Classification Appropriateness Recipient Help Non-Cash Food Using *Learning Vector Quantization (LVQ)* Method

Ayu Lestari[1]*, Anang Aris Widodo[2], Nanda Martyan Anggadimas[3]
[123]Faculty of Information Technology, Universitas Merdeka Pasuruan, Indonesia
[1]ayulestari121299@gmail.com*; [2]anang@unmerpas.ac.id; [3]nandama@unmerpas.ac.id
*corresponding author

ABSTRACT

Help Non-Cash Food is a program from the Government that is used to overcome poverty. The program is not functioning as well as it could because the procedure of receiving aid is not uniform, and individuals responsible for making choices are having trouble determining which families are qualified to receive the assistance. To overcome this problem, a classification system is needed to classify the eligibility of Non-Cash Food Assistance recipients so that the results are more efficient and accurate. This research uses the Learning Vector Quantization (LVQ) method with Python. This research aims to implement the LVQ method for the eligibility classification of non-cash food assistance recipients. System design is a stage that contains the process from start to finish of running this system which is described in the form of a flowchart, including system requirements that support this research, both software and hardware. In the process of analyzing the results and tests that are used as evaluation material in the process of finding a solution to a problem and making decisions in the process of planning activities, it is necessary to assess whether or not the LVQ approach is practicable to apply based on the findings of the research. In this study, 200 datasets were used with three epoch values and a learning rate of 0.1. The data set was randomly divided into a training portion of 80% and a testing portion of 20%. So that the results of this research using the LVQ method on the eligibility classification of recipients of Non-Cash Food Assistance obtain an accuracy of 97.5%.

Keywords: Non-Cash Food Classification; ML Classification; LVQ Method; Data Mining; Python.

Article History

Received : May, 20th 2023	Accepted : July, 06th 2023	Published : July, 25th 2023

I. INTRODUCTION

Plinggisan is a regional unit in the Kraton Pasuruan sub-district with a population of 3014 people. In overcoming poverty in its citizens [1], the Plinggisan Village government runs various government programs, including the Non-Cash Food Assistance Program, or BPNT. Beneficiary Families, commonly known as KPM, receive non-cash social assistance through electronic staple food utilization cards as Non-Cash Food Assistance [2]. This card is to purchase at supermarkets or e-warungs authorized by the Government to sell groceries [3]. A presidential regulation has regulated the basic food assistance distribution system [4]. The data used are 200 datasets, and the weights used are 14 parameters.

In running the program in the village area, the Plinggisan subdistrict Kraton still cannot hold optimally. There is no support system, so there is still not enough objective in the selection process decision For obtain help social. There are several complaining citizens to devise a village because citizens should be entitled to obtain the help that otherwise is no help. It makes the acceptance Tprocess help inaccurate and not evenly distributed regarding the acceptance process. The Government makes rule instructions technical and criteria for recipient help. There are 14 categories of poor or less people capable based on BPS standards [5]. The category that can be utilized to determine family or house possible ladder is classified as family, not wealthy enough, or low-income families. According to the Central Statistics Agency (2022), this assistance is obtained for each Beneficiary Family (KPM) worth IDR 110,000 at the beginning of the month. Meanwhile, the nominal value of staple foods from January to February was IDR 150,000, but from March to August, the total assistance received increased to IDR 200,000 until now [6].

In the application procedure, selecting candidate recipient Help Non-Cash Food is carried out manually with compare between registry files. Form such a selection very throw away time so that no effective and efficient [7]. Another obstacle is the party giver's difficulty deciding on a family that will obtain help. To make it easy party taker decisions about classification appropriateness, recipient Help Non-Cash Food is necessary to make something modeling technique classification using artificial intelligence and realize language programming Python.

DOI: http://dx.doi.org/10.25139/ijair.v5i1.6287

The classification machine learning approach created in effort settlement case categorization suitability recipient help for the resolution of the non-cash food issue using the alternative method [8] [9] such as Naive Bayes, K-Nearest Neighbour, K-Means, or Learning Vector Quantization (LVQ). The K-Nearest Neighbour approach, utilized in a previous study to categorize grant recipient villages using a dataset of 159 entries, was shown to have an accuracy level of 81.25% in terms of the criteria for eligibility [10]. Another study employed the LVQ method to classify the work of 80 datasets, achieving an accuracy of 93.78% [11]. Other studies use LVQ (Learning Vector Quantization), which has a level of accuracy of 98.84% and is applied to the classification of risk for hypertension totaling 100 data [12].

Other studies comparing the method LVQ (Learning Vector Quantization) with Naïve Bayes use credit rating to predict eligibility for gift credit at PT. BPR Lebak Sejahtera uses 97 data with 73.20% accuracy using Naïve Bayes, whereas using Learning Vector Quantization (LVQ) own level accuracy of 82.47% [13]. In another study on the implementation of the C4.5 algorithm to determine recipients of social assistance (Case study: Village Lurah poverty classification) with 95 datasets divided into two parts, namely 80% training data and 20% testing data, the resulting accuracy was 58.18% in this test [14].

II. METHODS

Beneficiary Families (KPM) is commonly given to families receiving Non-Cash Food Assistance (BPNT). Beneficiary Family data has been included in the Integrated Social Welfare Data (DTKS) and the Ministry of Social Affairs [15] [16]. However, even though it is contained in the DTKS, not all registered families automatically receive the benefit. The only exceptions to this rule are families that have met the standards outlined in Table I by the Government. The Central Bureau of Statistics has identified 14 characteristics associated with low-income families, which became the research parameters [17]. In research, planning and structured steps are needed so that research can run well [18]. The stages of the research conducted can be seen in Fig. 1.

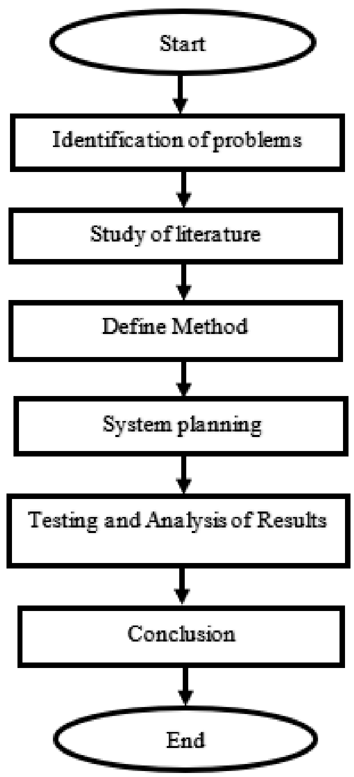

Fig. 1. Flowchart of Research Stages

A. Problem Identification

The first step in conducting this research is determining the problem. Identifying problems in research discusses what problems underlie research, then determines the process to be carried out in researching and solving these problems.

B. Literature Study

The second step in conducting this research is to conduct a literature study based on a predetermined research topic. This study's literature relates to the eligibility classification of Non-Cash Food Assistance recipients by implementing LVQ (Learning Vector Quantization). Residents of Plinggisan Village provided the data used in this study.

DOI: http://dx.doi.org/10.25139/ijair.v5i1.6287

C. Determine Research Methods

The third stage is to determine what method we will use, that the method must follow the topic to be studied, and the method used does not include the old method widely used in previous studies. This study uses the Learning Vector Quantization (LVQ) method.

D. System Design

System design is the fourth stage, which contains the process from start to finish of running this system which is described in the form of a flowchart, including the system requirements that support this research, both software and hardware.

E. Testing and Analysis of Results

The fifth step is the analysis of results and testing, which serves as evaluation material in solving a problem and determining a decision in the planning activity, to find out whether the LVQ (Learning Vector Quantization) method is suitable for use based on the results expected by the authors in this study. The last stage is to briefly state the research results and discuss the eligibility classification of Non-Cash Food Assistance (BPNT) recipients by applying Learning Vector Quantization (LVQ).

The method used in studying this is Learning Vector Quantization (LVQ), which is a method for classifying every existing unit based on the pattern. The output will display a class or category that has been determined before. The input distance will be determined class obtained. LVQ is another supervised training learning method that falls under the single-layer category. This means that only the owning layer's input and the layer's mutual output are related to something's weight. In the process of calculating the Learning Vector Quantization (LVQ) method, training is carried out which functions to create a picture, as shown in Fig. 2. The purpose of carrying out the training process for this method is to obtain new weights that are processed to get the best weights or final weights for the testing process [19].

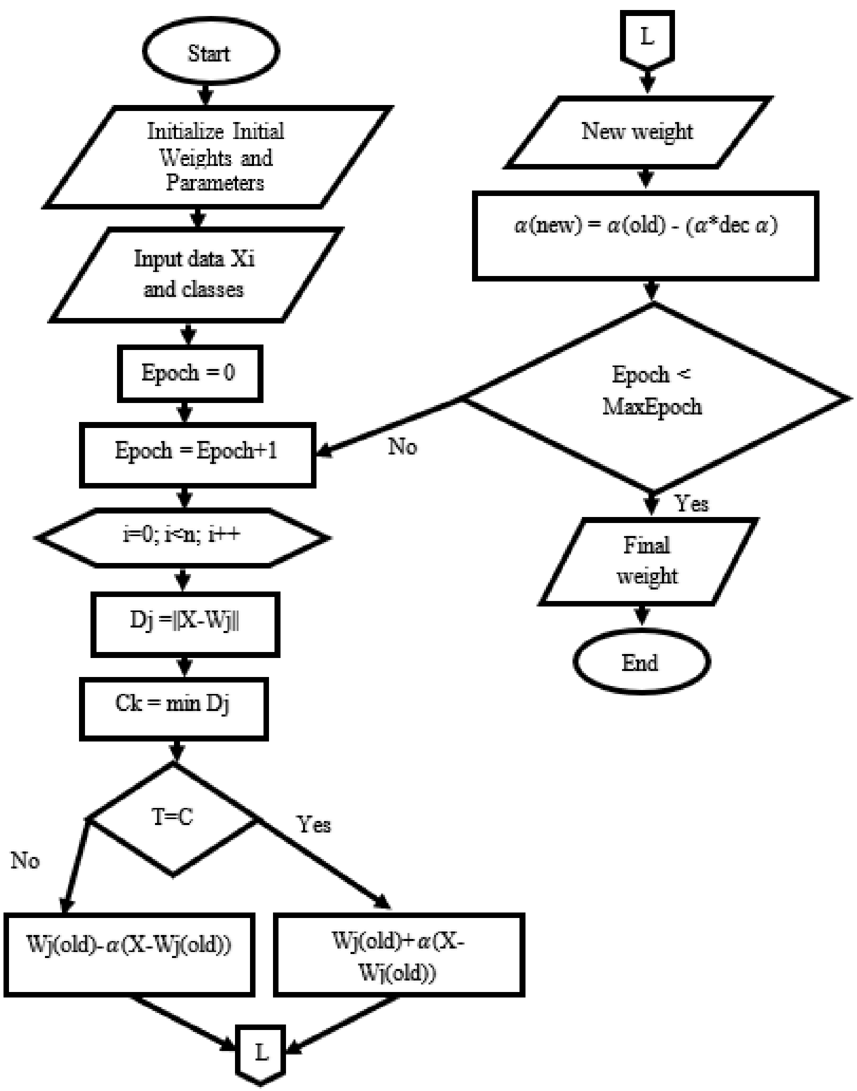

Fig. 2. Training Flowchart

Based on Fig. 2, it can be explained that the Learning Vector Quantization (LVQ) algorithm training process goes through the following stages:

- Initialized mark subtraction learning rate (Dec α), learning rate (α), value greatest epoch as well mark the smallest learning rate (Min α) will determine the weight start on each target or class.
- Upload input data and categories or target class.
- If ($epoch \leq$ maksimal $epoch\ and\ \alpha \geq$ minimal α) then *Epoch value = epoch + 1*
- For the most accurate determination of the shortest distance, utilize the Euclidean distance using Equation (1). Where the *D* variable is Euclidean, then x means input, and w means weight.

$$D_i = \sqrt{\sum_{j=1}^{n}(x_j - w_{ij})^2} \tag{1}$$

- Update weight To use repair mark weight in condition :
  if $T = C_j$ then $w_j(new) = w_j(old) + \alpha(x_i - w_j(old))$
  else If $T \neq C_j$ then $w_j(new) = w_j(old) - \alpha(x_i - w_j(old))$
- Do reduction in the value of α through way using $\alpha\ (new) = \alpha\ (old) - (\alpha * dec\ \alpha)$.
- The process will stop if the learning rate (α) value is minimal or Already produced the maximum epoch.
- After the training process, mark weight (*w*) will be obtained.

The process of calculating the Learning Vector Quantization (LVQ) training method has been completed, and it will go through the testing phase in Fig. 3.

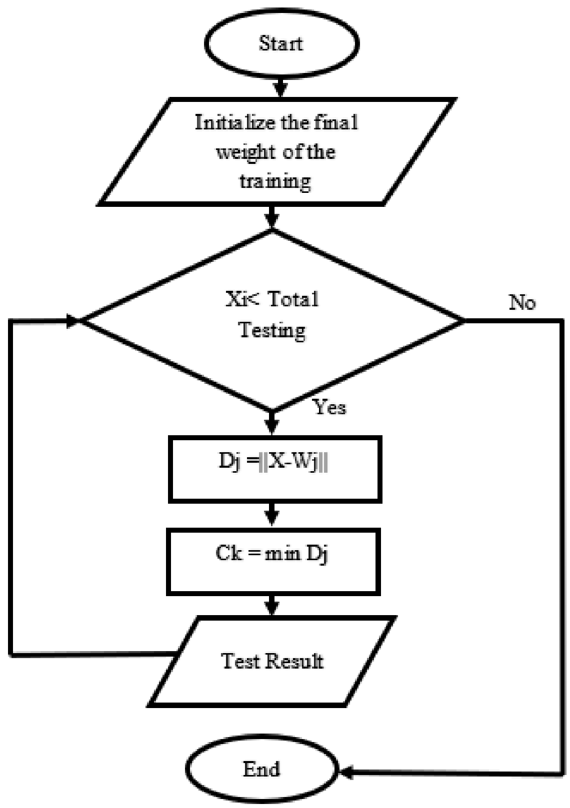

Fig. 3. Testing Flowchart

This illustrates that when starting the research, initializing the final weights of the training in the previous image, then initializing the initial weights & parameters then, inputting data Xi and class, then performing calculations using the Euclidean Distance formula to obtain test results regarding eligibility for families receiving Non-Cash Food assistance (BPNT), namely code 2 which is included in the Ineligible class and code 1 which is included in the Eligible class.

The classification results will be collected to find models from the training dataset capable of appropriately classifying or categorizing records. After getting the classification results from the testing phase, then carry out the calculation process to determine the data's accuracy level. The stages in calculating accuracy in Fig. 4. Accuracy value is to find the accuracy value in processing data so that performance is obtained in processing the data, in addition to finding the best performance of the algorithm that can be used.

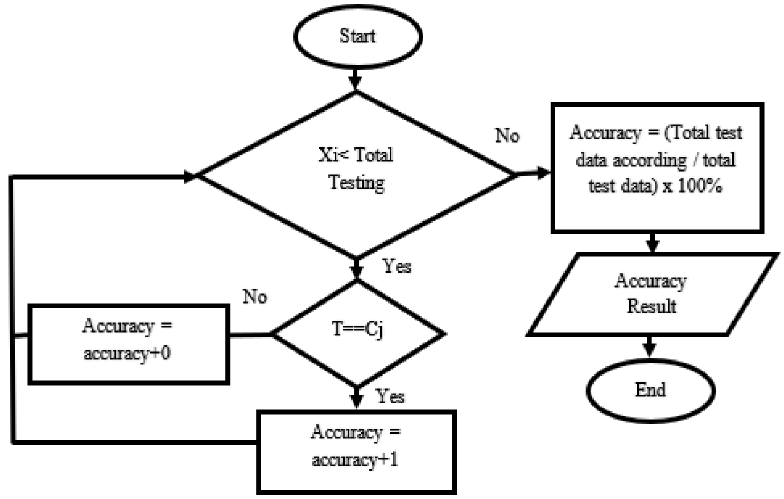

Fig. 4. Flowchart Accuracy

III. RESULTS AND DISCUSSION

The initial process is carried out by collecting the data needed as a reference in the implementation and the weight of each calculation process using the LVQ method for BPNT classification using the Python programming language. The data used were 200 datasets, and the weights used were 14 parameters. Based on Table I, the parameters used in this study are 14 parameters that can be initialized with the P-code, namely P1 to P14. So the data used in this study uses these parameters to carry out the LVQ calculation process. can be seen in Table II, which contains a sample of the data used in this study.

TABLE I
PARAMETERS

Code	Parameters
P1	The area of the base of the building where each person lives is less than 8 meters.
P2	The house's base is made of bamboo or dirt or cheap wood.
P3	The walls are not plastered and are made of low-quality bamboo or wood.
P4	There are no bath wash toilet facilities.
P5	There is no electricity as a source of household lighting.
P6	The main drinking water sources are unprotected wells, springs, rivers, or rainwater.
P7	Kerosene, firewood, and charcoal are sources of daily energy for cooking.
P8	In a week, eat meat or chicken or milk once.
P9	Within a year, have one new dress.
P10	Eat in a day only once or twice.
P11	Cannot afford medical expenses at the polyclinic.
P12	Source of income for the head of household: Farmers, fishermen, plantation workers, construction workers, and/or others who earn less than IDR 600,000 per month, including farmers with a land area of 0.05 ha.
P13	Head of household with the highest level of education: never attended school, did not finish elementary school, or only finished elementary school.
P14	Do not have savings or goods that can be easily sold for a minimum of IDR 500,000, such as gold, credit, non-credit motorcycles, boats, livestock, or other capital goods.

In Table II, codes 1 and 0 are in the input data in each parameter column. Code 1 means "Yes" and code 0 means "No". Then there is the Class column, which means that the 14 parameters produce an Eligible or Ineligible classification which is initialized to codes 1 and 2, namely code 2, which is included in the Ineligible class, and code 1, which is included in the Eligible class.

Implementation is the result of research on the system that has been created. The first thing to do is enter all the datasets into the Python Programming Language. The dataset in this study was made in Excel form, which will later be called using the Python programming language. The data used in this study are 200 datasets that contain the parameters of each input and its class. The dataset is divided into two parts: training data and testing data. The training data will be used as much as 80% of the 160 rows or data dataset. Then test data will be used, which is 20% of the total dataset totaling 40 rows of data. Data is distributed randomly.

E - ISSN : 2686-6269

TABLE II
DATASET

Number	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	Class
1.	0	1	0	1	0	1	0	1	1	0	1	1	0	0	2
2.	1	0	0	1	0	1	1	1	0	1	1	1	1	0	2
3.	1	0	1	1	0	0	1	1	1	1	1	0	1	1	1
4.	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1
5.	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1
6.	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1
7.	1	1	0	1	0	0	1	0	0	1	0	1	1	0	2
8.	1	0	1	0	0	1	1	0	1	0	1	0	1	1	2
9.	0	1	1	1	1	1	0	0	0	1	0	0	0	0	2
10.	1	1	1	1	0	1	0	1	0	1	1	0	1	1	1

The training data uses 160 datasets and 15 columns consisting of 14 columns for parameters and 1 column for classes. The training data will be used in the training calculation process to get the best or final weight for the testing calculation process. Then the data testing used in this study is 40 datasets selected randomly and 15 columns composed of 14 columns for parameters and 1 column for classes. The test data will be used in the test calculation process to get the classification results on the Learning Vector Quantization (LVQ) method. The initial weights are only 2 data taken randomly; 200 datasets are obtained because the first weight is taken, which represents vector 1 for class 1, which means feasible, and vector 2 for class 2, which means not feasible. The training process on the Learning Vector Quantization (LVQ) method is carried out after initializing the initial vector weights, initializing the max epoch value, and the learning rate value that will be used in the training process. The max epoch value used in this study uses three epochs and a learning rate of 0.1. Then perform the Learning Vector Quantization (LVQ) calculation process, which will stop if the learning rate (α) is minimum or has produced a maximum epoch. After all the training calculation processes are complete, the best vector weight or final sign weight (w) will be obtained, which will be used as the initial weight in the testing process to produce the classification value. Then calculate the accuracy value with the confusion matrix, the results of which can be seen in Fig. 5.

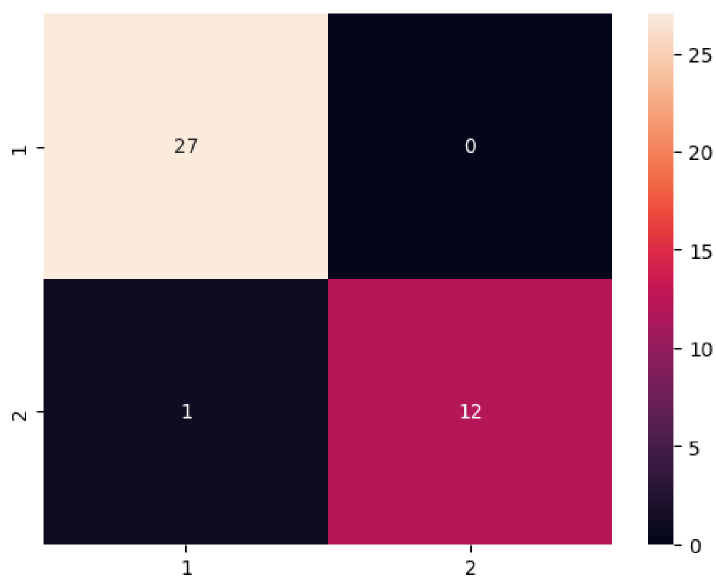

Fig. 5. Confusion Matrix

The confusion matrix measures problems in which machine learning output classifications can form two or more classes [20]. Fig 5 displays an image of the Confusion Matrix. In class 1 classification and actual class 1, there are 27 appropriate data visualized in pink. Then for class 2 classification, and actual class 1 is 0, or nothing is visualized in black. Class 1 and class 2, 1 data is not visualized with black sunny color. Then in class 2 and actual class 2, there are 12 appropriate data visualized in magenta. Accuracy value generated in the trial process using Equation (2).

$$Accuracy = \frac{The\ total\ test\ data\ is\ appropriate}{Total\ of\ all\ test\ data} x\ 100\% = \frac{39}{40} x\ 100\% = 97.5\%. \quad (2)$$

DOI: http://dx.doi.org/10.25139/ijair.v5i1.6287

E - ISSN : 2686-6269

So, got is known that mark accuracy was obtained using data testing as many as 40 datasets through a calculation using the LVQ method is 97.5%.

	precision	recall	f1-score	support
1	0.96	1.00	0.98	27
2	1.00	0.92	0.96	13
accuracy			0.97	40
macro avg	0.98	0.96	0.97	40
weighted avg	0.98	0.97	0.97	40

Fig. 6 Display of Accuracy Value

Based on the entire calculation process, the eligibility classification for non-cash food assistance recipients using the LVQ method, the accuracy in Equation (3) of the rounding number results is 0.97 or 97%. Fig.5 explains precision or accuracy. The ratio of positive data classified as True to the total number of positive classification results. To get precision, use Equation (4) for class one and Equation (5) for class two.

$$accuracy = \frac{(first\ class\ result + second\ grade\ result)}{2} = \frac{(0{,}98+0{,}96)}{2} = \frac{1{,}94}{2} = 0.97 \quad (3)$$

$$precision = \frac{True\ Positive\ (TP)}{True\ Positif\ (TP)+False\ Positif\ (FP)} = \frac{27}{(27+1)} = \frac{27}{28} = 0.96 \quad (4)$$

$$precision = \frac{True\ Positive\ (TP)}{True\ Positif\ (TP)+False\ Positif\ (FP)} = \frac{12}{(12+0)} = \frac{12}{12} = 1.00 \quad (5)$$

It can also be explained in the figure that the recall value called True Positive Rate or Sensitivity is the ratio of the number of correctly classified positive data (TP) to the number of data that is actually positive. To get the recall value, use Equation (6) for class one and Equation (7) for class two.

$$recall = \frac{\text{True Positive (TP)}}{\text{True Positif (TP)+False Negative(FN)}} = \frac{27}{(27+0)} = \frac{27}{27} = 1.00 \quad (6)$$

$$recall = \frac{\text{True Positive (TP)}}{\text{True Positif (TP)+False Negative (FN)}} = \frac{12}{(12+1)} = \frac{12}{13} = 0.92 \quad (7)$$

From these calculations, the precision and recall values (f1_score) can be generated using Equation (8) for class one and Equation (9) for class two.

$$f1\_score) = \frac{(precision\ value+recall\ value)}{2} = \frac{(0{,}96+1{,}00)}{2} = \frac{1{,}96}{2} = 0.98 \quad (8)$$

$$f1\_score) = \frac{\text{(precision value+recall value)}}{2} = \frac{(1{,}00+0{,}92)}{2} = \frac{1{,}92}{2} = 0.96 \quad (9)$$

IV. CONCLUSIONS

Based on the tests conducted on the system with the recommendation of using a learning rate value of 0.1, the amount of training data is 80%, and the testing data is 20%. The maximum epoch value is three times the value, and the accuracy of BPNT recipient eligibility classification using the LVQ method is 97.5%. Future research uses a method other than the LVQ method or compares two methods in one case study. For example, use LVQ and K-Nearest Neighbor methods. Optimize with other methods to study this to obtain a more effective and efficient information processing or presentation system. For example, the LVQ method is optimized with a genetic algorithm (GA) or Particle Swarm Optimization (PSO). Research that is expected to be further improved is the amount of data when testing the data to be used because the more data, the more accurate the results.

DOI: http://dx.doi.org/10.25139/ijair.v5i1.6287

REFERENCES

[1] M. Ahmad Yanda, "Penentuan Penerima Bantuan Pangan Non Tunai (BPNT) Menerapkan Metode Multi Objective Optimization on the Basis of Ratio Analysis (MOORA)," *Bulletin of Informatics and Data Science*, vol. 1 No 2, pp. 38-45, 2022.

[2] D. A. N. F. Khotim Fadhli, "Pengaruh Pendapatan, Pendidikan, dan Gaya Hidup Terhadap Kesejahteraan Keluarga Penerima Manfaat (KPM) Bantuan Sosial Covid-19," *Journal Education and Development,* vol. 9 No 3, pp. 118-124, 2021.

[3] E. Ermawati, "Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization untuk Prediksi Penerima Bantuan Pangan Non Tunai," *SISTEMASI : Jurnal Sistem Informasi,* vol. 8 No 3, pp. 513-528, 2019.

[4] Peraturan Presiden Republik Indonesia Nomor 63 Tahun 2017, Penyaluran Bantuan Sosial Non Tunai, 2022.

[5] Badan Pusat Statistik, Kemiskinan dan Ketimpangan Pendapatan, 2021.

[6] Badan Pusat Statistik, Penerima Bantuan Pangan Non Tunai (BPNT)/Program Sembako, Sistem Informasi Statistik, 2022.

[7] O. Nurdiawan, "Seleksi Penerima Bantuan Sosial Berdasarkan Sistem Pendukung Keputusan Dalam Upaya Mengurangi Siswa Rawan Putus Sekolah," *Jurnal Teknologi Informasi*, vol. 8 No 2, pp. 32-40, 2018.

[8] I. H. K. A. S. E. M. A. A. P. J. Aryo De Wibowo Muhammad Sidik, "Gambaran Umum Metode Klasifikasi Data Mining," *FIDELITY : Jurnal Teknik Elektro,* vol. 2 No 2, pp. 34-38, 2020.

[9] S. Wang Hoaxiang, "Big Data Analysis and Pertubation using Data Mining Algorithm," *Journal of Soft Computing Paradigm (JSCP),* vol. 3 No 1, pp. 19-28, 2021.

[10] M. H. W. E. P. F. F. W. W. G. Riyan Latifahul Hasanah, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (K-Nearest Neighbor)," *Jurnal TECHNO Nusa Mandiri*, vol. 16 No 1, pp. 1-6, 2019.

[11] M. T. F. C. D. Romlah Tantianti, "Implementasi Metode Learning Vector Quantization (LVQ) untuk Klasifikasi Persalinan," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 3 No 10, pp. 9701-9707, 2019.

[12] E. S. B. R. Ivan Agustinus, "Klasifikasi Risiko Hipertensi Menggunakan Learning Vector Quantization (LVQ)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2 No 8, pp. 2947-2955, 2018.

[13] R. R. H. I. Muhammad Rianto, "Penerapan Data Mining Dengan Metode Naive Bayes Dan Learning Vector Quantization Credit Rating Dalam Memprediksi Kelayakan Pemberian Kredit Oleh PT. BPR Lebak Sejahtera," *Jurnal Teknologi Informasi,* vol. 17 No 1, pp. 69-76, 2022.

[14] O. N. Washilaturrizqi, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Bantuan Sosial (Studi Kasus: Klasifikasi Kemiskinan Desa Lurah)," *JATI (Jurnal Mahasiswa Teknik Informatika),* vol. 7 No 1, pp. 373-377, 2023.

[15] Kementrian Sosial, Pedoman Bantuan Pangan Non Tunai, 2019.

[16] "Implementasi Program Bantuan Pangan Non Tunai (BPNT) Di Kecamatan Sitinjau Laut Kabupaten Kerinci," *Qawwam : The Leader's Writing,* vol. 1 No 1, pp. 38-43, 2020.

[17] H. F. Nugrahana Fitria Ruhyana, "Strategi Kebijakan Pemerintah Kabupaten Sumedang Dalam Penyaluran Bantuan Sosial Di Masa Pandemi Covid-19," *Khazanah Intelektual,* vol. 4 No 2, pp. 789-804, 2020.

[18] U. R. H. R. M. A. Amat Damuri, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer),* vol. 8 No 2, pp. 219-225, 2021.

[19] E. S. B. R. Ivan Agustinus, "Klasifikasi Risiko Hipertensi Menggunakan Metode Learning Vector Quantization (LVQ)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2 No 8, pp. 2947-2955, 2018.

[20] A. B. Mercury Fluorida Fibrianda, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naive Bayes Dan Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 2 No 9, pp. 3112-3123, 2018.

DOI: http://dx.doi.org/10.25139/ijair.v5i1.6287