FEASIBILITY TEST OF POOR RICE RECIPIENTS IN BENCOY SUKABUMI VILLAGE USING NAIVE BAYES

Taufik Hidayatulloh 1; Ardi Winardi 2; Lestari Yusuf 3*; Satia Suhada 4; Saeful Bahri 5

Sistem Informasi Kampus Kota Sukabumi 1
Universitas Bina Sarana Informatika
www.bsi.ac.id 12
taufik.tho@bsi.ac.id 1; ardiwinardiv@gmail.com 2
Sistem Informasi 3,4,5
Universitas Nusa Mandiri 3,4,5
www.nusamandiri.ac.id 3,4,5
lestari.lyf@nusamandiri.ac.id 3; satia.sqh@nusamandiri.ac.id 4; saeful.sel@nusamandiri.ac.id 5
(*) corresponting author

Abstract—A regional head must have a work plan every regional head must have a work plan which is sure to be of benefit to the community. Assisting is a definite work plan in every region. A lot of assistance is usually given from the government to the community and must be managed by the village government so that the aid gets to the right hands. And to improve food security, the people in each region have activities to distribute Poor Rice as a subsidy from the government. In the distribution method, sometimes there are constraints in data collection so that poor rice or what we usually call Raskin is not suitable for distribution. Because of this, a way is needed so that the distribution is appropriate or not in the community in accepting the Raskin so that government assistance can be delivered properly and on target. By using secondary data obtained from Bencoy Village, 205 data were obtained containing the attributes of the eligibility category of Raskin recipients, and 6 categories of attributes were found with the classification method of the Naïve Bayes algorithm. The accuracy value obtained is 96.59%, proving that the prediction using the Naive Bayes algorithm has a good performance. The next results obtained are in the form of AUC value which after being calculated results in an application which is an implementation with a flow that is adjusted to the calculation algorithm in the form of a web-based application.

Keywords: Raskin, Naïve Bayes, Data Mining

INTRODUCTION

Government assistance is a program that can help improve the welfare of the community, one of which is Poor Rice (Raskin), whose distribution is also a social assistance program (Tone, 2016). Problems that often occur in the distribution of poor rice assistance usually still use estimates only and there is no prediction at the time of distribution of the distribution of poor rice recipients. So that few or many people sometimes protest because the people...
who should have received assistance did not receive assistance (Suryeni et al., 2015).

Bencoy Village is one of the Five Villages in the Cireunghas District, and its address is on Jalan Raya Cireunghas KM 15, Sukabumi Regency. Bencoy Village has an area of 831,750 Km², in general, the topography of Bencoy Village also has 4 hamlets, 10 Rw and 54 Rt. The total population of Bencoy Village consists of 446 heads of households, 1,816 Poor households include: 4,110 people Male + female: 4,029 people = 8,139 people.

The distribution of rice distribution activities for the poor in Bencoy Village is which devoted to poor or insufficient families is certainly very far from what was expected because poor rice which was originally devoted to poor families, now well-off and high-income families also receive. Some poor families do not receive assistance, while families who can afford rice assistance are poor. The major obstacle encountered was the distribution of Raskin due to the mistargeting of Raskin recipients (Nasir, 2019).

One thing that can used in predicting a receipt of assistance is by using data mining. As it is well known that mining data is the decomposition of database findings using mathematical techniques, Intelligent artificial statistics, machine learning to extract information (Kaesman, 2016). To make decisions with a large enough amount of data, data mining is a technique used to get new patterns (Wulandari et al., 2020). While the data mining method used in this study is the Naïve Bayes algorithm. The naïve Bayes algorithm assumes a simple estimated probability classification method for calculating a set of probabilities by summing the frequencies and combinations of data (Hidayat et al., 2017).

Not only that but the assessment of the distribution of Raskin was also implemented using the top-down approach (Nisak, 2014). Then the previous researchers had also applied the product weighted method to determine recipients of Raskin assistance (Firdyana et al., 2017). Even in previous studies, the calculation of the distribution of Raskin rice was carried out using the transportation method (Simbolon et al., 2014).

The feasibility of Raskin recipients using the Naïve Bayes algorithm has previously been used to get an accuracy of 35.6% for a decent class and 36.32% for an unfeasible class (Fadlan et al., 2018). In previous research, Raskin recipients were determined by a decision-making system using the C.45 algorithm with the results concluding that the largest entropy for Raskin recipients was from the Farmer category (Ermawati & Hidayatulloh, 2016).

Datasets

By looking at previous research, the observation data used the Naïve Bayes algorithm, and then the test data were documented and with an accuracy of 75% (Waliyansyah & Fitriyah, 2019). With 1335 data, the accuracy rate of Naive Bayes even outperformed K-NN with an accuracy value of 83.83% while the accuracy of K-NN was 82.34% (Maricar & Dian Pramana, 2019).

So this study aims to use data mining classification to calculate the accuracy of the Naive Bayes algorithm in determining the eligibility of Raskin recipients and implementing it into a program.

MATERIALS AND METHODS

To solve the problem in this study, the following methods will used by the author, methods will describe in figure 1:

Datasets

By looking at previous research, the observation data used the Naïve Bayes algorithm, and then the test data were documented and with an accuracy of 75% (Waliyansyah & Fitriyah, 2019). With 1335 data, the accuracy rate of Naive Bayes even outperformed K-NN with an accuracy value of 83.83% while the accuracy of K-NN was 82.34% (Maricar & Dian Pramana, 2019).

So this study aims to use data mining classification to calculate the accuracy of the Naive Bayes algorithm in determining the eligibility of Raskin recipients and implementing it into a program.

MATERIALS AND METHODS

To solve the problem in this study, the following methods will used by the author, methods will describe in figure 1:

Datasets

By looking at previous research, the observation data used the Naïve Bayes algorithm, and then the test data were documented and with an accuracy of 75% (Waliyansyah & Fitriyah, 2019). With 1335 data, the accuracy rate of Naive Bayes even outperformed K-NN with an accuracy value of 83.83% while the accuracy of K-NN was 82.34% (Maricar & Dian Pramana, 2019).

So this study aims to use data mining classification to calculate the accuracy of the Naive Bayes algorithm in determining the eligibility of Raskin recipients and implementing it into a program.
categorized (Sugiharti et al., 2017). The steps to obtain the dataset this time are to analyze the poor rice distribution system, then conduct interviews with staff who entrusted with managing the poor rice distribution data. Next, with a population belonging to the Bencoy Village of 205 residents, the sample data dataset obtained as described in the following table 1:

<table>
<thead>
<tr>
<th>Raskin Reception Year</th>
<th>Feasible</th>
<th>Not feasible</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>116</td>
<td>89</td>
<td>205</td>
</tr>
</tbody>
</table>

Source: (Winardi, 2021)

Data analysis was carried out using data on the feasibility of poor rice which has 2 classes that are feasible and unfit. The data processing uses data mining techniques by applying the naïve Bayes algorithm and tested using the cross-validation method the evaluation of the an algorithm testing can to seen using the Confusion Matrix method to produce accuracy value while the ROC (Receiver Operating Characteristics) curve is used to measure. This research consists of six stages based on the CRISP-DM experimental model, along with the explanation:

**Bussines Understanding**
Based on the observations made. From this data, there is feasibility data with 6 attributes as predictor attributes and 1 result attribute. This research was conducted to determine the community that can receive assistance for receiving rice for the poor (Raskin). So it is necessary to have a study that applies a method to predict eligibility of acceptance which will make it easier for Bencoy Village officers to provide poor rice assistance.

**Data Understanding**
The data understanding stage is carried out in determining the purpose of a study, at this stage the process of collecting, analyzing, and evaluating the feasibility data is carried out such as understanding each predictor attribute, these attributes become parameters that can be seen in table 2.

**RESULT AND DISCUSSION**
The results of this study aim to determine the accuracy value of the naïve Bayes algorithm in the eligibility of receiving Raskin by processing sample data. The results of this study are in the form of a

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of House Floor</td>
<td>Board</td>
</tr>
<tr>
<td></td>
<td>Billik Bambu</td>
</tr>
<tr>
<td></td>
<td>Cement</td>
</tr>
<tr>
<td>Home Lighting</td>
<td>Personal electricity</td>
</tr>
<tr>
<td></td>
<td>Connect to other</td>
</tr>
<tr>
<td>Head of Family's work</td>
<td>Seller</td>
</tr>
<tr>
<td></td>
<td>Labor</td>
</tr>
<tr>
<td></td>
<td>PNS</td>
</tr>
<tr>
<td></td>
<td>Retired</td>
</tr>
<tr>
<td></td>
<td>Driver</td>
</tr>
<tr>
<td></td>
<td>Doctor</td>
</tr>
<tr>
<td></td>
<td>Farmer</td>
</tr>
<tr>
<td></td>
<td>Unemployment</td>
</tr>
<tr>
<td>Income</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Ownership of Asset</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

Source: (Winardi, 2021).

**Preparation**
At the Data Preparation stage, feasibility data processing is carried out, where the data is processed first with the help of sort & filter aims to simplify the process of calculating the probability later.

**Modeling**
This modeling stage implements the naïve Bayes algorithm, for processing the dataset, then using the 10 fold cross-validation method on the testing data to produce an evaluation model that is measured by the Accuracy and ROC Curve values.

**Evaluation**
Evaluation and validation use the cross-validation method as a testing method, confusion matrix as an evaluation of this research which will produce Accuracy, where this Accuracy will be a measure of the success of a method because the higher the accuracy value the better the method used. The method used this time is the naïve Bayes method. Meanwhile, for measuring the performance of the Naïve Bayes algorithm, the ROC curve is used.

**Deployment**
This deployment stage carries out a process of making a report on the results of the research that has been carried out and will implemented on a website.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of House Floor</td>
<td>Board</td>
</tr>
<tr>
<td></td>
<td>Billik Bambu</td>
</tr>
<tr>
<td></td>
<td>Ceramic</td>
</tr>
<tr>
<td></td>
<td>Cement</td>
</tr>
</tbody>
</table>
calculation process based on the naïve Bayes model and calculations using cross-validation, then the results of these calculations are implemented on a website.

Dataset

Before entering the Naïve Bayes algorithm process, we must first sort (sort & filter) the data, the amount of data collected is 205 records. From this data, we get 6 criteria, including the type of house walls, the lighting used, the work of the head of the household, the amount of income, and asset ownership.

Naïve Bayes Algorithm

After the data is sorted, the next step is to calculate the available data into Naïve Bayes. The results of the calculation of the probability value above can be made of the Naïve Bayes shown below:

\[
P(X|Ci) : P(X|result = feasible) \quad (1) \\
P(X|Ci) : P(X|result = unfeasible) \quad (2)
\]

And from the results of the total calculation that has been done, it can be seen that the P-value \(X | \) Eligible is 5.28148E-05, while the P-value \(X | \) Unfeasible is 0.00025338. This means that the value of \(P(X | \) Eligible) is smaller than the value of \(P(X | \) Not Feasible), so it can be concluded that for this case it will be classified as an Inadequate class.

Evaluation Model with Confusion Matrix

The results of the trials carried out are to produce the accuracy value and the AUC (Area Under Curve) value. The confusion matrix model will form a matrix consisting of true positive, true negative, false positive, and false negative. Based on the image above, there are details of the number of True Positives (TP) 83, False Negative (FN) 1, False Positive (FP) 6, and True Negative (TN) 115. These data can be calculated for the value of accuracy, precision, sensitivity, specificity, PPV. And NPV. The test results of the testing data using the Naïve Bayes algorithm produce the AUC value in the ROC graph which can be seen in the image below:

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Parameter</th>
<th>Feasible</th>
<th>unfeasible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of house floor</td>
<td>Board</td>
<td>0.0603444828</td>
<td>0.033707865</td>
</tr>
<tr>
<td>Type of house wall</td>
<td>Board</td>
<td>0.172413793</td>
<td>0.08988764</td>
</tr>
<tr>
<td>Home Lighting</td>
<td>Personal</td>
<td>0.8103444828</td>
<td>0.921348315</td>
</tr>
<tr>
<td>Head of Family's work</td>
<td>Seller</td>
<td>0.206896552</td>
<td>0.404494382</td>
</tr>
<tr>
<td>Income</td>
<td>Middle</td>
<td>0.155172414</td>
<td>0.516853933</td>
</tr>
<tr>
<td>Ownership of Asset</td>
<td>Yes</td>
<td>0.344827586</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: (Winardi, 2021)
application view of the calculations that have been done.

Source: (Winardi, 2021)
Figure 4. Implementation Model from Naïve Bayes Algorithm

Figure 4 is an implementation of the calculation using the naïve Bayes algorithm. This program is made using PHP language. This is the first testing to input the category by the attribute.

Source: (Winardi, 2021)
Figure 5. Implementation Model from Naïve Bayes Algorithm

Figure 5 shows the results obtained from testing on the application. Based on this test, it can be seen that the predictions generated by the application are the same as the data result’s probability.

Source: (Winardi, 2021)
Figure 6. Implementation Model from Algorithm

This application in figure 6, which was made to check the eligibility of receiving poor rice assistance, will be able to help Bencoy Village employees determine the eligibility of receiving poor rice assistance.

CONCLUSION

Based on the results of research using the Naïve Bayes algorithm, in predicting the feasibility of receiving poor rice with 205 datasets of testing data, an accuracy value of 96.59% is obtained, where the accuracy is obtained from the calculation of true positive values of 83 records, true negative of 115 records, false positive for 6 records, and negative for 1 record. While the resulting AUC value is 0.999 and when the curve number is rounded it will be 100%, then this accuracy has a very good classification level. And with the creation of an application that adapts to calculations that are adjusted to the probability, it is hoped that it will greatly facilitate the Bencoy village officers when they will assist. And last This study uses a dataset of 205 data and 6 predictor attribute variables and 1 result class attribute, it is recommended for further research to add to the predictor attribute variables and the addition of more datasets so that the results seen from the measurement of accuracy and the roc curve can be better. It is also hoped that further research can be developed again by comparing this naïve Bayes method with other data.

REFERENCE


