International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 442-446 www.ijemh.com

Income classification of MSMEs (Micro, Small and Medium Enterprises) affected by Covid-19 for receiving direct cash assistance using Naive Bayes

Ferdy Julianto, Yaddarabullah, Silvester Dian Handy Permana
Department of Informatics, Trilogy University, Jakarta, Indonesia

Date of Submission: 01-08-2021 Date of Acceptance: 14-08-2021

ABSTRACT, The Covid Virus pandemic or known as Covid-19 that has occurred throughout the world, including in Indonesia, has resulted in instability in the world and Indonesian economies, which has made Indonesian citizens restless because of the Covid-19 outbreak, with this condition the activities of citizens are very limited and have an impact on the economy. traders, especially in the MSME sector (Micro, Small and Medium Enterprises) who experienced a decline in income due to the Covid-19 outbreak. The study conducted a classification based on declining income data from several MSME actors obtained from data from the Jakarta Open Data Website. The method used is the Naïve Bayes Classifier method, which is one of the classification techniques in data mining. The test results using Naive Bayes which are proven to have decreased, this can be proven by the Naïve Bayes manual method using the Income Attribute in the training data table. The results of the Confusion Matrix test with Split Validation on the research object obtained 8 (eight) positive or correct ones to get direct cash assistance with the actual value in the testing data.

KEYWORDS: Covid-19, Classification, Naive Bayes, MSME

I. INTRODUCTION

The Corona Virus pandemic or known as Covid-19 that has occurred throughout the world and is no exception in Indonesia has resulted in instability in the world economy as well as in Indonesia. These problems have made citizens of the world and Indonesians nervous, because of the Covid-19 outbreak, many countries have implemented lockdowns or isolation to limit the spread of Covid-19, with this condition the activities of citizens are very limited and have an impact on the economy of traders, especially in the MSME sector (MSMEs). Micro, Small and Medium) which experienced a decrease in income due to the Covid-19 outbreak [1]. The current handling of this problem is with assistance from the government in the form of Direct Cash Assistance given to the community or MSMEs affected by the Covid-19 outbreak who meet the requirements as recipients of assistance. The classification data that will be used is the data of MSME owners based on the type of business, income and income data that has decreased or is stable. The research data is taken from the Jakarta open data website, and the method that will be used is the Naïve Bayes Classifier method, which is one of the classification techniques in data mining. [2]. The Naïve Bayes Classifier method was chosen because this method works better in terms of accuracy and application, which allows us to determine the exact value of declining or stable sales of MSMEs. [3]. The Naïve Bayes algorithm is a data mining classification algorithm that utilizes a probability and simple statistics discovered by the English scientist Thomas Bayes. Bayes assumes all attributes to be value variables of this class while assuming that itself rarely applies to world applications [4], hence the characterization is assumed to be simplicity but the algorithm tends to perform well and fast in various classification problems. The formulation of the research problem is to determine how to engineer a system for the classification of MSME incomes that have decreased during the Covid-19 pandemic in order to receive direct cash assistance from the government, using the Naïve Bayes Algorithm. [5]. The purpose of this study is the classification of MSME income to receive Direct Cash Assistance using the Naïve Bayes Algorithm and applying the Nave Bayes Algorithm [6].

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 442-446 www.ijemh.com

II. RESEARCH METHODOLOGY

The methods and procedures used in this study are:

A. Classification

Classification or function that describes and distinguishes a type of data class or concept that aims to predict an unknown class or class object [7]. The Naïve Bayes algorithm that is often widely used for classification is,Decision/classification trees, Bayesian classifiers/Naïve Bayes classifiers, Neural networks, Statistical Analysis, Genetic Algorithms, Rough sets, Memory based reasoning, and Support vector machines (SVM) [8].

B. Naïve Bayes Algorithm

Naïve Bayes classifiers is a statistical classification that can be used to predict a data member of a certain class [9]. Bayes method is a statistical approach to perform induction inference, first discussed the basic concepts and definitions in Bayes theorem, then use classification in data mining [10]. Naïve Bayes classifiers are proven to have high accuracy and speed to be applied in databases with large data. Naive Bayes has the following formula:

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

Information :
X : A data with an unknown class
H : Is a data hypothesis which is a specific class
P(H|X): A probability hypothesis H based on condition X (posterior probability)
P(H) : Hypothesis probability H (prior probability)
P(X|H): Probability of hypothesis X based on condition H
P(X) : Probability X

Tests carried out on the Naïve Bayes classification method use the confusion matrix and split validation techniques.

1. Confusion Matrix

Often called an error matrix, basically provides information on the results of the classification comparisons made by the system model with the actual results. The Confusion Matrix is in the form of a table depicted by a classification model on the data that has been tested which displays the true value [11].

The benefits of the Confusion Matrix:
1.Shows the model When it has predictions.
2.Shows the error made and also the type of error made.
3.Can represent instances of classes.
4.Can represent each actual class.

Table 1.
Confusion Matrix 2 class

	True Values	
	True	False
True	TP Correct result	FP Unexpecte d result
False	FN Missing result	TN Correct absence of result

TP (True Positive) : Correctly predicted positive data.
FP (False Positive) : Negative data but predicted as positive data.
FN (False Negative) : Positive data but suspected negative data.
TN (True Negative) : True negative data.

2. Split Validation

Split Validation is a validation technique that divides the data into two randomly, the two data are training data and testing data, using Split Validation a training data experiment will be carried out with a predetermined split ratio and the remaining training data will be considered as testing data [12]. While data training includes data that is used in conducting learning, while data testing is data that has never been used as learning material that will function as test data for true or false learning outcomes. [13].

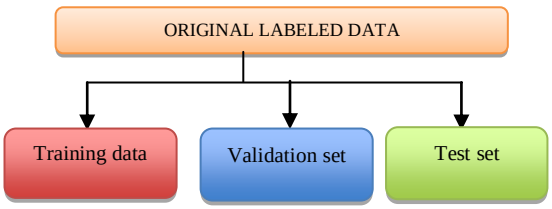

Figure 1. Illustration of Split Validation

3. Accuracy, Precision, and Recall

Accurary is how accurate the model can classify correctly, or it can be called accuracy is a positive prediction ratio [14]. While Precision is the level of accuracy between the

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 442-446 www.ijemh.com

data obtained with the results given by the model and a lot of data that is really positive [15].and Recall is a model in finding back an information which is a positive correct prediction compared to the whole correct data.

C. Research procedure

1.Identification of problems

The problem in this research is the impact of MSME income due to the Covid-19 outbreak. This study was conducted to classify MSME actors whose income is declining/stable to obtain social assistance funds from the government.

2.Data collection

Data collection was taken through the Jakarta open data website which contains 20 random data from various types of MSMEs in Jakarta.

Table 2.
Classification of Data Collection

No	Type Of Business	Income	Income Data during Covid 19
1	Foods	LOW	DOWN
2	Cosmetic	MEDIUM	DOWN
3	Fashion	MEDIUM	DOWN
4	Online Shop	HIGH	STABLE
5	Property	LOW	DOWN
6	Roadside stall	MEDIUM	DOWN
7	Shoes	MEDIUM	DOWN
8	Foods	LOW	DOWN
9	Online Shop	HIGH	STABLE
10	T-Shirt	LOW	DOWN
11	Roadside stall	MEDIUM	STABLE
12	Foods	MEDIUM	STABLE
13	Online Shop	HIGH	STABLE
14	Accecories	LOW	DOWN
15	Property	LOW	DOWN
16	Parfum	MEDIUM	STABLE
17	Foods	MEDIUM	STABLE
18	Online Shop	HIGH	STABLE
19	T-Shirt	LOW	DOWN
20	Foods	LOW	DOWN

3. Data processing

The data that has been collected is then processed using the Split Validation Technique with the Confusion Matrix. In the Naïve Bayes classification test using the Split Validation technique where the existing data will be presented in the classification table, 20 data collection will be divided into two parts, namely 90% of the remaining 10% which will be used as testing data. In the classification process, there is training data using the class attribute of business type, income and income data. For the use of the Naïve Bayes classification method, the accuracy, precision and recall values can be calculated.

III. Results and Discussion

A. Use of Classification Methods Using Data Training:

Based on the training data, it can be classified as declining or stable income if the following new data are given:
Type of business = Property
Income = Low
Then the classification of MSME income can be determined as follows:
Naive Bayes Process

1. Counting the number of classes

P (Decrease) = $\frac{12}{20}$ = 0.6

0.6 x 100% = 6%
"the amount of income data decreases in the training data divided by the total data multiplied by 100%"

P(Stable) = $\frac{8}{20}$ = 0.4

0.4 x 100% = 4%
"the number of stable income data in the training data divided by the total number of data multiplied by 100%"

2. Counting the same number of cases with the same class

State attribute probability:

P (Food - Down) = $3/20$ = 0.15
P (Food - Stable) = $2/20$ = 0.1

P (Roadside stall - Down) = $1/20$ = 0.05
P (Roadside stall - Stable) = $1/20$ = 0.05

P (Cosmetic - Down)= $1/20$ = 0.05

P (Fashion - Down) = $1/20$ = 0.05

P (Property - Down) = $1/20$ =0.05

P(Shoes - Down) = $1/20$ = 0.05

P(T-Shirt - Down) = $2/20$ = 0.1

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 2, Issue 4, pp: 442-446 www.ijemh.com

P(Accecories - Down) = $1/20$ = 0.05

P(Perfume - Stable) = $1/20$ = 0.05

P (Online Shop - Stable) = $4/20$ = 0.2

3. Multiply the result of the declining and stable income variable using the new data

P (Revenue Down)
= (0.15 x 0.05 x 6 x 0.1)
= 0.0045 (Decreased)

P (Stable Income)
= (0.1 x 0.05 x 0.05 x 0.2)
= 0.000005 (Stable)

The probability value of declining income is higher than the probability value of stable income, it is concluded that the new data above belongs to the category of declining income.

B. Classification Method Testing Using Tetsing Data:

The results of the classification process can be converted into the form of a Confusion Matrix table.
The following is the form of the Confusion Matrix test table:

Table 3.
Confusion Matrix Test

No	Income	IncomeData During Covid 19	Pretictic Class
1	LOW	DOWN	8(TP)
2	MEDIUM	DOWN	4(FP)
3	MEDIUM	STABLE	4(TN)
4	HIGH	STABLE	4(FN)

TP (True Positive): Positive data that is predicted correctly.
FP (False Positive): Negative data but predicted as positive data.
FN (False Negative): Positive data but suspected negative data.
TN (True Negative): True negative data.

Based on the Confusion Matrix table above, it can be concluded that the income of MSMEs affected by Covid-19 for receiving direct cash assistance is 8 (Eight) MSMEs who are entitled to receive assistance, because of the positive and correct data.

IV. CONCLUSION

The income classification system for MSMEs affected by Covid-19 for receiving direct cash assistance using Naive Bayes which is proven to have decreased, this can be proven by the Naïve Bayes manual method using the Income Attribute in the training data table. The results of the Confusion Matrix test with Split Validation on the research object obtained 8 positive or correct ones to get direct cash assistance with the actual value in the testing data. This means that most of the classifications are appropriate and the error in classification also depends on the dataset being tested. In this case, it is certain that the classification built can be maximized as input for decision makers.

REFERENCES

[1] T. Tambunan, MICRO, SMALL AND MEDIUM ENTERPRISES AND ECONOMIC GROWTH, 2016.

[2] M. H. S. R. M. Muhathir, "Analysis Naïve Bayes In Classifying Fruit by Utilizing Hog featur extraction," 2020.

[3] e. W. Eshetu Bekele, Women Entrepreneurship in Micro, Small and Medium Enterprises: The Case of Ethiopia, 2016.

[4] K. P. Murphy, Naive Bayes classifiers, 2016.

[5] h. L. C. Y. P. Z. a. Y. S. Xianda Feng, "Intrusion Detection using Naive Bayes Classifier with Feature Reduction," Prediction of Slope Stability using Naive Bayes Classifier, 2017.

[6] U. O. S. O. Mucahid Barstugan, Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods, 2020.

[7] D. R. I. M. S. E. H. R. R. B. A. N. A. A. N. Heru Agus Santoso, Hoax classification and sentiment analysis of Indonesian news using Naive, vol. 18, 2020.

[8] E. P. W. M. Ayu Pariyandani, "Analysis of the Naïve Bayes Method in Classifying Formalized Fish," 2020.

[9] R. A. M. A. M. Yofi Firdan Safri, K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor, 2018.

[10] D. Agyapong, Micro, Small and Medium Enterprises' Activities, Income Level and Poverty Reduction in Ghana – A Synthesis of Related Literature, 2010.

[11] A. R. Ratih Prasetya, Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbor Algorithm With Model, vol. 4, 2019.

[12] C. V. S. Delcea Cristian, Validation and Standardization of The Questionnaire for Evaluation of Paraphilic Disorders, 2020.

[13] D. S. M. V. S. B. C. B. a. K. T. Robert Häusler, Generating Content-Compliant Training Data in Big Data Education, vol. 1, pp. 104-110, 2020.

[14] d. powers, EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC,, 2019.

[15] F. R. Aszhari, Classification of thalassemia data using random, 2020.