones, M., & Poggio, T. (1995). Regularization Theory and Neural Networks Architectures. *Neural Computation*, *7*(2), 219–269. https://doi.org/10.1162/neco.1995.7.2.219

[16] Green, S., Dawson, B. T., Goodman, C., & Carey, M. F. (1994). Y-intercept of the maximal work-duration relationship and anaerobic capacity in cyclists. *European Journal of Applied Physiology and Occupational Physiology*, *69*(6), 550–556. https://doi.org/10.1007/BF00239874

[17] Halabi, K. N. M., & Hands, D. (2018). Identifying and Aligning Product 'Attributes' with Latent Consumer Purchase Motivations. *Asian Journal of Social Sciences and Management Studies*, *5*(1), 16–22. https://doi.org/10.20448/journal.500.2018.51.16.22

[18] Hidayati, N., & Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, *2*(2), 86–91. https://doi.org/10.21831/jeatech.v2i2.42777

[19] Hormann, A. M. (1964). Programs for machine learning. Part II. *Information and Control*, *7*(1), 55–77. https://doi.org/10.1016/S0019-9958(64)90259-1

[20] Hsu, D. (n.d.). *Linear regression Empirical risk minimization*. *2*(Coms 4771), 1–5.

[21] J.Ross Quinlan. (1993). *C4.5:programs for Machine Learning*. Morgan Kaufmann Publishers.

[22] K, M., Pradhan, R., & Sushan, S. (2010). Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix. *International Journal of Advanced Computer Science and Applications*, *1*(5). https://doi.org/10.14569/ijacsa.2010.010516

[23] Kang, M., & Tian, J. (2018). Machine Learning: Data Pre-processing. *Prognostics and Health Management of Electronics*, 111–130. https://doi.org/10.1002/9781119515326.ch5

[24] Kataria, A., & Singh, M. D. (2013). A Review of Data Classification Using K-Nearest Neighbour Algorithm. *International Journal of Emerging Technology and Advanced Engineering*, *3*(6), 354–360.

[25] Kecman, V. (2005). *Support Vector Machines – An Introduction*. *May 2005*, 1–47. https://doi.org/10.1007/10984697_1

[26] Könings, K. D., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2011). The match between students' lesson perceptions and preferences: Relations with student characteristics and the importance of motivation. *Educational Research*, *53*(4), 439–457. https://doi.org/10.1080/00131881.2011.625155

[27] Landwehr, N., Hall, M., & Frank, E. (2003). Logistic model trees. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *2837*, 241–252. https://doi.org/10.1007/978-3-540-39857-8_23

[28] Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

[29] Lubis, A. R., Lubis, M., & Al-Khowarizmi. (2020). Optimization of distance formula in k-nearest neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326–338. https://doi.org/10.11591/eei.v9i1.1464

[30] Luo, H., Husin, N. A., & Aris, T. N. M. (2023). ROME: A Graph Contrastive Multi-View Framework From Hyperbolic Angular Space for MOOCs Recommendation. *IEEE Access*, *11*(November 2022), 9691–9700. https://doi.org/10.1109/ACCESS.2022.3232552

[31] Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, *1*(4), 140–147. https://doi.org/10.38094/jastt1457

[32] Mohamed, W. N. H. W., Salleh, M. N. M., & Omar, A. H. (2012). A comparative study of Reduced Error Pruning method in decision tree algorithms. *Proceedings - 2012 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2012*, 392–397. https://doi.org/10.1109/ICCSCE.2012.6487177

[33] Moogan, Y. J., Baron, S., & Harris, K. (1999). Decision-making behaviour of potential higher education students. *Higher Education Quarterly*, *53*(3), 211–228. https://doi.org/10.1111/1468-2273.00127

[34] Neville, P. G. (1999). Decision Trees for Predictive Modeling. *Stats, August*, 1–24.

[35] Olivier, E., Archambault, I., De Clercq, M., & Galand, B. (2019). Student Self-Efficacy, Classroom Engagement, and Academic Achievement: Comparing Three Theoretical Frameworks. *Journal of Youth and Adolescence*, *48*(2), 326–340. https://doi.org/10.1007/s10964-018-0952-0

[36] Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. *Iarjset*, 20–22. https://doi.org/10.17148/iarjset.2015.2305

[37] Petre, E. G. (2009). A Decision Tree for Weather Prediction. *Seria Matematică-Informatică-Fizică*, *LXI*(1), 77–82.

[38] Powers, D. M. W. (2020). *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. 37–63. http://arxiv.org/abs/2010.16061

[39] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/bf00116251

[40] Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

[41] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–21. https://doi.org/10.1002/widm.1355

[42] Salman, R., & Kecman, V. (2012). Regression as classification. *Conference Proceedings - IEEE SOUTHEASTCON, March 2012*. https://doi.org/10.1109/SECon.2012.6196887

[43] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. *ACM International Conference Proceeding Series, Part F1481*, 7–11. https://doi.org/10.1145/3318396.3318419

[44] Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044

[45] Sutton, O. (2012). Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction. *Introduction to k Nearest Neighbour Classification*, 1–10.

[46] Twomey, P. J., & Kroll, M. H. (2008). How to use linear regression and correlation in quantitative method comparison studies. *International Journal of Clinical Practice*, *62*(4), 529–538. https://doi.org/10.1111/j.1742-1241.2008.01709.x

[47] Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, *106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

[48] Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. *ACM International Conference Proceeding Series*, *March*, 524–533. https://doi.org/10.1145/3375462.3375483

[49] Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, *6*(2), 53–69. https://doi.org/10.18608/jla.2019.62.4

[50] Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology*, *47*(6), 1304–1323. https://doi.org/10.1111/bjet.12324

[51] Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, *98*, 166–173. https://doi.org/10.1016/j.chb.2019.04.015