

## Implementation of the Naive Bayes and Simple Additive Weighting Methods in the Feasibility Selection of Village Cash Assistance Recipients

Bain Khusnul Khotimah<sup>1\*</sup>, Wahyu Indra Kustina<sup>2</sup>, Yeni Kustiyahningsih<sup>3</sup>, Devie Rosa Anamisa<sup>4</sup>, Fifin Ayu Mufarroha<sup>5</sup>

Department of Informatics Engineering, University of Trunojoyo Madura, East Java

**Corresponding Author:** Bain Khusnul Khotimah [bainkk@trunojoyo.ac.id](mailto:bainkk@trunojoyo.ac.id)

---

### ARTICLE INFO

*Keywords:* VF-DCA, Decision Support Systems, Methods, NB, SAW

*Received :* 15 March

*Revised :* 20 April

*Accepted:* 22 May

©2023 Khotimah, Kustani, Kustiyahningsih, Anamisa, Mufarroha: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



### ABSTRACT

Village Fund Direct Cash Assistance (VF-DCA) is a direct cash assistance policy in which funds come from the village. The VF-DCA policy is expected to ease the burden on the community, especially those with a low economy. Still, it is undeniable that there are opportunities for misuse of village funds by some VF-DCA organizers, intentionally or unintentionally. Problems that arise in implementing VF-DCA are mistakes in determining the beneficiary communities that are not on target so that the impact of VF can be manipulated for the interests of certain groups in several cases of social assistance programs. The solution to these problems is that this research creates a decision support system in selecting the eligibility of assistance recipients. This research used the Naïve Bayes (NB) method in selecting VF-DCA recipients and the Simple Additive Weighting (SAW) way to determine the order of eligibility scores for each candidate for VF-DCA recipients. In this study, the results of the classification process using the NB method obtained an accuracy rate of 99.36%. Hence, this is simplified in determining potential recipients of VF-DCA using the NB and SAW methods

## **INTRODUCTION**

Poverty is a severe problem that is still being experienced by several countries around the world. The Indonesian government itself makes many programs and policies to overcome the problem of poverty. One such program is the VF-DCA. Bandung Village is one of the villages in Konang District, Bangkalan Regency, which has benefited from BLT-DD. Providing VF-DCA by the government to people experiencing poverty is one way to counteract the impact of COVID-19 on rural communities. Both the Village Minister's PDT Instruction Number 1 of 2020 Concerning the Acceleration of VF-DCA Distribution and the Minister of Finance's Regulation Number 50/PMK.07/2020 Concerning the Second Amendment to the Minister of Finance's Regulation Number 205/PMK.07/2019 Concerning Village Fund Management, regulate direct assistance policies cash funds originating from village funds or VF Dana Desa. Even though the VF-DCA policy is intended to help people with financial difficulties, it is still possible that village funds will be misused by various actors involved in its implementation. Problems that arise in the performance of VF-DCA include errors in determining the beneficiary community that is not on target to the point that the impact of VF distribution is not as expected with predetermined goals so that it can be manipulated for the interests of certain groups in some cases of social assistance programs(Herdiana et al., 2021).

Compare the Fuzzy C-means Algorithm and the Naive Bayes Algorithm to identify beneficiaries with the lowest socioeconomic status. It can be concluded from this study that Naïve Bayes is more effectively used to determine beneficiary families(Saputra, 2021). The Naive Bayes method, when applied to accurate data, is generally only used for validation. Then, a decision support system is needed to produce a decision, one of which is by using the SAW method.

Decision support systems' use in the decision-making process for prospective VF-DCA recipients, according to a study conducted in 2021, which compared the SAW and Fuzzy methods used in VF decision support systems, the two ways obtained equally accurate results. However, the first method was more effective regarding algorithms than the second one(Ningtyas & Suyatno, 2021). In 2019, implementing the Naive Bayes and SAW methods on category data to select diabetes mellitus menus. The results of this study get the best alternative accuracy results of 85%, and the system is considered to give the best results(Nasution et al., 2019).

Based on this background, the researcher proposes to use the NB method in selecting VF-DCA beneficiaries and the SAW method to determine the ranking of prospective VF-DCA recipients based on the weighting of the eligibility criteria specified by the village head, with the type of data used in this study discrete and categorical data. NB is a technique from the field of Machine Learning (ML) that utilizes probability and statistical calculations developed by British scientist Thomas Bayes to make predictions; the advantages of NB are that it is easy to implement, requires a small amount of data for the classification process and faster in the computing process(Arfanda et al., 2021b). The SAW method was chosen as the development method from previous research. SAW itself is known as the weighted sum method. The SAW method has an advantage

over other decision support system methods because it can assign relative importance to different criteria and weights based on interest in the research. The SAW method is a decision-making approach that calculates the weight value of each bar in evaluating alternatives so that it can make the right and accurate decisions (Arfanda et al., 2021b).

## METHODS

### A. NB Method

From 1702-1761, Reverend Thomas Bayes popularized what became known as the NB method (Socrates et al., 2016). NB Method, or Bayes Theorem, can make predictions based on historical data. The advantages of NB are that it is easy to implement and is included as an ML method by utilizing probability and statistical calculations. To determine the classification, only the variance of a variable in a class is needed because it is assumed to be an independent variable (Syarli & Muin, 2016). The equation defined in Bayes' Theorem using Equation 1 (Arfanda et al., 2021a).

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Where X is Data of unknown class, C is the data hypothesis is class specific,  $P(C|X)$  is hypothesis probability C based on conditions X (posterior probability),  $P(C)$  is hypothesis probability C (prior probability),  $P(X|C)$  is the probability of a hypothesis X based on conditions C.

The Naïve Bayes calculation steps can be done as follows (Syarifudin, Hidayat and Fanani, 2018):

1. By using Equation 2 to determine the class mean, we can determine the prior values for each class.

$$P = \frac{X}{C}$$

Where P is Prior Value, X is the amount of data per class, C is Total data for all classes.

2. Determine the likelihood for each category using Equation 3.

$$P = \frac{A}{B}$$

Where P is Likelihood Value, A is Number of data features for each class, B is The total number of features of each class.

3. Using the existing class, calculate the posterior value using equation 4.

$$P(C|A) = P(C) \cdot P(A|C)$$

Where  $P(C)$  is prior value of each class,  $P(A|C)$  is likelihood value.

When using the NB method for classification, the results are compared with the posterior values for each class. Classification results are determined by the highest posterior value.

### B. SAW Method

The SAW method is a weighting criterion that considers benefits and costs. The concept of the SAW method is to calculate the weighted sum of the ratings given to each alternative on all attributes. The decision matrix (x) must be

normalized before the SAW method can be applied to compare one set of ratings with another (Pertwi, Fedinandus and Limantara, 2019).

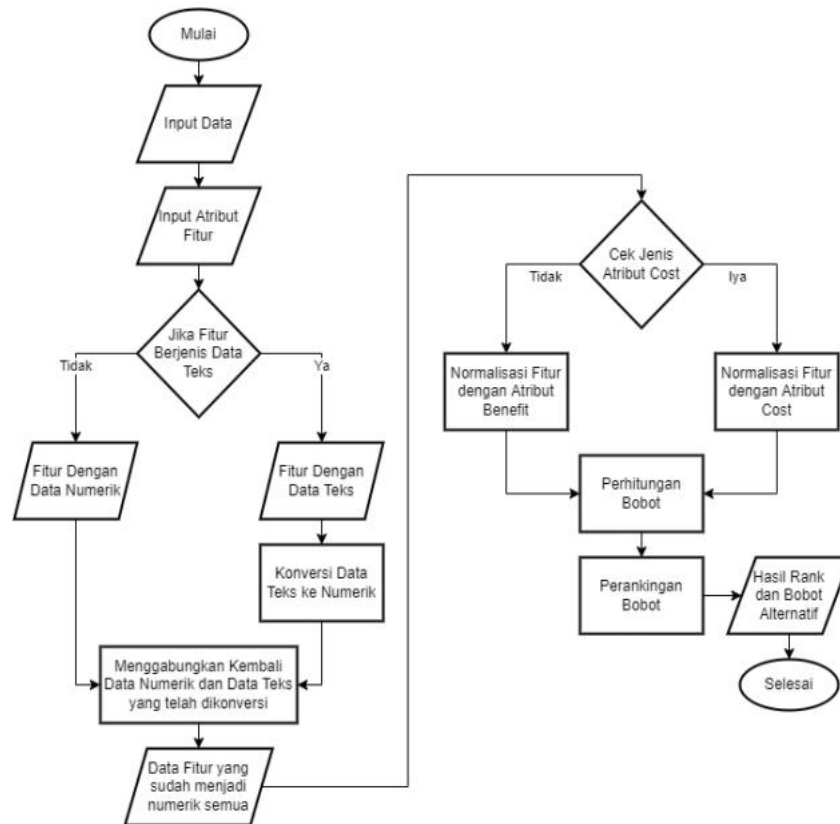


Figure 1. SAW Flowchart

There are three stages in performing calculations using the SAW method, including:

1. Analysis, at this stage determines whether the type of criterion is a benefit or cost, and changes all attribute values based on the values in the data crips. If the attribute does not have crips data, then the original data is directly entered.
2. Normalization, this stage is used to change the value of each attribute to a 0-1 range by focusing on the type of criteria, be it benefit / cost, you can use equation 5.

$$r_{ij} = \begin{cases} \frac{X_{ij}}{\text{Max}_i X_{ij}} & \text{if } j \text{ is the gain attribute} \\ \frac{\text{Min}_i X_{ij}}{X_{ij}} & \text{if } j \text{ is the cost attribute} \end{cases}$$

3. At this stage, all attributes are multiplied by the weight of the criteria for each alternative; you can use Equation 6.

$$V_{ij} = \sum_{j=1}^n w_j r_{ij}$$

Where  $V_{ij}$  is ranking of each alternative,  $w_j$  is weight value for each criterion,  $r_{ij}$  is normalized performance rating values.

The SAW method is more capable of performing an accurate evaluation because it is based on the criteria values and the required weights. Furthermore,

the SAW method can select alternatives and then carry out a ranking process from the selection results based on the total weight value of all criteria.

## RESULTS AND DISCUSSION

Testing for accuracy at this stage is testing the model, which begins with classification and continues with ranking.

### A. Dataset

The dataset used for eligibility selection for VF-DCA recipients is data obtained from the Bandung Village Hall, Konang District, Bangkalan Regency in April 2020 totaling 623. The dataset has two classes, namely those that meet the requirements as VF-DCA recipients and do not meet the requirements as recipients of VF-DCA, with several criteria used. Description of each of these criteria are:

1. Gender Criteria is a variable with two categories, namely male and female. The head of the household with the female category is the husband or the male head of the household who has died with the aim of being able to prioritize female family heads who are the backbone of the family.
2. Criteria for Many Family Members: Number of dependents which is the number of family members financed.
3. Total Criteria for Poor People: This is the criteria for poor families according to the Ministry of Social Affairs.
4. Social Safety Net Recipients: Divided into four groups, namely Family Hope Program beneficiaries, Non-Cash Government Assistance beneficiaries, Cash Social Assistance beneficiaries, and non-Social Safety Net beneficiaries (non-beneficiaries).
5. Not Received Social Safety Net is a variable with the status of a family condition with four categories: loss of livelihood, unreported, chronic illness, and having no criteria.

Table 1. Alternative Criteria

Code	Criteria	weight
A01	gender	0.1
A02	number of family members	0.2
A03	poor family category	0.3
A04	group of recipients of social security	0.5
A05	not a group receiving social security	0.3
A06	recommendation	0.2

A. System Architecture

In Figure 3, a system design flow explains the system's running process, starting with the data input process for prospective VF-DCA recipients. Furthermore, the data will enter the preparation stage and be classified using Naïve Bayes. The classification process will display the output data of the classification results. The classification model created will enter the model testing phase and display the accuracy level output. Finally, the ranking process is carried out using SAW and will show the results of the data output given a rank.

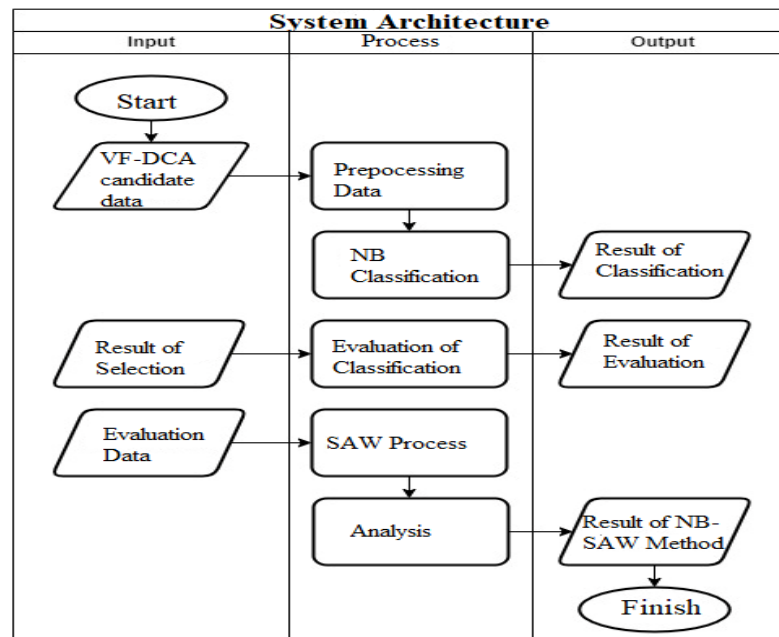


Figure 2. Combined Architecture of NB Classification and SAW Ranking

B. Naïve Bayes Accuracy Test Results

Testing for accuracy at this stage is testing the model using a system created using the VF-DCA beneficiary dataset obtained from Bandung Village. As many as 623 data will be processed using the NB algorithm. The dataset will be divided into two types: training data and testing. The training data will be used for learning from the NB algorithm. In contrast, the Testing data will be used as trial data to determine whether the classification results are appropriate. Based on the test results made using 5-fold Cross-Validation and the Confusion Matrix, the accuracy value for each fold is obtained, summarized in Table 1.

Table 2. Results 5 Fold Cross Validation

<b>Fold</b>	<b>Parameter</b>	<b>Hasil %</b>
Fold 1	Accuracy	99.2
Fold 2	Accuracy	99.2
Fold 3	Accuracy	99.2
Fold 4	Accuracy	99.2
Fold 5	Accuracy	100
<b>Overall Accuracy</b>		<b>99.36</b>

From Table 2, the highest accuracy is obtained in the 5th fold with an accuracy value of 100%, while the accuracy of folds 1 - 4 gets an accuracy value of 99.2%. Overall accuracy obtained from the Confusion Matrix on 5-Fold Cross Validation has an accuracy value of 99.36%.

The value of the Confusion Matrix results is also obtained from the total of all folds, which records the results of the classification – a confusion Matrix can be used to determine the classification errors' location from the Naive Bayes model. Table 2 describes the Confusion Matrix obtained from the evaluation results of the Naive Bayes model. Then calculate the accuracy can be seen in Table 3.

Table 3. Confusion Matrix Classification Results

		<b>TARGET</b>	
		<b>Eligible</b>	<b>Not Eligible</b>
<b>Prediction</b>	<b>Class</b>		
	Eligible		107
Not Eligible		4	512

Table 4. The Final Value of TP, FP, TN and FN

<b>Mark</b>	<b>Amount</b>
TP	107
FP	0
TN	512
FN	4

Where TP (True Positive) is data that meets the conditions that are predicted to be correct by 107, FP (False Positive) is data that does not meet the requirements that are expected to be wrong by 0, TN (True Negative) is data that does not meet the needs that are predicted to be correct by 512, FN (False Negative) is data that meets the expected conditions do not meet the requirements.

The ranking results were carried out using the SAW method with the number of alternative VF-DCA recipient data of 107 data obtained from the results of the classification process, so there is a ranking value from the highest to the lowest. The alternative with the highest ranking value is alternative 82 = 1.00, which is named Misnatun, and the lowest is alternative 443 = 0.6106, Marsiyah. Based on the data from alternative calculations carried out with predetermined criteria, the order or rank value of the alternative VF-DCA recipients is obtained.

## **CONCLUSIONS AND RECOMMENDATIONS**

Based on the results of the discussion that has been done, it can be concluded in this study, namely:

1. The performance of the Naive Bayes method can be applied well to determine the eligibility selection of prospective VF-DCA recipients in Bandung Village. Testing the validity of the Naive Bayes method using 5-fold cross-validation obtains an accuracy value of 99.36%.
2. Based on the results of applying the SAW method in determining the order value of VF-DCA recipients, an alternative with the highest rank value is alternative 82 = 1.00, while the lowest rank value is alternative 443 = 0.6106 as VF-DCA recipients. The results of system calculations get the same results as manual calculations.

In this study, there are still some shortcomings. Therefore, the authors suggest further research that can develop a decision support system for selecting the eligibility of VF-DCA recipients with more features. This simple system prioritizes the core of the Naive Bayes method and the SAW calculation process.

## REFERENCES

- Arfanda, I., Ramdhan, W. and Yusda, R.A. (2021a) 'Digital Transformation Technology ( Digitech ) | e-ISSN: 9999-9999 Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai Digital Transformation Technology ( Digitech ) | e-ISSN: 9999-9999', 1(1), pp. 9-16.
- Arfanda, I., Ramdhan, W. and Yusda, R.A. (2021b) 'Naive Bayes Dalam Menentukan Penerima Bantuan Langsung Tunai', *Digital Transformation Technology*, 1(1), pp. 9-16. Available at: <https://doi.org/10.47709/digitech.v1i1.1091>.
- Herdiana, D. *et al.* (2021) 'Implementasi Kebijakan Bantuan Langsung Tunai ( BLT ) Dana Desa Bagi Masyarakat Terdampak COVID-19 di Kabupaten Sumedang : Isu dan Tantangan Implementation of Village Fund Direct Cash Assistance Policy for Villagers Affected by COVID-19 in Sumedang Regenc', *Jurnal Inspirasi*, 5(June), pp. 1-16. Available at: <https://doi.org/10.35880/inspirasi.v1i1.175>.
- Nasution, S.R., Andreswari, D. and Wahyu, T. (2019) 'Implementasi Naive Bayes Classifier Dan Simple Additive Weighting (Saw) Untuk Pemilihan Menu Diet Penyakit Diabetes Mellitus', *Jurnal Rekursif*, 7(1), pp. 1-9.
- Ningtyas, R.P.A.C. and Suyatno, D.F. (2021) 'Sistem Pendukung Keputusan Bantuan Langsung Tunai Pada Masa Pandemi Covid 19 Menggunakan Perbandingan Metode Simple Additive Weighting dan Fuzzy Berbasis Website (Studi Kasus: Desa Krisik, Kecamatan Gandusari, Kabupaten Blitar)', *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 2(2), pp. 56-65.
- Pertiwi, I.P., Fedinandus, F. and Limantara, A.D. (2019) 'Sistem Pendukung Keputusan Penerima Program Keluarga Harapan (PKH) Menggunakan Metode Simple Additive Weighting', *CAHAYAtech*, 8(2), p. 182. Available at: <https://doi.org/10.47047/ct.v8i2.46>.
- Saputra, P.S. (2021) 'Perbandingan Algoritma Fuzzy C-Means Dan Algoritma Naive Bayes Dalam Menentukan Keluarga Penerima Manfaat (Kpm) Berdasarkan Status Sosial Ekonomi (Sse) Terendah', *JST (Jurnal Sains dan Teknologi)*, 10(1), pp. 1-8. Available at: <https://doi.org/10.23887/jstundiksha.v10i1.23340>.
- Socrates, I.G.A. *et al.* (2016) 'Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio', *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 7(1), p. 22. Available at: <https://doi.org/10.24843/lkjiti.2016.v07.i01.p03>.
- Syarifudin, A., Hidayat, N. and Fanani, L. (2018) 'Sistem Pakar Diagnosis Penyakit Pada Tanaman Jagung Menggunakan Metode Naive Bayes Berbasis Android', 2(7).

Syarli, S. and Muin, A. (2016) 'Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)', *Jurnal Ilmiah Ilmu Komputer*, 2(1), pp. 22-26.