Journal of Electrical Engineering and Computers (JEECOM)
Vol. 7, No. 1 (2025), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Implementation of the C4.5 Decision Tree Algorithm in Determining PIP Recipient Students Based on Poverty

Heri Susanto[1], Dwi Yanto[2], Jamal[3], Choirul Anam[4], Thohirotun Nafisah[5]
[1245]Information Systems, Taruna Informatics and Computer Management Academy, Probolinggo, Indonesia
[3]Accounting Information System, Taruna Informatics and Computer Management Academy, Probolinggo, Indonesia

Article Info

Article history:

Received April 14 2025
Revision April 20 2025
Published April 23 2025

Keywords:

Decision Tree
Aid Recipients
Smart Indonesia Program
SDN Kedungjajang 02

ABSTRACT

Abstract— The *Program Indonesia Pintar* (PIP) is a government assistance in the form of a certain amount of cash given directly to students who are vulnerable to poverty according to the established criteria. However, in its implementation, there are still many recipients of PIP funds who are still not on target. Several results of evaluations and ongoing studies on the implementation of PIP show weaknesses in this program, namely related to the accuracy of determining the target of PIP fund recipients, where it was found that there were still many non-poor households who received this PIP fund assistance. The purpose of this study is to produce a decision support system to determine students who are truly worthy of being recommended to receive PIP fund assistance. The method used in this study is the Decision Tree C4.5 algorithm. The data used were 300 datasets, by selecting several relevant attributes. For processing using the Rapidminer tool, one of the popular software used in data mining processing. Meanwhile, the evaluation method uses a confusion matrix by calculating the accuracy value. From the test results in determining PIP acceptance at SD Negeri Kedungjajang 02 using the C4.5 algorithm, an accuracy of 97.33% was obtained, with the accuracy criteria of good classification.

Corresponding Author:
Mr. Heri Susanto,
Taruna Informatics and Computer Management Academy, Jl. Raya Leces No.A3 Leces, Probolinggo 67202, Indonesia
E-mail:herisusanto@amik-taruna.ac.id

1. INTRODUCTION

The *Program Indonesia Pintar* (PIP) is a government assistance in the form of a certain amount of cash given directly to students according to the established criteria. In an effort to equalize opportunities to obtain education and quality of education. To reduce the number of school dropouts, the government is expanding access to higher quality and more equitable education by paying greater attention to the poor.[1] The purpose of this program is to help school-age children from underprivileged families to experience fair and equal education services[2]. The target recipients of PIP are students from families participating in the Family Hope Program, from families holding Prosperous Family Cards, orphans/half-orphans from schools/social institutions/orphanages, students affected by natural disasters, students who are not in school (drop out) who are expected to return to school, students who have physical disabilities, victims of disasters, from parents who have been laid off, in conflict areas, from families of convicts, in correctional institutions, have more than 3 (three) siblings living in the same house[3].This attention is in the form of providing funds for the Program Indonesia Pintar (PIP).[4] However, in its implementation, there are still many PIP fund recipients who are not on target.

Several results from evaluations and ongoing studies on the implementation of PIP show weaknesses in this program, namely related to the accuracy of determining the targets for recipients of PIP funds, where it was found that there were still many non-poor households who received PIP funds and the number of scholarships was inadequate.[5]

The school selects students to be recommended for PIP in previous decision-making using a manual system. In recommending students for PIP, the school must collect student data to determine which students are eligible to be recommended for PIP. This makes it a little difficult for teachers to make decisions. In the 2021/2022 school year there were 110 students, in the 2022/2023 school year there were 114 students and in the 2023/2024 school year there were 109 students

Based on the facts of the problem, this study realizes a system that makes it easier for teachers to determine aid recipients. Using a decision support system with the Decision Tree C4.5 method. It is hoped that this study can produce a decision support system to determine students who are truly worthy of being recommended.

2. LITERATURE STUDY

2.1 Data Mining

According to (Han, J., et al, 2012: 8), data mining is the process of finding patterns and knowledge from large amounts of data. Data sources are the basic things that must be present to carry out the data mining process. The database in this study is a database with a relational model, where the data is in the form of tables consisting of a number of rows and columns that show certain attributes.

Data Mining is a knowledge in the Knowledge Discovery from Data (KDD) process, where the process of analyzing or processing a lot of data is done to obtain output in the form of knowledge of patterns or tendencies, which output will later be used to determine policies or actions.[6]

2.2 RapidMiner

RapidMiner is a software for data processing. Using data mining principles and algorithms, RapidMiner extracts patterns from large data sets by combining statistical methods, artificial intelligence and databases [7]. Rapidminer is an independent software used to analyze data and data mining engines, which can be easily integrated with various programming languages [8]. RapidMiner is open source software created using the Java programming language so that it can be accessed by all operating systems. RapidMiner can be used as a solution in analyzing data mining, using descriptive and predictive techniques provided to users so that they can make the best decisions [9].

RapidMiner has several properties as follows: 1. Written in Java programming language so it can be run on various operating systems; 2. The knowledge discovery process is modeled as operator trees; 3. Internal XML representation to ensure standard data exchange formats; 4. Scripting language allows for large-scale experiments and automation of experiments; 5. Multi-layer concept to ensure efficient data display and ensure data handling; 6. Has GUI, command line mode, and Java API that can be called from other programs.[10]

3. METHOD

This study cannot be separated from the framework of previous research that has been done, so it can be used as a reference and basis for further analysis. Several previous studies relevant to this study used different algorithms including Naïve Bayes, K-Nearest Neighbor, and decision trees as listed in Table 1. This study used the C4.5 algorithm with the intention of presenting research results, compared to using the previous algorithm.

Table1. Relevant Research

No.	Title	Method	Results
1	Data Mining Implementation Using Methods Naive Bayes for Determining Recipients of *Program Indonesia Pintar* (PIP) Assistance (Case Study: State Elementary School 9 Air Kumbang).[11]	Naive Bayes	The background of this study is that SDN 9 Air Kumbang is one of the elementary schools that received funds from the Program Indonesia Pintar (PIP), but the PIP target at SDN 9 Air Kumbang is still not on target due to the lack of criteria for the number of dependents. Therefore, the criteria for the number of dependents were added in this study. This journal discusses the application of data mining methods using the Naive Bayes algorithm to determine recipients of the *Program Indonesia Pintar* (PIP) at SDN 9 Air Kumbang. This study has 5 attributes, namely Parental Occupation, Parental Income, Number of Dependents, KIP Recipients, and KPS Recipients. The test results show that the Naive Bayes method has an accuracy of 90.00%.

ISSN:2715-6427

2	Information System for Selecting Participants in the *Program Indonesia Pintar* (PIP) Using the K-Nearest Neighbor Method at Pejuang V State Elementary School, Bekasi City.[12]	K-Nearest Neighbor	The background of this journal is that the current PIP participant selection system at SD Negeri Pejuang V does not meet the current technological development standards, because it is still done manually and uses supporting files. This causes a long time and human error in the selection process. Therefore, this study aims to optimize technological development by designing a PIP participant selection information system using the K-Nearest Neighbor method. This study involved student data from 77 students and succeeded in achieving an accuracy of 90.90% using a K value = 10 based on attribute criteria of 4 indicators such as parental occupation, parental status, income and number of dependents
3	Implementation of the C4.5 Decision Tree Algorithm for the Model for Determining Recipients of the *Program Indonesia Pintar* (PIP) Scholarship. Case Study of State Senior High School 3 Timang Gajah.[13]	Decision Tree C4.5	The results of the study showed an accuracy of 72% in determining PIP eligibility and there were 3 attributes that were most influential, namely the recipient of assistance, type of residence and parents' occupation.

This research framework is a method for collecting data regularly so that it can avoid the validity of data information, in this research framework it can be designed efficiently and in detail in collecting data. In data collection, namely by using qualitative methods, namely by conducting direct observations and interviews so that it can organize the data needed in the study. [14] The following is a description of the research framework.

The framework for this research is as follows:

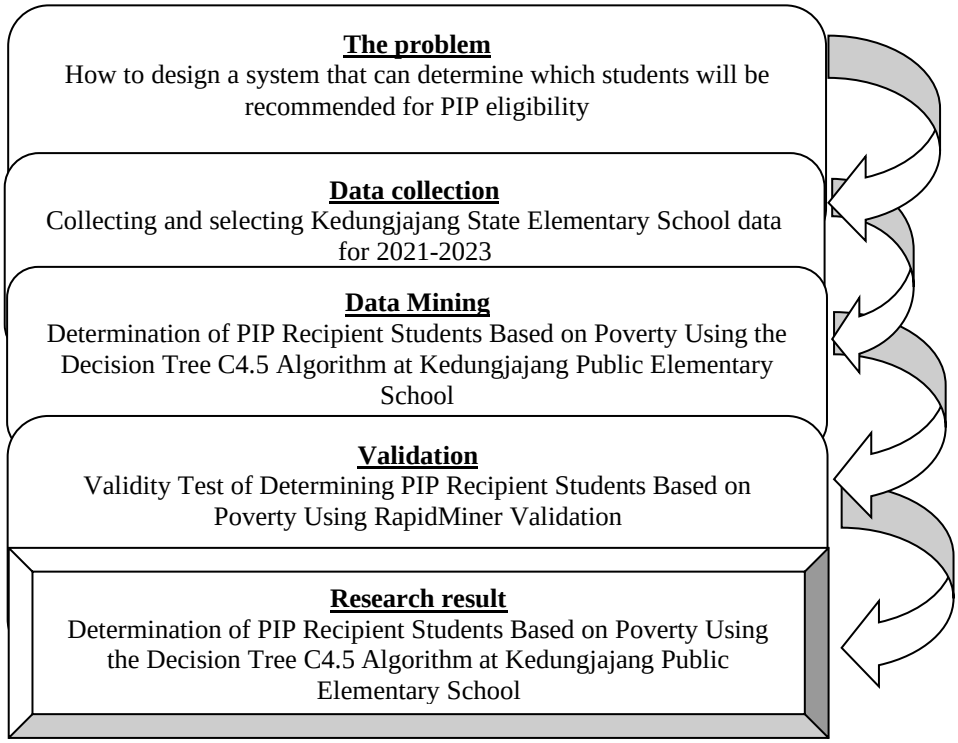

Figure 1. Research Framework

The flow of stages in this research is as follows:

1. Selection
   The initial dataset was taken from the data of Kedungjajang 02 State Elementary School for the years 2021-2023 obtained from the SIPINTAR Website. From the existing school data, it was collected (collection) to then be selected (selection) and the data needed for this study was combined (selected data). In this study, the indicators used were parental occupation, parental status, parental income, having a SKTM, and the number of parental dependents.
2. Preprocessing
   This stage is used to prepare the dataset to have better quality and be more effective before being modeled. School dataset with simple features with attributes: Parental occupation, parental status, parental income, having SKTM, number of parental dependents, and recipient of social assistance. In the pre-processing stage, we use filtering techniques to organize student data based on categories relevant to PIP receipt. Information about parental occupation helps in identifying the socio-economic background of students, parental income, family dependents, and social assistance recipient status are also important factors in the data categorization process. With this filtering process, the dataset will be more structured and have a better representation of the PIP receipt criteria.
3. Decision Tree C4.5
   At this stage is the dataset classification stage. [15]
   The classification steps of the C4.5 decision tree algorithm are as follows:

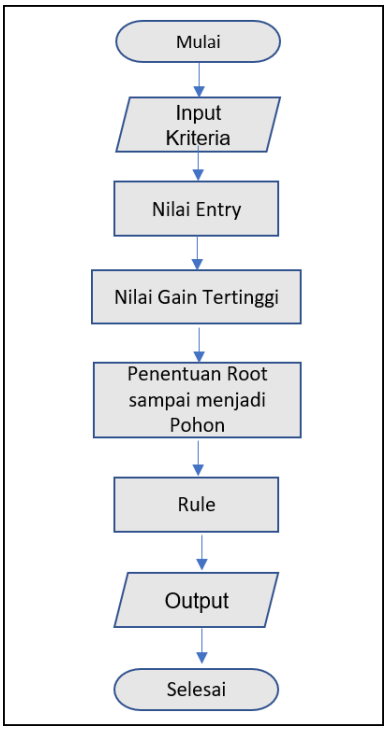

Figure 2 Steps in implementing Algorithm C.45

4. RESULTS AND DISCUSSION

1. Selection Stage
   As explained previously, the research dataset was taken from the data of Kedungjajang 02 State Elementary School for 2021-2023. From several school data tables (data collection) needed for this study.
   a. Data Collection
   b. Not all variables of the data that have been cleaned are used in this study, therefore data selection needs to be carried out, so that only data that meets the criteria will be analyzed, as many as 300 data and the samples displayed are as many as 27 data as table 2 follows:

ISSN:2715-6427

Table 2. Sample Dataset

NAMA	PEKERJAAN ORANG TUA	STATUS ORANG TUA	PENGHASILAN ORANG TUA KURANG DARI 600RB DAN MEMILIKI SKTM	JUMLAH TANGGUNGAN ORANG TUA LEBIH DARI TIGA	PENERIMA PROGRAM BANTUAN SOSIAL	PIP
Ahmad Fahimmudin	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Ahmad Wildan	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Ashadiyah Layla Arindita	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Bintang Samudra	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Lailatul Badriyyah	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Lailatul Fitria	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Mauna Tasya Putri	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
M. Indra Kurniawan	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
M. Fathan Al Maisan Zhafar	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Muhammad Jihan Albian	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Nabigha Faruq Dzhurnurrain	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Naila Muazara Ulfa	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Naurin Azhalia	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Rafa Al Fariski	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Ramadhan Febriyansyah	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Ranfi Prastian	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Riski	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Siti Amelia Putri	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Syakila Azzahra Romadhona	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Afifah Afkarina	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Ahmad Nur Aditya Putra	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Aisyah	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Alfin Abdillah Pratama	BURUH TANI	YATIM/PIATU	IYA	TIDAK	IYA	DAPAT
Almira Pricilia Putri	BURUH TANI	LENGKAP	IYA	TIDAK	TIDAK	DAPAT
Ana Tasya Azahra	BURUH TANI	YATIM/PIATU	IYA	TIDAK	TIDAK	DAPAT
Arindita Keiza Az Zahra	KARYAWAN SWASTA/WIRASWASTA/WIRAUSAHA	LENGKAP	TIDAK	TIDAK	TIDAK	TIDAK DAPAT
Indah Ayu Marsela	KARYAWAN SWASTA/WIRASWASTA/WIRAUSAHA	LENGKAP	TIDAK	TIDAK	IYA	TIDAK DAPAT

2. Preprocessing

This stage is used to prepare the dataset to have better quality and be more effective before being modeled.

3.3 Implementation of Decision Tree C4.5 and Calculation Results

Implementation of Classification of determining PIP recipient students based on poverty with the Decision Tree C4.5 algorithm was carried out in testing, namely school datasets from 2021-2023. In this process, the attributes processed consist of 5 attributes used in this study, namely parental occupation, parental status, parental income, SKTM and number of parental dependents.

Below are the steps for calculating the C4.5 method.

a. Calculating Entropy Value

The first process of the C4.5 algorithm is to determine the entropy value. The first step is to determine the total entropy of the case first from the data, here is the calculation:

$$Entropy(S) = \left(-\left(\frac{Dapat}{total} x\ Log_2\left(\frac{Dapat}{total}\right)\right)\right) + \left(-\left(\frac{Tidak\ Dapat}{total} x Log_2\left(\frac{Tidak\ Dapat}{total}\right)\right)\right)$$

$$\text{Entropy(S)} = \left(-\left(\frac{118}{300} x\ log_2\ \left(\frac{118}{300}\right)\right) + \left(-\left(\frac{182}{300} x\ log_2\ \left(\frac{182}{300}\right)\right)\right)\right)$$

$$\text{Entropy(S)} = 0{,}966917009$$

Penghitungan Nilai *Entropy* Total

ATRIBUT	NILAI	TOTAL	DAPAT	TIDAK DAPAT	ENTROPY
TOTAL (S)		300	118	182	0.996117009

Figure 3. Entropy Formula and Calculation

b. Calculating Gain Value

Once the total entropy value is known, the next step is to find the gain and entropy values of each attribute. To find the gain and entropy values for the five attributes of the data, use the following formula: [16]

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy\,(S_i)$$

Figure 4. Gain Value Formula

3.4 Classifier Performance Measurement

A system that performs classification is expected to be able to classify all data sets correctly. But the performance of a system cannot work 100% correctly. Therefore, the performance of the classification system must be measured, the way to measure the classification performance used is the confusion matrix. To calculate the accuracy, the formula used is the number of data predicted correctly divided by the number of predictions made [17].

Dataset processing using RapidMiner software with the Decision Tree C4.5 method, as shown in Figure 5.

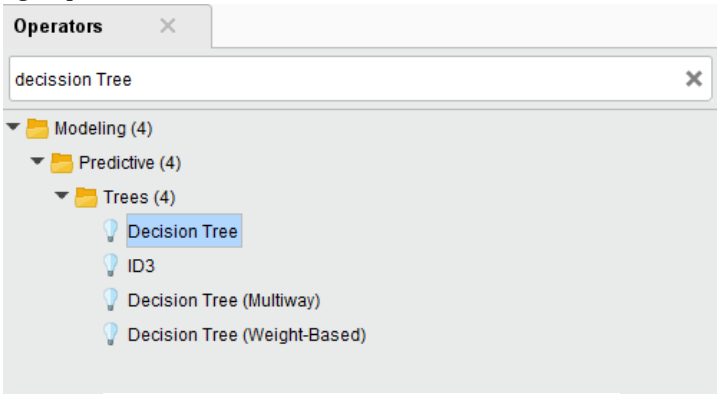

Figure 5. Decision Tree Algorithm

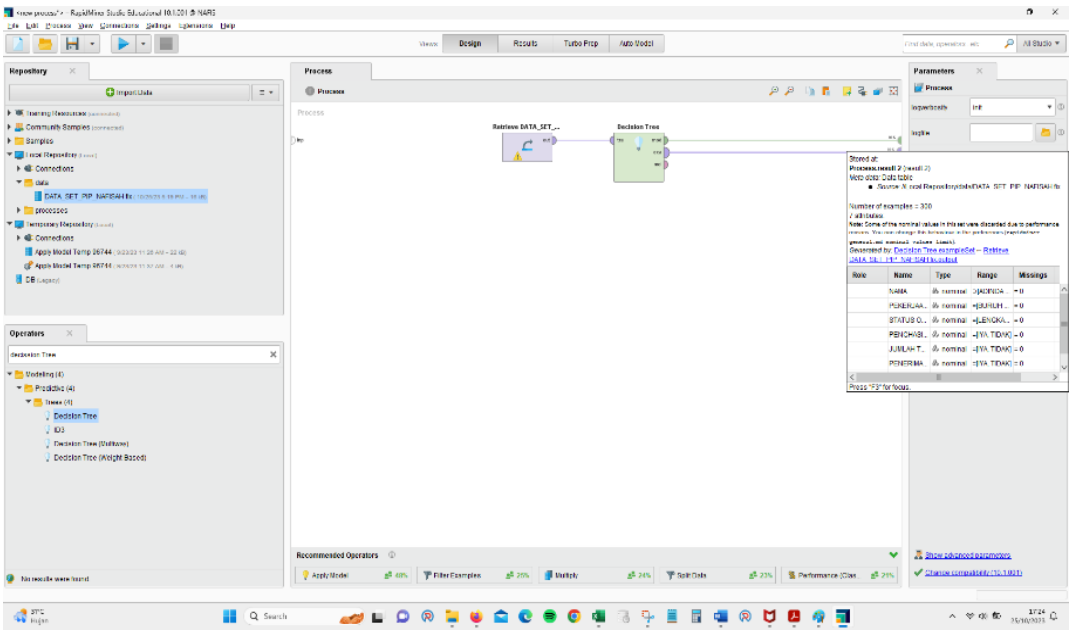

Figure 6. Processing Using Decision Tree On RapidMiner

ISSN:2715-6427

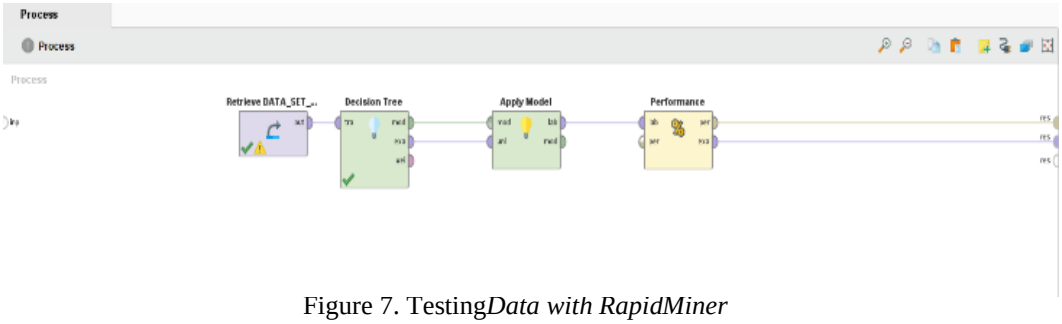

Figure 7. Testing*Data with RapidMiner*

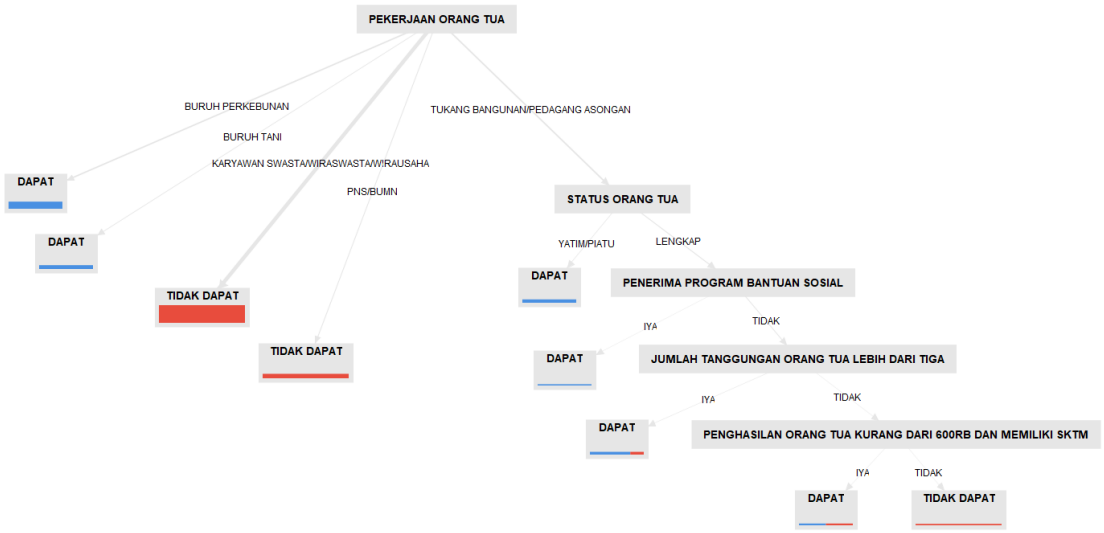

Figure 8. Decision Tree Results of Processing with RapidMiner

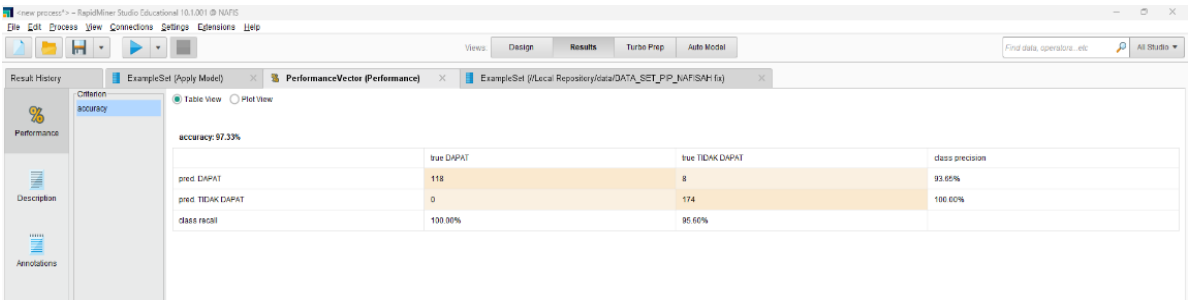

Figure 9. Confusion Matrix Processing Results

The results of processing using Rapidminer software displayed through a confusion matrix obtained an accuracy of 97.33%.

5. CONCLUSION

The results of the research that has been done, it can be concluded that pre-processing is an important step in data mining. From the results of testing and implementing data mining in determining PIP acceptance at SD Negeri Kedungjajang 02 using the C4.5 algorithm, an accuracy of 97.33% was obtained with the criteria for

good classification accuracy. There is an increase in accuracy compared to previous research conducted by Aprilyani with an accuracy of 72%.

ACKNOWLEDGEMENTS

Our deepest gratitude to the Head Master of Kedungjajang Elementary School, and all parties who have helped in completing this research. We sincerely express our gratitude to the leaders, colleagues who have provided valuable input and constructive criticism in completing this research.

REFERENCE

[1] Lugas, Jihan (2018) Decision Support System for Recommendation of *Program Indonesia Pintar* (PIP) Fund Recipients Using Analytical Hierarchy Process (AHP) Method. Other thesis, Universitas Islam Riau.
[2] Ministry of Education and Culture. (2020). Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 10 of 2020 concerning the *Program Indonesia Pintar*.
[3] Bona, FM, and Markus JS (2018, April 18). There Are Still Problems in KIP Distribution. Berita Satu.https://www.beritasatu.com/nasional/488872/masih-ada-persoalan-dalam-penyaluran-kip
[4] Ministry of Education and Culture. (2020). Regulation of the Minister of Education and Culture of the Republic of Indonesia Number 10 of 2020 concerning the Smart Indonesia Program.
[5] Bona, F. M., and Markus J. S. (2018, April 18). There Are Still Problems in the Distribution of KIP. Berita Satu. https://www.beritasatu.com/nasional/488872/masih-ada-persoalan-dalam-penyaluran-kip
[6] Han, Jiawei, et al. "Data Mining Concepts and Techniques." Edition 3. Morgan Kaufmann, USA. 2012
[7] S. Sumarlin and D. Anggraini, "Implementation of K-Nearest Neighbord in Rapidminer for Predicting Student Graduation," High Educ. Organ. Arch. Qual. J. Teknol. Inf., vol. 10, no. 1, pp. 35–41, 2018, doi: 10.52972/hoaq.vol10no1.p35-41.
[8] M. Nasution, A. A. Ritonga, and A. P. Juledi, "Implementation of Rapidminer in Classifying Democracy Index," J. Comput. Sci. Inf. Technol., vol. 3, no. 3, pp. 99–106, 2022.
[9] A. Apriyadi, M. R. Lubis, and B. E. Damanik, "Application of C5.0 Algorithm in Determining the Level of Student Understanding of Online Learning," Komputa J. Ilm. Comput. and Inform., vol. 11, no. 1, pp. 11–20, 2022, doi: 10.34010/komputa.v11i1.7386.
[10] Baginda, Baginda. (2023). Rapid Miner Analysis of the Purpose of the Expertise Approach to the Desire to Leave with Intrinsic Motivation as a Moderator Variable. Jurnal Minfo Polgan. 12. 10.33395/jmp.v12i1.12926.
[11] O. Rini and S. O. Kunang, "Implementation of Data Mining Using the Naive Bayes Method for Determining Recipients of the Smart Indonesia Program (PIP) (Case Study: Sd Negeri 9 Air Kumbang)," Bina Darma Conf. Comput. Sci., vol. 3, no. 4, pp. 714–722, 2021, [Online]. Available: https://conference.binadarma.ac.id/index.php/BDCCS/article/view/2450
[12] Prihatin, Sandy & Atika, Prima & Herlawati, Herlawati. (2021). Information System for Selecting Participants of the Smart Indonesia Program (PIP) Using the K-Nearest Neighbor Method at Sd Negeri Pejuang V, Bekasi City. 2. 165-176. 10.31599/jsrcs.v2i2.911.
[13] Aprilyani, Nova & Zulfa, Ira & Syahputra, Hendri. (2022). Application of the C4.5 Decision Tree Algorithm for the Determination Model for Recipients of the Smart Indonesia Program (Pip) Scholarship Case Study of Sma Negeri 3 Timang Gajah. Journal of Informatics and Electrical Engineering. 5. 96-109. 10.55542/jurtie.v5i1.452.
[14] Yakin, Ipa & Supriatna, Uus & Rusdian, Suca & Global Akademia, Mia. (2023). QUALITATIVE RESEARCH METHODS.
[15] Hermawati, Fajar Astuti. (2013). Data Mining. Yogyakarta: Andi Offset
[16] Ratna, W., Oktapia, S., Sumarno, S., & Nasution, ZM (2022). Classification of C4 Algorithm . 5 for Determining Recipients of the Program Indonesia Pintar on MIS Al-Khoirot. 1(1), 65–76. https://doi.org/10.55123/jomlai.v1i1.165
[17] H. Susanto, D. Yanto, N. Rusdiana, and D. T. Rahayu, "Prediction of Heart and Blood Vessel Disease Risk Using K-Nearest Neighbor (KNN) Algorithm," Joined J. (Journal of Informatics Educ., vol. 6, no. 2, pp. 148–158, 2023.