Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Decision Tree Algorithm to Improve the Learning Discipline Classification Model of Group Guidance Students at MTs Darul Mutta'alimin

Widina Khoeru Nissa[1]*, Ade Irma Purnamasari[2], Agus Bahtiar[3], Kaslani[4]

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

khoerunissawidina21@gmail.com[1*], irma2974@yahoo.com[2], Agusbahtiar038@gmail.com[3], kaslani123@gmail.com[4]

Abstract

This study aims to find patterns of student learning discipline at MTS Darul Mutta'alimin. In addition, this research also seeks to identify the main factors that influence discipline and build an analysis model that can be used by teachers and mentors to effectively improve student discipline. In its implementation, this research uses the Knowledge Discovery in Databases (KDD) method which consists of several stages, namely data collection, preprocessing, transformation, data mining, and evaluation of results. The data used includes test scores, student participation in class, and various other behaviors related to discipline. The results showed that the Decision Tree model developed had a high level of accuracy, reaching 98.36%. The main factors found to influence discipline are "Test Grades" and "Class Participation". In the process of classifying students into disciplined or undisciplined categories, the confusion matrix shows excellent model performance with a very low error rate. These results can be applied in education to assist teachers in monitoring students who need special attention. This study proves that the Decision Tree algorithm is very effective for finding patterns in data. For further development, it is recommended to add non-academic factors and test other algorithms to improve accuracy and broader generalization of the model.

Keywords: *Data Mining, Discipline Student, Decision Tree, Knowledge Discovery in Databases (KDD), dan Education Analysis*

1. Introduction

Many educational institutions, including MTS Darul Muta'allimin, deal with student discipline issues. Decision Tree algorithms have been widely used in various educational research, Decision tree algorithms are used for classification in education and data management.[1] developed a decision tree algorithm based on Ant Colony Optimization to handle imbalanced data problems. This algorithm produces better models in various conditions.[2] rioritize system requirement development criteria using the decision tree algorithm and the BOCR approach. This is relevant for educational planning.[3] about how the Decision Tree (DT) and Support Vector Machine (SVM) algorithms classify students who take remedial programs (her). The purpose of this study is to reduce the number of students who take remedial courses by creating a model that can predict students who need remedial courses based on historical data.[4] emphasizes the ability of this algorithm to solve problems with multivariate data. Although the decision tree algorithm has many successful applications outside of education, only a few studies have specifically used it to assess student discipline in group tutoring.[5] the effect on student learning motivation on learning motivation as a supporting variable. As a branching variable. The results of the study showed that discipline has a significant influence on students' learning motivation, both directly and through increasing learning motivation. In addition, learning motivation is also important to increase students' learning motivation. Therefore, learning motivation is also very important in increasing students' learning motivation. Thus, disciplined teaching methods, such as providing timely materials and creating a conducive learning environment, can increase students' motivation and enthusiasm for learning, which ultimately contributes to improving the quality of teaching.[6] Discipline and self-regulation have a significant impact on students' learning outcomes. learning objectives by fostering their creativity, perseverance, and emotions, while discipline that is carried out well can increase motivation and consistency in learning. Therefore, it is believed that teachers can apply more creative and focused teaching methods to help students improve their learning skills,[7][8] This approach aims to shape the character of students who are moral, disciplined, honest, and have a strong love for their country, so that they can face the challenges of globalization without losing their identity and noble values of the nation. Students who have a high level of discipline and supporting facilities have more positive self-control.

This study aims to apply the decision tree algorithm in improving student learning discipline. at MTs Darul Mutta'alimin through the development and implementation of a group guidance system supported by the Decision Tree algorithm. In addition, this study aims to fill the knowledge gap by offering a data-based discipline analysis method, which has not been widely studied in the context of secondary education

2. Literature Review

2.1. Data Mining

Data mining is the process of extracting useful information from large data sets using statistical analysis, machine learning, and human judgment. Data mining is the process of extracting useful information from large data sets using statistical analysis, machine learning, and human judgment. In this study, data mining is used to create a classification model for student learning using the Decision Tree algorithm.

2.2. Classification

Classification can be used to predict or make decisions based on patterns or characteristics identified in a particular group.[9] Student data is classified using the Fuzzy Tsukamoto method based on factors such as home ownership, achievement, parental income, dependents, and average report card grades.

2.3. Decision Tree

A decision tree is a flowchart consisting of trees used in the classification and development of decisions. The diagram illustrates the possible outcomes of various decisions, with each branch representing a choice and each leaf representing the final outcome. for example, about the use of the REPTree decision tree algorithm to optimize rule generation for the Tsukamoto fuzzy inference system[10].

3. Research Methods

All paragraphs must be justified alignment. With justified alignment, both sides of the paragraph are straight.

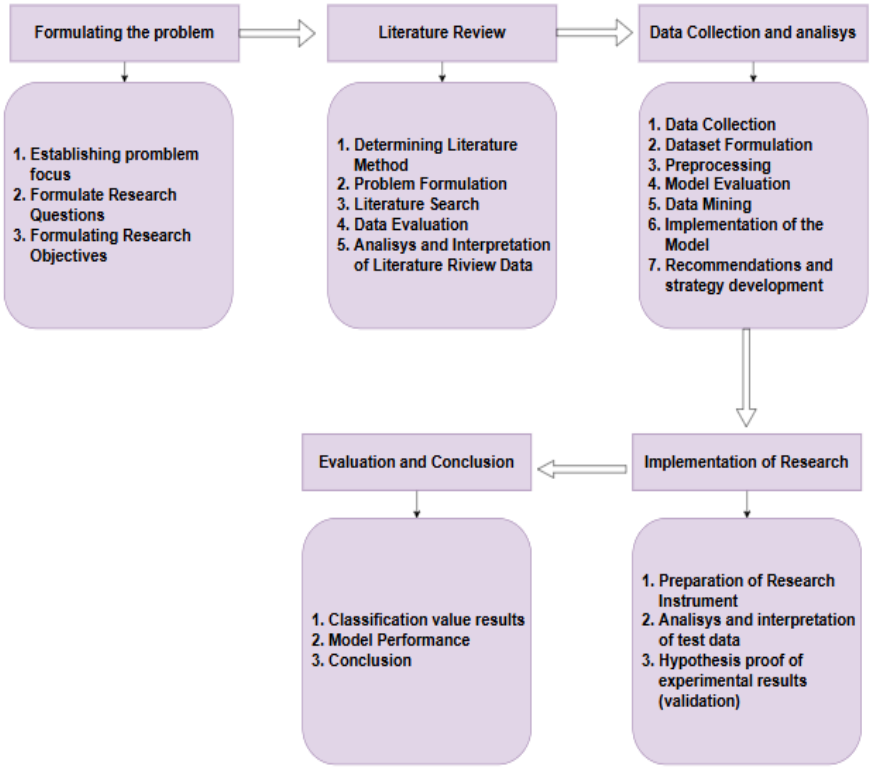

Fig.1: Description of Research Method Activities

Table 1: Description of Research Method Activities

Stages	activity	activity description
Formulating the Problem	Establishing Problem Focus	Using the Decision Tree algorithm to determine the research objectives and key issues related to analyzing student learning discipline to develop the analysis model.
	Formulate Research Question	Formulate clear and specific research questions.
	Formulate Research Objectives	Formulate specific, measurable and achievable research objectives based on the research questions with available resources and time. Therefore, the objectives set were as follows: 1.Apply the Decision Tree algorithm to identify patterns of student learning discipline in group guidance at MTS Darul Mutta'alimin.

		2.Identifying the dominant factors that influence student learning discipline based on the results of the Decision Tree analysis. 3.Analyzing the student learning discipline model
Literature Review	Determining Literature Method	Choosing the right method for literature searches such as searching for suitable keywords for research such as decision tree, classification, and student learning discipline.
	Problem Formulation	Develop research questions by reviewing the literature reviewed to determine the next research process.
	Literature Search	Collecting relevant literature according to the research topic taken by searching keywords or others, from various sources such as publish or perish, google scholar, Scopus.
	Data Evaluation	To assess the quality and relevance of data found in the literature to ensure validity and reliability.
	Analysis and Interpretation of Literature Review Data	Analyze the data from the literature collected and interpret the findings to identify patterns, findings and existing research gaps.
Data Collection and analysis	Data Collection	Data was obtained directly from MTs Darul Mutta'alimin. Student Learning Discipline Case Data contains some information, such as student participation in class, test scores, and attendance.
	Dataset Formulation	The research data above is the initial data collected from 122 students of MTs Darul Mutta'alimin. The data includes attributes such as name, class type, attendance, assignment scores, test scores, tardiness rate, and class participation rate. The label or target variable is the student's learning status (Discipline or Undiscipline).
	Preprocessing	The dataset will undergo a cleaning process or called Data Cleaning.
	Model Evaluation	The evaluation model in this study is carried out by dividing the dataset into training and testing data, then measuring the performance of the model using evaluation metrics such as accuracy, precision, recall, and confusion matrix.
	Data Mining	Apply data mining modeling techniques using the decision tree algorithm to classify student learning discipline data at mts darul mutta'alimin.
	Implementation of the model	By using the Decision Tree algorithm to analyze data and find patterns related to student discipline.
	Recommendations and strategy development	Recommendations and development strategies from the above research include the application of the Decision Tree model as a tool to identify students who have potential discipline problems, so that teachers and counselors can immediately design targeted interventions..
Implementation of Research	Preparation of Research Instrument	The development of research instruments is done by utilizing comprehensive data collection tools to analyze key variables such as attendance, test scores, and class participation which are key indicators of student learning discipline.
	Analysis and interpretation of test data	Test data processing is carried out by separating the dataset that has been transformed into training data and testing data, where the testing data includes sample data that is not used in the training model.
	Hypothesis proof of experimental results (validation)	The use of the Decision Tree algorithm to categorize student learning will produce a model with a high level of accuracy, at least 98%, and shows that the repetition and class participation variables are the main predictors in determining student discipline status.
Evaluation and Conslusion	classification value results	Presenting results using metrics such as accuracy, precision, recall, and confusion matrices.

	Model Performance	The performance model in this study is demonstrated by its very high accuracy, which reached 98.36%, meaning that most of the student data was successfully classified correctly into the "Disciplined" and "Undisciplined" categories.
	conclusion	Summarize the results, provide recommendations for future research, and discuss limitations.

4. Result and Discussion

4.1. Result

The results of this study present data on the number of students 122 at MTs Darul Mutta'alimin. The next step is to group the data using the decision tree algorithm by applying the Knowledge Discovery in Database (KDD) method.

4.1.1. Data Selection

The data used in this study is the data of MTs Darul Mutta'alimin students. The data collected were 122 students at MTs Darul Mutta'alimin, this data contains name, gender, attendance, assignment scores, test scores, tardiness, class participation and learning discipline. From the data that will be used as attributes, namely attendance, test scores and class participation, while learning discipline is used as a label. The dataset table can be seen below.

NO	Nama Siswa	Jenis Kelamin	kehadiran (%)	Tugas Selesai (%)	Nilai Ulangan (%)	keterlambatan (%)	Partisipasi kelas (%)	Disiplin belajar
1	ADITYA RESTU	L	90	85	80	5	50	Disiplin
2	AJUN KOMARA KHOERUDIN	L	70	60	65	20	88	Tidak Disiplin
3	ASSYIFA NURFADILAH	P	95	90	92	2	40	Disiplin
4	BUNGA AULIA LESTARI	P	98	78	70	1	30	Disiplin
5	CANDRA MUKTI	L	85	80	75	10	30	Disiplin
6	DESY NURFADILLAH YUSUF	P	100	80	78	4	45	Disiplin
7	DHIKA AL-GHANI HAKIM	L	92	88	85	4	45	Disiplin
8	DIAN ARDIANSYAH	L	98	70	75	4	43	Disiplin
9	DIAN FIRMANSYAH	L	90	85	88	5	60	Disiplin
10	DIKRA AGNIYA VILLAH	L	96	78	75	2	25	Disiplin
11	EUIS MARFUAH	P	88	80	83	8	52	Disiplin
12	FEBY PEBRIANTI	P	95	78	70	0	50	Disiplin
13	GILANG RAMADAN	L	95	92	91	3	47	Disiplin
14	IMAM MANDALA PUTRA	L	100	70	70	6	28	Disiplin
15	IZHIKA NUR AZHARA	P	85	80	78	7	65	Disiplin

Fig.2: Data Selection

4.1.2. Preprocessing Data

The second stage after data selection is data preprocessing. This process aims to eliminate missing values, inconsistent data, and data noise. At this stage, researchers use 2 operators, namely the Set Role and Select attribute operators.

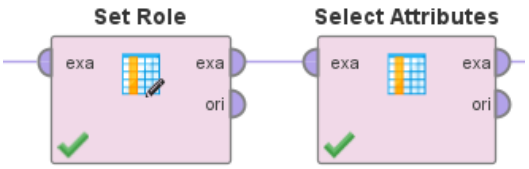

Fig.3: Operator Preprocessing Data

Set Role: Used to change the role attribute on this operator, researchers only change the Discipline attribute as a label.

Table 2: Set Role result

No.	Parameter	Isi
1.	attribute name	Disiplin belajar
2.	target role	label

Select Attributes: The second step of the select attributes operator is used to select several attributes based on the attribute name list, this is used to remove attributes that are not used in this study. The dataset generated after using this operator will be smaller and will only consist of the selected columns. From the results of reading the select attributes operator, the following information is obtained.

Row No.	Disiplin belaj...	prediction(D...	confidence(...	confidence(...	NO	Nama Siswa	Jenis Kelamin	kehadiran (%)	Tugas
1	Disiplin	Disiplin	0.988	0.012	1	ADITYA RESTU	L	90	85
2	Tidak Disiplin	Disiplin	0.988	0.012	2	AJUN KOMA...	L	70	60
3	Disiplin	Disiplin	0.988	0.012	3	ASSYIFA NU...	P	95	90
4	Disiplin	Disiplin	0.988	0.012	7	DHIKA AL-GH...	L	92	88
5	Disiplin	Disiplin	0.988	0.012	10	DIKRA AGNIY...	L	96	78
6	Disiplin	Disiplin	0.988	0.012	11	EUIS MARFU...	P	88	80
7	Disiplin	Disiplin	0.988	0.012	16	KANIA MADANI	P	100	96
8	Disiplin	Disiplin	0.988	0.012	26	SANDIKA	L	90	88
9	Disiplin	Disiplin	0.988	0.012	28	SATRIA RAM...	L	92	90
10	Disiplin	Disiplin	0.988	0.012	29	SEPIA NURS...	P	100	88
11	Disiplin	Disiplin	0.988	0.012	31	SYAKIRA LAT...	P	95	93
12	Disiplin	Disiplin	0.988	0.012	34	MEGA HARTA...	P	97	80
13	Disiplin	Disiplin	0.988	0.012	37	NURUL SRI ...	P	92	90
14	Tidak Disiplin	Disiplin	0.988	0.012	38	RAMDANI	L	68	62

Fig. 4: Select Attributes result

4.1.3. Transformation Data

transformation is to prepare the data so that it is ready to be used to build a Decision Tree model. In this study, data was imported from an Excel file using the Read Excel operator, then certain relevant attributes, such as "Attendance Level", "Assignment Value", and "Participation", were selected using Select Attributes. Furthermore, the Set Role operator is used to assign the role of the attribute in the analysis, for example to determine the target attribute.

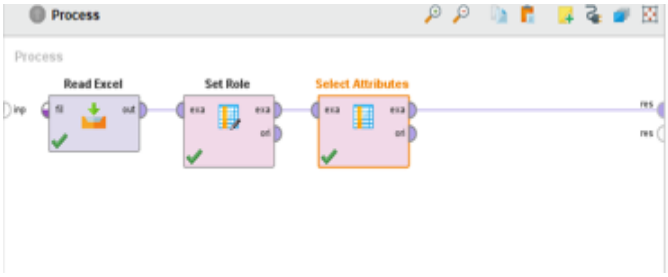

Fig. 5: Operator process models up to the Transformation step

4.1.4. Data Mining

At the data mining stage, researchers use a decision tree algorithm that involves several operators such as Read Excel, Set Role, Select Attributes, decision tree, apply model, and Performance.

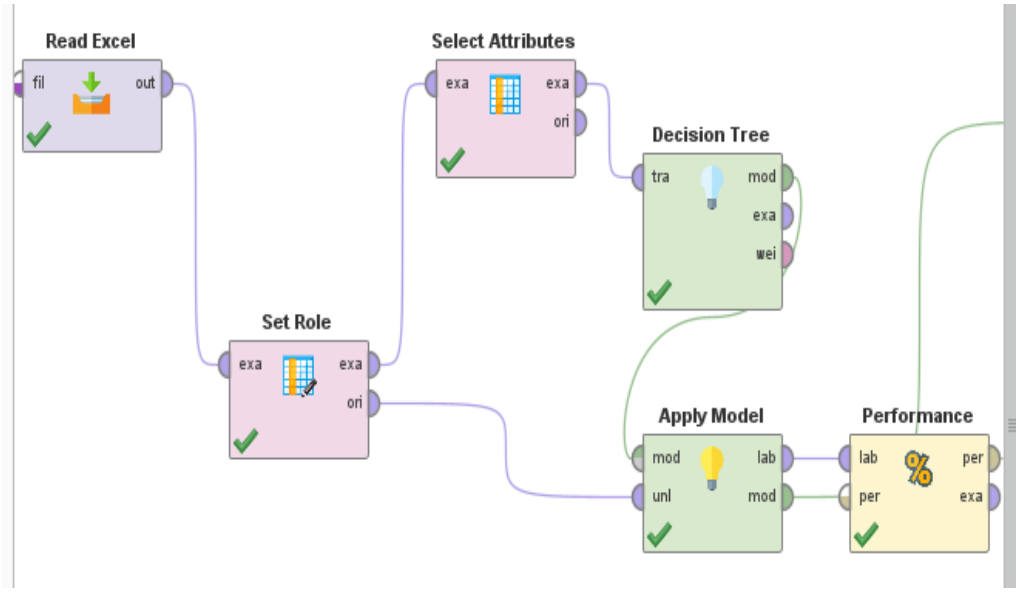

Fig. 6: Decision tree algorithm model

4.1.5. Evaluation

The results of the experimental activities on the dataset were evaluated, and the results are as follows.

a. Model accuracy; With an accuracy of 98.36%, this model shows an extraordinary ability to classify data.

accuracy: 98.36%

	true Disiplin	true Tidak Disiplin	class precision
pred. Disiplin	114	1	99.13%
pred. Tidak Disiplin	1	6	85.71%
class recall	99.13%	85.71%	

Fig. 7: Acurracy

b. Confusion Matrix.; The performance of the classification model in distinguishing students based on the level of discipline is shown by the evaluation matrix. With an accuracy of 98.36%, the model shows excellent prediction ability, and most of the test data are correctly classified. Depicting the distribution of prediction results against the original labels, the chaos matrix shows the distribution of prediction results against the original labels. In this case, the model correctly identified 114 students as disciplined, only one undisciplined student was incorrectly considered disciplined, six undisciplined students were also correctly predicted, and one disciplined student was incorrectly considered undisciplined.

Table View Plot View

weighted_mean_precision: 92.42%, weights: 1, 1

	true Disiplin	true Tidak Disiplin	class precision
pred. Disiplin	114	1	99.13%
pred. Tidak Disiplin	1	6	85.71%
class recall	99.13%	85.71%	

Fig. 8: Table View

c. Precision dan Recall; In this assessment, the precision and recall of each class are very important. The "Discipline" class has a precision and recall of 99.13% each, indicating that they are very accurate in predicting disciplined students. In contrast, the "Undisciplined" class has a precision and recall of 85.71% each, indicating that the model still has room for improvement in detecting undisciplined students. This shows that Overall, the classification model used, which is based on the Decision Tree algorithm, works well to analyze student discipline. However, to ensure more comprehensive and accurate analysis results, further optimization is needed for the "Undisciplined" element in the minority class.

4.1.6. Discussion

a. Application of decision tree algorithm for student learning discipline patterns in group guidance at Mts. Darul Mutta'alimin; The results of the study show that the Decision Tree-based classification model that was built has a very high accuracy rate, which is 98.36%, which means that this model can accurately predict student discipline in most situations. Based on the results of the decision tree, the main factors in categorizing students as "Disciplined" or "Undisciplined" are test scores (%) and class participation (%); tree analysis shows that students with test scores of more than 66.5% tend to be categorized as "Disciplined", especially if their class participation is more than 6%. From an educational perspective, these results indicate that academic achievement and student participation in class play an important role in shaping student discipline behavior. This makes sense because students who are more focused on learning and actively participate in class tend to have better discipline attitudes. Thus, schools can concentrate on increasing student participation in class and encouraging students to achieve better test scores.

b. Results of identification of dominant factors influencing learning discipline based on the results of decision tree analysis; The results of the analysis show that the main factors influencing student discipline are "Test Scores", followed by "Class Participation". This interpretation is in line with the literature stating that students' disciplinary attitudes are often indicated by academic achievement; class participation is also proven to be an important indicator because students' active involvement in the learning process contributes to improving their discipline. This finding supports the theory that academic engagement and motivation are important factors determining students' disciplinary behavior.

c. Results of analysis of student learning discipline models; The results of the analysis of the created model show that they can easily find and predict student discipline. The model is not only accurate but also reliable in classifying students into the category of "Discipline" or "Undisciplined", as indicated by evaluation metrics such as precision, recall, and F1 score. In addition, the classification rules created can be used by educators to create more targeted guidance programs, such as increasing student participation and attendance.

5. Conclusion and Suggestions

5.1. Conclusions

The results of the study show that:

1. The high accuracy of 98.36% of the developed Decision Tree algorithm model shows that it is able to classify data very well to determine the level of student discipline.
2. Test scores and class participation are two main factors that affect student learning discipline. Students with test scores of more than 66.5% tend to be categorized as "Disciplined", especially students with class participation of more than 6%.
3. From the results of the metric evaluation, the "Disciplined" class has a high level of precision and recall (99.13%), but the "Undisciplined" class has a lower level of precision and recall of 85.71%.

5.2. Sugestions

To improve the generalizability and accuracy of the model, this study can be expanded by including non-academic elements such as students' psychological conditions and their social environment. In addition, the research data can be expanded to other schools. In order for the results of the model analysis to be used effectively in student guidance programs, regular training and coordination are needed to improve collaboration between teachers, counselors, and management. In addition, management must involve parents through clear communication about the predicted results and planned interventions so that schools and families work together. To ensure smooth and sustainable implementation, the budget allocated for staff training and technology development must also be considered. Schools can use this method to create data-based policies and improve the quality of education and discipline.

References

[1] M. H. Bin Mohd Razali, R. Bin Saian, Y. B. Wah, and K. R. Ku-Mahamud, "A class skew-insensitive ACO-based decision tree algorithm for imbalanced data sets," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 1, pp. 412–419, 2021, doi: 10.11591/ijeecs.v21.i1.pp412-419.

[2] T. Amelia and R. Mohamed, "A decision tree approach based on BOCR for minimizing criteria in requirements prioritization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 2, pp. 1094–1104, 2023, doi: 10.11591/ijeecs.v32.i2.pp1094-1104.

[3] J. J. Purnama, H. M. Nawawi, S. Rosyida, R. Ridwansyah, and R. Risnandar, "Klasifikasi Mahasiswa HER Berbasis Algoritma SVM dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1253, 2020, doi: 10.25126/jtiik.0813080.

[4] I. Mansoury, D. El Bourakadi, A. Yahyaouy, and J. Boumhidi, "A novel decision-making approach based on a decision tree for micro-grid energy management," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 1150–1158, 2023, doi: 10.11591/ijeecs.v30.i2.pp1150-1158.

[5] M. Yany, D. Soelistya, E. Decembrianita, R. A. Samtoso, and M. Sholichah, "The role of lecturer discipline on student interests: motivation to learn as intervening variable," *JPPI (Jurnal Penelit. Pendidik. Indones.*, vol. 9, no. 1, p. 61, 2023, doi: 10.29210/020221982.

[6] J. Siska, M. Afrina, O. L. Agusta, R. N. Sasongko, and M. Kristiawan, "Self regulation and discipline development to improve independence students in English course," *JPPI (Jurnal Penelit. Pendidik. Indones.*, vol. 8, no. 2, p. 329, 2022, doi: 10.29210/020221640.

[7] A. Surip, M. A. Pratama, I. Ali, A. R. Dikanda, and A. I. Purnamasari, "Penerapan Machine Learning menggunakan algoritma C4.5 berbasis PSO dalam Menganalisa Data Siswa Putus Sekolah," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 2, pp. 147–155, 2021, doi: 10.51211/itbi.v5i2.1530.

[8] E. Ernawati, "Fashion design education students' ability to create fashion patterns: investigating the effect of antecedent factors," *JPPI (Jurnal Penelit. Pendidik. Indones.*, vol. 8, no. 2, p. 312, 2022, doi: 10.29210/020221733.

[9] R. Akbar and S. 'Uyun, "Penentuan Bantuan Siswa Miskin Menggunakan Fuzzy Tsukamoto Dengan Perbandingan Rule Pakar dan Decision Tree (Studi Kasus : SDN 37 Bengkulu Selatan)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 651–662, 2021, doi: 10.25126/jtiik.0813191.

[10] Tundo and S. Uyun, "KONSEP DECISION TREE REPTREE UNTUK MELAKUKAN OPTIMASI RULE DALAM FUZZY INFERENCE SYSTEM TSUKAMOTO," *J. Teknol. Inf. dan IlmuKomputer*, vol. 9, no. 3, pp. 513–522, 2021, doi: 10.25126/jtiik.202292601.