Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Application of Decison Tree Algorithm to Improve Student Learning Pattern Classification Model at Sempoa Sip Perjuangan

Aulia Agustina Sri Maharani[1], Nana Suarna[2], Irfan Ali[3], Dodi Solihudin[4]

[1,2,3,4]STMIK IKMI Cirebon

auliaagustina2618@gmail.com[1] , st_nana@yahoo.com[2] , irfanaali0.0@gmail.com[3] , dodisolihudin@gmail.com[4]

Abstract

This study aims to analyze the relationship patterns between attributes in the dataset using the *Decision Tree* model on the *RapidMiner* Studio platform. This model is used to identify various factors that affect the performance of participants in solving visual and auditory-based questions. The research stages include data preprocessing, model building, and performance evaluation based on accuracy, precision, and recall metrics. The results show that the Decision Tree is able to intuitively divide the data with attributes such as gender and test-taking method as the main factors. The final node of the decision tree provides a prediction of the number of questions that can be solved correctly. Model evaluation showed good accuracy, although there were indications of overfitting that required pruning. This research supports previous literature that highlights the influence of individual characteristics and learning methods on participant performance. The results can be used to design more personalized and effective learning strategies. Further studies are recommended to use larger datasets and other machine learning algorithms for comparison.

Keywords: *RapidMiner, Decision tree, Study Pattern*

1. Introduction

The development of informatics has brought significant impact in education, especially through the application of data-based technology and algorithms to improve learning effectiveness. The main challenge faced is the heterogeneity of student abilities, so it is necessary to classify learning patterns to ensure targeted learning strategies. Decision Tree algorithm, as a supervised learning method, can be used to classify student learning patterns based on data such as academic grades, attendance, and assignment time. At Sempoa SIP Perjuangan, this algorithm is applied to improve the quality of learning with a more personalized and adaptive approach. Despite its great potential, the utilization of machine learning algorithms in education is still limited due to lack of understanding and complexity of implementation. This research aims to fill the gap by developing a classification model of student learning patterns based on structured data, which is expected to help educational institutions manage learning more effectively, provide appropriate interventions, and optimally utilize data mining technology. In addition to enriching the literature on machine learning in non-formal education, this research also has a practical impact in the development of data-based educational technology to support digital transformation in the education sector.

Research according to [1] discusses the prediction of postpartum hemorrhage using the Decision Tree (DT) and Random Forest (RF) algorithms. This research is entitled "Implementation of Decision Tree and Random Forest Algorithms in Postpartum Hemorrhage Prediction" with a focus on the problem of high maternal mortality due to postpartum hemorrhage and the need for accurate prediction tools. With Random Forest and Decision Tree methods evaluated using accuracy, AUC, and other metrics, the results showed that the Random Forest algorithm performed better than Decision Tree, with an accuracy of 0.830 and AUC of 0.74.
Research according to [2] on the topic of predicting Type 2 Diabetes Mellitus using the Decision Tree algorithm. This research is entitled "Implementation of Decision Tree Method for Diabetes Mellitus Type 2 Prediction" which aims to overcome the high number of cases of Type 2 Diabetes Mellitus through effective early detection. The method used is Decision Tree with certain parameters, validated using K-Fold Cross Validation. The results showed that the model accuracy reached 92%, with precision, recall, and f1-score values of 0.92, 0.915, and 0.915, respectively.

Furthermore, research according to [3] examines the level of student satisfaction with online learning media using the C4.5 algorithm. This research is entitled "Classification of the C4.5 Algorithm in Applying Student Satisfaction Levels to Online Learning Media" and discusses

the low student interest in online learning media. With the C4.5 algorithm-based classification method and questionnaire data, this study shows that the model used is able to provide a fairly high level of accuracy to help evaluate the effectiveness of learning media.
According to [4], it discusses the classification of family welfare levels in Tiga Dolok Village using the C4.5 algorithm. This research is entitled "Application of the C4.5 Decision Tree Algorithm for the Classification of Family Welfare Levels in Tiga Dolok Village" and raises issues related to the distribution of government subsidies that are often not on target. With survey data analyzed using RapidMiner, the results show that the resulting decision tree is able to classify welfare levels with high accuracy, so that it can support more precise subsidy distribution.

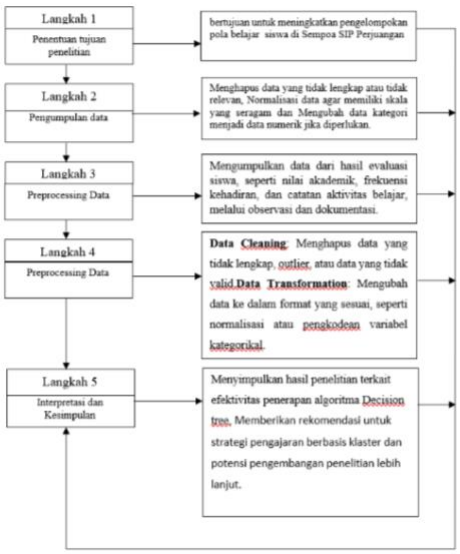
Research from [5] entitled "Detection of Learning Styles of High School Students in Virtual Based Learning Environment (VBLE) with Decision Tree C4.5 and Naive Bayes" discusses the identification of learning styles of high school students in virtual-based learning environments. The algorithms used are Decision Tree C4.5 and Naive Bayes. The problem faced is the difficulty of detecting students' learning styles in a complex virtual learning environment. Data is collected through questionnaires, analyzed using Python, and the performance is measured with parameters such as accuracy, precision, and others. The results show that Naive Bayes algorithm has 98% accuracy, slightly better than Decision Tree with 96% accuracy.

Previous research from [6] entitled "Implementation of Decision Tree Algorithm for Classification of Best-Selling Products", analyzes sales data to determine the best-selling products using the C4.5 Decision Tree algorithm. This research was conducted because trading companies have difficulty determining best-selling products manually. Stock and sales data are processed using Decision Tree C4.5, which provides an accuracy of 90% with an AUC value of 0.709, making it an effective method for product classification.
Furthermore, research according to [7] entitled "Analysis of Responses to PSBB in Indonesia with Decision Tree Algorithm on Twitter" uses the Decision Tree algorithm to analyze public sentiment related to PSBB in Indonesia based on Twitter data. The main problem is the difficulty of manually categorizing public opinion due to the large amount of data. This research resulted in an accuracy of 84.78%, with a precision of 84.78% and a recall of 100%, showing the effectiveness of Decision Tree in sentiment analysis.
According to [8] a study entitled "Sentiment Analysis of Public Opinion about Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decision Tree" compares the Naïve Bayes algorithm and Decision Tree in processing Facebook comment data about COVID-19 vaccination. The problem faced is the inability of the system to automatically identify positive or negative opinions. The results showed that Naïve Bayes excelled with 100% accuracy, while Decision Tree only reached 50.39%, making it a better choice for this sentiment analysis.

2. Research Method

The research method used is Knowledge Discovery in Databases (KDD), with systematic stages to improve the classification model of student learning patterns at SEMPOA SIP Perjuangan. These stages include Selection, to select relevant data such as test results and teacher evaluations; Preprocessing, to clean data from duplicates and inconsistencies; Transformation, to convert data into a suitable analysis format, such as normalization; Data Mining, where the Decision Tree algorithm is applied to identify and classify student learning patterns; and Interpretation/Evaluation, to assess accuracy and interpret model results. In the Modeling stage, the algorithm is optimized with a training and test data split of 80:20, and validation using k-fold cross-validation. Evaluation is done using metrics such as accuracy, precision, recall, and F1-score. Finally, in the Deployment stage, the model is applied in the learning system to provide recommendations for adaptive learning strategies, such as material or schedule adjustments, to improve learning effectiveness in the institution. . The following is an example of table 1 regarding the stages of the research method:

Table 1: Research Workflow



4. Result and Discussion

In this section, it is explained that the dataset used in this research is data that is relevant to the research objective of analyzing students' learning patterns at SEMPOA SIP Perjuangan. The data is collected through various methods, such as surveys or experiments, and contains important information related to students' learning preferences, for example whether they prefer the "Listen Count" (Auditory) or "See Count" (Visual) method.

4.1. Dataset

The dataset contains columns that indicate students' learning preferences, personal information, and responses to several questions related to the learning methods used in SEMPOA. For example, the "Name" column contains the student's identity, while the "Level" column describes their learning level (e.g. Foundation or Intermediate). Other columns provide information such as students' age, gender, as well as their willingness to participate in the study.

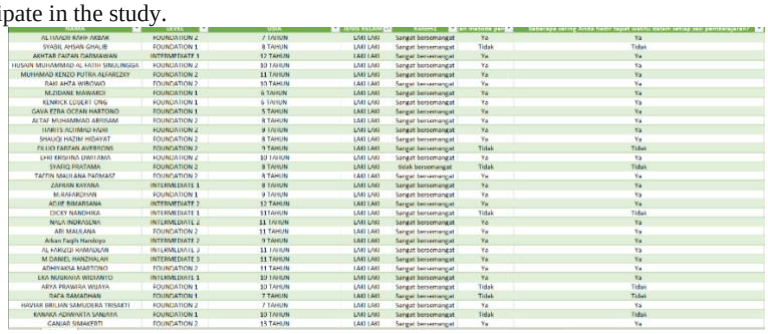


Fig. 1 Dataset

4.2. Data Selection

In the data selection stage, data relevant to the research is selected based on the purpose of the analysis. This process ensures that only data that can support the formation of classification models that match student learning patterns are used. This data selection is based on attributes that are considered important in determining students' preferences for learning methods

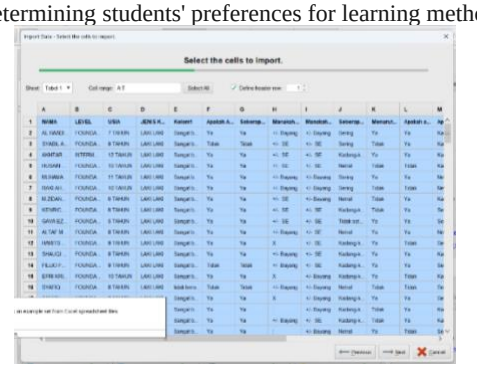


Fig. 2 Data Selection

4.3. Data Pre-Processing

In the pre-processing stage, initial processing is carried out to prepare the data before it is analyzed using data mining methods. Pre-processing includes removing incomplete data, removing duplicate data, and handling missing values. Table 4.3 illustrates the data before is pre-processed. The columns contained in the dataset will be prepared so that they are ready to be used in making classification models.

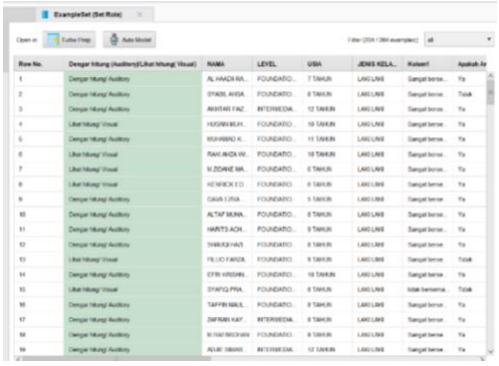


Fig. 3 Data Pre-Processing

4.4. Data Transformation

The processed data is then converted into a format that is more easily understood by the model, for example by separating the target attributes and input attributes. In this research, the classification method applied uses Decision Tree to predict students' learning method preferences, either "Listen Count" (Auditory) or "See Count" (Visual). The columns containing the original data are combined with the model predictions as well as the confidence value of the prediction results, which will be used to evaluate the accuracy of the model.

Fig. 4 Data Transformation

4.5. Data Mining

Data mining is performed using the Knowledge Discovery in Databases (KDD) method, which provides a systematic framework for data analysis. RapidMiner was used to apply the Decision Tree algorithm, which involves several steps: import data, set roles for attributes, split data into training and test data, and apply the Decision Tree model. Each of these steps plays an important role in ensuring the data is processed correctly to produce an optimal model. Once the model is built, evaluation is performed using evaluation metrics such as accuracy, precision, recall, and F1-score.

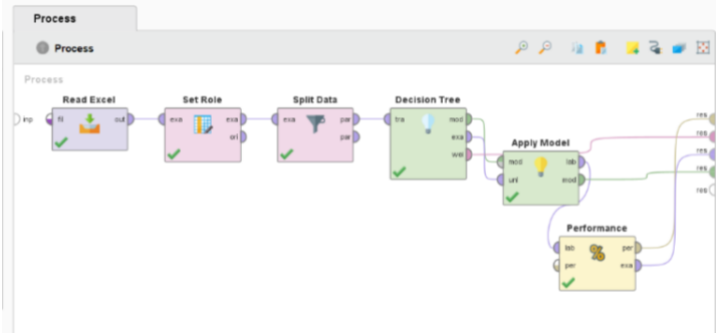


Fig. 5: Data Mining

4.6. Interpretation/Evaluation

In the evaluation stage, the results of the analysis using the Decision Tree model showed that the decision tree successfully split the data based on relevant attributes, such as "Gender," "Difficulty in Recording Numbers," and preference for learning methods. The decision tree split the data in an intuitive and logical way, separating the dataset based on the answers to certain questions. However, this model runs the risk of overfitting, especially if the tree is too complex. Therefore, pruning is necessary to improve the generalizability of the model to new data.

Evaluation results using performance metrics show that this Decision Tree model has a fairly high accuracy rate, around 90.21%. However, this model requires further evaluation to optimize its performance and ensure that it can be used on more diverse data.

attribute	weight
Menurut anda apakah lebih mudah mengerjakan soal dengar hitung daripada soal lihat hitung	0.335
Seberapa efektif menurut Anda metode dengar hitung dalam membantu mengasah kemampuan berhitung cepat?	0.013
Apakah Anda merasa kesulitan ketika merekam angka yang ditampilkan selama 3 detik dalam metode lihat hitung?	0.007
Manakah di bawah ini yang lebih anda sukai saat mengerjakan soal lihat hitung	0.198
JENIS KELAMIN	0.011
Apakah anda semakain percaya diri setelah mengerjakan soal dengar hitung +/-	0.035
Manakah di bawah ini yang lebih anda sukai saat mengerjakan soal dengar hitung	0.146
Dari total 20 soal lihat hitung, berapa soal yang berhasil Anda jawab dengan benar?	0.256

Fig. 6: Interpretation/Evaluation

4.7. Visualization Results

The results of the decision tree visualization, as shown in Figure 4.7, provide an overview of how each attribute affects the final decision of the model. Attributes that are more important in predicting students' learning method preferences are given higher weights.

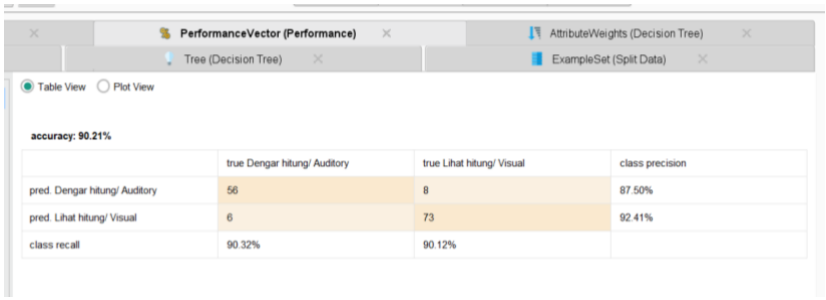


Fig. 7: Visualization Results

5. Conclusion And SuggestionZ

5.1. Conclusion

Previous research by Aidi Saputra et al. (2020) used the C4.5 algorithm to classify the level of student satisfaction with online learning media, showing that the model is effective with high enough accuracy to evaluate the effectiveness of learning media. In another context, analysis using the Decision Tree model in RapidMiner Studio demonstrated the algorithm's ability to identify patterns and significant relationships between attributes such as gender, confidence level, and learning method preference (auditory or visual), which affect student performance. The results support the importance of demographic variables and self-confidence in determining performance, while providing new insights into the influence of task difficulty. At Sempoa SIP Perjuangan, the application of Decision Tree successfully categorizes student learning patterns based on attributes such as learning method preference, attendance, and task evaluation results, with the main variables being confidence level and learning method preference. Although accurate, this model has limitations such as the risk of overfitting due to limited datasets, so development opportunities include using larger datasets and exploring other algorithms to improve classification performance.

5.2. Suggestions

Based on the results of this study, there are several development suggestions for further implementation. First, the research can be expanded by using a larger and more diverse dataset and adding other variables, such as stress level or learning style, to enrich the findings and increase the validity of the results. Second, the Decision Tree model can be applied to other fields, such as user behavior analysis or educational evaluation, to expand its practical benefits. Third, model optimization can be done by comparing other algorithms, such as Random Forest or Gradient Boosting, and applying pruning techniques to the Decision Tree to prevent overfitting and improve model generalization.

References

[1] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, 2023, doi: 10.57152/malcom.v3i1.591.

[2] M. F. Aditya, A. Pramuntadi, D. P. Wijaya, and Y. Wicaksono, "Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1104–1110, 2024, doi: 10.57152/malcom.v4i3.1284.

[3] K. Aidi Saputra, J. Tata Hardinata, M. Ridwan Lubis, S. Retno Andani, and I. Syahputra Saragih, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online," *Media Online)*, vol. 1, no. 3, pp. 113–118, 2020, [Online]. Available: https://djournals.com/klik

[4] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *J. Krisnadana*, vol. 1, no. 2, pp. 21–32, 2022, doi: 10.58982/krisnadana.v1i2.108.

[5] A. Kisnu Darmawan and M. Makruf, "KLIK: Kajian Ilmiah Informatika dan Komputer Deteksi Gaya Belajar Siswa SMA pada Virtual Based Learning Environment(VBLE) dengan Decision Tree C4.5 dan Naive Bayes," *Media Online*, vol. 3, no. 5, pp. 532–544, 2023, [Online]. Available: https://djournals.com/klik

[6] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.

[7] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 2, no. 2, pp. 91–97, 2021, doi: 10.37859/coscitech.v2i2.2851.

[8] A. Harun and D. Putri Ananda, "Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve bayes dan Decission Tree," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 58–64, 2021, doi: 10.57152/malcom.v1i1.63.