Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Optimizing the Social Assistance Recipient Model in CangkringVillage Using the Naïve Bayes Algorithm

Rotika[1]*, Nana Suarna[2], Irfan Ali[3], Dendy Indriya Efendi[4]

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

rotika.rotika7102001@gmail.com [1]*, nana.ikmi@gmail.com [2], irfanali0.0@gmail.com [3], dendy.ikmi@gmail.com[4]

Abstract

Social assistance is one of the methods used by the government to help the underprivileged. Cangkring Village is a village in Cirebon Regency that has inaccurate data on recipients of social assistance or underprivileged people. The Naive Bayes algorithm is one of the most effective techniques in machine learning for classifying data, in determining the eligibility of recipients of social assistance. The method works with a probabilistic approach to analyze data efficiently and accurately, can group data based on attributes and produce high accuracy. The problem in Cangkring Village, namely the accuracy of data on recipients of social assistance, is still a problem that requires special attention. This inaccuracy not only reduces the effectiveness of social assistance programs but also creates injustice for people in need. Invalid and inappropriate data causes the distribution of social assistance to be suboptimal. The purpose of this study is to optimize the accuracy model of social security recipients using the Naive Bayes algorithm, which can help improve the accuracy in determining eligible recipients.The method used in the study is secondary data processing taken from social assistance recipient data in Cangkring Village. This process includes data preprocessing stages, training and testing data distribution, and implementation or application of the Naive Bayes algorithm to perform classification. The results of the study show that the Naive Bayes algorithm is able to increase the accuracy of the classification of social assistance recipients with an accuracy rate of 90%, compared to the conventional method used previously. This study contributes to providing a more efficient and targeted method in selecting social assistance recipients, so that it can improve the social assistance distribution system in the future. Thus, the Naive Bayes algorithm can be an effective method for data-based decision making in the context of social policy.

***Keywords**: Naive Bayes algorithm, social assistance, data accuracy, data classification, decision making*

1. Introduction

The Naive Bayes algorithm is a machine learning technique that is widely used in various fields, including data classification. The use of this algorithm is increasingly popular because of its ability to handle classification problems with a fairly high level of accuracy[1] In the context of social assistance in Cangkring Village, the main problem faced is the inaccuracy of targeting aid recipients, which is caused by the lack of validity and accuracy of recipient data[2]. This condition shows the need for methods that can increase the accuracy of determining aid recipients so that distribution is more targeted.[3] In this digital era, the rapid development of information technology has had a significant impact on various aspects of life, including providing solutions to social problems. One of the important developments in the field of Informatics is the use of machine learning algorithms for more accurate and efficient decision making.[4]

The application of the Naive Bayes algorithm is expected to be able to overcome the problem of classifying social assistance recipients, so that the distribution of assistance can be more optimal and right on target. The main problem in determining recipients of social assistance in Cangkring Village is the inaccuracy of target recipients, which is often caused by a lack of validity and accuracy of data.[5] In many cases, the data used to determine the eligibility of aid recipients does not reflect actual conditions on the ground, resulting in an unequal distribution of aid.[6]

In addition, the many variables that influence the eligibility of aid recipients, such as income level, number of family members, and socio-economic conditions, make the decision-making process complex.[7] Gaps in the literature indicate that despite the variety of methods that have been implemented, challenges remain in increasing the accuracy of classification of social assistance recipients.[8] This research focuses on the application of the Naïve Bayes algorithm for address this classification problem, in the hope of improvingaccuracy and effectiveness in determining social assistance in villages cangkring.[9] This research aims to optimize accuracy recipients of social assistance in Cangkring Village by implementing Naive Bayes algorithm. This research is expected to provide significant contribution in managing social data in rural areas with provides a more accurate model for determining recipients help. The results of this research can increase system efficiency distribution of aid, reducing errors in determining recipients, and provide recommendations for internal policy makers use of technology for social assistance programs. Study can provide technology-based solutions to improve justice and efficiency in the distribution of social assistance, this research can

providing practical benefits in the form of a more accurate and faster system in selecting recipients of social assistance, which will ultimately improve community welfare. This research effort increase the accuracy of classification of social assistance recipients in the Village Cangkring using the Naive Bayes algorithm. This algorithm It is hoped that it will be able to optimize the accuracy model more precisely who is worthy of receiving help, rather than method other conventional. Apart from that, research to measure the effectiveness of Naive Bayes in overcoming the problem of data imprecision
aid recipients in the region.

1.1. Naïve Bayes

The Naive Bayes algorithm is a learning method that is increasingly being used in various fields, including data classification. The use of this algorithm is increasingly popular because of its ability to handle classification problems with a fairly high level of accuracy[8] The application of the Naive Bayes algorithm is expected to be able to overcome the problem of classifying social assistance recipients, so that the distribution of assistance can be more optimal and right on target. application of the naïve Bayes algorithm to identify significant patterns in data.[6].

1.2. Knowledge Discovery Database

To analyze [i]data in the data mining process, the Knowledge Discovery in Databases (KDD) stage process is used which consists of, Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge, From the literature review or research literature study, the application of the Naïve Bayes Classifier method has the potential to determine the probability for classifying whether the recipient is classified as "eligible" or "not eligible". The following are the steps for applying the Naïve Bayes Classifier method [2]

2. Research Method

2.1. Research Method

The methodology employed in this research is Knowledge Discovery in Databases (KDD). The workflow or sequence of steps utilized throughout the study is depicted in Figure 1 below, outlining the systematic approach taken for data analysis and interpretation.

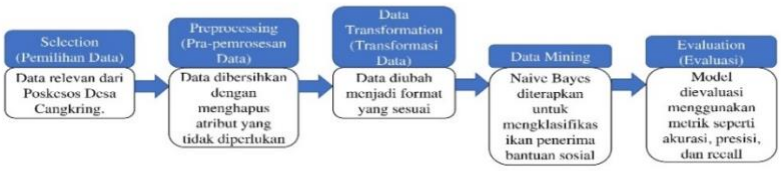


Figure 1: Research Method

3. Result and Discussion

3.1. Selection

At this stage, data relevant to the research was selected from the main data source, namely the Cangkring Village Social Health Post. The data taken must include information that is relevant to analyzing the eligibility of social assistance recipients, such as information on marital status, type of work, and other data. Is the selection of data that will be used to become an attribute. Below is the Data Selection process.

Table 1: Selection Data

NO	ATRIBUT	TIPE ATRIBUT	KETERANGAN
1.	NIK	Real	Nomor Induk Kependudukan sebagai ID unik
2.	PROGRAM BANSOS	Polynominal	Jenis program bantuan sosial yang diterima.
3.	NAMA	Polynominal	Nama penerima bantuan.
4.	TEMPAT LAHIR	Polynominal	Tempat kelahiran penerima.
5.	JENIS KELAMIN	Binominal	jenis kelamin: Laki-laki atau Perempuan.
6.	JENIS PEKERJAAN	Polynominal	Jenis pekerjaan penerima.
7.	STATUS KAWIN	Polynominal	Status pernikahan: Menikah, Belum Menikah, dll.
8.	ALAMAT	Polynominal	Alamat lengkap penerima.
9.	RT	Integer	Nomor RT tempat tinggal.
10.	RW	Integer	Nomor RW tempat tinggal.
11.	STATUS DISABILITAS	Polynominal	Apakah penerima memiliki disabilitas atau tidak.
12.	STATUS RUMAH	Polynominal	Status kepemilikan rumah: Milik sendiri, sewa, dll.

13.	JUMLAH TANGGUNGAN	Integer	Jumlah orang yang menjadi tanggungan dalam keluarga
14.	PENGHASILAN	Real	Total penghasilan penerima per bulan.
15.	USIA	Integer	Usia penerima dalam tahun.
16.	JUMLAH KELUARGA	Integer	Total anggota keluarga dalam satu rumah
17.	KONDISI LANTAI	Polynominal	Kategori kondisi lantai rumah: Tanah, Keramik, dll.
18.	KONDISI DINGDING	Polynominal	Kategori kondisi dinding rumah: Kayu, Tembok, dll.
19.	STATUS PENERIMA	Binominal	Apakah penerima layak (1) atau tidak layak (0) bantuan.

.

3.2. Data Cleaning or prepocessing

After the data collection stage, the next stage will be Data Cleaning or data cleaning so that there is no duplication of data, checking for inconsistent data and correcting errors in the data such as printing errors, so that the data can be processed and entered into the data mining process. After all the required data has gone through the data cleaning stage. Next, the data will be processed into training data and testing data which are ready to be entered into the software used in this research. Below is the data cleaning process after the data selection process.

Table 2: Data Cleaning

NO	NIK	NOKK	NAMA	TEMPAT LAHIR	IBU KANDUNG	JENIS KELAMIN	JENIS PEKERJAAN	…	……	RT	RW	STATUS DISABILITAS
1	3209363105080003	3209361905090029	MUHAMAD RAYHAN	CIREBON	ROSE SANTY	L	PELAJAR/MAHASISWA	…	…	006	002	TIDAK
2	3209360204860003	3209361306110004	YOGI SUTRISNA	CIREBON	KURIAH	L	WIRASWASTA	…	……	006	002	TIDAK
3	3209362116600004	3209361205230003	TALO	CIREBON	SARWI	L	WIRASWASTA	…	….	011	004	TIDAK
4	3209366702220001	3209360404120003	PRISKA PRIYANKA	CIREBON	SISKA WATI	P	BELUM/TIDAK BEKERJA	…	…	009	003	TIDAK
5	3209365410210002	3209360611150001	ELVINA NINDI LIVIANI	CIREBON	HENI NURANI	P	BELUM/TIDAK BEKERJA	…	…	005	002	TIDAK
6	3203011304040012	3209361301200001	DEDEN APRIANSYAH	CIREBON	DIAH KARYANINGSIH	L	PELAJAR/MAHASISWA	…	……	005	002	TIDAK
7	3203012701100010	3209361301200001	ALFAN FIRMANSYAH	CIREBON	DIAH KARYANINGSIH	L	PELAJAR/MAHASISWA	…	…	005	002	TIDAK
	.....	.....	.....	...	.....	...	.....	...	...	...	...	...
524	3209364808710001	3209360210070071	SA'ANAH	CIREBON	PATONA	P	MENGURUS RUMAH TANGGA	…	…	007	003	TIDAK

3.3. Transformation

Data Transformation is the stage of processing data into a form suitable for processing in data mining. In this research, the data will be processed from Microsoft Excel which will be used for processing in Rapidminner software, such as recipient status which will be converted into labels. For the data type, some attributes are said to be polynomial because they have more than 2 choices, such as the type of work which contains students, self-employed, not yet or not working, private employees, taking care of the household, agricultural or plantation workers, casual daily workers, carpenters and traders, except for the gender attribute, and the recipient status is of the Binomial type because it only has 2 options in the record. For example, the recipient status attribute only has 2 options, namely Eligible to Receive and Not Eligible to Receive. The transformation data can be seen in the table 3.

Table 3: Transformation Data

NO	NIK	PROGRAM BANSOS	JENIS PEKERJAAN	STATUS KAWIN	STATUS DISABILITAS	STATUS RUMAH	JUMLAH TANGGUNGAN	PENGHASILAN				STATUS PENERIMA
1	3209363105080003	NONBANSOS	PELAJAR/MAHASISWA	BELUM KAWIN	TIDAK	MILIK SENDIRI	5	500.000	…	…	…	LAYAK MENERIMA
2	3209360204860003	NONBANSOS	WIRASWASTA	KAWIN	TIDAK	KONTRAK	6	3000.000	…	…	…	TIDAK LAYAK MENERIMA

3	3209362211660004	NONBANSOS	WIRASWASTA	KAWIN	TIDAK	MILIK SENDIRI	2	2500.000	…	…	…	LAYAK MENERIMA
4	3209366702220001	NONBANSOS	BELUM/TIDAK BEKERJA	BELUM KAWIN	TIDAK	KONTRAK	4	100.000	…	…	…	TIDAK LAYAK MENERIMA
5	3209365410210002	NONBANSOS	BELUM/TIDAK BEKERJA	BELUM KAWIN	TIDAK	MILIK SENDIRI	1	100.000	…	…	…	LAYAK MENERIMA
7	3203012701100010	NONBANSOS	PELAJAR/MAHASISWA	BELUM KAWIN	TIDAK	MILIK SENDIRI	5	50.000	…	…	…	LAYAK MENERIMA
8	3209360109820004	NONBANSOS	KARYAWAN SWASTA	KAWIN	TIDAK	KONTRAK	6	3500.000	…	…	…	TIDAK LAYAK MENERIMA
10	3209362203130001	NONBANSOS	PELAJAR/MAHASISWA	BELUM KAWIN	TIDAK	KONTRAK	4	1500.000	. .	…	…	TIDAK LAYAK MENERIMA
…	….	..	…….	….	…...	……	…..	……	…	…	. .	…….
524	3209362005930002	NONBANSOS	KARYAWAN SWASTA	BELUM KAWIN	TIDAK	KONTRAK	4	100.000	…	…	…	TIDAK LAYAK MENERIMA

3.4. Data Mining or rapidminer software testing

In this stage, the research focuses on model development using the Naïve Bayes method. The implementation involves detailed steps, including determining optimal parameters for the Naïve Bayes, such as the node selection criteria and the maximum tree depth. Below is the application of the data mining process performance on Rapidminer using the Naïve Bayes method.

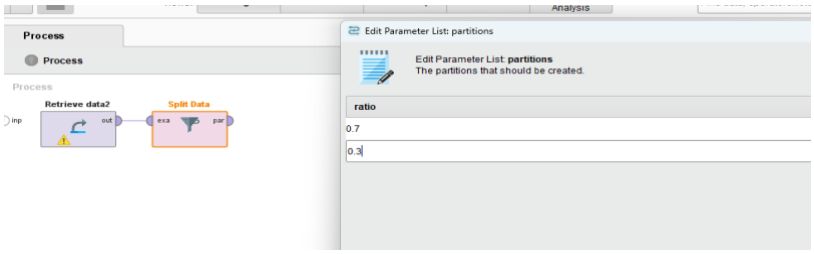


Figure 2: Split the data into training and testing sets

In the next stage, the Split Data operator is an operator to divide Training Data and Testing Data, in the parameter section Traning Data is test data or training data, here the training data is 70% and Testing Data is data that is not for training, the testing data is 30% so training data and testing data total 100%. Below is the Split Data process. Below is the Split Data operator.

in the image 3 are the operators that will be used for the classification process. The next stage, Entering the Naïve Bayes process, Apply model, and performance will be connected to each other.

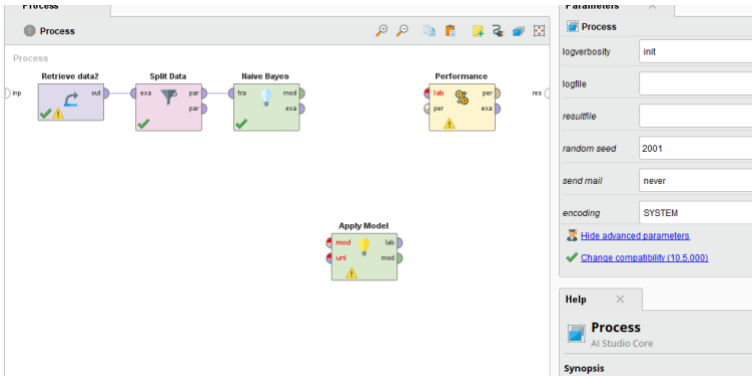


Figure 3: Data & Model Insertion Process

Figure 3 explains the Drag&Drop parameters of Naïve Bayes. In the parameter section, there is no choice because it is automatic, where the algorithm used to carry out the training is the Naïve Bayes method. (Classification) in this parameter section is automatically clicked, namely accuary only.

Next stage, Enter the Naïve Bayes process, Apply model, and performance which will later be connected to each other. Then connect the training data with the Naïve Bayes model, then connect to the Apply Model which functions to run the classification model, from the apply model the mod is connected to the result to find out the naïve Bayes modeling, then it is also connected to performance (classification) which functions to create and obtain accuracy and performance values. One thing that needs to be explained is that testing data does not need to be connected to the Naïve Bayes model because the data is not data that needs to be trained but data that needs to be tested. Below is the process of connecting all operators.

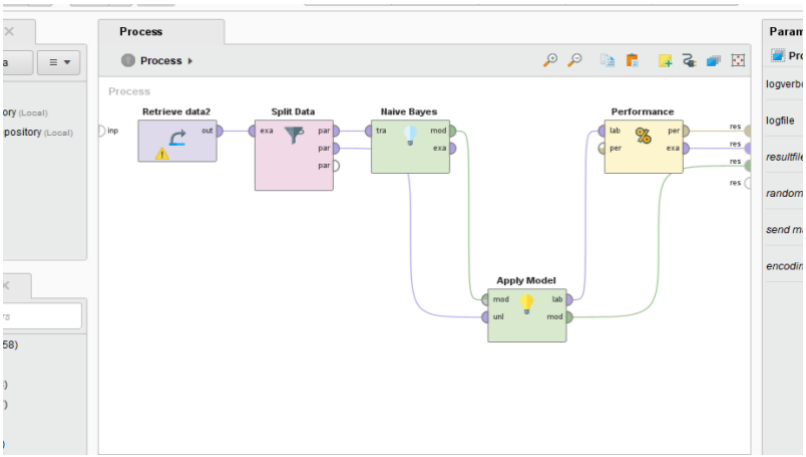
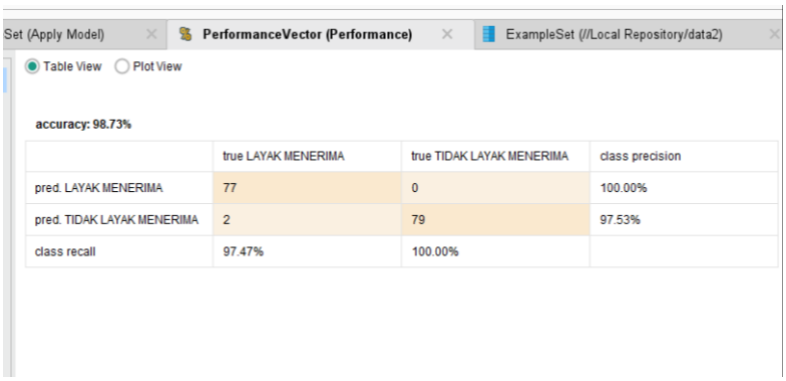


Figure 4: Process of Linking Data and Models



	true LAYAK MENERIMA	true TIDAK LAYAK MENERIMA	class precision
pred. LAYAK MENERIMA	77	0	100.00%
pred. TIDAK LAYAK MENERIMA	2	79	97.53%
class recall	97.47%	100.00%	

Figure 5: View Performance Results (classification

The results showed that there were 2 names of recipients who were said to be worthy of receiving it, but after using the Naïve Bayes method, the names of the recipients were said to be not worthy of receiving it and worthy of receiving this social assistance program. On the other hand, the Class Recall in the data displays a true result capable of around 97.73% and true is not worthy of receiving 100%, which means that the recall of the data after the data has been carried out by the Naïve Bayes process goes well and it is said that the data is data that has good quality so produces a high level of accuracy in helping determine recipients of the Social Assistance Program.

3.5. Evaluation

PerformanceVector

```
PerformanceVector:
accuracy: 98.73%
ConfusionMatrix:
True:   LAYAK MENERIMA  TIDAK LAYAK MENERIMA
LAYAK MENERIMA: 77      0
TIDAK LAYAK MENERIMA:   2       79
```

Figure 6: View Performance Vector

The explanation in Figure 6 shows the Simple Distribution results of the Naive Bayes model on the Recipient Status label attribute in a classification process. The Acceptable Class (0.500) indicates that the Acceptable class has a prior probability of 0.5 or 50%. This means that before considering other features in the data, this class is considered to have a 50% probability based on the initial distribution in the dataset. Class Not Eligible to Accept (0.500) Likewise, the class Not Eligible to Accept also has an initial probability of 0.5 or 50%. 18 distributions indicates that there are 18 attributes or features in the dataset which are used to develop probability distributions for each class (Eligible to Accept and Not Eligible). Each feature will have a distribution that helps the model determine the probability of whether a piece of data falls into one of the two classes.
After the result run, we can visualize the simple distribution (Naïve Bayes) and can be seen below.

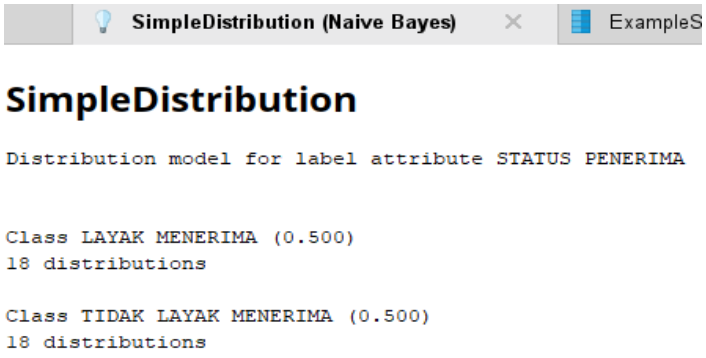


Figure 7: Naïve Bayes

4. Conclusion

In this research, the Naïve Bayes algorithm was implemented to form a classification model, which aims to optimize the Naïve Bayes classification model for social assistance recipients. Based on the results of research carried out with the Naïve Bayes algorithm using Rapidminer or Ai Studio software, the model succeeded in achieving an accuracy of 98.73% showing excellent ability to classify data on social assistance recipients with a high level of precision and recall in both classes, namely worthy of receiving. and it is not feasible to accept the results that the naïve Bayes model can recognize patterns in data attributes accurately enough to determine the appropriateness of social assistance recipients. From several previous literature studies, research [8]. The results of this research support several previous literature which shows that the Naive Bayes algorithm is effective for classifying social assistance eligibility, as well as research conducted[10] which found that socio-economic attributes such as type of work and living conditions play an important role in classifying recipient eligibility. social assistance, research conducted[9]which shows that the Naive Bayes algorithm is effective in capturing patterns of socio-economic variables to predict eligibility for assistance, research conducted[11] shows that marital status is not significant in the classification of aid eligibility, whereas in this study, marital status (especially married and unmarried status) seems to play a role in increasing the chances of receiving aid. This difference could be caused by differences in context or additional variables used in the research dataset. Thus, this research is expected to make a significant contribution in improving the optimization of the accuracy model for social assistance recipients.

References

[1] N. Attamami, A. Triayudi, and R. T. Aldisa, "Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 2, pp. 262–269, 2023.

[2] D. Utami and P. A. R. Devi, "Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Metode Weighted Naïve Bayes Dengan Laplace Smoothing," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1373–1384, 2022.

[3] D. Kurniadi, F. Nuraeni, and M. Firmansyah, "Klasifikasi Masyarakat Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Naïve Bayes dan SMOTE," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 309–320, 2023.

[4] J. Homepage, D. Pramudita, Y. Akbar, and T. Wahyudi, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of the Indonesian Smart College Card Program on Social Media X Using the Naive Bayes Algorithm Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah Pada Med," *Malcom* , vol. 4, no. October, pp. 1420–1430, 2024.

[5] N. Rosanti, "Jurnal Teknologi Terpadu PENERAPAN MODEL MACHINE LEARNING UNTUK MENENTUKAN," vol. 8, no. 2, pp. 127–135, 2022.

[6] M. M. Alfitri, N. Nurahman, M. Minarni, and D. Rusda, "Evaluasi Performa Algoritma Naïve Bayes Dalam Mengklasifikasi Penerima Bantuan Pangan Non Tunai," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1433, 2023.

[7] N. Huda, M. Hasbi, and T. Susyanto, "Seleksi Penerima Bantuan Pangan Non Tunai di Desa Menggunakan Metode Naïve Bayes dan Simple Additive Weighting," *J. Ilm. SINUS*, vol. 19, no. 1, p. 39, 2021.

[8] N. Riyanah and F. Fatmawati, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 206–213, 2021.

[9] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020.

[10] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 219, 2021.

[11] H. Harliana and F. N. Putra, "Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 165–173, 2021.