

Classification Analysis Of The Eligibility Of Recipients Of Non-Cash Food Assistance And Family Hope Programs In The City Of Sukabumi Using The Naïve Bayes Classifier Algorithm

Ariski Muhammad Nazmi^{1*}, Prajoko², Agung Pambudi³

^{1,2,3} Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H. No.50 Sukabumi, Indonesia
Ariskimuhammad_ummi.ac.id¹, Prajoko-ti@ummi.ac.id², Agungpambd@ummi.ac.id³

Abstract

Providing social assistance is the government's effort to improve the welfare of the underprivileged. Non-Cash Food Assistance (BPNT) and the Family Hope Program (PKH) are two social assistance programs provided by the Indonesian government. BPNT is a food assistance program that is provided non-cash through electronic cards, while PKH is a cash social assistance program provided to poor families with certain criteria. Both programs aim to help the poor meet their food and education needs. To evaluate the effectiveness and efficiency of social assistance programs, a method is needed that can process and analyze data quickly and accurately. One method that can be used is the Naïve Bayes Classifier, which is a probabilistic classification method based on Bayes' theorem. This method can be used to classify data into certain categories based on its probability. In this study, researchers used the Naïve Bayes Classifier method to analyze social assistance data obtained from the BPNT and PKH programs. Data from the Sukabumi City Social Service was used to classify the eligibility of beneficiaries using the Naïve Bayes Classifier algorithm. Out of 5,183 data, 31.2% were classified as "Eligible" and 68.8% as "Ineligible". The algorithm showed 98.77% accuracy in eligibility classification. These results indicate the effectiveness of the Naïve Bayes Classifier algorithm in analyzing social data, providing new insights for better decision-making by relevant agencies in the development of more targeted and efficient social assistance policies.

Keywords: Social Assistance, Non-Cash Food Assistance (BPNT) and Family Hope Program (PKH), Naïve Bayes Classifier, Knowledge Discovery in Database

1. Introduction

Indonesia, as a developing country, is tackling poverty and improving societal welfare. Poverty, influenced by high unemployment due to a large population and limited job opportunities, is a significant issue. Social assistance programs like Non-Cash Food Assistance (BPNT) and the Family Hope Program (PKH) target vulnerable groups, providing necessities like food, shelter, and healthcare. The Social Services Department oversees these programs, focusing on the basic needs of individuals, families, or groups. Challenges in social assistance include managing a vast number of recipients, budget limitations, and distribution errors. BPNT provides electronic cards for food purchases to poor families, while PKH offers broader cash assistance for education and health, aiming to break the poverty cycle. Proper recipient selection is crucial for effective assistance. Classification methods, like the Naïve Bayes Classifier, analyze data to determine eligibility. Research indicates that the Naïve Bayes algorithm is highly accurate in this context, suggesting its effectiveness in social assistance programs.

2. Research Methods

Knowledge discovery methods in databases refer to techniques and processes used to extract useful, previously unknown, and potentially valuable information from large data sets. These methods are often used in data mining, machine learning, and artificial intelligence applications. Here are some common methods used for knowledge discovery in databases. The method applied in this eligibility classification research is the Knowledge Discovery in Databases (KDD) method using the Naïve Bayes Classifier algorithm, one of which is problem solving in processing the data. The stages consist of Selection, Preprocessing, Transformation, Data Mining, Evaluation, and Knowledge as shown below.

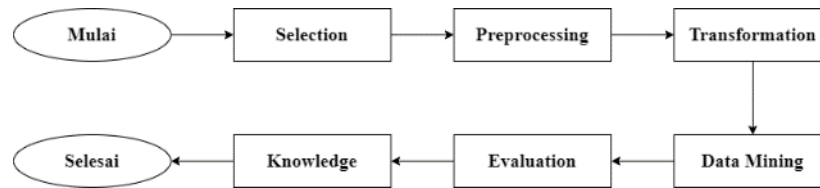


Figure 1: Research Stages of KDD Method

Naïve Bayes Classifier is a simple and probabilistic machine learning algorithm used for classification tasks. It is based on Bayes' Theorem, which relates the conditional probability of an event occurring given some evidence to the prior probability of that event. The "Naïve" part of the name comes from the assumption of independence between features, which simplifies calculations and allows efficient processing.

3. Result And Discussion

3.1. Data Selection

In collecting this data set, researchers collected information in interviews with supervisors from the Sukabumi City Social Service. The data set used in conducting research comes from one data source, namely the Sukabumi city P3KE community data in 2023. The number of original data sets of this community data is 5,183 data sets. Of all the attributes, only 12 attributes will be used in the KDD process. Here are some of the attributes in the data set, including.

ID	Indicates the code or ID of the Community
Provinsi	Show the location of the Province
Kota	Show the location of the City
Kecamatan	Shows the location of the Subdistrict
Desa/Kelurahan	Shows the location of the village
Desil Kesejahteraan	Indicates the level of household welfare
Padan Dukcapil	Demonstrate the accuracy of population data
Jenis Kelamin	Indicates Gender
Tanggal Lahir	Indicates Date of Birth
Pekerjaan	Indicates Job
Pendidikan	Indicates Education
Kepemilikan Rumah	Indicates Home Ownership
Harta	Indicates having money/jewelry/animal/other savings
Jenis Atap	Indicates the type of roof in the residence
Jenis Dinding	Indicates the type of wall in the residence
Jenis Lantai	Indicates the type of floor in the residence
Sumber Penerangan	Indicates the type of lighting source in the residence
Bahan Bakar Memasak	Indicates the type of cooking fuel at home
Sumber Air Minum	Indicates the type of drinking water source at the place of residence
Memiliki Fasilitas Buang Air Besar	Indicates the type of defecation facilities in the residence
Resiko Stunting	Showing the risk of stunting from the community
Hasil	Indicates eligibility results as a recipient

3.2. Preprocessing Data

This stage shows the results of the data preprocessing process. Where the data that has been selected from the selected data undergoes a preprocessing stage, namely eliminating irrelevant or inconsistent data, overcoming missing values, cleaning duplicate data, and deleting unnecessary data. Some attributes or data that were not used included "Provinsi", "Kota", "Kecamatan", "Desa/Kelurahan", "Desil Kesejahteraan", "Padan Dukcapil", "Jenis Kelamin", and "Tanggal Lahir". This is done because these data or attributes are not taken as a reference in this classification process. The results are as in the following table.

3.3. Transformation Data

In this step, it is the transformation stage of several types of data or attributes that have been selected, converting them into the desired form in the Input-Process-Output (IPO) model. Thus, this type of data will be suitable for the Data Mining process according to the analysis needs that the researcher examines. The Transformation stage in the Knowledge Discovery in Databases (KDD) process is where creativity emerges and depends heavily on the pattern or patterns of information to be identified in the Database.

3.4. Data Mining

In this research, applying Data Mining using the Naïve Bayes Classifier algorithm to classify data regarding the eligibility of BPNT & PKH social assistance recipients. The process includes 2 main stages, namely the training stage, where the data is trained as artificial intelligence for classification, and the Testing stage, where the data is thoroughly tested to determine the results that match the trained data. This research involves 5,183 data from the existing dataset, using the Naïve Bayes Classifier classification application program developed by the researcher, to facilitate the classification process. The following is the data that will be used in the data mining process.

3.4.1. Training Stage

From the entire existing data set, there are 32 data that have been selected as Training Data, which have been processed manually. The following is training data taken from data on people who are eligible and ineligible to receive BPNT & PKH social assistance.

Pendidikan	Kepemilikan Rumah	Pekerjaan	Jenis Atap	Jenis dinding	Harta	Jenis Lantai	BBM	Sumber penerangan	Sumber air minum	Fasilitas BAB	Risiko Stunting	status
Tinggi	Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Rendah	Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Konvensional	Non-Konvensional	Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Non-Konvensional	Risiko Sedang	Layak
Rendah	Tidak Milik Sendiri	Nonaktif	Konvensional	Non-Konvensional	Tidak	Non-Konvensional	Konvensional	Non-Konvensional	Konvensional	Non-Konvensional	Risiko Rendah	Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Konvensional	Non-Konvensional	Non-Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Milik Sendiri	Nonaktif	Konvensional	Konvensional	Ya	Konvensional	Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Risiko Rendah	Tidak Layak
Tinggi	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Non-Konvensional	Konvensional	Non-Konvensional	Risiko Sedang	Layak
Rendah	Tidak Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Ya	Non-Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Risiko Rendah	Layak
Tinggi	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Rendah	Milik Sendiri	Nonaktif	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Non-Konvensional	Konvensional	Risiko Rendah	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Non-Konvensional	Risiko Sedang	Tidak Layak
Rendah	Milik Sendiri	Nonaktif	Konvensional	Konvensional	Ya	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Non-Konvensional	Risiko Rendah	Tidak Layak
Rendah	Milik Sendiri	Mandiri	Konvensional	Konvensional	Ya	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Non-Konvensional	Risiko Sedang	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Risiko Sedang	Layak
Tinggi	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Tinggi	Milik Sendiri	Nonaktif	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Tinggi	Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Rendah	Layak
Tinggi	Tidak Milik Sendiri	Nonaktif	Konvensional	Konvensional	Ya	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Tinggi	Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Risiko Tinggi	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Risiko Tinggi	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Non-Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Tinggi	Tidak Layak
Rendah	Milik Sendiri	Mandiri	Non-Konvensional	Non-Konvensional	Ya	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Tinggi	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Ya	Konvensional	Non-Konvensional	Non-Konvensional	Konvensional	Konvensional	Risiko Tinggi	Tidak Layak
Tinggi	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Konvensional	Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Tidak Milik Sendiri	Nonaktif	Konvensional	Non-Konvensional	Ya	Non-Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Non-Konvensional	Risiko Rendah	Layak
Rendah	Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Milik Sendiri	Mandiri	Konvensional	Konvensional	Ya	Non-Konvensional	Non-Konvensional	Konvensional	Non-Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Ya	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Layak
Rendah	Milik Sendiri	Nonaktif	Konvensional	Konvensional	Tidak	Non-Konvensional	Non-Konvensional	Konvensional	Non-Konvensional	Non-Konvensional	Risiko Rendah	Layak
Rendah	Tidak Milik Sendiri	Mandiri	Non-Konvensional	Konvensional	Tidak	Konvensional	Konvensional	Konvensional	Non-Konvensional	Non-Konvensional	Risiko Sedang	Tidak Layak
Rendah	Tidak Milik Sendiri	Mandiri	Konvensional	Konvensional	Tidak	Konvensional	Non-Konvensional	Konvensional	Konvensional	Konvensional	Risiko Sedang	Tidak Layak

Fig. 2: Data Training

After transforming or changing the data in the training data, the next step is to search or calculate the Dependent Data from the Training Data to facilitate the Bayes Theorem Data Mining process in classifying the Naïve Bayes Classifier algorithm.

Table 2: Data Dependen

Pekerjaan					Pendidikan				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Mandiri	11	13	11/16	13/16	Tinggi	7	2	7/16	2/16
Nonaktif	5	3	5/16	3/16	Rendah	9	14	9/16	14/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Kepemilikan Rumah					Harta				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Milik Sendiri	8	7	8/16	7/16	Ya	4	6	4/16	6/16
Tidak Milik Sendiri	8	9	8/16	9/16	Tidak	12	10	12/16	10/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Jenis Atap					Jenis Dinding				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Konvensional	12	12	12/16	12/16	Konvensional	13	15	13/16	15/16
Non-Konvensional	4	4	4/16	4/16	Non-Konvensional	3	1	3/16	1/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Jenis Lantai					Bahan Bakar Memasak				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Konvensional	4	9	4/16	9/16	Konvensional	1	4	1/16	4/16
Non-Konvensional	11	7	11/16	7/16	Non-Konvensional	15	12	15/16	12/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Sumber Penerangan					Sumber Air Minum				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Konvensional	10	11	10/16	11/16	Konvensional	13	11	13/16	11/16
Non-Konvensional	6	5	6/16	5/16	Non-Konvensional	3	5	3/16	5/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Fasilitas Buang Air Besar					Resiko Stunting				
	L	TL	P(L)	P(TL)		L	TL	P(L)	P(TL)
Konvensional	9	11	9/16	11/16	Risiko Rendah	5	3	5/16	3/16
					Risiko Sedang	9	9	9/16	9/16
Non-Konvensional	7	5	7/16	5/16	Risiko Tinggi	2	4	2/16	4/16
Total	16	16	100%	100%	Total	16	16	100%	100%
Hasil									
Hasil								P(L)/P(TL)	
Eligible								16/32	
Ineligible								16/32	
Total								100%	

3.4.2. Testing Stage

Testing data or test data is used as an experiment to evaluate data on the eligibility of BPNT & PKH social assistance recipients in Sukabumi City. This data is taken as a sample of the training data. then, calculate the probability value by entering each sample data that has been taken through the eligibility classification application. This classification is done using the Naïve Bayes Classifier algorithm that has been built, so that the system can automatically perform the calculation. The trick is to access the test data page, then enter each data sample to be tested.

Fig. 3: Application Data Testing

After the sample data is entered, the classification results will be displayed as shown below. This calculation is based on the process previously described.

Kelas Efektif = 0.0002490367601106

Kelas Tidak Efektif = 0.00022150363071236

Jumlah Data Kelas (Efektif) Kelas (Tidak Efektif)

321616

---- Probabilitas Prior ----

Kelas (Efektif) Kelas (Tidak Efektif)

0.50.5

---- Probabilitas Data Uji ----

	Pendidikan	Kepemilikan Rumah	Pekerjaan	Jenis Atap	Jenis Dinding	Jenis Simpanan Perhiasan	Jenis Jenis Lantai	Jenis Bahan Bakar Memasak	Jenis Sumber Penerangan	Jenis Sumber Air Minum	Jenis Fasilitas Buang Air Besar	Jenis Risiko Stunting	Hasil
PC1 (Efektif)	0.44	0.5	0.69	0.75	0.81	0.75	0.75	0.06	0.63	0.81	0.56	0.56	0.0002490367601106
PC0 (Tidak Efektif)	0.13	0.44	0.81	0.75	1	0.63	0.44	0.25	0.69	0.69	0.69	0.56	0.00022150363071236

Dapat disimpulkan bahwa Data Uji tersebut mendapatkan hasil **Efektif**

Fig. 4: Result Classification Data

The system calculates based on the training data using the numbers obtained from the dependent training data, with the calculation method shown in the following table.

Table 3: Probability Calculation of Test Data

Probability Eligible	
$P(H L)$	$= P(X L).P(L)$
	$= 0.44 * 0.5 * 0.69 * 0.75 * 0.81 * 0.75 * 0.75 * 0.06 * 0.63 * 0.81 * 0.56 * 0.56$
	$= 0.0002490367601106$
Ineligible Probability	
$P(H TL)$	$= P(X TL).P(TL)$
	$= 0.13 * 0.44 * 0.81 * 0.75 * 1 * 0.63 * 0.44 * 0.25 * 0.69 * 0.69 * 0.69 * 0.56$
	$= 0.00022150363071236$

From a total of 5,183 datasets obtained from the Sukabumi City Social Service, data mining was conducted using the Naïve Bayes Classifier algorithm on each data and attribute. The final result shows that 31.2% or 1617 data are classified as "Eligible", while 68.8% or 3556 data are classified as "Ineligible". All data has gone through this Data Mining process can be found on the data attachment page.

3.5. Evaluation (Accuracy)

The process of calculating the accuracy value in the Rapidminer Tool begins with uploading the classification result data from Data Mining. This research successfully obtained an accuracy value of **98.77%**. This shows that the Naïve Bayes Classifier algorithm has a high level of accuracy in performing Data Mining classification on the Classification of Eligibility for Recipients of Non-Cash Food Assistance and Family Hope Program in Sukabumi City based on datasets from the Sukabumi City Social Service.

accuracy: 98.77% +/- 0.72% (micro average: 98.77%)

	true Layak	true Tidak Layak	class precision
pred. Layak	795	18	97.79%
pred. Tidak Layak	14	1765	99.21%
class recall	98.27%	98.99%	

Fig. 5: Accuracy Value Results

3.6. Knowledge

This research involves the application of knowledge that has been developed with the construction of a Classification Application system for Eligibility Recipients of Non-Cash Food Assistance and Family Hope Program in Sukabumi City. This application is designed using the Hypertext Preprocessing (PHP) programming language and using the MySQL Database. In addition, this application is implemented using the Naïve Bayes Classifier algorithm. The knowledge that has been collected and compiled during the research is used to evaluate and determine the results of the Non-Cash Food Social Assistance Recipients and the Family Hope Program in Sukabumi City that have been implemented by researchers.

4. Conclusion

This study uses a dataset from the Sukabumi City Social Service provided to researchers to classify the Eligibility of Non-Cash Food Assistance and Family Hope Program Recipients in Sukabumi City. This classification is based on a dataset from the Sukabumi City Social Service, using the Naïve Bayes Classifier algorithm for data mining. From a total of 5,183 data, 31.2% (1,617 data) were classified as "Eligible" and 68.8% (3,566 data) were classified as "Ineligible". The accuracy of the Naïve Bayes Classifier algorithm in classifying the Eligibility of Non-Cash Food Assistance and Family Hope Program Recipients in Sukabumi City is **98.77%**. After conducting a series of analysis and research, several conclusions can be drawn as follows.

1. The use of the Naïve Bayes Classifier algorithm in the Classification of Eligibility for Non-Cash Food Assistance Recipients and the Family Hope Program in Sukabumi city proved to be Effective. The algorithm was able to classify the data with a high accuracy rate of 98.77%, demonstrating its potential in practical applications in the social field.
2. The Data Mining process conducted provides new insights into the patterns that exist in the Non-Cash Food Assistance and Family Hope Program data, which can be used as a basis for better decision making by relevant agencies.

References

- [1] R. A. Pratama, M. I. Fasa, and S. Suharto, "Sosialisasi Penyaluran Bantuan Sosial (Bansos) Pada Era New Normal Di Desa Ciamis Kecamatan Sungkai Utara Kabupaten Lampung Utara," *Jurnal Pengabdian ...*, 2022, [Online]. Available: <https://jurnal.stie.asia.ac.id/index.php/jpm/article/view/1183>
- [2] H. Putri, A. I. Purnamasari, and ..., "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," ... and Science (BITS), 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/bits/article/view/1093>
- [3] H. N. F. Fikrillah, S. Hudawiguna, and ..., "Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes," *JATISI (Jurnal Teknik ...*, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/3624>
- [4] M. R. Firdaus, A. Latif, and W. Gata, "Klasifikasi Kelayakan Calon Pendorong Darah Menggunakan Neural Network," *Sistemasi: Jurnal Sistem ...*, 2020, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/840>
- [5] M. A. Senubekti and L. A. P. Dewi, "Prinsip Klasifikasi Dan Data Mining Dengan Algoritma C4. 5," *NUANSA INFORMATIKA*, 2022, [Online]. Available: <http://journal.uniku.ac.id/index.php/ilkom/article/view/5834>
- [6] W. Habibulloh and S. Topiq, "Klasifikasi Kelayakan Kredit Menggunakan Algoritma Naive Bayes Pada Ksp Mekar Jaya Maleber," *Jurnal Responsif: Riset Sains dan ...*, 2021, [Online]. Available: <http://ejurnal.ars.ac.id/index.php/jti/article/view/440>
- [7] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4. 5," ... *Informatika Penelitian Bidang Komputer Sains dan ...*, 2017, [Online]. Available: <https://ejournal.upgrisba.ac.id/index.php/eDikInformatika/article/view/1465>