Inside | Jurnal Sistem Informatika Cerdas

Volume 1 Nomor 2 | Desember 2023 | halaman 62-66

Application of Data Mining to Classify Receiving Social Assistance Using the Naïve Bayes Method

Abdul Wahid[1*], Fauzan Azim[2], M. Faiz Firdausi [3]
Institut Teknologi dan Sains Mandala, Jember, Indonesia[1,2,3]

Corresponding Author:
Abdul Wahid, Institut Teknologi dan Sains Mandala, Grujugan Kidul Kec Grujugan, Bondowoso and 68261, Indonesia
Email: abdulwahid@itsm.ac.id

Abstract

The application of Naïve Bayes in classifying potential recipients of Non-Cash Food Assistance (BPNT) in Grujugan Kidul Village was made to solve problems in the process of distributing non-cash food assistance programs that had not run optimally and had not been on target. The purpose of this study is to find out how the Naïve Bayes Algorithm is applied in classifying Non-Cash Food Assistance Recipients in Grujugan Kidul Village so as to obtain optimal results. The method used in this study is the Naïve Bayes classification method. The Naïve Bayes algorithm is proven to have good performance in predictions, and produces high accuracy values. The stages of data analysis were carried out based on the CRISP-DM method while the algorithm testing was carried out on the RapidMiner 9.10.001 software as a comparison between manual calculations and software calculations. The results of the accuracy in this study were 89.19% from 8:2 Comparison of 80% test data, 20% test data.

Keywords : Naïve Bayes Algorithm, Acceptance of Social Assistance, RapidMiner.

1 INTRODUCTION

Poverty is a condition of being economically incapable of meeting the homogeneous standard of living of the people in an area. This condition of incapacity is characterized by a low level of income ability. someone in meeting their basic needs both clothing, food and boards. In terms of the right to food, the state has an obligation to meet the people's need for affordable and adequate food and nutrition. As stated in Law Number 13 of 2011 concerning Handling the Poor and Government Regulation Number 63 of 2013 concerning Implementation of Efforts to Handle the Poor Through a Regional Approach (State Gazette of the Republic of Indonesia of 2013 Number 157). To overcome this poverty by referring to this regulation, the government has implemented a poverty assistance program for the population, one of which is in the aspect of fulfilling food, namely Non-Cash Food Assistance (BPNT).

Grujugan Kidul Village is a village administration located in the Grujugan sub-district which has 6 RWs and 26 RTs with a total population of 6,053 as of 2021. Based on the data tabulation, it was identified that in Grujugan Kidul Village the number of residents who had a livelihood was 59.84%. Of these, their lives depend on the agricultural sector, there are 43.49% of the total population

BPNT is one of the government assistance programs in the form of groceries in Nanjung Mekar Village to fulfill food needs for beneficiary families (KPM). However, in its distribution, the assistance was considered not optimal because the proposals for potential beneficiaries were submitted based on direct recommendations from each RW where there were frequent complaints from a large portion of the community because they did not receive assistance, while some communities who were considered capable actually received the assistance..

The problems that occur in the process of receiving non-cash food assistance in Grujugan Kidul Village include: submitting prospective recipients who are recommended directly by each RT without prior selection so that the non-cash food assistance program has not run optimally and has not been on target, as well as identification of recipient eligibility the assistance did not meet the criteria for beneficiaries, because it did not go through a document selection process.

The Naïve Bayes algorithm is one of the algorithms that is widely used in the world of machine learning for classification cases (Annur, 2018). The purpose of this study is to find out how the application of the Naïve Bayes Algorithm is used in classifying prospective BPNT social assistance recipients in Grujugan Kidul

Village in order to obtain optimal and targeted results and to obtain results of identification of the eligibility of recipients of assistance according to the criteria for beneficiaries.

2 RESEARCH METHOD

The method used in this study is the Naive Bayes Algorithm. Naïve Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset [1]. The Naïve Bayes algorithm is a classification algorithm based on the Bayesian theorem in statistics [2]. The Naive Bayes method is a method for generating parameter estimates by combining information from samples and other previously available information (Nugraha & Rahayu, n.d.). The Naïve Bayes algorithm can be used to predict the probability of membership in a class [3].

Naive Bayes is based on the simplifying assumption that attribute values are conditionally independent when given output values. In other words, given the output value, the probability of observing together is the product of the individual probabilities [4]. The advantage of using Naive Bayes is that this method only requires a small amount of training data to determine the parameter estimates needed in the classification process [5]. Naive Bayes often works much better in most complex real world situations than expected [6].

In Bayes' theorem, the probability or conditional probability is stated as :

$$\mathbf{P(H|X)} = \frac{\text{p (X |H) . P (H)}}{\text{P (X)}}$$

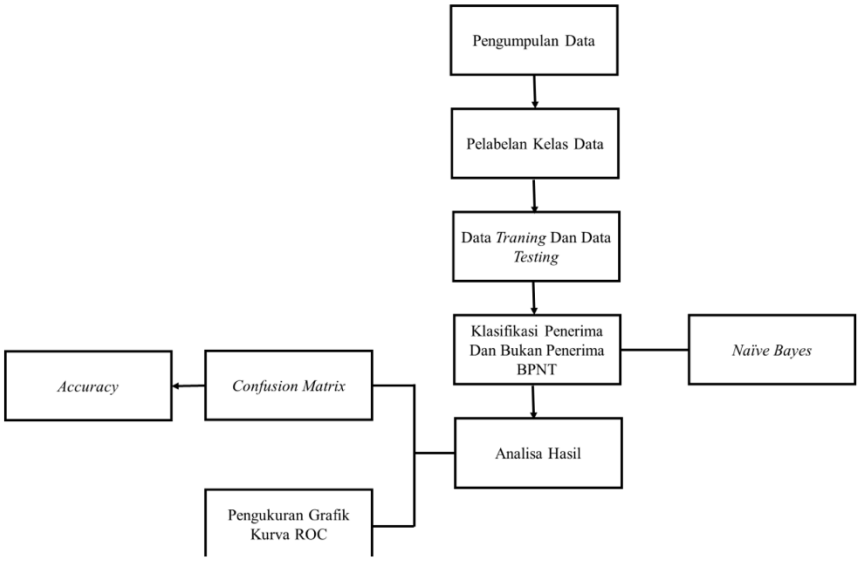
Then it can be simplified into the following formula:

$$\mathbf{P(H|X) = P(X|H)\ P(X)}$$

Dimana:

X	: Data dengan class yang belum diketahui
H	: The data hypothesis X is a class specific.
P(H|X)	: The probability of hypothesis H based on condition X (posteriori probability)
P(H)	: The probability hypothesis H (prior probability)
P(X|H)	: The probability of X is based on the conditions in the H hypothesis
P(X)	: The probability of X

The flow of the Naive Bayes method can be seen in Figure 1 as follows:



Gambar 1 *Naïve Bayes Method Workflow*

.

Based on Figure 1, it can be explained as follows:

1. Data collection was carried out by conducting interviews with the Kaur or village operator of Grujugan Kidul, namely by asking for data so that it could be re-analyzed.
2. After getting the data then labeling the data for classification. The labels used are "Accepting" and "Not Accepting".
3. Training data and testing data are determined using a purposive sampling technique.
4. The classification used uses the Naïve Bayes algorithm by calculating the probability on the training data.

After doing the classification then analyzing the data using the confusion matrix technique to get the level of accuracy of the classification results. The type of data used is the type of secondary data obtained from related agencies through reports, books and others related to research issues [7]. Then do the selection of variables according to research needs. The software used in this research is Rapidminer version 9.10.001 and Microsoft Excel 2016.

The data collection method used in this study is as follows:

1. Observation
   This method is carried out by coming directly to Grujugan Kidul Village and then looking for the desired data and observing the research object so that you can get the things needed according to research needs.
2. Interview (Interview)
   In an effort to get an explanation of the problems that exist, the researchers visited Grujugan Kidul Village and spoke directly with the relevant employees.
3. Library Studies (Library Research)
   This research uses literature study by using data collection techniques such as searching, reading and studying from various sources, literature (books) or documents related to the procedure for implementing the distribution of non-cash food assistance (BPNT) which will later be used as a reference..

3 RESULTS AND ANALYSIS

3.1 Data Collectioon

The total data of beneficiaries and non-recipients of BPNT assistance in the village of Grujugan Kidul in 2023 obtained in this study is 250 data. The criteria used are 5 of them: Income (X1), Occupation (X2), Number of Children (X3), Status of DTKS (X4).

3.2 Naïve Bayes Calculations

Classification is done by calculating the prior probability categories of Accepting and Not Accepting and the prior probabilities of all criteria based on their respective categories in the training data. Probability results on training data are then used to calculate probability on data testing [4].

3.3 Data Training

The purposive sampling technique obtained the amount of data receiving and not receiving BPNT assistance in Grujugan Kidul village in 2023 as training data of 185 data. Comparison of 80% test data, 20% test data.

To determine the data to be analyzed using the Naive Bayes method, the first step is to read the training data. The training data used can be seen in the table 1

Table 1 *Data Training*

No	NAMA	NIK	PENDAPATAN	PEKERJAAN	JUMLAH ANAK	STATUS DTKS
1	AHMAD BYRON PRAMATA	3511062806090001	1.000.000 - 1.500.000	Petani	Dua	Tidak
2	SULASTRI	3511064505730001	1.500.000 - 2.000.000	PNS	Satu	Ya
3	ULFA SAFITRI	3511065011040001	500.000 - 1.000..000	Buruh Tani	Tiga	Ya

4	YANTO	3511061711830001	500.000 - 1.000..000	Ibu Rumah Tangga	Satu	Ya
5	AKIB	3511061505650011	1.000.000 - 1.500.000	Buruh Tani	Satu	Ya
6	MUZAIYANAH	3511065609710001	500.000 - 1.000..000	Ibu Rumah Tangga	Satu	Tidak
7	NUR KHOLISAH	3511066408990001	1.500.000 - 2.000.000	Wiraswasta	Satu	Tidak
8	SARAH EVELYNA	3511066705060002	> 2.000.000	Wiraswasta	Tiga	Tidak
9	SUHARTONO	3511061201730001	> 2.000.000	PNS	Tiga	Tidak
10	TEGAR	3511061102040002	1.000.000 - 1.500.000	Wiraswasta	Tiga	Tidak
11	UMYANA	3511064207780001	0 - 500.000	Buruh Tani	Satu	Tidak
12	AHMAD ARDIS ABDUL MUGHNI	3511060909190002	1.000.000 - 1.500.000	Petani	Tiga	Ya
13	MOHAMMAD MAHMUD	3511062912940003	1.500.000 - 2.000.000	Wiraswasta	Dua	Ya
14	YULIATUL HASANAH	3511067012950001	1.000.000 - 1.500.000	Buruh Tani	Dua	Tidak
15	FATHORROZI	3511060108920001	1.500.000 - 2.000.000	PNS	Satu	Tidak
16	NURUL KOMARIYAH	3511156110030001	1.000.000 - 1.500.000	Petani	Dua	Tidak
17	WITA ZERLINA	3511064412190001	500.000 - 1.000..000	Ibu Rumah Tangga	Satu	Tidak
18	ASMINI	3511064505650003	1.000.000 - 1.500.000	Ibu Rumah Tangga	Tiga	Tidak
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
181	ROHAMA	3511065510690001	500.000 - 1.000..000	Buruh Tani	Tidak	Tidak
182	SATIP	3511060107600006	1.000.000 - 1.500.000	Petani	Ya	Ya
183	EKA NUR INTAN	3511066803030002	1.500.000 - 2.000.000	Wiraswasta	Ya	Ya
184	HANINA	3511065207600006	1.000.000 - 1.500.000	Petani	Tidak	Tidak
185	INDRA DWI KURNIAWAN	3511062803100001	> 2.000.000	PNS	Ya	Ya

3.4 Data Analysis

The results of the classification of recipient and non-BPNT assistance data in Nanjung Mekar Village for the 2021 period with a total of 10 data testing data, 8 data are classified correctly while 2 data are classified incorrectly. To test the level of accuracy of the data from the results of the calcification in this study using the confusion matrix and RappidMiner techniques.

Calculation of accuracy is done by using the technique of confusion matrix. RappidMiner is used to compare the accuracy results obtained in the confusion matrix. The level of accuracy obtained is as much as 80%. With the results of the classification calculation, there are 8 data that are classified correctly and 2 data that are classified incorrectly. To get the accuracy value is done in a way where the amount of data that is correctly classified in data testing is divided by the entire data testing. The formula for obtaining the accuracy value can be seen in Figure 2.

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN}$$
$$= \frac{6+2}{6+2+0+2} = 0{,}8 * 100 \sim 80\%$$

.

Accuracy : 89.19%

	True Tidak	True Ya	Class precision
Pred. Tidak	16	2	88.89%
Pred. Ya	2	17	89.47%
Class recall	88.99%	89.47%	

Tabel 2 *Image Accuracy in RapidMiner Tools*

The results of the accuracy in this study were 89.19% of 8:2 Comparison of 80% test data, 20% test data.

4 CONCLUSION

Based on the results and discussion, it can be concluded that according to the results of performance vector testing, the accuracy value obtained in this study was 89.19% from 8:2 Comparison of 80% test data, 20% test data. With the level of accuracy obtained in this study, it is included in the Medium category with an excellent classification. This algorithm model can optimize the distribution of non-cash food assistance so that it is right on target and runs optimally. The application of the naïve Bayes algorithm classification can provide new understanding for Grujugan Kidul Village officials who will provide eligibility for non-cash food assistance (BPNT) for those who are entitled to receive assistance according to predetermined criteria. By carrying out the document selection process, assistance can be provided with the right target.

As for the suggestions for Grujugan Kidul Village, it is expected that for proposing prospective beneficiaries/candidate KPM in the next period so that the assistance program can run optimally and is right on target and in accordance with the criteria for beneficiaries, strategic steps are implemented by setting quotas for prospective beneficiary families (KPM) in each RT with the same amount so that the proposed KPM candidates are even. Each KPM candidate submitted by the RT is required to fill out a beneficiary eligibility questionnaire, as well as the KPM candidates submitted by each RT, are verified and re-validated by the village. To facilitate verification and validation, it can be carried out according to the formulas/calculations in this study or using the help of Rapidminer analysis software. A document selection team was also formed at the village level to facilitate eligibility for prospective beneficiary families (KPM) so that they were right on target.

REFERENCES

[1] M. F. Nugraha and S. B. Rahayu, "Penerapan Naïve Bayes dalam Mengklasifikasi Calon Penerima Bantuan Pangan Non Tunai di Desa Nanjung Mekar".

[2] M. Hermansyah, "ANALISIS SENTIMEN TWITTER UNTUK MENGETAHUI KESAN MASYARAKAT TENTANG PELAKSANAAN POMPROV JAWA TIMUR TAHUN 2022 DENGAN PERBANDINGAN METODE NAÏVE BAYES CLASSIFIER DAN DECISION TREE BERBASIS SMOTE," *J. Inform. Dan Tekonologi Komput. JITEK*, vol. 2, no. 3, pp. 249–255, Nov. 2022, doi: 10.55606/jitek.v2i3.551.

[3] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifer Dengan Seleksi Fitur Dan Boosting," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 3, no. 2, pp. 227–232, Aug. 2019, doi: 10.29207/resti.v3i2.1042.

[4] J. Sulaksono, R. H. Irawan, and I. N. Fahmi, "Penerapan Metode Naive bayes Terhadap Bantuan Sosial Keluarga PraSejahtera," vol. 3, no. 2.

[5] A. Muliawan, T. Badriyah, and I. Syarif, "Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula," *Technomedia J.*, vol. 7, no. 2, pp. 231–247, Sep. 2022, doi: 10.33050/tmj.v7i2.1893.

[6] M. F. Nugraha and M. Furqon, "Perancangan Data Warehouse Sistem Pendaftaran Mahasiswa Menggunakan Online Analytical Procesing (OLAP) di Universitas Ma'soem," *Intern. Inf. Syst. J.*, vol. 4, no. 1, pp. 35–40, Jun. 2021, doi: 10.32627/internal.v4i1.283.

[7] D. A. Fauziah, A. Maududie, and I. Nuritha, "Klasifikasi Berita Politik Menggunakan Algoritma K-nearst Neighbor," *Berk. SAINSTEK*, vol. 6, no. 2, p. 106, Dec. 2018, doi: 10.19184/bst.v6i2.9256.