COMPARISON OF NAIVE BAYES ALGORITHM AND C.45 ALGORITHM IN CLASSIFICATION OF POOR COMMUNITIES RECEIVING NON CASH FOOD ASSISTANCE IN WANASARI VILLAGE KARAWANG REGENCY

Yuris Alkhalifi[1]; Ainun Zumarniansyah[2]; Rian Ardianto[3]; Nila Hardi[4]; Annisa Elfina Augustia[5]

[12345] Ilmu Komputer
[12345] STMIK Nusa Mandiri Jakarta
[12345] www.nusamandiri.ac.id
14002360@nusamandiri.ac.id[1]; 14002362@nusamandiri.ac.id[2]; 14002391@nusamandiri.ac.id[3];
14002358@nusamandiri.ac.id[4]; 14002344@nusamandiri.ac.id[5]

Abstract— *Non-Cash Food Assistance or Bantuan Pangan Non Tunai (BPNT) is food assistance from the government given to the Beneficiary Family (KPM) every month through an electronic account mechanism that is used only to buy food at the Electronic Shop Mutual Assistance Joint Business Group Hope Family Program (e-Warong KUBE PKH) or food traders working with Bank Himbara. In its distribution, BPNT still has problems that occur that are experienced by the village apparatus especially the apparatus of Desa Wanasari on making decisions, which ones are worthy of receiving (poor) and not worthy of receiving (not poor). So one way that helps in making decisions can be done through the concept of data mining. In this study a comparison of 2 algorithms will be carried out namely Naive Bayes Classifier and Decision Tree C.45. The total sample used is as much as 200 head of household data which will then be divided into 2 parts into validation techniques is 90% training data and 10% test data of the total sample used then the proposed model is made in the RapidMiner application and then evaluated using the Confusion Matrix table to find out the highest level of accuracy from 2 of these methods. The results in this classification indicate that the level of accuracy in the Naive Bayes Classifier method is 98.89% and for the accuracy level in the Decision Tree C.45 method is 95.00%. Then the conclusion that in this study the algorithm with the highest level of accuracy is the Naive Bayes Classifier algorithm method with a difference in the accuracy rate of 3.89%.*

Keyword: Data Mining, BPNT, Naive Bayes, C.45

Intisari—Bantuan Pangan Non Tunai (BPNT) adalah bantuan pangan dari pemerintah yang diberikan kepada Keluarga Penerima Manfaat (KPM) setiap bulannya melalui Mekanisme akun Elektronik yang digunakan hanya untuk membeli pangan di Warung Elektronik Gotong Royong Kelompok Usaha Bersama Program Keluarga Harapan (e-Warong KUBE PKH) atau pedagang bahan pangan yang bekerja sama dengan Bank Himbara. Dalam pendistribusiannya, BPNT masih terdapat masalah yang terjadi yang dialami oleh aparatur desa khususnya aparatur Desa Wanasari dalam pengambilan keputusan yakni mana yang layak menerima (miskin) dan tidak layak menerima (tidak miskin). Maka salah satu cara yang membantu dalam pengambilan keputusannya dapat dilakukan melalui konsep *data mining*. Pada penelitian ini akan dilakukan perbandingan dari 2 algoritma yakni *Naive Bayes Classifier* dan *Decision Tree C.45*. Total sampel yang digunakan adalah sebanyak 200 data kepala keluarga yang kemudian akan dibagi menjadi 2 bagian kedalam teknik validasi yakni data latih sebanyak 90% dan data uji sebanyak 10% dari total sampel yang dipakai kemudian dibuatkan model usulannya dalam aplikasi RapidMiner dan akan dievaluasi menggunakan tabel *Confusion Matrix* untuk mengetahui tingkat akurasinya tertinggi dari 2 metode tersebut. Hasil dalam klasifikasi ini menunjukkan bahwa tingkat akurasi pada metode *Naive Bayes Classifier adalah* sebesar 98,89% dan untuk tingkat akurasi pada metode *Decision Tree C.45* adalah sebesar 95,00%. Maka dihasilkan kesimpulan bahwa dalam penelitian ini algoritma dengan tingkat akurasi tertinggi adalah metode algoritma *Naive Bayes Classifier* dengan perbedaan tingkat akurasi sebesar 3,89%.

Kata Kunci: *Data Mining*, BPNT, Naive Bayes, C.45

INTRODUCTION

The problem of Poverty has now become a very fundamental problem and is the center of attention by the state government both in rural and urban areas (Septiyana Firdyana, Dedy Cahyadi, 2017). The causes of poverty vary, among others, due to economic, environmental and existing policies. The population in Indonesia will increase every year. In 2018 the population in Indonesia will reach 265 million (Katadata Indonesia, 2018). And in 2019 the population in Indonesia will increase until it reaches 267 Million (Katadata Indonesia, 2019). There is a very fast comparison

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Naive Bayes...*
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018**

of increasing numbers in the population every year. However, the increase in the population was not accompanied by economic growth in Indonesia due to the large number of Human Resources so that the availability of jobs was limited.

Supporting aspects to overcome poverty include the government having population data in various cities or villages so that in assisting will be right on target. One of the improvements in the welfare of the poor is increasing human resource employment and carrying out programs on assistance that meet the conditions and criteria set by the government. However, in receiving the assistance provided so far it has not been right on target so many people who should have received assistance did not even get their rights, namely receiving assistance.

To maintain stability, the government now has a program plan to help people who have a below-average income. One of them is a government program that helps the community in Non-Cash Food Assistance (BPNT), which is now a development of food aid Literature Subsidies (Kementerian Sosial Republik Indonesia, 2017).

Non-Cash Food Assistance (BPNT) is food assistance from the government given to the Beneficiary Families or Keluarga Penerima Manfaat (KPM) every month through an electronic account mechanism that is used only to buy food at the Gotong Royong Electronic Stall Joint Business Group Joint Family Hope Program (e-Warong KUBE PKH) or food traders working with Himbara Bank

Previous research to determine the acceptance of non-cash food aid was carried out by one method, the Decision Tree C4.5 algorithm, which is one of the most effective decision tree algorithms for classification. In this study using the optimization method namely Particle Swarm Optimization (PSO). PSO is one of the evolutionary computational techniques that is based on the search for an algorithm that starts with a random population called particle (Zuhair et al., 2019). Then this study uses 1109 data of household heads consisting of 292 poor family heads and 817 non-poor family heads. Testing in this study uses the RapidMiner application and Graphical User Interface (GUI) to test the selected algorithm rule is PHP. Based on the results of the discussion, an accuracy rate of 98.59% was obtained without using the PSO optimization method and an accuracy rate of 98.92% was obtained by using the PSO optimization method. (Ermawati, 2019).

As for other studies in determining the classification of eligibility of people receiving Non-Cash Food Assistance using the Naive Bayes Classifier algorithm method. Method This method is to predict the probability of class membership.

This research was conducted in the Main Village, Cimahi using data from 626 heads of households. Based on the results of the discussion using Cross Validation validation and produces an accuracy rate of 58.29%(Sugianto, Maulana, & Mining, 2019).

The purpose of this study is to compare the classification algorithm that produces the most precise accuracy which will be used to create a system that can predict and determine the eligibility of non-cash food aid recipients (BPNT). This research hopes to be able to assist local governments in determining BPNT recipients.

MATERIALS AND METHODS

1.   Materials

The research material in the form of a dataset in this study was taken using a secondary data source approach that is sourced from the head of the Neighborhood Association or Rukun Tetangga (RT) in Wanasari Village. Also, to get the existing problems, we need information from experts. Then in this study conducted interviews with the Government Section or Bagian Pemerintahan, Public Welfare Sector or Kesejahteraan Masyarakat (Kesra) and the Central Statistics Agency or Badan Pusat Statistik (BPS) Karawang Regency. The criteria used are referring to the criteria of the Central Statistics Agency (BPS) about the criteria of the poor which totaled 14 criteria including the area of the house, the floor of the house, the wall of the house, toilet facilities, the main lighting source, source of drinking water, cooking fuel, eating meat/fish/eggs for a week, buying clothes at least once a year, eating frequency, being able to go to the health center, income for work, education and assets. These criteria will then be the variables in the study. Then the samples taken were 200 records which were the data of the Head of the Family in Wanasari Village. The data will be processed and analyzed to provide results from the two methods comparing the Naïve Bayes algorithm and the C45 Algorithm.

2.   Methods
a.   Research Method Models

In a study, there are models of research methods. The research methodology model provides an overview of the set of steps that will be carried out from the beginning of the study to the end of the study. The research method model in this study was made as a flowchart of this study which will then be processed to facilitate the research. The research method model can be seen in Figure 1 below:
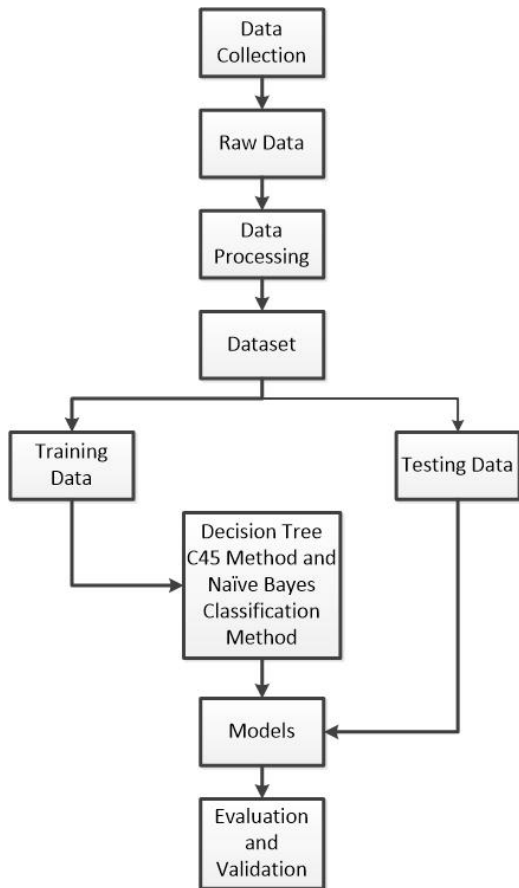
Source: (Alkhalifi et al., 2020)

Figure 1 Research Method Models

b.   Data Processing

After the research model is created, the next step is processing data from the research to be conducted. From the dataset that has been obtained will then be processed using a supporting application that uses the RapidMiner 9.1 application in calculating the results of the method to get the Confusion Matrix table and measuring the accuracy of the two methods which will then be a comparison of the methods in this study.

c.   Proposed Model

The model in this study uses the concept of data mining. Where data mining is a process of finding meaningful relationships, patterns, and trends by examining in a large set of data stored in storage by using pattern recognition techniques such as statistical and mathematical techniques (Husin & Mulyaningsih, 2015), Other opinions suggest that Data Mining is a process of analyzing data from different perspectives and summarizing it into important information that can be used to increase profits, reduce costs, or even both (Yunita, 2018)

The proposed method for predicting the classification of the poor people receiving Non-

Cash Food Assistance (BPNT) in the Village of Wanasari uses the Naive Bayes Classifier and Desicion Tree C.45 methods which will then be compared for accuracy to get the best results.

1.   Naive Bayes Classifier Method

The Naive Bayes Classifier method or commonly abbreviated as NBC is a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. (Saleh, 2015). The stages of the Naive Bayes method are as follows (Bustami, 2014):
a.   Read training data
b.   Calculate Amount and probability, but if numeric data then:
    1.   Find the mean and standard deviation of each parameter which is a numeric data.
    2.   Find the probability value by calculating the appropriate amount of data from the same category divided by the amount of data in that category.
c.   Get values in the mean, standard deviation and probability tables.

The model proposed in the Naive Bayes Classifier algorithm method in the RapidMiner application can be seen in Figure 2 as follows:
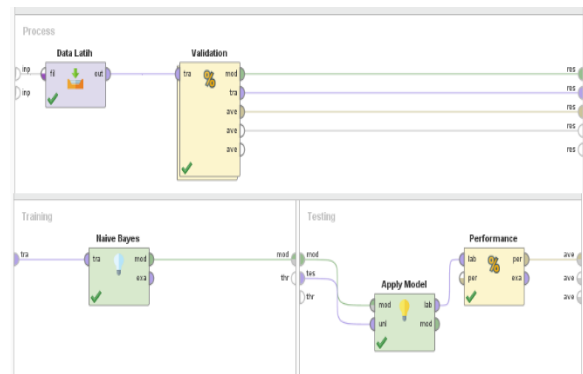
Source: (Alkhalifi et al., 2020)

Figure 2 Proposed Model Naïve Bayes Classifier Method

2.   *Decision Tree C.45 Method*

Decision Tree C.45 method is C4.5 algorithm and decision tree are two inseparable models, because to build a decision tree C4.5 algorithm is needed. The concept of the C4.5 algorithm is to transform data into decision trees and decision rules. C4.5 is an algorithm that is suitable for classification and data mining problems, C4.5 maps attribute values into classes that can be applied to new classifications. (Ermawati, 2019). The stages of the Decision Tree C.45 method are as follows (Apandi, Maulana, Piarna, & Vernanda, 2019):

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Naive Bayes…*
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018**

a. Set up *data training*.
b. Determine the roots of the tree.
c. Calculate the *Gain.*

To choose the attribute as the root, based on the highest Gain value of the existing attributes. To calculate the gain can use the following formula:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entrophy(Si) \dots (1)$$

Explanation :
S : The Set
A : Attribute
n : Number of Attribute Partition *A*
| Si | : Number of Cases on The Partition to-i
| S | : Number of Cases in *S*

d. Repeat the second step until each branch is fulfilled. Meanwhile, for calculating Entropy values, you can use the following formula:
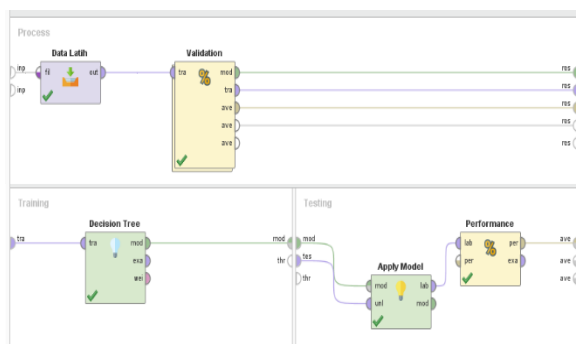
$$Entrophy(S) = \sum_{i=1}^{n} - \text{Log}^2 \text{pi} \dots (2)$$

Explanation :
S : Case Set
A : Features
n : Number of Cases s
pi : Proportion of *Si* to *S*

e. The decision tree participation process will stop when all branches in node N get the same class.

The model proposed in the Decision Tree C.45 algorithm method in the RapidMiner application can be seen in Figure 3 as follows :

Source: (Alkhalifi et al., 2020)
Figure 3. The Proposed Model Decision Tree C.45 Method

RESULTS AND DISCUSSION

A. Test Result

From the total sample used in this study, there are 200 head of household data which can then be called 200 datasets, then the dataset is divided into 2 parts namely training data (training data) and training data (training data). The amount

of data that is divided into training data and test data is 180 training data (90%) and 20 test data (10%) of the dataset. Data will be tested using the RapidMiner application.

1. Test Result of *Naive Bayes Classifier*

Based on the dataset tested using the RapidMiner support application for the Naive Bayes Classifier algorithm method, an accuracy rate of 98.89% was obtained. The results of calculations using the RapidMiner application such as table 1 below :

Table 1. Naive Bayes Classifier Calculation Results

Accuracy : 98,89%			
	true Poor	*true* Not Poor	*Class Prediction*
Prediction Poor	69	0	100,00%
Prediction Not Poor	2	109	98,20%
Class Recall	97,19%	100%	

Source: (Alkhalifi et al., 2020)

It is known from table 1 above, the number of True Positive (TP) is 69 records classified as POOR and False Negative (FN) as many as 2 records classified as POOR but NOT POOR. Next 109 records for True Negative (TN) are classified as NOT POOR, and 0 False Positive (FP) records classified as POOR turned out POOR. Based on Figure 4 also shows that the level of accuracy using the Naive Bayes Classifier is 98.89%.

2. Decision Tree Test Results C.45

Then for the dataset tested using the RapidMiner support application for the Decision Tree C.45 algorithm method, an accuracy rate of 95% was obtained. The results of calculations using the RapidMiner application as in table 2 below :

Table 2. Decision Tree Calculation Results C.45

| Accuracy : 95,00%			
	true Poor	*true* Not Poor	Class Prediction
Prediction Poor	62	0	100,00%
Prediction Not Poor	9	109	92,37%
Class Recall	87,32%	100%	

Source: (Alkhalifi et al., 2020)

It is known from table 2 above, the number of True Positive (TP) is 62 records classified as POOR and False Negative (FN) as many as 9

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Naive Bayes...*
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018**

records classified as POOR but NOT POOR. Next 109 records for True Negative (TN) are classified as NOT POOR, and 0 False Positive (FP) records classified as POOR turned out POOR. Based on Figure 5 also shows that, the level of accuracy using the Naive Bayes Classifier is 95.00%.

B.   Comparison of Test Results
Based on the entire data tested by the two methods above namely Naive Bayes Classifier and Decision Tree C.45 in the prediction of the classification of the poor recipients of Non-Cash Food Aid. The comparison of the 2 methods can be seen in Figure 6 below:
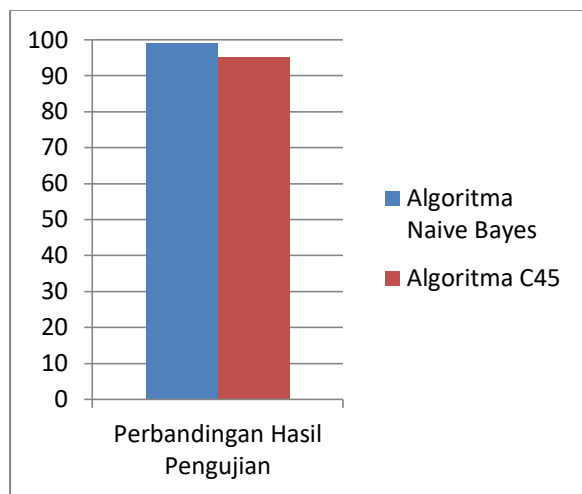
Source: (Alkhalifi et al., 2020)
Figure 4 Comparisons of Test Results

From Figure 4 above shows that testing between the 2 algorithm methods tested, the method that has the greatest accuracy is the Naive Bayes Classifier algorithm which has a higher accuracy rate of 3.89% than the Decision Tree C.45 algorithm.

CONCLUSION

Based on the results of the research that has been carried out, it can be concluded that the sample data of 200 households is then divided into 2 parts namely training data as much as 90% and as much as 10% test data resulting in an accuracy rate on the Naive Bayes Classifier algorithm having an accuracy rate of 98.89% and the accuracy of the Decision Tree C.45 algorithm produces an accuracy rate of 95.00%. From the 2 accuracy levels it is known to have a difference of 3.89%. In other words, in this research the algorithm method that has the highest accuracy rate in the prediction of the classification of the poor recipients of Non-Cash Food Assistance (BPNT) in Wanasari Village is the Naive Bayes Classifier algorithm in the hope that

the comparison of these 2 algorithms can be a solution for village apparatus, especially the Village apparatus Wanasari in making the classification decision of the poor recipient of Non-Cash Food Assistance (BPNT) in Wanasari Village.
As for suggestions for the continuation of this study is expected in subsequent studies, more records are used so that the comparison of the accuracy can be better. It is also hoped that further research can be developed with different methods or develop this research with optimization methods from compared algorithms such as Particle Swarm Optimization or other optimization methods..

REFERENCES

Alkhalifi, Y., Zumarniansyah, A., Ardianto, R., Hardi, N., & Annisa Elfina Augustia. (2020). *Laporan Akhir Penelitian Mandiri*. Jakarta.

Apandi, T. H., Maulana, R. B., Piarna, R., & Vernanda, D. (2019). Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung). *Jurnal Techno Nusa Mandiri*, *16*(2), 93–98. https://doi.org/10.33480/techno.v16i2.659

Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika (Yogyakarta)*, *8*(1), 884–898. https://doi.org/10.26555/jifo.v8i1.a2086

Ermawati, E. (2019). Ermawati, Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai 513. *Sistemasi*, *8*(September), 513–528.

Husin, A. I., & Mulyaningsih, F. (2015). Penerapan Metode Data Mining Analisis Terhadap Data Penjualan Pakaian Dengan Algoritma Apriori. *Sniptek*, 45–56.

Katadata Indonesia. (2018). 2018, Jumlah Penduduk Indonesia Mencapai 265 Juta Jiwa.

Katadata Indonesia. (2019). Jumlah Penduduk Indonesia 2019 Mencapai 267 Juta Jiwa.

Kementerian Sosial Republik Indonesia. (2017). *Bantuan Pangan Non Tunai (BPNT)*. Jakarta Pusat.

Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Naive Bayes…*
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018**

Penggunaan Listrik Rumah Tangga. *Citec Journal*, *2*(3), 207–217. https://doi.org/doi.org/10.24076/citec.2015v2i3.49

Septiyana Firdyana, Dedy Cahyadi, I. F. A. (2017). Penerapan Metode Weighted Product Untuk Menentukan Penerima Bantuan Beras Masyarakat Miskin (Raskin). *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, *2*(1), 1–7.

Sugianto, C. A., Maulana, F. R., & Mining, D. (2019). Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama). *TechnoCom*, *18*(4), 321–331.

Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri). *Jurnal Sistemasi*, *7*(September), 238–249.

Zuhair, A., Suyono, H., Muslim, M. A., Elektro, J. T., Teknik, F., & Brawijaya, U. (2019). Optimasi Injeksi Photovoltaic Distributed Generation Menggunakan Metode Ant Colony Optimization Continuous Domain dan Improved Particle Swarm Optimization, *13*(3), 145–149.

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Naive Bayes...*
Techno Nusa Mandiri : Journal of Computing and Information Technology
As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018**