

## Analysis of Covid-19 Cash Direct Aid (BLT) Acceptance Using K-Nearest Neighbor Algorithm

Ahmad Ari Aldino<sup>\*1</sup>, Ryan Randy Suryono<sup>2</sup>, Riyama Ambarwati<sup>3</sup>

<sup>1</sup>Department of Informatics, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

<sup>2</sup>Department of Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Indonesia

<sup>3</sup>Department of Mathematics Education, Faculty of Tarbiyah and Teacher Training, UIN Raden Intan Lampung, Indonesia

e-mail: <sup>\*1</sup>aldino@teknokrat.ac.id, <sup>2</sup>ryan@teknokrat.ac.id, <sup>3</sup>riyama@radenintan.ac.id

### Abstrak

Di masa pandemi COVID-19, Pembatasan Sosial Berskala Besar (PSBB) diberlakukan pemerintah guna mengurangi atau memperlambat penyebaran COVID-19. Hal ini, menyebabkan masyarakat tidak dapat bekerja seperti biasa, bahkan tidak sedikit masyarakat yang kehilangan pekerjaan. Hal ini memicu pemerintah untuk meluncurkan program bantuan langsung tunai (BLT) Covid-19. Salah satu daerah yang terkena dampak PSBB ini adalah Desa Batu Ampar yang dalam penyaluran BLT dinilai kurang efektif oleh warga setempat karena terdapat BLT yang tidak tepat sasaran. Penyebab tidak efektifnya penyaluran bantuan ini dinilai karena tidak sinkron nya data, data baru sulit diverifikasi dan divalidasi kebenarannya karena luasnya wilayah dan jumlah warga kurang mampu yang terus berubah. Untuk mengatasi permasalahan tersebut, Untuk mengatasi hal tersebut diperlukan suatu model untuk memprediksi penerima BLT Covid-19 ini. Penelitian ini menggunakan algoritma K-Nearest Neighbor (K-NN) dan tools RapidMiner dalam melakukan prediksi kemudian divalidasi menggunakan Cross-Validation. Data yang digunakan berjumlah 711 baris dengan proporsi 474 data training dan 237 data testing menghasilkan akurasi sebesar 89.68% untuk data training dan 88.61% untuk data testing.

**Kata kunci**— COVID-19, Data Mining, K-NN, dan Cross Validation.

### Abstract

During the COVID-19 pandemic, the government imposed Large-Scale Social Restrictions (PSBB) to reduce or slow down the spread of COVID-19. This causes people to be unable to work as usual, and not even a few people have lost their jobs. This prompted the government to launch the Covid-19 direct cash assistance (BLT) program. One of the areas affected by the PSBB is Batu Ampar Village, which distributing BLT is considered less effective by residents because there are BLTs that are not well-targeted. The cause of the ineffectiveness of the distribution of aid was assessed because the data was out of sync; it was difficult to verify and validate the new data due to the size of the area and the constantly changing number of underprivileged residents. To overcome these problems, a model is needed to predict the recipients of this Covid-19 BLT. This study uses the K-Nearest Neighbor (K-NN) algorithm and RapidMiner tools to make predictions and validate using Cross-Validation. The data used are 711 lines with 474 training data and 237 testing data resulting in an accuracy of 89.68% for training data and 88.61% for testing data.

**Keywords**— COVID-19, Data Mining, K-NN, Cross-Validation

## 1. INTRODUCTION

Currently, the government has implemented various types of poverty and disaster reduction programs. However, the assistance that reaches the community is not as expected. The discrepancy in receiving aid funds occurs because the determination of the status of low-income families as recipients of aid has not been optimal, so in the provision of poverty assistance, there are still those not on target. One of them is the COVID-19 Direct Cash Assistance (BLT) program [1][2].

The COVID-19 BLT program was designed as a substitute for improving living standards during a pandemic due to Large-Scale Social Restrictions (PSBB). Therefore, the amount of BLT is calculated as an increase in the standard of living of the poor caused by price increases (inflation) caused directly or indirectly. Judging from the government's program, poverty alleviation efforts in the State of Indonesia are attractive to the public. So that people are busy praising the government for the BLT program. Not surprisingly, then the people at the level of the able category also made themselves the target of the COVID-19 special BLT. However, this government program is considered ineffective, so it must be predicted to provide COVID-19 aid funds [3][4].

Prediction is an art and science form of predicting future events. This can be done by taking historical data and projecting it into the future in a mathematical model. Alternatively, it can also be done using mathematical models adapted to the best judgment. Prediction plays a critical role in business. Currently, research on prediction has developed. Several methods, such as KNN, have been used for document classification [5], traffic detection [6], Heart Disease Prediction [7], stock price prediction models [8], and others. The results showed that the KNN method had accuracy in predicting. This algorithm performs data classification based on learning data (train data sets) taken from the k closest neighbors (nearest neighbors). This is an opportunity to explore KNN to predict decision-making in providing direct COVID-19 assistance funds in Batu Ampar Village.

Batu Ampar Village is one of the villages that have the regulation of Constitution No. 6 of 2014 on Villages following the good mandate of the village has the implementation of government, development, and community. Batu Ampar Village in BLT distribution has been assigned to provide COVID-19 funds directly to the community. Giving BLT is currently less effective because it is related to the availability of appropriate data for low-income family data that is on target. The availability of this data is often unverified and validated due to the region's vastness with a vast and frequently changing number of low-income families. The status of low-income families obtained is often inaccurate data regarding the number and location of its spread. Often, there is a shortage of families receiving assistance in a region, while there is an over-receiving aid in other areas. The existence of appropriate data can help plan the right program.

Based on the above problem, the author needs to study the prediction of BLT COVID-19 in Batu Ampar Village using the K-Nearest Neighbor Algorithm (K-NN). The K-NN method can be used to select decent communities and does not receive funds for the help of low-income families. This method performs the classification of objects based on learning data closest to the object. The K-NN method serves to classify or group criteria from the data tested [9].

## 2. METHODS

### 2.1 Data Mining

According to [10], the rapid development of data mining cannot be separated from the development of information technology that allows large amounts of data to accumulate. However, the rapid growth of data accumulation has created a condition called the "rich of data but poor of information" because the collected data cannot be used properly. Not infrequently, the data set is left alone to create "data tombs" (data graves).

Data mining is also known as KDD (knowledge discovery in the database). However, in 1995, an International KDD Conference in Montreal successfully defined that KDD is a process of recognizing information or a new truth and is practical and recognizes discernible data patterns. The primary purpose of the KDD process is to predict the valuable values of existing variables or find patterns of a data cluster that humans can interpret. Following these objectives, recognizing new information and discovering such patterns needs to be applied with data mining. So data mining is a part that cannot be separated from the KDD process [11].

Please note that data mining is one of the fields that is quite widely supported by other branches of science in information technology, namely statistics, database technology, machine learning, expert systems, parallel algorithms, genetic algorithms, pattern recognition, data visualization, and others [12][13].

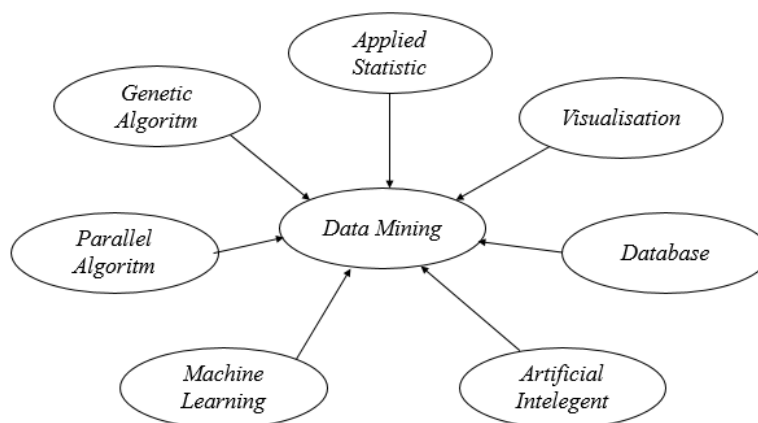


Figure 1 Data mining is a multidisciplinary field.

Several factors are the main reasons why using data mining:

1. The amount of data collected takes a long time and enough experts to analyze it.
2. Computers have become one of the leading choices because of their speed, accuracy, never tiredness, and ease of operation.
3. Pressure from business competition strengthens so that information becomes critical and must be owned immediately.
4. It was able to find an unthinkable pattern at all.

According to [14], data mining is one of the activities in software that can provide a high ROI (Return of Investment). The thing to note is that data mining is different from query tools. Query and data mining are two things that complement each other. The existence of data mining is not to replace queries but to add some meaningful additions. At a simple question, the information can be accessed about 80% of the data in the database, while the other 20% will be hidden information that requires special techniques to access [15][16].

Because data mining is a series of processes, it is divided into several stages, among others:

1. Data cleaning: to dispose of inconsistent data and noise.
2. Data integration: to combine data from multiple sources.
3. Data transformation: to convert data into a form suitable for mining.
4. Application of data mining techniques.
5. Evaluation of the patterns found: to find interesting or valuable information.
6. Presentation of knowledge with visualization techniques.

### 2. 2 *K-Nearest Neighbor (K-NN)*

The K-Nearest Neighbor (K-NN) algorithm is a method that uses a supervised algorithm. K-NN includes the instance-based learning group. This algorithm is also one of the lazy learning techniques. K-NN is done by searching for groups of k objects in training data that are closest (similar) to objects in new data or data testing [17].

The KNN algorithm is a method of classifying new objects based on (K) their nearest neighbors. KNN includes supervised learning algorithms, where the results of a new instance query are classified based on most of the categories on the KNN. The class that appears the most will be the class resulting from the classification.

In general, to define the distance between two objects x and y, the Euclidean distance formula is used in equation 2.2.

$$dyx = \sqrt{\sum_{f=1}^n (x_1 - y_1)^2} \dots\dots\dots(2.2)$$

Information:

dyx = Jarak *Euclidean*

$X_l$  = Test Data

$Y_l$  = Training Data

### 2. 3 *Rapid Miner*

Rapid Miner is open-source software. Rapid Miner is a solution for data mining, text mining, and predictive analysis. Rapid Miner uses various descriptive and predictive techniques to provide users with the best decisions. There are approximately 500 data mining operators owned by Rapid Miner, including operators for input, output, data preprocessing, and visualization. A Rapid Miner is a stand-alone software for data analysis and a data mining machine that can be integrated into its products [19]. Rapid Miner is written using the Java language to work across all operating systems.

Rapid Miner has the following properties:

1. It is written in the Java programming language to run on various operating systems.
2. Multi-layer concept to ensure efficient data display and ensure data handling.
3. It has a GUI, command-line mode, and Java API that can be invoked from other programs.

Some of the features of Rapid Miner, among others:

1. Many data mining algorithms include decision trees and self-organization maps.
2. Sophisticated graphical forms, including Histogram diagrams, Tree Diagrams, and 3DScatter Plots.
3. There are many variations of plugins, such as text plugins, to perform text analysis.
4. Provide data mining and machine learning procedures, including ETL (extraction, transformation, loading), data preprocessing, visualization, modeling, and evaluation.
5. The data mining process comprises nestable operators, described with XML and created with a GUI.
6. Integrates Weka data mining project and R statistics.

### 2. 4 *Cross-Validation*

Validation and testing are done to determine whether all the functions are working correctly or not. Validation is done with 10-fold CrossValidation. Tenfold cross-validation is performed by dividing the data set into ten segments of equal size by randomizing the data. Validation and testing were carried out to determine the accuracy, precision, and recall of the

classification prediction results. Accuracy is the percentage of records classified correctly in the test of the data set. Precision is the percentage of data belonging to a good model, which is also good. The recall is a measurement of the actual positive recognition rate [20].

### 2.5 Research Framework

The framework for thinking about the researcher's picture can be seen in Figure 2 below.

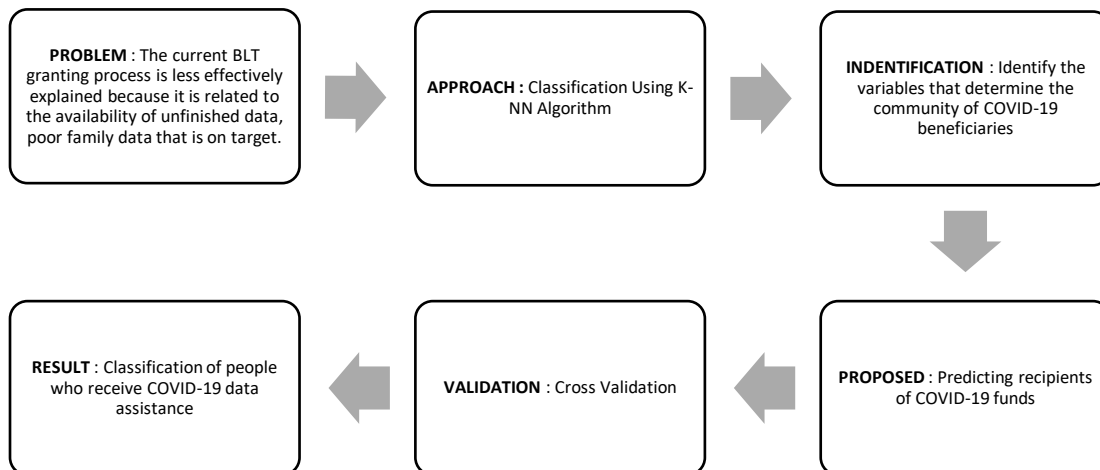


Figure 2 Research Framework

### 2.6 Data Collection Methods

In the preparation of this study, to get the data and information needed, the methods used in the data collection process are carried out as follows:

1. Literature Review  
It is the study of data by collecting references through books and journals during the analysis period.
2. Interview  
It is a data collection technique by holding a question and answer directly to the village head and staff of the Batu Ampar Village office.
3. Documentation  
It is a technique of collecting documentation and data evidence taken at the Batu Ampar Village Office.

## 3. RESULTS AND DISCUSSION

### 3.1 Exploratory Data Analysis

This chapter will discuss the analysis carried out on the data obtained. The data obtained amounted to 711 in 2020, along with a comparison of the criteria described using a graph for each variable. Figure 3 illustrates the occupational categories, Figure 4 depicts the land area (hectare), Figure 5 describes the income category, Figure 6 represents the number of children, and Figure 7 illustrates the recipients of aid funds.

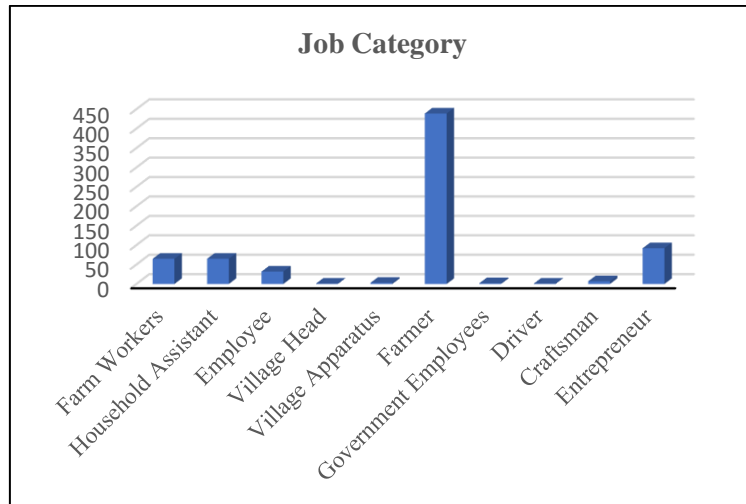


Figure 3 Job Category

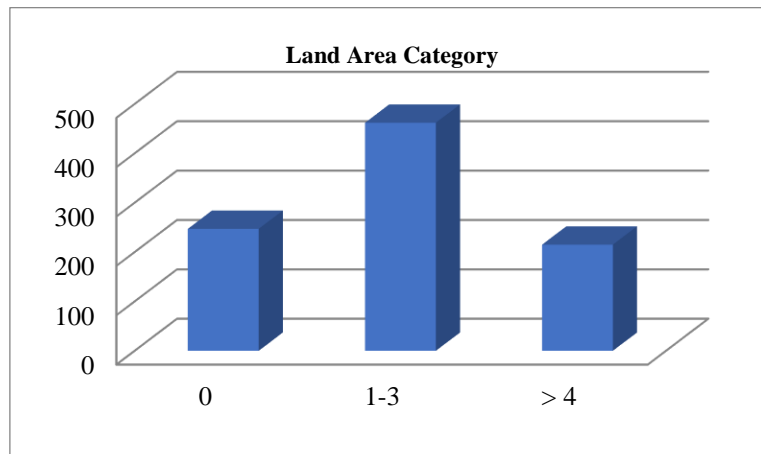


Figure 4 Land Area Category

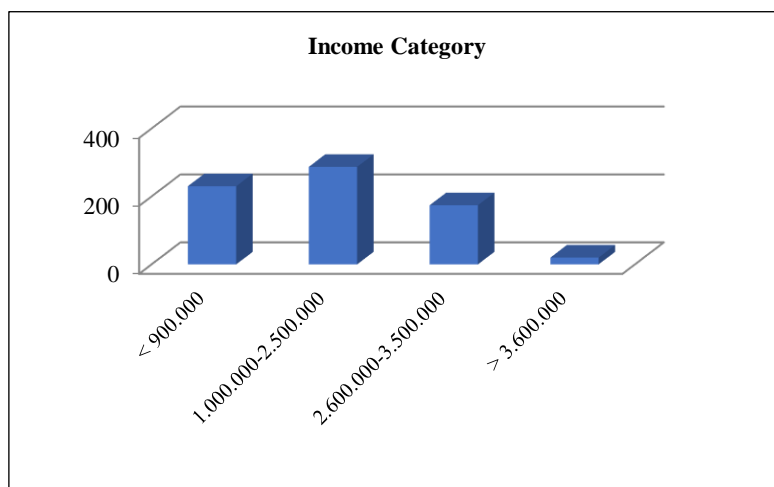


Figure 5 Income Category

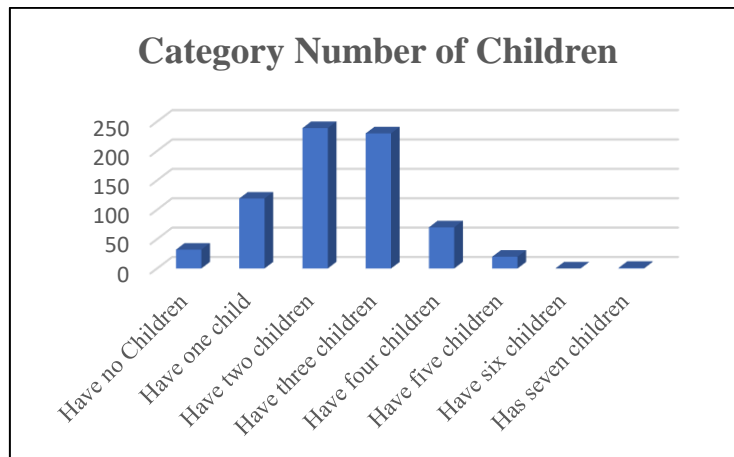


Figure 6 Number of Children Category

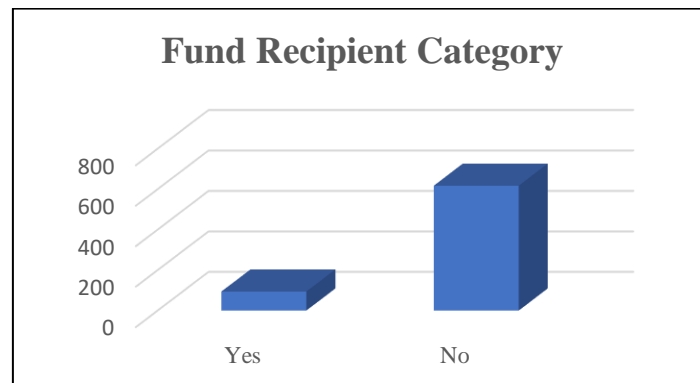


Figure 7 Fund Recipient Category

### 3. 2 Encoding Variable

In this encoding process, the data will be converted to numeric. Variables that must be changed are Job and Aid Recipient Category. In the variables of work, several types of jobs must be changed, such as:

FarmWorkers "0"

IRT "1"

Private Employee "2"

Village Chief "3"

Village Device "4"

Farmer "5"

PNS "6"

Driver "7"

Handyman "8"

Self-employed "9"

As for the variable category of aid recipients, that must be changed, namely.

YES "1" and

TIDAK "0".

The results of the encoding table can be seen in the following image:

Table 1 Encoding Results

| Name          | Job | Land Area (Ha) | Monthly Income | Dependents | Aid Recipient |
|---------------|-----|----------------|----------------|------------|---------------|
| AAN ANDRIANTO | 9   | 3              | 3500000        | 3          | 0             |
| ABDUL BAHRUN  | 9   | 1,5            | 2500000        | 2          | 0             |
| ABDUL GANI    | 5   | 2,75           | 3000000        | 2          | 0             |
| ABDUL MUKTI   | 0   | 0              | 300000         | 1          | 1             |
| ABDUL RAHMAN  | 5   | 4,5            | 4000000        | 4          | 0             |
| ...           | ... | ...            | ...            | ...        | ...           |
| ZAENUDIN      | 5   | 1              | 1500000        | 3          | 0             |
| ZAINUJI       | 5   | 2              | 1500000        | 1          | 1             |
| ZULHAN        | 5   | 2              | 1500000        | 3          | 0             |
| ZULTONI       | 5   | 5              | 3000000        | 3          | 0             |
| ZUNAINAH      | 1   | 2              | 2000000        | 2          | 0             |

3. 3 Test Results with Rapid Miner Studio application

Process the KNN Algorithm by dragging and dropping data and the operators to be used in the process directed in Figure 8.

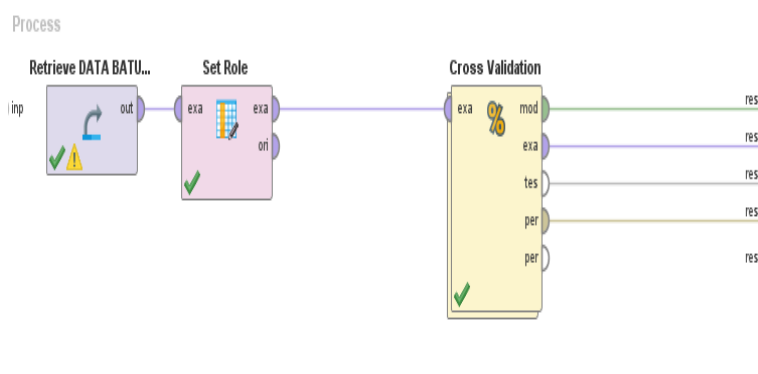


Figure 8 Rapid Miner Process

The testing process in Figure 9 is to break down and drop the KNN Algorithm operator in the Cross-Validation process, enter 10-Cross Validation values, and perform the Drag and Drop data process and the operators to be used by clicking the cross-validation operator drop.

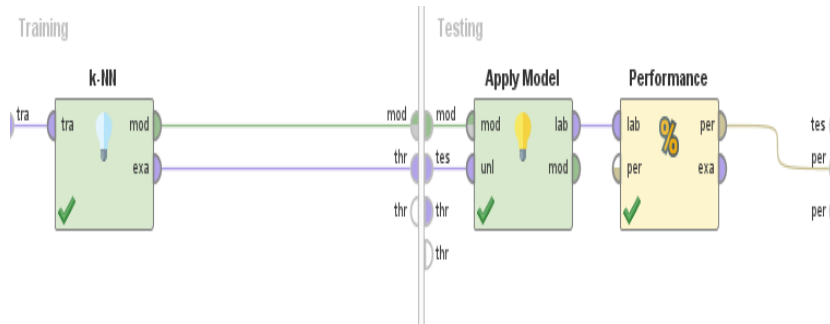


Figure 9 Rapid Miner Process



After entering cross-validation and dragging and dropping the operator, connect the KNN to the Result, and click the process button to generate an accuracy value. Based on the validation calculations carried out using KNN with 474 training data, an accuracy of 89.68% can be seen in Table 2.

Table 2 Accuracy of Training Data

| Accuracy: 89.68% +/- 3.36% (micro average: 89.68%) |                    |         |          |                 |
|----------------------------------------------------|--------------------|---------|----------|-----------------|
|                                                    | True Aid Recipient | True No | True Yes | Class Precision |
| Pred. Aid Recipient                                | 0                  | 0       | 0        | 0.00%           |
| Pred. No                                           | 1                  | 400     | 32       | 92.38%          |
| Pred. Yes                                          | 0                  | 16      | 26       | 61.90%          |
| Class Recall                                       | 0.00%              | 96.15%  | 44.83%   |                 |

Based on the training data process results above, as many as 400 citizens are classified "No" in getting financial aid, and 26 citizens are classified "Yes" in obtaining financial assistance, according to calculations using rapid miner tools. It can be known that:

1. True Positive (TP): a case where citizens are predicted (Positive) Yes, and the decision is actual (True) Yes, there are 26 citizens
2. True Negative (TN): the case where citizens are predicted (Negative) No, and in the decision is actual (True) No, there are 400 citizens
3. False Positive (FP): the case where the citizens are predicted Yes, it turns out no, so the prediction is wrong (False) there are 16 citizens
4. False Negative (FN): a case where the predicted citizens are not (Negative), but it turns out (True) Yes, there are 32 citizens.

Testing 237 data resulted in an accuracy of 88.61%. The table of data accuracy test results can be seen in Table 2 below.

Table 3 Accuracy of Data Testing

| Accuracy: 88.61% +/- 5.39% (micro average: 88.61%) |                    |         |          |                 |
|----------------------------------------------------|--------------------|---------|----------|-----------------|
|                                                    | True Aid Recipient | True No | True Yes | Class Precision |
| Pred. Aid Recipient                                | 0                  | 0       | 0        | 0.00%           |
| Pred. No                                           | 1                  | 192     | 17       | 91.43%          |
| Pred. Yes                                          | 0                  | 9       | 19       | 67.86%          |
| Class Recall                                       | 0.00%              | 95.52%  | 52.78%   |                 |

The results of the training data process above, then for the data, are as many as 237 citizens classified "No" in getting financial aid is 192 and 19 citizens classified "Yes" in obtaining financial assistance, according to calculations using rapid miner tools. It can be known that:

1. True Positive (TP): a case where citizens are predicted (Positive) Yes, and the decision is actual (True) Yes, there are 19 citizens
2. True Negative (TN): the case where citizens are predicted (Negative) No, and in the decision is actual (True) No, there are 192 citizens
3. False Positive (FP): the case where the citizens are predicted Yes, it turns out No, so the prediction is wrong (False). There are nine citizens

4. False Negative (FN): a case where the predicted citizens are not (Negative), but it turns out (True) Yes, there are 17 citizens.

#### 4. CONCLUSIONS

Based on the results of the refraction that has been done, it can be concluded as follows:

1. Analysis of BLT COVID-19 acceptance predictions using *the K-Nearest Neighbor* algorithm in Batu Ampar Village with rapid miner studio application and validation used *cross-validation*.
2. The data used in this test is based on job variables, land area, income, number of dependents, and assistance category.
3. The data used were 711 consisting of 474 training data and 237 test data. The tests that have been done resulted in an accuracy of 89.68% for *training* data and 88.61% for data *testing*.
4. The model produced in this study can predict other CASES of BLT by proving the accuracy obtained quite well.

#### ACKNOWLEDGEMENTS

We would like to thank the Universitas Teknokrat Indonesia for providing financial support to carry out this research.

#### REFERENCES

- [1] U. Lestari and M. Targiono, "Sistem Pendukung Keputusan Klasifikasi Keluarga Miskin Menggunakan Metode Simple Additive Weighting (Saw) Sebagai Acuan Penerima Bantuan Dana Pemerintah (Studi Kasus: Pemerintah Desa Tamanmartani, Sleman)," *Jurnal TAM (Technology Acceptance Model)*, vol. 8, no. 1, pp. 70–78, 2017.
- [2] R. Wowiling, "PERAN PEMERINTAH DESA DALAM PENYALURAN BANTUAN LANGSUNG TUNAI PADA MASYARAKAT DI MASA PANDEMI COVID-19 DI KECAMATAN ...," *JURNAL POLITICO*, 2021.
- [3] D. Herdiana, I. Wahidah, N. Nuraeni, and A. N. Salam, "Implementasi Kebijakan Bantuan Langsung Tunai (BLT) Dana Desa Bagi Masyarakat Terdampak COVID-19 di Kabupaten Sumedang: Isu dan Tantangan," *Jurnal Inspirasi*, vol. 12, no. 1, 2021.
- [4] B. Iping, "PERLINDUNGAN SOSIAL MELALUI KEBIJAKAN PROGRAM BANTUAN LANGSUNG TUNAI (BLT) DI ERA PANDEMI COVID-19: TINJAUAN PERSPEKTIF EKONOMI DAN SOSIAL," *JURNAL MANAJEMEN PENDIDIKAN DAN ILMU SOSIAL*, vol. 1, no. 2, 2020, doi: 10.38035/jmpis.v1i2.290.
- [5] S. Du and J. Li, "Parallel Processing of Improved KNN Text Classification Algorithm Based on Hadoop," *2019 7th International Conference on Information, Communication and Networks, ICICN 2019*, pp. 167–170, 2019, doi: 10.1109/ICICN.2019.8834973.
- [6] M. Dixit, R. Sharma, S. Shaikh, and K. Muley, "Internet traffic detection using naïve bayes and K-Nearest neighbors (KNN) algorithm," *2019 International Conference on Intelligent*

- Computing and Control Systems, ICCS 2019*, no. Iccics, pp. 1153–1157, 2019, doi: 10.1109/ICCS45141.2019.9065655.
- [7] C. Chethana, "Prediction of heart disease using different KNN classifier," *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, no. Iccics, pp. 1186–1194, 2021, doi: 10.1109/ICICCS51141.2021.9432178.
- [8] Q. Yunneng, "A new stock price prediction model based on improved KNN," *Proceedings - 2020 7th International Conference on Information Science and Control Engineering, ICISCE 2020*, pp. 77–80, 2020, doi: 10.1109/ICISCE50968.2020.00026.
- [9] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)," *Jurnal Techno Nusa Mandiri*, vol. 16, no. 1, 2019, doi: 10.33480/techno.v16i1.25.
- [10] I. Arfanda, W. Ramdhan, and R. A. Yusda, "Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai," *Digital Transformation Technology*, vol. 1, no. 1, 2021, doi: 10.47709/digitech.v1i1.1091.
- [11] A. Pratama, F. Ali, I. Ade, and R. Rinaldi, "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Data Science & Informatika (JDSI)*, vol. 1, no. 1, 2021.
- [12] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *Jdmsi*, vol. 2, no. 1, pp. 31–37, 2021.
- [13] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021.
- [14] I. G. Mustafa, "Studi Tentang Pemberian Insentif Dalam Meningkatkan Kinerja Pegawai Di Sekretariat Daerah Provinsi Kalimantan Timur," *Jurnal Paradigma (JP)*, vol. 1, no. 3, pp. 373–388, 2017.
- [15] D. Alita, I. Sari, A. R. Isnain, and S. Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *Jurnal Data Mining Dan Sistem Informasi*, vol. 2, no. 1, pp. 17–23, 2021.
- [16] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *Jurnal Ilmiah Fakultas Ilmu Komputer*, vol. 9, no. 1, pp. 25–34, 2020.
- [17] Aprilla Dennis, "Belajar Data Mining dengan RapidMiner," *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, vol. 5, no. 4, pp. 1–5, 2013, doi: 10.1007/s13398-014-0173-7.2.
- [18] I. G. Mustafa, "Imam, Studi Tentang Pemberian Insentif Dalam Meningkatkan Kinerja Pegawai ... 373," pp. 373–388.
- [19] C. A. Sugianto and F. R. Maulana, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai ( Studi Kasus Kelurahan Utama )," *Techno.Com*, vol. 18, no.

- 4, 2019, doi: 10.33633/tc.v18i4.2587.
- [20] H. Sulistiani and A. A. Aldino, "Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia)," *Edutic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.