VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 8 No. 1 (2026), 1-7

DOI: 10.35580/variansiunm418

Application of the *Naive Bayes* Algorithm for Classification of Family Hope Program (FHP) Assistance Recipients

Nunung Marlika[1*], Suwardi Annas[2], & Aswi[3]

Statistics Study Program, Faculty of Mathematics and Natural Sciences, Makassar State University, Indonesia

Keywords: Naive Bayes, Family Hope Program, Accuracy.

Abstract:

One of the classification methods commonly used to determine the eligibility of recipients of the Familiy Hope Program (FHP) is the *Naive Bayes Algorithm*, also often called the *Naive Bayes Classifier*. This method is probability to classify data quickly and efficiently for eligibity analysis in social assistance programs. *Naive Bayes* is classification that uses a probabiity and statistical approach to group data. In this study, the *Naive Bayes Algorithm* was applied in classifying recipients of the Family Hope Program and to determine the level of accuracy, recall and precision of the *Naive Bayes* method. The results of this study are the accuracy, recall and precision of the *Naive Bayes* methode. The results of this study are the accuracy value produced by the *Naive Bayes* methode of 90% in the division of training and testing data 60%:40% the accurqacy value of 93% in the division of training and testing data 70%:30%, and the accuracy values of 90% ih the division of training and testing data 80%:20%.

1. Introduction

Data mining has been around since the 1990s as a correct and precise way to extract patterns and information used to find relationships between data to group them into one or more clusters so that objects in one cluster will have high similarities with each other (Mai et al., 2022). Data mining is the process of extracting information from data sets through the use of algorithms and techniques involving the fields of statistics, machine learning, and database management systems. Data mining is used to extract important hidden information from large datasets (Winarti et al., 2021). Where the process of predicting involves the use of classification methods.

In this context, after understanding the meaning of *Naive Bayes*, the next step is to apply the classification method. Classification is the process of dividing objects into one of several complementary and mutually exclusive categories known as classes. Classification methods are systematic approaches to building classification models based on input data. One common classification method used to determine eligibility for recipients of the Family Hope Program (FHP) is the *Naive Bayes Algorithm*, often called the *Naive Bayes Classifier*. This method utilizes probability to quickly and efficiently classify data for eligibility analysis in social assistance programs.

Naive Bayes is a fairly effective and efficient algorithm for classification in data mining. Therefore, this algorithm is highly reliable for classifying the Family Hope Program (Nurchaidir & Prasetya Adhi, n.d.). *Naive Bayes* is often applied in data mining and text mining due to its ability to handle large data with simple calculations. According to (Darwis et al., n.d.), the *Naive Bayes* algorithm is based on Bayes' theorem, which states that all activities provide an equally important or mutually independent contribution to the selection of a particular class. *Naive Bayes* is used for

* Corresponding author.
E-mail address: nunungmarlika119@gmail.com

various purposes, including document classification and the classification of eligibility or ineligibility for Family Hope Program assistance.

Social assistance programs (bansos) are a key issue requiring government attention to improve public welfare (Entini et al., 2023). The Family Hope Program (FHP) is a conditional social assistance program from the government aimed at poor families as beneficiaries. This program aims to help Beneficiary Families (BF) escape the cycle of poverty sustainably. This study aims to predict the eligibility classification of Family Hope Program recipients, thus assisting village governments in determining eligible residents and achieving faster, more accurate, and more targeted results for recipients.

In this context, after understanding the definition of *Naive Bayes*, the next step is to apply classification methods. Classification is the process of dividing objects into one of several complementary and mutually exclusive categories known as classes. Classification methods are systematic approaches to building classification models based on input data.

2. Literature Review

2.1. Naïve Bayes

The *Naive Bayes Classifier* is a classification method that uses probability and statistical approaches to group data. This algorithm is based on a theory proposed by British scientist *Thomas Bayes*, who predicts the probability of a future event by relying on historical data (Indah Saputri, 2021). *Bayes' theorem* is used to calculate the probability of an event occurring by considering information from existing observations. (Rennie et al., n.d.) states that errors in the independence assumption do not always have a significant impact on the model's ability to effectively separate classes.

2.2. Naïve Bayes Method Equation.

a. *Bayes Theorem*
The equation for Bayes' theorem according to (Rahmadanid et al., 2023) is:

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

dimana:
X : data with unknown class
H : hypothesis on data X which is a special class
$P(H|X)$: probability value of hypothesis H based on condition X
$P(H)$: probability value of the hypothesis H
$P(X|H)$: the probability value of X based on the condition H
$P(X)$: probability value on data X

2.3. Naïve Bayes Method Algorithm

Bayes' theorem is the basis of Naive Bayes. *Bayes' theorem* is a statistical theorem used to calculate the probability of a class from any group with existing characteristics and determine which class is the most optimal (Pratiwi et al., 2024). The *Naive Bayes Classifier* method algorithm is:

a. Grouping variables. In the *Naive Bayes Classifier* method, the first step is to group the data of the Family Hope Program (FHP) recipients. Variable grouping is based on the classification of the Family Hope Program (FHP) assistance program, which is first done by grouping discrete and continuous data variables. From the data obtained, it can be seen that there are 5 discrete data variables and 1 continuous data variable, including:
    1) Discrete Data
    Discrete data is a type of numeric data that can only take on separate and limited values.

2) Continuous Data
Continuous data is a type of numeric data that can take values within a certain range including decimal numbers and fractions.

b. Find the mean and standard deviation values for parameters that are numerical data. Calculating the mean and standard deviation can only be applied to continuous variables, or variables whose values are continuously related to each other according to their attributes, such as the income variable. Therefore, the calculation of the mean and standard deviation can be seen in the following equation:

Mean

Mean is the middle value or value that represents all the data, where the way to calculate it is by adding up all the values and then dividing it by the number of data.

$$\mu = \frac{x_1 + x_2 + \cdots + x_3}{n}$$

dimana:
μ : mean
x : sample value
n : total of samples

Standard Deviation

Standard deviation is a measure used in statistics to show how varied or spread out the data is in a set of values.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

dimana:
σ : standard deviation
x_i : sample value
μ : mean
n : total of samples

c. For the prior probability value in the Family Hope Program (FHP) classification determination category and the probability for each category itself (Rahmadani et al., n.d.-a). Then it is formulated:

$$P(Ci) = \frac{\sum Ci}{n}$$

dimana:
$P(Ci)$: label probability on Ci
$\sum Ci$: number of data with labels in class Ci
n : total of samples

d. Calculation of probability values for each class in each data (Rahmadani et al., n.d.-b):

$$P(C_i|X) = P(C_i) \prod\nolimits_{i=1}^{n} \times P(X_i|C_i)$$

dimana:
$P(C_i|X)$: likelihood of instance C_i for class X
$P(C_i)$: probability of class label C_i
$P(X_i|C_i)$: the probability that feature X_i has class label C_i

2.4. Training Data and Testing Data

Training data is used to train the model to recognize patterns or relationships between label features (classes). Meanwhile, testing data is used to assess the model's performance after training. In this study, the *training and testing data* ratios were divided into three scenarios: 60%:40%, 70%:30%, and 80%:20%.

2.5. Confusion Matrix

To evaluate the performance of a classification method, a *confusion matrix* is used. A *confusion matrix* is a table used to assess the accuracy and effectiveness of a model generated from a classification model for classifying and predicting attributes from testing data. This method was developed for evaluating machine learning algorithms applied to classification problems (Markoulidakis et al., 2021). A confusion matrix contains False Negatives (FN), False Positives (FP), True Negatives (TN), and True Positives (TP). The following is a table of the *confusion matrix* (Luque et al., 2019):

Table 1. *Confusion Matrix*

Class Aktual	**Class Prediktion**	
	Positive	***Negative***
Positive	TP	FP
Negative	FN	TN

The results of the classification of eligibility status for Family Hope Program (FHP) recipients using the *Naive Bayes Classifier* method using R software will then be compared with actual data and the performance of the classification system used will be tested, including an accuracy test. To determine the accuracy of the data, the following formula is used:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

Keterangan:

TP (*True Positive*) : the number of data that are true positive and classified as positive by the model.
TN (*True Negative*) : the number of data that are truly negative and classified as negative by the model.
FP (*False Positive*) : the amount of data that is actually negative but is classified as positive by the model.
FN (*False Negative*) : the number of data that are actually positive but are classified as negative by the model.

Accuracy is the ratio of correct predictions to the total data. Recall is the ratio of correct positive predictions to the total data. Precision is the ratio of correct positive predictions to the total data predicted as positive.

2.6. Classification

Classification is the process of assessing data objects to be placed into a particular class from a number of available classes. Classification involves two main tasks: building a model to be stored in memory, and using that model to recognize/classify/predict other data objects to determine which class the data object belongs to in a model that is easy to store. Classification cases can be distinguished based on the types of problems commonly encountered, namely:

a. Classification 1 class
b. Classification 2 class (binary)
c. Classification of more than 2 class (multiclass)

2.7. Family Hope Program (FHP)

The Family Hope Program has two main goals. First, it provides regular financial assistance to families in need, with conditions such as ensuring adequate access to education and healthcare for children.

3. Research Methods

The type of research used is research with a quantitative approach, namely by taking or collecting the necessary data and conducting analysis using *Naive Bayes Classifier* modeling to apply the classification of recipients of the Family Hope Program (FHP) assistance.

This study uses secondary data. It consists of two variables: five independent variables: head of household income, number of dependents, housing conditions, education, and gender. There is one dependent variable: aid eligibility status.

4. Results and Discussion

4.1. Data Description

The research data consisted of 51 respondents who received Family Hope Program (FHP) assistance in Takalar Regency in 2023, consisting of 6 variables, namely 5 independent variables (total income of the head of the family, number of dependents, house condition, education and gender) and 1 dependent variable (assistance eligibility status).

Descriptive results show that most respondents have low incomes (average Rp. 2,494,118), the average number of dependents is 3 people, and the majority live in uninhabitable houses (74.5%). The level of education of respondents is relatively low with 72.5% having an education of < Junior High School and the characteristics of poor households that are targeted as recipients of the Family Hope Program (FHP) assistance.

4.2. Naïve Bayes Classification Results

The data was divided into three *training* and *testing* data scenarios: 60%:40%, 70%:30%, and 80%:20%. The evaluation results using a confusion matrix produced accuracy, recall, and precision values as shown in Table 2.

Table 2. *Naive Bayes* Model Evaluation Results

Evaluation	Distribution of Training and Testing Data		
	60% : 40%	70% : 30%	80% : 20%
Accuracy	90%	93%	90%
Recall	100%	100%	100%
Presicion	50%	75%	50%

Based on Table 2, it can be seen that the highest accuracy value is found in the training and testing data division, namely 70%: 30% with a level of 93%, which means that the percentage value illustrates how strong the model is in classifying correctly. Meanwhile, the recall value of 100% in the data division of 60%: 40%, 70%: 30% and 80%: 20% which illustrates the success of the model in re-determining information. And the best precision value is found in the training and testing data division, namely 70%: 30% with a value of 75%.

4.3. Discussion

The results of this study indicate that the *Naive Bayes* algorithm is effective in generating eligibility criteria for Family Hope Program (FHP) recipients with a high level of accuracy. A recall value of 100% indicates this indicates that all

eligible households were successfully identified. However, lower precision in the 60%:40% and 80%:20% scenarios indicates that some households were predicted as eligible but were not.

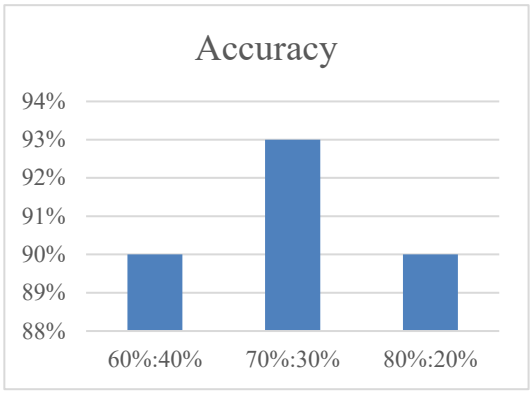

Figure 1. Accuracy Comparison of Each Division *Training Data* and *Testing Data*

Figure 1 shows the accuracy values for three different data sharing scenarios, namely 60%:40% and 80%:20%, which both produce an accuracy value of 90%. Meanwhile, in training and testing, 70%:30% produces an accuracy of 93%, which is the highest accuracy value among the three scenarios.

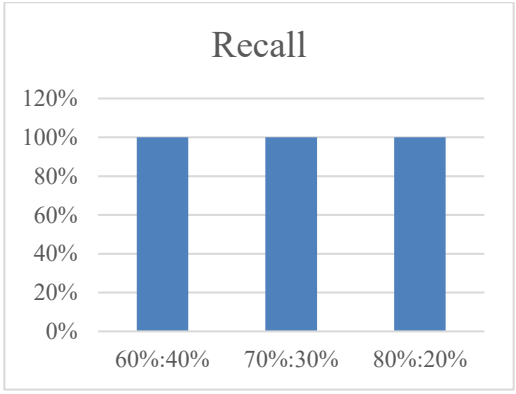

Figure 2. Comparison of Recall from Each Division of *Training Data* and *Testing Data*

From the diagram in **Figure 2** it can be seen that the recall value is 100% for all tested data split scenarios. A recall value of 100% means that the model successfully identified all positive instances in the test data for each different data split.

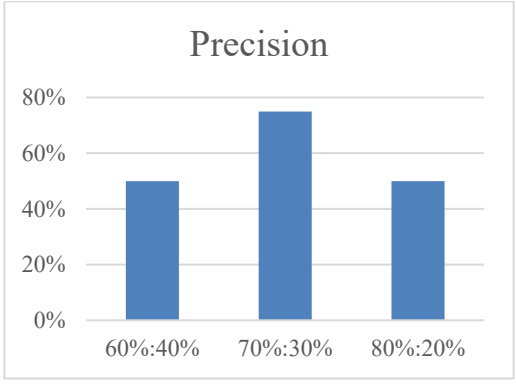

Figure 3. Comparison of Precision of Each Division *Training Data* and *Testing Data*

From the diagram in **Figure 3**. It shows that the *training* and *testing* data 70%: 30% has the highest precision value of the three scenarios above with a precision level of 75%.

5. Conclusion

Based on the results of research regarding the classification of the eligibility status of recipients of the Family Hope Program assistance, it can be concluded that:

a. The Naïve Bayes method was applied to classify the eligibility of recipients of the Family Hope Program assistance based on independent and dependent variables, namely the head of the family's income, number of dependents, housing conditions, a person's education, and gender. This model was built by dividing the data into several training and testing data scenarios, namely 60%: 40%, 70%: 30%, and 80%: 20% to evaluate the classification performance.
b. The 70% : 30% data split produced the best overall performance, particularly in accuracy values of 93% and precision of 75%. Although all data splits achieved 100% recall, meaning the model successfully identified all relevant information and achieved the highest precision, the 70% : 30% training and testing data split showed that this model had a higher proportion of correct positive predictions compared to other training and testing data splits.

References

Darwis, D., Siskawati, N., & Abidin, Z. (n.d.). *Application of Naive Bayes Algorithm for Sentiment Analysis Review of National BMKG Twitter Data*. *15*(1).

Entini, A., Raja, L., & Handoko, K. (2023). Implementation of Data Mining with the Naive Bayes Algorithm for Classifying the Eligibility of Basic Food Assistance Recipients. In *Jurnal Comasie*.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Mai, P., Tarigan, S., Tata Hardinata, J., Qurniawan, H., Safii, M., Winanjaya, R., Studi, P., Informasi, S., Tunas, S., & Pematangsiantar, B. (2022). *Implementation of Data Mining Using the Apriori Algorithm in Determining Inventory of Goods (Case Study : Toko Sinar Harahap)* (Vol. 12, Issue 2). https://jurnal.umj.ac.id/index.php/just-it/index

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, *9*(4). https://doi.org/10.3390/technologies9040081

Nurchaidir, T., & Prasetya Adhi, B. (n.d.). *Music Genre Classification Using Naive Bayes Classifier Algorithm For Youtube Streaming Service*.

Pratiwi, I. M., Fauzi, A., Lestari, S. A. P., & Cahyana, Y. (2024). Application of the Naïve Bayes Algorithm for Employee Recruitment Prediction. *Journal of Information and Computer Engineering*, *7*(1), 236. https://doi.org/10.37600/tekinkom.v7i1.1282

Rahmadani, N., Dwi Sena, M., Informatika, M., Royal, S., & Computer, S. (n.d.-a) College of Informatics and Computer Management. Application of the *Naïve Bayes* Algorithm in Determining the Eligibility of Recipients of the Family Hope Program Assistance. (*Journal of Computer Technology and Information Systems) August*, *2023*(2), 40–48. http://jurnal.goretanpena.com/index.php/teknisi

Rahmadani, N., Dwi Sena, M., Informatika, M., Tinggi Manajemen Informatika dan Komputer Royal, S., & Komputer, S. (n.d.-b). Application of the Naïve Bayes Algorithm in Determining the Eligibility of Recipients of the Family Hope Program Assistance. (*Journal of Computer Technology and Information Systems) Agustus*, *2023*(2), 40–48. http://jurnal.goretanpena.com/index.php/teknisi

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (n.d.). *Tackling the Poor Assumptions of Naive Bayes Text Classifiers*.

Winarti, D., Kom, M., Revita, E., Yandani, E., Lintas Sumatera, J., 18 Koto, K. M., Dharmasraya, B., & Barat, S. (2021). Application of Data Mining for Crime Rate Analysis Using the Association Rule Algorithm FP-Growth Method. *Journal SIMTIKA*, *4*(3).