



## CLASSIFICATION OF ASSISTANCE RECIPIENTS USING THE ALGORITHM C4.5

Diana Santi<sup>1\*</sup>, Siti Aisyah<sup>2</sup>, Asep Afandi<sup>3</sup>, Dwi Marisa Efendi<sup>4</sup>

<sup>1,2,3,4</sup>Information Systems Study Program, Institut Teknologi Bisnis dan Bahasa Dian Cipta Cendikia Kotabumi

<sup>1,2,3,4</sup>HOS Cokroaminoto Street, No. 3, Kotabumi, Cempedak, North Lampung, Lampung

e-mail: [dianadcc88@gmail.com](mailto:dianadcc88@gmail.com)<sup>1\*</sup>, [siti45449@gmail.com](mailto:siti45449@gmail.com)<sup>3</sup>, [asepafandi189@gmail.com](mailto:asepafandi189@gmail.com)<sup>3</sup>,

[dwi.marisa@dcc.ac.id](mailto:dwi.marisa@dcc.ac.id)<sup>4</sup>

### Article history:

Received: November 12, 2023

Revised: December 14, 2023

Accepted: December 30, 2023

Corresponding authors

\*[dianadcc88@gmail.com](mailto:dianadcc88@gmail.com)

### Keywords:

Algorithm C4.5;

Data Mining;

Python;

Family Hope Program.

### Abstract

Poverty is a condition that results in an inability to meet basic needs such as housing, food, education and health. Based on data from the Central Statistics Agency (BPS) that the number of poor people in Indonesia in September 2019 was 24.79 million people. In order to minimize social welfare problems, especially the problem of increasing poverty, the Government of Indonesia issued policies related to empowering poor families. The government policy is the Family Hope Program (PKH), the Implementation of the Family Hope Program (PKH) has not been implemented on target. This has resulted in social jealousy between residents in the community, one of the villages where PKH assistance distribution is still not on target is Gilih Suka Negeri Village. Gilih Suka Negeri Village is one of the areas in Lampung Province, North Lampung Regency, South Abung District, which has a total of 7 RTs (RTs) with a population of approximately 1,776 people. Algorithm method C4.5 (Decision Tree) was chosen by the author because it can be used as a solution to determine the classification of beneficiaries. Algorithm C4.5 (Decision Tree) will classify beneficiaries of the Family Hope Program (PKH) based on how much they earn, where they live, and how many people they have in their household.



This is an open access article under the CC-BY-SA license.

### I. INTRODUCTION

A person living in poverty cannot meet basic needs such as housing, food, education and health care. The Central Statistics Agency (BPS) reported that there were 24.79 million poor people in Indonesia in September 2019. In addition, a country's development is slow and underdeveloped due to the high poverty rate. Removing poverty is certainly not something easy, but poverty can be suppressed or limited. In order to limit the problem of government social assistance, especially the ever-growing issue of poverty, Indonesian community experts provide plans that relate to less fortunate families who cannot afford the costs. The public authority's strategy is the Family Hope Program (PKH)[1].

The Family Hope Program (PKH) provides cash assistance to Very Poor Households (RTSM). In return, RTSM is required to fulfill requirements related to health and education related to efforts to

improve the quality of human resources. In fact, PKH's overall goal is to make it easier for low-income families to gain access to services that can improve their quality of life, such as health care, education, and social welfare. In the long term, PKH is expected to break the chain of poverty and reduce the financial burden on low-income families in the short term[2]. The Family Hope Program (PKH) has not been implemented as planned. Social jealousy among members of society developed as a direct result of this. This is due to a lack of an electronic framework that could make it easier to determine the qualifications of future PKH beneficiaries. Gilih Suka Negeri Village is one of the villages that has not been effective in PKH assistance[3].

One of the areas in Lampung Province, North Lampung Regency, and South Abung Regency is Gilih Suka Negeri Village which has a total of 7 RTs (RTs) with a population of approximately 1776

people. The Integrated Social Welfare Data (DTKS) for 2022 from the North Lampung Social Service recorded a total of 922 poor people (Gakin) in Gilih Suka Negeri Village. 82 families are among the poor receiving PKH benefits, according to the data. PKH beneficiaries who are not synchronized are the result of this problem.

Communities who are entitled may not get it, but PKH assistance is actually given to people who are not poor. This will reduce the asymmetry of PKH beneficiaries and also make it easier for officers to distribute PKH assistance if PKH beneficiaries are clearly documented and the eligibility requirements for PKH assistance recipients are well documented. Utilizing technology, increasing knowledge in the field of PC and innovation has now and again developed in various parts of life. As a result of this growth, new knowledge about how to present information to meet demand has emerged. Today, information technology is an important part of life. Work can be facilitated by humans with information technology. Consequently, information technology can be applied in a variety of settings. Data mining is one area where technology can be applied [4].

The process of extracting or understanding a database that was not previously understood is known as data mining and is used to make significant decisions. By searching for data that contains rules, sequences, or patterns, the goal of data mining is to find knowledge in large databases [5]. The C4.5 algorithm has several advantages, the most important being that it can handle discrete and numeric attributes, produces models in the form of rules or trees that are easy to understand, and has an acceptable level of accuracy. The decision tree principle was used in the development of the ID3 algorithm, which is called the C4.5 algorithm. The decision tree method is used to classify the data using it, the Decision Tree based model resulting from learning from the training data is used in Algorithm C4.5 [6][7].

Looking at the description of the problems above, it tends to be seen that in overcoming the problems PKH beneficiaries are not concurrent in Gilih Suka Negeri Village. Therefore, a method will be developed for this research in order to classify individuals who are eligible and not eligible to receive PKH assistance. PKH beneficiaries can be right on target with the data classification of the poor.

## II. RESEARCH METHODS

The research was conducted in Gilih Suka Negeri Village, South Abung District, North Lampung Regency from February to May 2023 which was used as the study population size. Sample is a population unit taken through a certain sampling procedure. Using an arbitrary sample for research purposes when sampling in the form of 20 data (2021) and samples in the form of 35 data (2022) for testing and 120 training data. To assess whether the Recipients of the Family Hope Program (PKH) Assistance deserve

assistance, the validity of the training data will be compared with the results of the system [8].

### 2.1. Thinking Framework

Research Framework This research framework is to serve as a guide in the form of actions that will be taken by researchers to analyze this research [8].

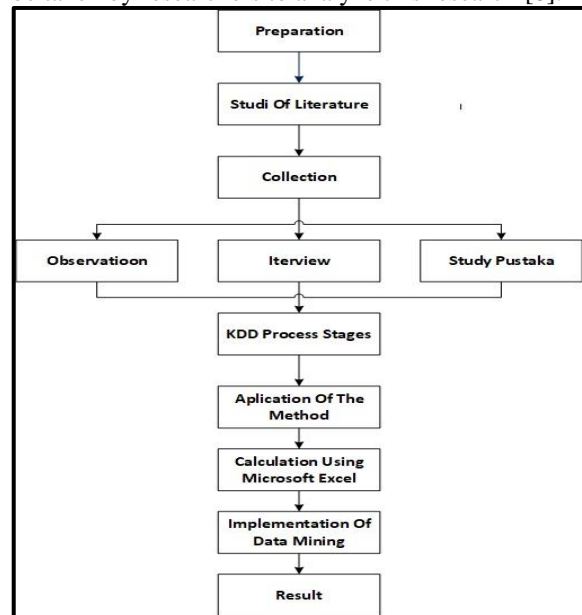


Figure 1. Thinking Framework

### 2.2. Data Mining

Data mining includes data collection and utilization of historical data to describe regularities, patterns and relationships in large data sets. Another definition of data mining is "a process that uses one or more computer learning methods to analyze and extract knowledge automatically" or "a series of processes to explore some additional value in a data set in the form of knowledge that is not yet known manually [9] and extract information from large amounts of data.

Most of the time, information mining is a logical field whose main goal is to find, investigate, or mine information from the data or information we have. Knowledge Discovery in Databases (KDD) or data mining in some circles. KDD is a method of collecting and analyzing large datasets of historical data to find regularities, patterns or relationships [10].

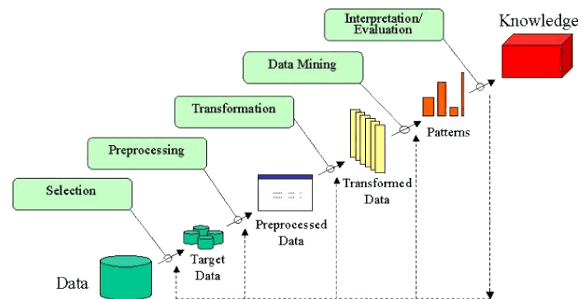


Figure 2. KDD Process Stages [11]–[13]

### 2.3. Algorithm Method C4.5

Classification problems in machine learning and data mining can be solved using the C4.5 algorithm

[14]. Calculations used for grouping information are memorized for calculations C4.5. Many of the advantages of making this approach popular The ID3-based C4.5 algorithm has been improved. Improvements made to C4.5 include, among other things, the ability to handle missing values, continuity, and truncation. How to find out the Gain and Entropy values for each criterion that contains Eligible and Ineligible information is as follows [8]. In the C4.5 algorithm, the first step is to find the entropy value. First, make sure the Entropy value for the entire case Take advantage of the following formula [8]:

$$\text{Entropy (S)} = \sum - p_i * \log_2 p_i \quad (1)$$

Information:

S : case set

A : attribute

n : number of partitions S

p<sub>i</sub> : the proportion of S<sub>i</sub> to S.

After calculating the entropy value, the gain value is calculated to identify the root of the decision tree using the formula below [8]:

$$\text{Gain (S, A)} = \text{entropy(S)} - \sum_{i=1}^n \frac{S_i}{S} * \text{entropy (S}_i) \quad (2)$$

S: case set

A: attribute

n: number of attribute partitions A

|S<sub>i</sub>|: the number of cases on the to partition - i

|S|: number of cases in S.

The following are the stages of the C4.5 Algorithm as depicted in picture 3[8] :

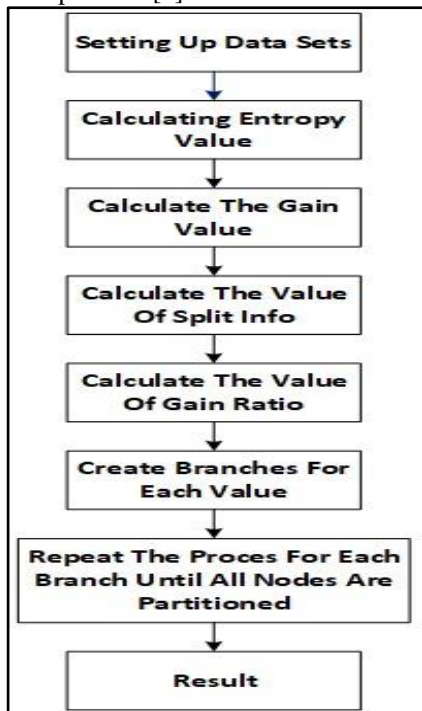


Figure 3. Algorithm Stages C4.5

#### 2.4. Decision Tree

One of the commonly used Information Mining techniques is a selection tree. A decision tree is a

structure for a flowchart that includes a tree [9]. The process of turning table data into a tree model is what a decision tree does. The tree model will create rules and rearrange them [15]. Task classification and predictions can be made with the help of Decision Trees, a predictive modeling technique. The problem search space is divided into problem sets using the "divide and conquer" method by decision trees. The stages in the calculation of making a tree (decision tree) are as follows:

1. Make attribute as root
2. For each value, create a branch
3. Partition the case into branches
4. The procedure must be repeated for each branch until all cases in that branch have the same class.

#### 2.5. Python

Python is a well-known programming language that works with multiple operating system platforms, including PC, Macintosh, and UNIX, and offers many advantages for object-oriented programming. The following are some of the benefits of the Python programming language [16]:

- a. Programming is done quickly and with less coding;
- b. Python is easy to learn;
- c. It supports multiple platforms;
- d. It has an automatic memory management system; And
- e. Python is Object Oriented Programming [17].

### III. RESULTS AND DISCUSSION

#### 3.1. Testing Data Analysis

Data testing was carried out on 20 (twenty) data testing data which will be used as a sample for entropy and gain calculations. Calculation of data testing on the 4th attribute assessment by using the average formula is:

$$A1 = \frac{\text{Number of Variable Range Values}}{\text{Many data ranges}} \quad (3)$$

Next define the Terms Eligibility, namely Eligibility Provisions = the average number of variable range values The following results are obtained as in the following table:

Table 1. Variable Average

No	Rating Attribute		Rating Value		Average
	Code	Ket	Range	Value	
1	A1	Profesi	Enterpreneur	4	2.5
			Farmer/Gardeners	3	
			Farm Workers	2	
			Housewife	1	
2	A2	Income	Low	1	1.5
			High	2	
3	A3	Home Status	Stay With Other People	1	1.5
			Own House	2	
4	A4	Total Dependents	Nothing	3	2
			a few	2	
			a lot	1	
					7.5

Table 2. Data Analysis

NO	Category	Classification	Amount
1	Not feasible	>= 7,5	12
2	Worthy	<7,5	8

From 20 testing data, it was found that 8 people were declared eligible to receive assistance and 12 people were declared unfit for it receive help.

**Table 3.** Data Testing Results

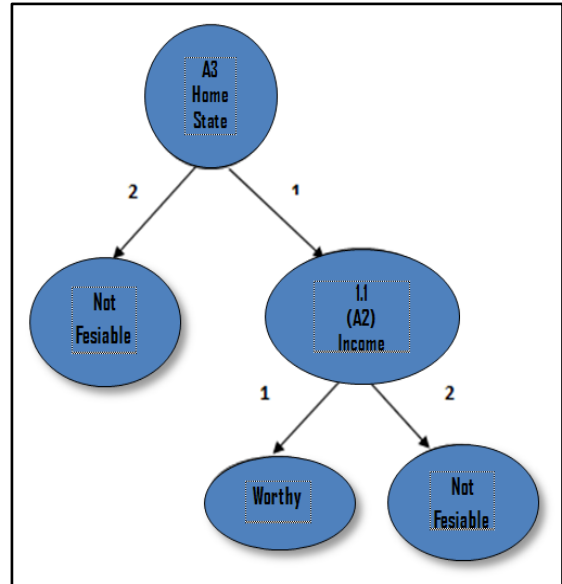
Knot	Amount	Recommendation		Entropy	Gain Information
		Worthy	No Worthy		
Total Recommendation	20	8	12	0,970950594	
A1 Profession					0
Entrepreneur	3	0	3	0	
Farmer/Gardeners	9	0	9	0	
Farm Workers	5	5	0	0	
Take care of the Household	3	3	0	0	
A2 Income					0,744484344
Low	9	8	1	0,503258335	
High	11	0	11	0	
A3 Home Status					0,753524817
Stay With Other People	8	7	1	0,543564443	
Own House	11	0	8	0	
A4 Total Dependents					0,744484344
Nothing					0,744484344
a few	9	8	1	0,503258335	
a lot	0	0	0	0	

Based on the calculation above, it shows that the highest gain is in the Home Status attribute, then the Home Status is used as the Root Node, the home status has 2 criteria, namely Own and Own House, for the home status attribute "Own Home" Has earned its class, namely "Not Eligible" while for the attribute the status of the "hung-over" house has not yet received its class, so a recalculation must be carried out.

**Table 4.** Calculation Results

Knot	Amount	Recommendation		Entropy	Gain Information
		Worthy	No Worthy		
Total Recommendation	9	8	1	0.503258335	
A1 Profession					0
Entrepreneur	0	0	0	0	
Farmer/Gardeners	1	0	1	0	
Farm Workers	5	5	0	0	
Take care of the Household	3	3	0	0	
A2 Income					0.503258335
Low	8	8	0	0	
High	1	0	1	0	
A4 Total Dependents					0
Nothing					0
a few	9	8	1	0.503258335	
a lot	0	0	0	0	

Based on the calculation table above, the decision tree is obtained as follows:



**Figure 4.** Decision Tree

**3.2. Accuracy**

The following formula is used to determine the percentage accuracy of the processed data:

$$\text{Accuracy Percentage} = \frac{\text{Number of Correct Prediction Result Data}}{\text{Number of Predictions Made}} \times 100\%$$

The accuracy of the calculation of 16 PKH beneficiary data and the presentation of the data test is as follows:

$$\text{Accuracy Percentage} = \frac{19}{20} \times 100\% = 95\%$$

**Table 5.** Confusion Table

Class	Predictions	
	Worthy	Not Feasible
Worthy	8	1
Not Feasible	0	11

Given the calculations from Table 5 above, it is very likely to be reasonable that the accuracy of the information. The test with 20 information has an accuracy of 95% and based on the decision tree above, the criteria that most influence in predicting the eligibility of PKH beneficiaries in Gilih Suka Negeri, North Lampung shows that the information gain in Criterion A3 (Home Status) is 0.753524817 greater than the other criteria.

**3.3. Data Analysis Using Python**

The author also performs predictive calculations on the eligibility data for Family Hope Program Recipients (PKH) using the Python Programming Language using data received through Ms. Excel and Algorithm C4.5. Following are the results of calculations performed using Python which can be seen in

```

+ Kode + Teks
y_pred = tree_dataset.predict(xtest)
cm = confusion_matrix(ytest, y_pred)
print("Confusion Matrix")
print(cm)
akurasi = classification_report(ytest,y_pred)
print("Tingkat Akurasi Algoritma C4.5")
print("Akurasi : ", akurasi)
akurasi = accuracy_score(ytest, y_pred)
print("Tingkat Akurasi: %d persen" % (akurasi*100))

Confusion Matrix
[[ 0  2]
 [ 0 21]]
Tingkat Akurasi Algoritma C4.5
Akurasi :
precision recall f1-score support
Layak      1.00    0.82    0.90     11
Tidak Layak 0.91    1.00    0.95     21
accuracy   0.96    0.91    0.94     32
macro avg  0.96    0.91    0.93     32
weighted avg 0.94    0.94    0.94     32
Tingkat Akurasi: 93 persen
    
```

Figure 5. Python Accuracy Results

After carrying out several stages of processing the data that has been imported into the Python programming language, the prediction results from the Python programming language which are known in the image above obtain an accuracy of 93%. The next stage is to obtain a Decision Tree to determine the eligibility requirements for beneficiaries of the Family Hope Program (PKH), which is depicted in Figure 6 below:

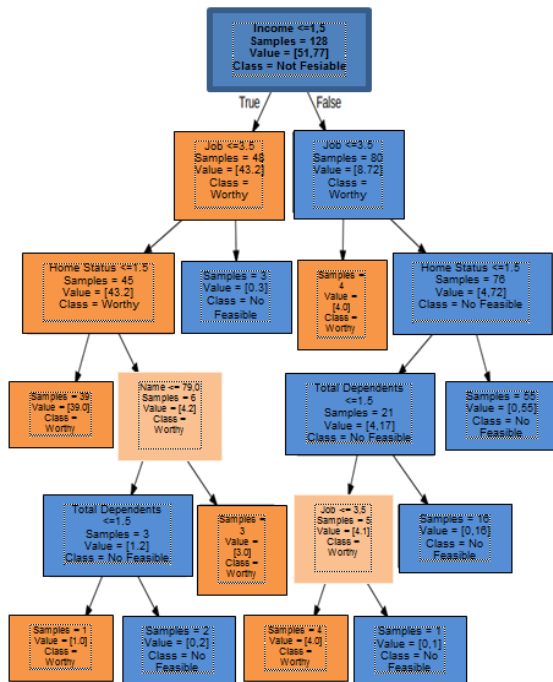


Figure 6. Decision Tree

Based on the results of data processing that has been carried out to determine the Decision Tree in the Python 120 programming language, the predicted data from the training data test, among the attributes of work, income, home status and number of dependents, it can be seen that the "income" criterion can have an impact on whether or not they are eligible to receive assistance from Family Hope Program (PKH). If using the C4.5 algorithm from Microsoft Excel, namely the "House Status" criterion, this is shown from the Gain value of 0.753524817

So if the manual method uses the Microsoft Excel Algorithm C4.5 method, the results are

obtained with an accuracy rate of 95% and the criteria for the status of the house can influence the status of being eligible or not eligible to receive PKH assistance, whereas if using the Python programming language, the results are obtained for an accuracy rate of 93% and the income criteria can affect the status of Eligible or not eligible to receive PKH assistance.

#### IV. CONCLUSION

The discussion in the previous chapters can be summarized as follows the eligibility of recipients of Family Hope Program Assistance can be predicted accurately using the C4.5 classification algorithm. It has been proven by Google Colab with an accuracy of 93%, and has an accuracy of 95% from the results of manual calculations performed in Microsoft Excel using the C4.5 Algorithm.

#### REFERENCES

- [1] S. A. Ruli Utami, Ferry Andhika Primadana, "Aplikasi Customer Relationship Management Untuk Klasifikasi Pelanggan Menggunakan Algoritma C4.5," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 3, pp. 1426–1434, 2022.
- [2] A. A. Hana Athifa Ryantika, Supriyanto, "Application of the K-Means Method for Clustering Best Selling Products in Ice Cream Sales," *J. TAM (Technology Accept. Model. Vol.*, vol. 14, no. 1, pp. 114–119, 2023.
- [3] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sist. J. Sist. Inf.*, vol. 9, no. 1, pp. 103–115, 2020.
- [4] I. M. Septia Dian Anggriani, Muhammad Syahril, "Data Mining Algoritma C4 . 5 Untuk Menganalisa Penduduk Penerima Program Keluarga Harapan ( PKH ) ( Studi Kasus : Kelurahan Tualang V )," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 1, no. 1, pp. 1–15, 2020.
- [5] A. I. Waspah *et al.*, "Expectation Maximization Algorithm Memprediksi Penjualan Susu Murni Pada Pt . Sewu Primatama Indonesia Lampung," *JUTIM (Jurnal Tek. Inform. Musirawas)*, vol. 7, no. 1, pp. 27–38, 2022.
- [6] W. Lidysari, H. S. Tambunan, and H. Qurniawan, "Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4.5 (Kasus Kantor Kelurahan Martoba)," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 3, no. 1, pp. 53–61, 2022.
- [7] K. Faradila Ilena Putri, Retno Damayanti, "Penerapan Algoritma K-Means Untuk Mengelompokan Kecamatan Di Kabupaten Gunungkidul Berdasarkan Program Keluarga Harapan," in *Prosiding Seminar Nasional*

- Matematika, Statistika, dan Aplikasinya*, 2022, pp. 408–418.
- [8] R. A. Islahudin, S. Rahmatullah, A. Afandi, and ..., "Algoritma C4. 5 Untuk Memprediksi Kelayakan Penerima Bantuan Pangan Non Tunai," *J. ...*, vol. 22, no. 02, pp. 147–159, 2022.
- [9] S. D. Anggriani, S. E. M. Syahril, and ..., "Data Mining Algoritma C4. 5 Untuk Menganalisa Penduduk Penerima Program Keluarga Harapan (PKH)(Studi Kasus: Kelurahan Tualang V)," *J. Cyber ...*, no. x, 2022.
- [10] S. Muhamad, Agus Perdana Windarto, "Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 753–760, 2019.
- [11] S. Mukodimah, M. Muslihudin, D. R. Mustofa, and D. Susianto, "Naive Bayes Classifier Method Analysis and Support Vector Machine ( SVM ) Student Graduation Prediction," *NEUROQUANTOLOGY*, vol. 20, no. 12, pp. 3522–3533, 2022.
- [12] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *Sinkron*, vol. 7, no. 2, pp. 671–678, 2022.
- [13] D. A. C, D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan RapidMiner*. 2013.
- [14] C. Algoritma, "Prediksi Kepuasan Tenan Pada Gedung Wisma Keiai Menggunakan," vol. 2, no. 2, pp. 238–243, 2020.
- [15] R. H. Pambudi and B. D. Setiawan, "Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," vol. 2, no. 7, pp. 2637–2643, 2018.
- [16] O. Kinerja, E. C. U. Study, K. Mobil, and A. Dan, "Implementasi Algoritma K-Means Dan Algoritma Apriori," vol. 1, no. 2, pp. 81–88, 2021.
- [17] V. S. Ginting, K. Kusriani, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 36–44, 2020.