Al-Ishlah: Jurnal Pendidikan
Vol.17, 2 (June, 2025), pp. 3639-3652
ISSN: 2087-9490 EISSN: 2597-940X, DOI: 10.35445/alishlah.v17i2.6433

Enhancing Mathematics Comprehension: A Decision Tree Analysis Using Orange Data Mining

Ariyanti Jalal[1], Dadan Dasari[2]

[1] Universitas Pendidikan Indonesia, Bandung, Indonesia; ariyantijalal.new@gmail.com
[2] Universitas Pendidikan Indonesia, Bandung, Indonesia; dadan.dasari@upi.edu

ARTICLE INFO

Keywords:

Educational Data Mining;
Mathematics Comprehension;
Student Learning Pattern;
Decision Tree;
Orange Data Mining

Article history:

Received 2024-12-01
Revised 2025-05-04
Accepted 2025-06-29

ABSTRACT

Comprehension skills are essential in mathematics learning, as students' understanding is influenced by various internal and external factors. Recognizing these factors is crucial for educators to design effective teaching strategies. This study aims to classify and predict students' mathematical comprehension based on gender, attitude, learning styles, and self-confidence. A total of 53 eleventh-grade students from SMA Negeri 8 Ternate participated. Primary data were analyzed using data mining techniques—specifically, classification and prediction using the Decision Tree method via Orange Data Mining software. The analysis identified learning style as the most influential factor in students' mathematical comprehension. The Decision Tree's root node represented comprehension data from 31 students, of which 19 students (61.3%) were classified as having understood the material. The internal node revealed two branches: students with an auditory learning style (8 students) showed a 100% understanding rate, whereas students with kinesthetic or visual styles (11 students) demonstrated a 47.8% understanding rate. The model's prediction accuracy based on the four attributes was 65%. Findings highlight the significance of tailoring instruction to students' learning styles. The relationship between visual, auditory, and kinesthetic learning preferences—when considered alongside gender, attitude, and self-confidence—can offer valuable insights into learning patterns. This study provides a practical reference for educators in developing effective and personalized teaching methods. By leveraging insights into learning styles and associated factors, instructional approaches can be optimized for improved mathematical comprehension.

Corresponding Author:
Dadan Dasari
Universitas Pendidikan Indonesia, Bandung, Indonesia dadan.dasari@upi.edu

1. INTRODUCTION

Understanding mathematical concepts is a basic ability that is very important for students to have. Mathematical understanding is a basic competency in mathematics (Herawaty et al., 2019). NCTM states that understanding mathematical concepts is important in the principles of mathematics learning. (Adawiyah et al., 2022). Mathematics learning is said to be successful if students are able to absorb and understand the concepts given well. This means that in learning, the role of the teacher becomes one of the important factors in developing and improving students' understanding abilities. In addition to teachers, there are several factors that also contribute to students' understanding of mathematical concepts including gender, student attitudes, learning styles, student self-confidence, and so on.

Some studies such as those conducted by Dewi and Liny show that gender differences affect the understanding of mathematical concepts (Dewi Rosikhoh, 2024), Wigati and Heni's research found that there was a significant effect of learning style on students' mathematics comprehension ability (Sari & Pujiastuti, 2020), and in Pang and Roger's research, it was found that the influence of students' attitudes and self-confidence on achievement was significant (Irvine, 2020). Students' attitudes towards mathematics are influenced by beliefs about these learning activities that will produce satisfactory learning outcomes (Sajiman et al., 2022). Satisfactory learning outcomes indicate that there is student understanding of mathematical concepts.

This study does not focus on seeing the influence of one of these factors on student understanding but aims to investigate which factors are most influential by making classifications and predictions of student mathematics understanding based on gender, attitude, learning style and student confidence. Therefore, in this study, researchers did not use statistical data processing or hypothesis testing as in previous studies, but used the concept of data mining. Data mining and statistics are two fields that involve data analysis, but they differ in their approach and purpose. Traditional statistical methods, such as regression analysis, usually start with a model based on hypotheses and assumptions about data distribution. However, data mining algorithms connect variables and determine the functional form of the model, which can help researchers determine distributions and find patterns (Collier & Sukumar, 2024).

The concept of data mining is a process of automatically finding useful information in large data stores. Data mining techniques are used to examine large databases as a way to find new and useful patterns (Anggada Maulana, 2018). Data mining is the process of applying methods with the aim of uncovering hidden patterns in large data sets (Koul, 2020). In the concept of data mining, there is a new field that has emerged, namely Educational Data Mining (EDM). EDM is the development of knowledge discovery methods from data derived from educational environments. The goal is to classify and predict students' future learning, study the effects of support, and advance scientific knowledge about learning (Koul, 2020). Classification tries to predict the goal with the highest precision (Charbuty & Abdulazeez, 2021). Shlomo and Sitte explained that classification is a way to put objects into categories that best fit their characteristics. Classification is a two-step process, the first of which builds a classifier by clearly examining a training set containing attributes and associated class labels (Jena & Dehuri, 2020).

Several studies have been conducted to predict student understanding in schools. The first research from Azhari, about predicting high school students' understanding of mathematics. The results of his research show the value of entropy and gain obtained 18 rules for student understanding decisions in mathematics subjects with nine rules with understanding status and 9 rules with non-understanding status. Classification modelling with the C4.5 algorithm on Rapidminer obtained an accuracy of 95.19% (Azhari et al., 2022). The second research study by Sofia and Oktaviawati examined the application of the C4.5 algorithm for understanding vocational students. The results of the study show that in the classification process using the C4.5 Algorithm for student understanding interest in competency subjects at SMK Guna Dharma Nusantara produce learning media indicators that have the highest gain so as to making them important indicators in the student's understanding interest. The application used for testing the classification process is RapidMiner with an accuracy value of 97.22% (Dewi & Oktaviawati, 2022).

The classification method that researchers use is Decision Tree with the help of Orange Data Mining application. The Decision Tree method was chosen because it is the simplest classification algorithm and is easier to use and interpret than other algorithms. Sawant explained that Decision Tree is a well-established tool, and one of the most powerful with a relatively small learning curve to interpret, and is applied regularly in various environments such as image processing, machine learning, data mining and pattern identification (Lee et al., 2022).. According to Sarker, Decision Tree is ranked the easiest to interpret compared to other supervised machine learning algorithms such as Naive Bayes (NB), Logistic regression (LR), Support Vector Machine (SVM) and Random Forest (RF), thus justifying simple math without even requiring statistical knowledge and no complicated formulas (Lee et al., 2022).

The Decision Tree classifier is widely recognized as one of the most prominent and effective methods for data classification. It is commonly used across various domains, including machine learning, image processing, and pattern recognition (Charbuty & Abdulazeez, 2021). A Decision Tree outlines a series of alternatives that aid in problem-solving by identifying relevant influencing factors and guiding the selection of the most suitable option (Ramdan et al., 2023). Among supervised machine learning algorithms, Decision Trees are especially popular due to their simplicity and interpretability (Maçãs et al., 2024). Their ability to represent knowledge in an intuitive and accessible format makes them valuable to experts and general users (Elouedi et al., 2000).

In this study, the Decision Tree method was implemented using the Orange data mining application. Orange is a beginner-friendly and expert-accessible tool that offers a visual, widget-based interface, eliminating the need for complex coding or calculations. It provides a comprehensive range of widgets for tasks such as classification, data preprocessing, visualization, and model evaluation. The findings from this research highlight the importance of understanding student characteristics to enhance their mathematical comprehension. This insight can help educators design more targeted, effective, and efficient teaching strategies.

2. METHODS

2.1 Research Procedure

This research employs a quantitative approach using data mining techniques. Data mining is the process of identifying patterns in large datasets to perform tasks such as classification, estimation, prediction, association, and clustering (Setio et al., 2020). The specific data mining techniques applied in this study are classification and prediction, with the Decision Tree method selected as the primary analytical tool.

The dataset used in this research is considered small due to several constraints encountered during the data collection process, including limited time, researcher fatigue, and scheduling difficulties with teachers and students. These limitations mean the study's findings cannot be fully generalized to broader populations. However, the Decision Tree method was still deemed appropriate, as it does not require assumptions about data distribution—such as normality—making it suitable for small sample sizes.

The Decision Tree algorithm operates by recursively partitioning data into regions based on feature values, enabling different outcomes to be predicted within each region (Talekar, 2020). As described by Sawant, a Decision Tree is a straightforward supervised classification tool that categorizes data records by applying specific rules or conditions (Lee et al., 2022). Prediction is closely related to classification, but unlike classification, prediction estimates values that occur in the future. Many techniques used in classification are also applicable to prediction tasks (Setio et al., 2020).

The data mining process in this study follows a structured sequence: data collection, data cleaning (preprocessing), transformation, data mining, and evaluation. These stages are explained in detail below.

2.1 Data Collection

The data collected in this study are primary data, namely a collection of questionnaire data on attitudes, learning styles, self-confidence, and students' understanding of mathematics from 53 subjects from class XI students of SMA Negeri 8 Ternate City. The attributes in this study are factors that influence students' understanding of mathematics consisting of gender, student attitudes, learning styles, and self-confidence. The description of primary data for the four attributes is stated in the following table.

Table 1. Primary Data

	Attributes								
	Gender		**Attitude**		**Learning Style**			**Self-confidence**	
Category	Male	Female	Medium	High	Visual	Auditory	Kinesthetic	Medium	High
Total	25	28	36	17	10	13	30	21	32

2.2 Data cleaning (preprocessing)

The second stage in this process is data cleaning or data preprocessing. The goal is to remove irrelevant data, missing data, outliers, input errors, data duplication, and inconsistent data. This stage is important because it can have an impact on the accuracy of the data mining analysis and the performance of the data mining system.

2.3 Transformation

In this process, the data will be converted into a scale or format that is suitable for the data mining method used. Classification is a form of data analysis that extracts models that describe important data classes, where models or classifiers are created to predict categorical class labels such as "safe" or "at risk", "yes" or "no" (Ha et al., 2011). Therefore, numerical math comprehension data was converted into categorical form with the label "understand" or "do not understand". Data transformation aims to change the data format to a form that is suitable for the classification method used. In the Decision Tree classification method, the *target* data *for* mathematics understanding in the form of numerical data is converted into categorical data, namely understanding and not understanding, with the provisions of understanding if the student's score is > 60, and not understanding if the student's score is ≤ 60.

2.4. Data Mining

The data mining stage is the determination of what algorithm or application technique will be used to optimize the results in accordance with the specifications of the data mining technique requirements. In this research, the techniques used are classification and prediction with the Decision Tree method. The classification technique can be seen in the tree viewer menu which displays the Decision Tree classification graph. The graph illustrates and describes the tree structure or flowchart of testing each attribute.

2.5 Evaluation

The final stage of this research is to evaluate one or more of the models used to ensure quality and effectiveness before the model is implemented.

2.6 Decision Tree

The Decision Tree is a straightforward supervised learning algorithm used to classify data into predefined categories by applying a set of decision rules (Lee et al., 2022). Structurally, a Decision Tree resembles a flowchart, where each internal (branching) node represents a test on an attribute, each branch denotes the outcome of that test, and each terminal (leaf) node corresponds to a class label. The root node, positioned at the top, initiates the decision-making process (Ha et al., 2011).

According to Berry and Gordon, a Decision Tree can effectively divide a large dataset into smaller subsets by sequentially applying decision rules (Setio et al., 2020). It is categorized under supervised learning and is among the most widely used data mining techniques for building classification models (Ramasamy & Sundar, 2022).

Statistically, the Decision Tree model evaluates outcomes by segmenting the dataset into subgroups based on predictor variables. These subgroups are formed through a series of binary splits, creating a tree-like structure (Battista et al., 2023). A Decision Tree continues to divide the data recursively until each terminal node (or leaf) contains only one class label, making the decision clear. Within the structure, two types of nodes exist: decision nodes, which perform attribute testing, and leaf nodes, which represent the final classification outcomes (Kaul et al., 2022).

2.7 Orange Data Mining

In this research, data mining techniques were implemented using the Decision Tree method through the Orange Data Mining application. Orange is an open-source software platform designed for data analysis and mining using a visual programming approach. It enables users to build data workflows through a graphical interface, making it highly accessible for both beginners and experienced analysts.

The Decision Tree method in Orange combines two well-known algorithms: C4.5 and ID3. The C4.5 algorithm supports both categorical and numerical attributes, while ID3 is limited to categorical data. In the Orange environment, these algorithms are unified under the component simply named Tree.

Orange was selected for this study due to its intuitive, visual-based design process. Its major strength lies in its widget-based interface, which allows users to construct data workflows without writing complex code (Dobesova, 2024). These widgets cover a wide range of functionalities—from basic data visualization and preprocessing to advanced tasks like model evaluation and predictive analytics (Hosseini & Sardo, 2021).

Orange supports both supervised and unsupervised learning techniques. With the "Test and Score" widget, users can evaluate multiple models and assess performance metrics such as accuracy, precision, and computation time for both training and testing phases (Hikmah & Yasa, 2022). The platform is built on Python, enhancing its capabilities in data visualization, subset selection, predictive modeling, and empirical analysis (Verma et al., 2019).

3. FINDINGS AND DISCUSSION

3.1 Data collection and cleaning (preprocessing)

This research was conducted in class X SMA Negeri 8 Kota Ternate with 53 students as research subjects. Data cleaning consists of cleaning outlier data and missing data or what is known as preprocessing. Outlier data cleaning in the Orange Data Mining application can be shown in the figure below.

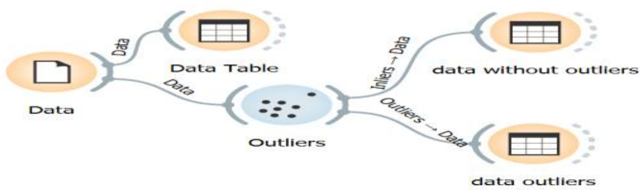


Figure 1. Outlier data cleaning

In the process of checking outlier data, 2 outlier data were found, namely in respondents R39 (understands mathematics) and R49 (not understand mathematics) who have the same attribute criteria, namely female gender, medium attitude, visual learning style and high self-confidence, so that both data will not be included in the data mining implementation process. The next stage checks for missing data with the *preprocess* method as follows. The next stage checks for missing data with the *preprocess* method as follows.

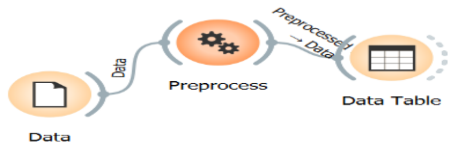
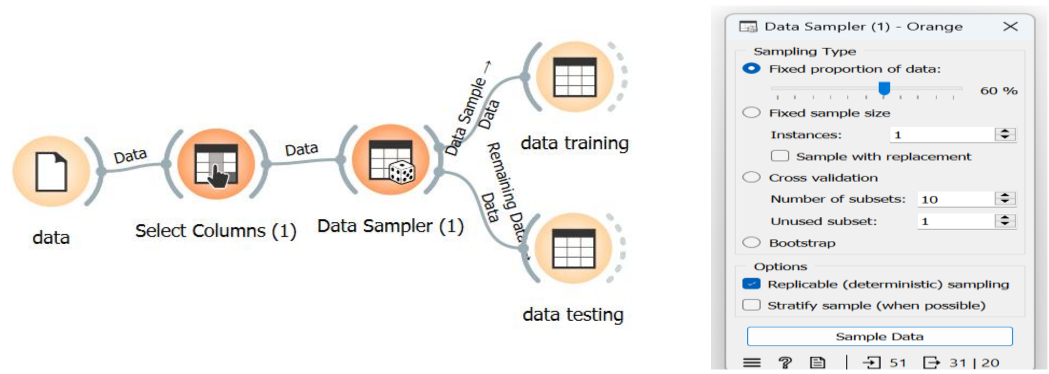


Figure 2. Data Preprocessing

In checking for missing data, no missing data was found so the data *preprocessing* stage produces 51 data points that will be used in the implementation of data mining.

3.2 *Data Transformation*

Transformation or selection of training and testing data is carried out using the following process.



(a) (b)

Figure 3. Data Sampler Stage: (a) Data selection process and (b) Amount of training data and testing data

In the Orange application, the process of selecting training data and testing data uses a data sampler with a ratio of 60: 40. The process produces 31 data points (60%) for training data and 20 data (40%) for testing data.

3.3 *Data Mining Implementation*

The next stage is the implementation of data mining using the Decision Tree classification method using the Orange application as shown below.

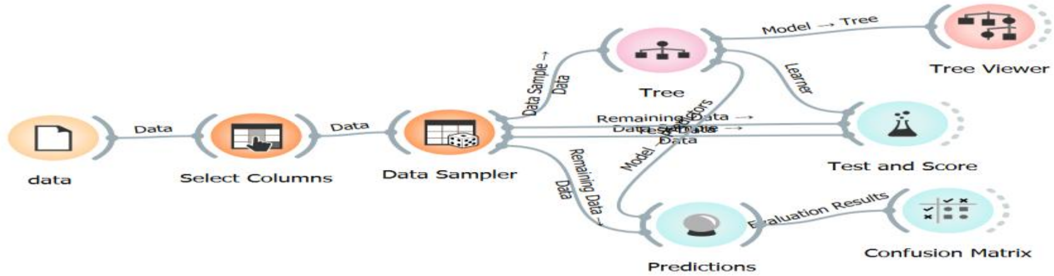


Figure 4. Decision Tree Classification

The results and interpretation of the Decision Tree can be distinguished into two data mining techniques as follows.

3.3.1 Classification

The results of data mining classification can be shown with the following tree structure.

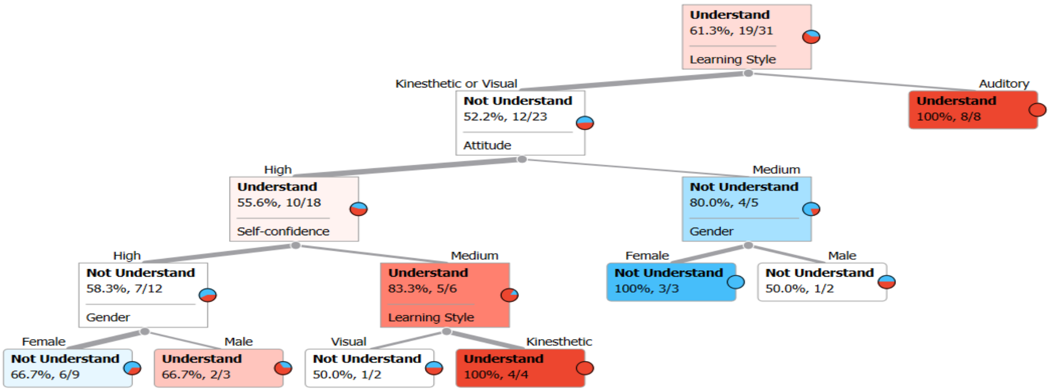


Figure 5. Decision Tree Classification

The classification results above can be interpreted as follows.

1. The topmost node, or what is called the Root Node, displays student understanding data from 31 students for training data as many as 19 students or 61.3% understand math and 12 students or 38.7% do not understand math.
2. Internal Node (branch node) is a learning style that is divided into two branches, namely Auditory learning style with 8 students or 100% understanding math, and kinesthetic or visual learning style with 12 students or 52.2% not understanding and 11 students or 47.8% understanding math.
3. From kinesthetic or visual learning styles, there are branch nodes of attitude attributes. In the high category attitude with kinesthetic or visual learning styles, there are 10 or 55.6% students who understand math and 8 or 44.4% students who do not understand math.
4. For kinesthetic or visual learning styles with medium category attitudes, 1 student (20%) understands math and four students (80%) do not understand math.
5. From the kinesthetic or visual learning learning style with a high category attitude, there is a branch node of the self-confidence attribute. For high category self-belief, 5 students (41.7%) understand and 7 students (58.3%) do not understand math. As for the medium category of self-confidence, 5 students (58.3%) understand and one student (16.7%) does not understand math.
6. For kinesthetic or visual learning learning styles with medium category attitudes, there is a gender branch node. For the male gender, 1 student (50%) understands and 1 student (50%) does not understand math. Meanwhile, for the female gender, 3 students (100%) did not understand math.
7. The last branch node for kinesthetic or visual learning styles with high attitudes and confidence is the gender attribute. For the male gender, 2 students (66.7%) understand and one student (33.3%) does not understand math. While in the female gender, 3 students (33.3%) understand, and 6 students (66.7%) do not understand math.
8. For the last branch node of the high attitude and medium category of self-confidence, there is a learning style attribute. For kinesthetic learning styles, 4 students (100%) understand and for visual learning styles, 1 student (50%) understands and 1 student (50%) does not understand math.

3.3.2 Prediction

The second technique in implementing data mining is prediction. In the Orange application, the prediction results of the testing data can be seen in the prediction or confusion matrix menu. The data can be explained as follows.

Table 2. Confusion Matrix

		Predicted		
		Not Understand	**Understand**	**Sum**
Actual	Not Understand	2	1	3
	Understand	6	11	17
	Sum	8	12	20

The confusion matrix table can be explained as follows.

1. In the true positive (TP) prediction result, there are 11 students who understand and are actually correct; there are also 11 students who understand out of 17 students.
2. For true negatives (TN), the model results predicted 2 students who did not understand and were actually correct.
3. In the False Positive (FP) model, it predicts that there are 6 students who do not understand, but in the reality, 3 of the 6 students actually understand.
4. For False negatives (FN), 1 student was predicted to understand, but the student did not actually understand.

3.4 The Influence of Learning Styles on Students' Understanding of Mathematics

The results of the interpretation above show that learning style features are the most dominant factors in influencing students' understanding of mathematics. Learning style is not only related to what students need to learn, but also how students learn most effectively (Cardino & Ortega-Dela Cruz, 2020). Students who learn according to their learning style will encourage students to receive information and materials provided by the teacher more quickly. The material received by students is more meaningful because students tend to remember the material they learn. For example, kinesthetic students remember material better if it is taught through direct practice or demonstration. Students with a visual learning style will understand faster if the material is presented in the form of pictures, graphs, or videos. Likewise, auditory students are more helped by verbal presentation or discussion.

Learning methods based on students' learning styles can improve students' focus, motivation and concentration so that they can also improve students' understanding. According to Sagitasari, students who learn using their dominant learning style, when doing tests will achieve much higher scores compared to if they learn in a way that is not in line with their learning style (Nasruddin et al., 2023). Students who are capable of identifying and optimizing their learning styles, as well as employing learning strategies that are suitable for their learning styles, are likely to achieve superior mathematics learning outcomes in comparison to those who acquire information or learn to employ strategies that are not conducive to their learning styles (Palobo et al., 2020). In the study, Wigati and Heni found that there was a significant influence of learning styles on students' mathematical understanding abilities. In accordance with the facts that occur because the higher or lower the level of learning style, the more it will affect students' understanding abilities in understanding mathematical problems (Sari & Pujiastuti, 2020). The results of research from Palobo et al also show that there is an influence of learning styles on students' mathematics learning outcomes at Urumb Merauke State Middle School (Palobo et al., 2020).

3.5 Evaluation

Evaluation is carried out to see the level of accuracy of the classification model used. In the Orange application. Evaluate the accuracy of the model on the Test and Score menu. The results are as follows.

Table 3. Test and Score Values

Evaluation Result for Target						
Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.621	0.613	0.618	0.643	0.613	0.239

The table above can be interpreted as follows.

1. The AUC value = 0.621, indicating that the model has a good performance of 0.621 to predict students' understand or not.
2. The CA value = 0.613 indicates the classification accuracy, meaning the model correctly predicted 61.3% of all data.
3. The value of F1 = 0.618, indicating harmonious data (data balance).
4. The value Prec = 0.643, measures the positive prediction or comprehension of students. This means that 64.3% of all comprehension predictions were correct.
5. The Recall value = 0.613, which measures how well the model finds all true positive cases. This indicates 61.3% of math-savvy students.
6. The MCC value = 0.239, which indicates a measure of the strength or weakness of the correlation between the predicted and actual data. This indicates 23.9% poor correlation.

The confusion matrix can also determine the accuracy value using the following formula.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{11 + 2}{11 + 2 + 6 + 1} = 0{,}65$$

The accuracy value of 0.65 indicates that 65% of the model can predict whether or not students understand based on gender attributes, student attitudes, learning styles, and self-belief. The results of this calculation are the same as the accuracy data in the Orange Data Mining application below, which shows a CA value of 0.65. This value indicates poor model accuracy (Adiwandra, 2019).

Table 4. Prediction Accuracy Value

Evaluation Result for Target						
Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.637	0.650	0.699	0.817	0.650	0.229

3.6 Comparison of Model Accuracy Based on Depth Level

The accuracy of the classification model is low because the dataset is too small or the training data selected is too small, which is a case of overfitting. Improving the accuracy performance of the model can be done by reducing overfitting, such as increasing the amount of training data. In addition, if the data is too small, data pruning can be done or the depth of the tree. In the application of Orange data with the Decision Tree method, it can increase the amount of training data by selecting the depth level, namely 70: 30, 80: 20, and 90: 10. The following is a comparison of the accuracy level based on the depth level.

Table 5. Comparison of accuracy based on depth level

Depth Level	60: 40	70: 30	80: 20	90: 10
CA Value	0.613	0.75	0.683	0.674
CA Value (Predicted)	0.65	0.60	0.90	0.80
AUC value	0.621	0.756	0.677	0.691

In a study conducted by Rizki et al. (2023) on the use of Decision Trees to classify students' independent and group learning methods based on their learning resources, model accuracy was evaluated by comparing performance across different tree depths. The highest accuracy was achieved using a data split ratio of 90:10. In that study, the best model performance in terms of classification accuracy (CA) was observed at a 70:30 data split, yielding a CA value of 0.75 and an AUC value of 0.756. Meanwhile, the highest prediction accuracy was attained with an 80:20 split, achieving a CA value of 0.90.

In contrast, the accuracy in the current study is relatively low due to the limited size of the dataset. The small data volume significantly impacts the model's ability to classify effectively or detect meaningful relationship patterns. In data mining, having a sufficiently large dataset is critical to improving model accuracy and overall performance.

Discussion

This study implemented data mining techniques using the Decision Tree method for classification and prediction to analyze the characteristics of students at SMA Negeri 8 Kota Ternate in understanding mathematics, based on variables such as gender, student attitudes, learning styles, and self-confidence. The classification results revealed that the most influential factor in students' mathematical understanding was their learning style. Specifically, the tree-based structure indicated that students with an auditory learning style showed a significantly higher level of understanding compared to those with kinesthetic or visual learning styles. This finding aligns with the actual data, which showed that out of 13 students with an auditory learning style, only one did not understand the mathematics content. These results demonstrate the ability of the Decision Tree method to generate precise and accurate insights in testing attribute-based influences on learning outcomes.

The performance of Decision Tree classifiers is widely recognized in data mining due to their high precision, optimized splitting criteria, and effective pruning techniques (Talekar, 2020). Their relevance continues in contemporary machine learning applications, maintaining competitive performance alongside more recent algorithms (Kaul et al., 2022). This is supported by findings from Devisetty and Kumar (2023), who compared Decision Tree and Support Vector Machine (SVM) classifiers in bradycardia prediction and found the Decision Tree outperformed SVM with a classification accuracy of 92.62% versus 82.35%. Similarly, research by Lestari and Lestari (2024) demonstrated the Decision Tree method's accuracy in identifying key factors influencing student understanding, with the most significant being facilities and infrastructure, followed by learning methods, student interest, teacher strategies, and media use, achieving a classification accuracy of 94.44%.

The prediction results from testing data in this study further supported the classification findings. All students with auditory learning styles demonstrated mathematical understanding. Additional insights obtained through prediction included: "male students with kinesthetic learning styles and high levels of attitude and self-confidence tend to understand mathematics"; "kinesthetic learners with high attitudes are more likely to understand mathematics than those who do not"; and "students with kinesthetic or visual learning styles combined with high attitudes are more likely to achieve understanding." Furthermore, the analysis identified student attitude as the second most influential attribute affecting understanding. The predictions were consistent with actual data, where students classified with high attitudes generally displayed strong mathematical comprehension.

This information is valuable for educators when designing instructional strategies. The classification outcomes provide insights into student learning characteristics, which can serve as a reference for teachers in developing effective and personalized learning experiences. Understanding these characteristics is essential for educators in tailoring instructional methods, models, and materials to student needs (Septianti & Afiani, 2020).

Educational Data Mining (EDM) is a growing discipline that focuses on utilizing large-scale and unique datasets—often generated through digital platforms—to understand student learning progress and the educational environment more deeply (Sarıtaş et al., 2022). EDM supports institutional decision-making and enhances the ability to predict student outcomes. With these insights, educators and institutions can focus not only on what content to teach, but also on how to teach it most effectively (Koul, 2020).

The results of this research offer valuable information about student learning patterns in mathematics, specifically for Grade XI students at SMA Negeri 8 Kota Ternate. These findings can serve as a foundation for teachers to develop more targeted learning strategies, including the selection of appropriate teaching models, methods, media, resources, and instructional techniques. For instance, among students with kinesthetic or visual learning styles, the number of students who did not understand mathematics exceeded those who did. To address this, educators might adopt collaborative teaching approaches, such as project-based learning, where kinesthetic learners engage in hands-on experiments while visual learners contribute through data visualization. Additionally, incorporating engaging, technology-based audio-visual materials and manipulatives may support all learning styles, including auditory learners who benefit from audio-enhanced content.

Although this study primarily involved students with high and moderate levels of attitude and confidence, some students with moderate attitudes and self-belief still struggled to understand mathematics. This indicates that while these attributes do influence learning outcomes, their impact may be less significant compared to learning style. Nonetheless, it remains crucial for educators to consider students' psychological traits when designing lessons. Creating an engaging and supportive learning environment can help foster positive attitudes and confidence, ultimately enhancing students' motivation and interest in mathematics.

4. CONCLUSION

The study concludes that the application of data mining using the Decision Tree method can provide useful insights into classifying and predicting factors that influence students' understanding of mathematics, particularly based on gender, attitudes, learning styles, and self-confidence. The model achieved an accuracy rate of 65%, indicating limited effectiveness in classification. While the confusion matrix demonstrated reasonable True Positive (TP) predictions, the presence of false positives suggests areas for improvement in model performance. A key limitation of this research is the small sample size, which significantly reduced the model's ability to generate strong classification patterns. Additionally, the absence of cross-validation testing restricted the evaluation of the model's generalizability to new data. The lack of data on students with low levels of attitude and self-confidence also limited the scope of analysis. These factors collectively contribute to the modest predictive accuracy. Future research should involve larger datasets with more balanced attribute distributions to enhance model reliability. Incorporating a wider range of variables—such as student interest, perceptions, teaching methods, and learning environments—would also provide deeper insights. Implementing cross-validation techniques is strongly recommended to improve the robustness and accuracy of predictive models. These advancements could significantly aid educators and policymakers in designing more effective instructional strategies and education policies.

Acknowledgements: Thank you to the Mathematics Teachers at SMA Negeri 8 Ternate City who have assisted researchers in collecting research data. Thank you also to the Mathematics Education Doctoral Study Program at the University of Mathematics Education for facilitating the author in conducting research both in terms of exploratory studies, monitoring and evaluation, and reporting research results.

Conflicts of Interest: The authors declare that they have no conflict of interest.

REFERENCES

Adawiyah, R., Meiliasari, M., & Aziz, T. A. (2022). The role of prior mathematical knowledge and interest in mathematics on mathematical concept understanding ability in senior high school students. *Journal of Innovative Mathematics Learning*, *5*(4), 196–204. https://doi.org/10.22460/jiml.v5i4.15397

Adiwandra, T. Y. (2019). Perbandingan kinerja model klasifikasi Decision Tree, Bayesian Classifier, Instance Base, Linear Function Base, Rule Base pada 4 dataset berbeda. *SATIN: Sains dan Teknologi Informasi*, *5*(1). http://jurnal.stmik-amik-riau.ac.id

Anggada Maulana. (2018). *Konsep dasar data mining* (Vol. 1, pp. 1–16).

Azhari, D. Z., Damanik, I. S., & Suhendro, D. (2022). Penerapan algoritma C4.5 untuk klasifikasi tingkat pemahaman siswa pada mata pelajaran matematika. *1*(1), 11–20.

Battista, K., Diao, L., Patte, K. A., Dubin, J. A., & Leatherdale, S. T. (2023). Examining the use of decision trees in population health surveillance research: An application to youth mental health survey data in the COMPASS study. *Health Promotion and Chronic Disease Prevention in Canada*, *43*(2), 73–86. https://doi.org/10.24095/hpcdp.43.2.03

Cardino, J. M., & Ortega-Dela Cruz, R. A. (2020). Understanding of learning styles and teaching strategies towards improving the teaching and learning of mathematics. *LUMAT*, *8*(1), 19–43. https://doi.org/10.31129/LUMAT.8.1.1348

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(1), 20–28. https://doi.org/10.38094/jastt20165

Collier, Z. K., & Sukumar, J. (2024). Discovering educational data mining: An introduction. *NSF Award Abstract*. https://www.nsf.gov/awardsearch/showAward?AWD

Devisetty, G., & Kumar, N. S. (2023). Prediction of bradycardia using decision tree algorithm and comparing the accuracy with support vector machine. *E3S Web of Conferences*, *399*, 09004. https://doi.org/10.1051/e3sconf/202339909004

Dewi Rosikhoh, L. M. (2024). *PRAGAMATIK: Jurnal Ilmiah Ilmu Pendidikan*, *1*(1), 42–49. (Note: Missing article title and DOI)

Dewi, S., & Oktaviawati, O. (2022). Penerapan algoritma C4.5 untuk pemahaman siswa SMK pada pelajaran kompetensi keahlian. *INTERNAL (Information System Journal)*, *5*(2), 116–125. https://doi.org/10.32627/internal.v5i2.626

Dobesova, Z. (2024). Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics. *Computer Applications in Engineering Education*, *32*(4), 1–18. https://doi.org/10.1002/cae.22735

Elouedi, Z., Mellouli, K., & Smets, P. (2000). Decision trees using the belief function theory. In *Information Processing and Management of Uncertainty*, *1*, 141–148.

Ha, J., Kambe, M., & Pe, J. (2011). *Data mining: Concepts and techniques*. https://doi.org/10.1016/C2009-0-61819-5

Herawaty, D., Sarwoedi, S., Marinka, D. O., Febriani, P., & Wirne, I. N. (2019). Improving student's understanding of mathematics through ethnomathematics. *Journal of Physics: Conference Series*, *1318*(1), 012080. https://doi.org/10.1088/1742-6596/1318/1/012080

Hikmah, I. R., & Yasa, R. N. (2022). Perbandingan hasil prediksi diagnosis pada Indian Liver Patient Dataset (ILPD) dengan teknik supervised learning menggunakan software Orange. *Jurnal Telematika*, *16*(2), 69–76. https://doi.org/10.61769/telematika.v16i2.402

Hosseini, S., & Sardo, S. R. (2021). Data mining tools: A case study for network intrusion detection. *Multimedia Tools and Applications*, *80*(4), 4999–5019. https://doi.org/10.1007/s11042-020-09916-0

Irvine, J. (2020). Positively influencing student engagement and attitude in mathematics through an instructional intervention using reform mathematics principles. *Journal of Education and Learning*, *9*(2), 48. https://doi.org/10.5539/jel.v9n2p48

Jena, M., & Dehuri, S. (2020). Decision tree for classification and regression: A state-of-the-art review. *Informatica (Slovenia)*, *44*(4), 405–420. https://doi.org/10.31449/INF.V44I4.3023

Kaul, S., Fayaz, S. A., Zaman, M., & Butt, M. A. (2022). Is decision tree obsolete in its original form? A burning debate. *Revue d'Intelligence Artificielle*, *36*(1), 105–113. https://doi.org/10.18280/ria.360112

Koul, R. (2020). Overview of data mining. *International Journal of Trend in Scientific Research and Development (IJTSRD)*. www.ijtsrd.com

Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: Decision tree. *Advances in Decision Sciences*, *26*(1), 1–29. https://doi.org/10.47654/V26Y2022I1P1-30

Lestari, D., & Lestari, S. (2024). Penerapan data mining klasifikasi tingkat pemahaman siswa pada kegiatan belajar mengajar dengan metode decision tree (Studi kasus SDN Malaka Jaya 11 Duren Sawit). *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, *5*(2), 1260–1268. https://doi.org/10.35870/jimik.v5i2.662

Maçãs, C., Campos, J. R., Lourenço, N., & Machado, P. (2024). Visualisation of random forest classification. *Information Visualization*, *23*(4), 312–327. https://doi.org/10.1177/14738716241260745

Nasruddin, Jahring, Rendy Prakasa, & Lutfiana Tarida. (2023). Pengaruh kepercayaan diri dan gaya belajar terhadap hasil belajar matematika siswa. *JKP: Jurnal Khasanah Pendidikan*, *1*(1), 51–56. https://doi.org/10.58738/jkp.v1i1.76

Palobo, M., Pagiling, S. L., & Nur'aini, K. D. (2020). Analysis of effect of learning style on mathematics learning outcomes. *Proceedings of the International Conference on Social Sciences*, *473*, 484–488. https://doi.org/10.2991/assehr.k.201014.106

Ramasamy, U., & Sundar, S. (2022). An illustration of rheumatoid arthritis disease using decision tree algorithm. *Informatica (Slovenia)*, *46*(1), 109–119. https://doi.org/10.31449/inf.v46i1.3269

Ramdan, D., Valentino, F., & [Author missing]. (2023). Penerapan data mining terhadap prediksi mahasiswa drop out pada kampus STMIK Widuri Jakarta dengan metode decision tree C4.5. *Jurnal Bidang Pendidikan dan Informatika*, 141–150. https://ejournal.kreatifcemerlang.id/index.php/jbpi/article/view/113

Rizki, Mumtaza, N. A., Dewi, M. F., Syahlafandi, C. F., & Fendiyanto, D. (2023). Implementasi pohon keputusan untuk klasifikasi cara belajar mahasiswa mandiri dan berkelompok berdasarkan sumber belajarnya. *Prosiding Seminar Nasional Pendidikan Matematika Universitas Mulawarman*, *3*, 101–106. https://jurnal.fkip.unmul.ac.id/index.php/psnpm

Sari, W. I., & Pujiastuti, H. (2020). The influence of students' learning style towards students' mathematics knowledge. *Matematika dan Pembelajaran*, *8*(1), 36–46. https://doi.org/10.33477/mp.v8i1.1274

Sarıtaş, M. T., Börekci, C., & Demirel, S. (2022). Quality assurance in distance education through data mining. *International Journal of Technology in Education and Science*, *6*(3), 443–457. https://doi.org/10.46328/ijtes.396

Septianti, N., & Afiani, R. (2020). Pentingnya memahami karakteristik siswa sekolah dasar di SDN Cikokol 2. *As-Sabiqun*, *2*(1), 7–17. https://doi.org/10.36088/assabiqun.v2i1.611

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi dengan pohon keputusan berbasis algoritme C4.5. *PRISMA: Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Talekar, B. (2020). A detailed review on decision tree and random forest. *Bioscience Biotechnology Research Communications*, *13*(14), 245–248. https://doi.org/10.21786/bbrc/13.14/57

Uki Sajiman, S., & Hasbullah, H. (2022). The effect of students' attitude in mathematics lessons on understanding the concept of mathematics for high school students. *International Journal of Advanced Research*, *10*(2), 582–585. https://doi.org/10.21474/ijar01/14242

Verma, K., Bhardwaj, S., Arya, R., Ul Islam, M. S., Bhushan, M., Kumar, A., & Samant, P. (2019). Latest tools for data mining and machine learning. *International Journal of Innovative Technology and*

Exploring Engineering, *8*(9 Suppl.), 18–23. https://doi.org/10.35940/ijitee.I1003.0789S19