E-Komtek Journal
Vol. 8 No. 2 (2024) pp. 275-286
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN: 2580-3719 e-ISSN : 2622-3066

Implementation of Random Forest Algorithm for Classification of Eligibility For Social Assistance Recipients In Information Systems

Mita Trianda Putri[1], Triase[2]
[1-2]Information System, UIN Sumatra Utara, Indonesia, 20235

mitatriandaputri@gmail.com

https://doi.org/10.37339/e-komtek.v8i2.2197

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
21-08-2024
Revised:
23-12-2024
Accepted:
23-12-2024
Online first :
26-12-2024

Abstract

This study aims to develop a web-based information system for classifying the eligibility of social assistance (BANSOS) recipients using the Random Forest algorithm in the Bagan Batu Kota Subdistrict. The system is designed to assist local authorities in identifying BANSOS recipients more accurately and efficiently, minimize errors, and enhance distribution fairness. A quantitative research method was employed, with data collection techniques including observation, interviews, literature review, and document analysis. The dataset consists of 1,100 samples with features such as income, family size, and housing conditions. The Random Forest algorithm was implemented by building a classification model based on training and testing data. The evaluation showed a system accuracy rate of 97%, with a classification error of only 3%. The system provides features for recipient data management, field validation, and automated reporting, supporting more precise decision-making. The results of this study are expected to offer a solution for more effective and transparent social assistance distribution.

Keywords: *Random Forest, Classification, Social Assistance, Data Mining, Bagan Batu Subdistrict*

Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem informasi berbasis web untuk mengklasifikasikan kelayakan penerima bantuan sosial (BANSOS) menggunakan algoritma Random Forest di Kecamatan Bagan Batu Kota. Sistem ini dirancang untuk membantu pemerintah daerah dalam mengidentifikasi penerima BANSOS secara lebih akurat dan efisien, meminimalisir kesalahan, dan meningkatkan keadilan dalam pendistribusian. Metode penelitian kuantitatif digunakan, dengan teknik pengumpulan data yang meliputi observasi, wawancara, studi literatur, dan analisis dokumen. Dataset terdiri dari 1.100 sampel dengan fitur-fitur seperti pendapatan, ukuran keluarga, dan kondisi rumah. Algoritma Random Forest diimplementasikan dengan membangun model klasifikasi berdasarkan data pelatihan dan pengujian. Evaluasi menunjukkan tingkat akurasi sistem sebesar 97%, dengan kesalahan klasifikasi hanya 3%. Sistem ini menyediakan fitur untuk manajemen data penerima, validasi lapangan, dan pelaporan otomatis, sehingga mendukung pengambilan keputusan yang lebih tepat. Hasil penelitian ini diharapkan dapat memberikan solusi untuk penyaluran bantuan sosial yang lebih efektif dan transparan.

Kata-kata kunci: Random Forest, Klasifikasi, Bantuan Sosial, Data Mining, Kecamatan Bagan Batu

1. Introduction

Advances in information technology have increased the volume of data generated in various sectors such as industry, healthcare, and government. Data mining techniques are used to extract valuable information from this big data. Data mining is a data collection process used to find patterns in data [1]. Data mining utilises statistics, mathematics, artificial intelligence, and machine learning to identify patterns and knowledge from large databases [2]. This data mining has been widely applied, such as the Application of Data Mining for Classification of Power Plant Equipment Performance Using the C4.5 Algorithm by Rendy Arianda Edwin (2022), and also widely applied to urban villages, such as the Application of Data Mining Clustering to the Economy in Wonorejo Village, Marpoyan Damai District, Riau Province using the K-Means Algorithm Method by Egye Santoso (2022). One data mining that will be applied is Bagan Batu Kota Village.

Bagan Batu Kota Urban Village is one of the urban villages in Bagan Sinembah District, Rokan Hilir Regency, Riau Province. This Kelurahan is one of the areas with a relatively high level of need for social assistance, meaning that many poor people need social assistance. Poverty is defined as the inability of individuals to fulfil their basic needs, such as food, clothing, shelter, education, and health [3]. The poor have an average monthly per capita expenditure below the poverty line [4]. Handling poverty requires comprehensive policies involving all relevant parties [5]. Based on the results of the pre-research of the location, this Kelurahan has a population of 3,375 families (Family Card). From this data, it was recorded that 1 100 families received social assistance per year.

Social assistance provided by the local government to the community in the form of funds or necessities of life is selectively chosen and adjusted to the financial capacity of the region. Aims to support local government programs and activities in improving community welfare with the principles of justice, togetherness, and public interest [6]. The state is obliged to guarantee the welfare of every resident, and every resident has the right to obtain welfare [7]. The types of assistance distributed by the kelurahan include BPNT (Non-Cash Basic Assistance), PKH (Family Hope Program), Elderly Assistance, and BLT (Direct Cash Assistance) [8]. The duration of receiving social assistance varies, as BLT is given every month, and PKH is given 4 times a year and is intended for people who are categorised as poor or poor. Inaccurate distribution of social assistance can be utilised by parties who are not entitled, resulting in financial losses for the state [9].

To ensure social assistance is properly on target, an information system is needed to more accurately and efficiently assist in the classification process. An information system can be defined as a system that collects, processes, stores, analyses, and disseminates information for specific purposes [10]. An algorithm is needed to maximise the performance of information systems. The above problem can be solved with an algorithm, namely *Random Forest*. Random Forest is an advanced form of decision tree in machine learning [11]. The *Random Forest* algorithm uses decision trees built from random samples [12]. Each tree in this model uses a subset of features to find the best data division, and the results are interpreted based on the majority of votes to determine the correct class [13]. *Random* Forest algorithms can classify large amounts of *imbalanced* data by providing good performance results and fast execution time [14]. *Random Forest* classification is done through tree merging by training, which is done to generate classification trees with many versions, which then combine them to obtain the final result, then in *Random Forest* the randomisation process to form a classification tree is not only done for sample data but also in taking predictor variables [15]. By implementing this algorithm, the information system can identify social assistance recipients who meet specific criteria to facilitate the decision-making process in providing social assistance. This system can produce more accurate estimates and can help reduce errors in classifying social assistance recipients.

The goal of the research is a system that can assist the village in classifying the eligibility of social assistance recipients for people experiencing poverty to minimise the inequality of social assistance recipients with a *web-based* information system equipped with automatic classification features using the *Random Forest* algorithm, print social assistance recipient data, validation from field officers, document management, to announcements and reports of social assistance recipients.

2. Method

In this study, the authors used quantitative methods. Quantitative research is a systematic scientific method for studying phenomena and causal relationships by collecting and analysing measured data. It is defined as a systematic investigation of phenomena by collecting data, emphasizing the use of statistical, mathematical, and computational methods in the development of science [16].

The data collection methods used in this research are as follows: (1) Observation: At this stage, data collection is carried out through direct observation of the research subject to take a

closer look at the activities carried out to complete this data. This observation was done by visiting the Bagan Batu Kota Urban Village Office. (2) Literature Study: this stage is carried out by collecting information from books, e-books, and journal articles related to the research. (3) Interview: At this stage, data is collected by asking the kelurahan and the community to provide the necessary information. By conducting interviews, the author can discover the problems and formulate the need to build an information system for classifying the eligibility of social assistance recipients. (4) Document Analysis: this stage is carried out to review the documentation provided. The documents obtained for analysis are data for determining social assistance, such as family size, age, income, type of house, house condition, number of school children, and others.

2.1 Random Forest Algorithm

The authors created a system in this study by implementing the *random forest* algorithm. Random Forest is an advanced form of decision tree in machine learning. [11] . *The Random Forest* algorithm uses decision trees built from random samples. [12] . Each tree in this model uses a subset of features to find the best data division, and the results are interpreted based on the majority of votes to determine the correct class. [13] . *Random* Forest algorithms can classify large amounts of *imbalanced* data by providing good performance results and fast execution time. [14] *Random forest classification is done by combining trees through training, which generates classification trees with many versions, which are then combined to obtain the final classification. Then, in a Random Forest, the randomisation process to form a classification tree is done for sample data and predictor variables [15].* The application of Random Forest consists of several stages: data collection, dataset sharing, decision tree formation, random forest formation, and classification using random forest.

a) Data Collection Stage

   The data used in this research is data on prospective social assistance recipients from Bagan Batu Kota urban village. The data consists of 1100 records comprised of several features such as name, gender, family size, income, type of house, house condition, and number of school children. This data will create a random forest model to classify prospective beneficiaries.

b) Dataset Sharing Stage

   The data collected at this stage will be divided into two parts: learning data and testing data. The learning data will be used to form random forest models for classifying prospective social assistance recipients, and the testing data will help test the accuracy of the random forest model. This study's dataset division consists of 80% learning data and 20% testing data.

c) Decision Tree Formation Stage

Several decision trees will be formed from existing data in a random forest. In this study, the authors will divide the learning data into four parts to create four different decision trees, each formed from 100 data. The author uses the CART method to make a decision tree by selecting the root node. In selecting the root node in the CART method, we will use Gini Impurity, where the Split with the smallest Gini Impurity value will be used as the Root Node in the decision tree. The formula for Gini Impurity is as follows.

$$\text{Gini impurity} = 1 - \sum_{i=0}^{c-1} p_i^2$$

Description:

Pi = the proportion of items with class label i in a node.

After knowing the impurity value of each sub-set, then calculate the impurity value of each split using the formula:

$$Gini(Split) = \sum_{0}^{k-1} \left(\frac{ni}{n}\right) Gini(SubSplit)$$

After knowing all the gini values of each split, the next step is to find the smallest gini value to become the root node. Next, the Gini values of each split are looked at in the previous split value until only one classification is left to receive social assistance.

2.2 System Development Method

In this study, the information system was developed using the *Waterfall* method. Application development with this method is carried out sequentially (sequentially) following five stages: requirements analysis, design, implementation, testing and verification, and maintenance **[17]**. To avoid repetition of stages, each step must be completed first so that the system development can achieve the desired results **[18]**.

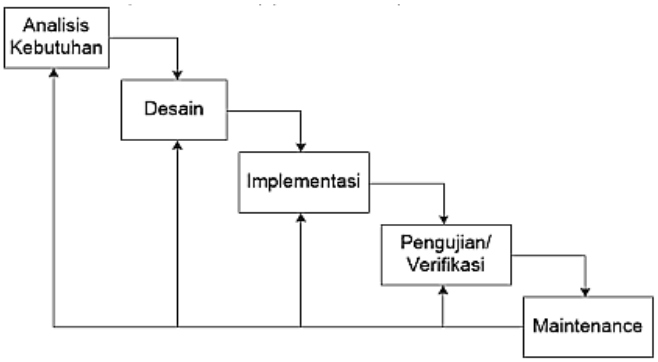

Figure 1. Stages of the Method **[19]**

3. Results and Discussion

In this regard, previous studies become the author's reference in this study, namely Andri

Nova Riswanto and Derman Janner Lubis's (2024) Implementation of Random Forest Algorithm for Optimizing One-Time Password (OTP) Authentication Security. Then another journal article by Aldiyansyah, Ade Irma Purnamasari and Irfan Ali (2024) Comparison of Accuracy Levels of Decision Tree and Random Forest Algorithms in Classifying Recipients of BPNT Social Assistance in Slangit Village and research by Ilham Kurniawan, Duwi Cahya Putri Buani, Abdussomad, Widya Apriliah, and Rizal Amegia Saputra (2023) Implementation of Random Forest Algorithm to Determine Raskin Recipients. The three reference journals show that Random Forest performs well, but the article does not develop information systems. The research only focused on one type of assistance in the second and third reference articles. So, in this study, the author aims to build an information system by implementing the Random Forest algorithm in different contexts and variables, such as the addition of types of social assistance so that it can be known whether those who receive social assistance have previously received social assistance or not at all.

3.1. System Design

System design provides an overview of the system and the information to be built. The following system design is designed.

Usecase Diagram

From the use case described, the design system will have two actors: the admin and the officer. Admins can manage data such as profile and user data using the system and data on prospective social assistance recipients. The other actor is the officer; the officer functions to verify the suitability of field data with data reported by residents. Citizen data that must be verified will be in the recipient review data.

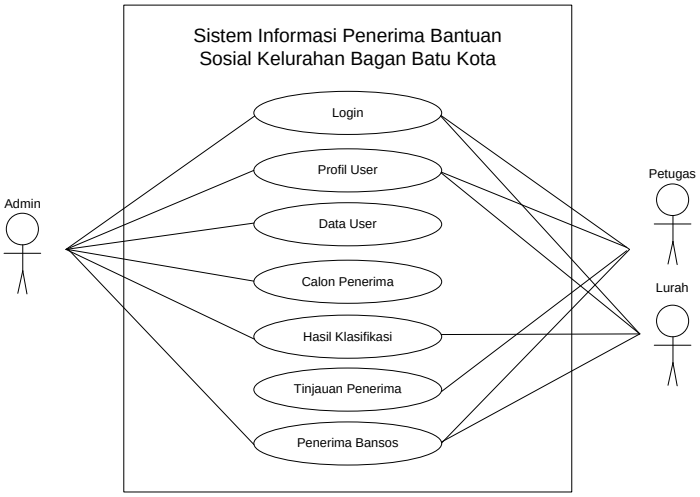

Figure 2. System Usecase Diagram

Activity Diagram

The activity diagram shows the process flow of interaction between actors and the system in carrying out a process. Here are some process flows in the system that are described using *activity diagrams.*

a. *Activity Diagram of* Prospective Recipient

In the Activity Diagram below, the process of determining the classification of prospective recipients of social assistance is known; at the beginning of the process, the admin will enter the system as an admin and register residents who apply for social assistance. After the prospective recipient data is entered into the system, the system will automatically classify the prospective recipient data as recipients or not recipients. Then, the system displays the updated prospective recipient data.

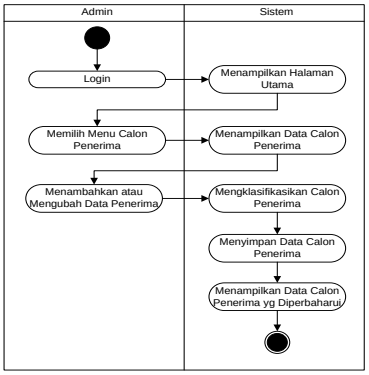

Figure 3. *Activity Diagram of* Prospective Recipients

- Review Activity Diagram

In the Activity Diagram below, the officer will check the citizen data registered in the system to see whether the data is correct; if the data registered with the admin is not suitable, then the officer will cancel the recipient. If the data registered is correct, the system will automatically determine the type of social assistance that matches the recipient's social assistance situation.

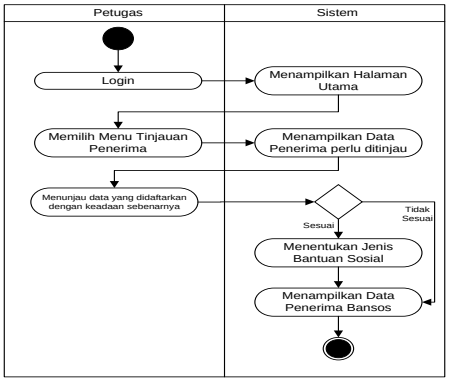

Figure 4. Receiver Review Activity Diagram

Class Diagram

The following class diagram shows that the design system will have three classes: user, user data, and prospective recipient. The user class will handle all processes involving user information in the system. In contrast, user data will handle processes involving user personal information data, and the prospective recipient class will handle the registration process, verification, and determination of prospective social assistance recipients' results.

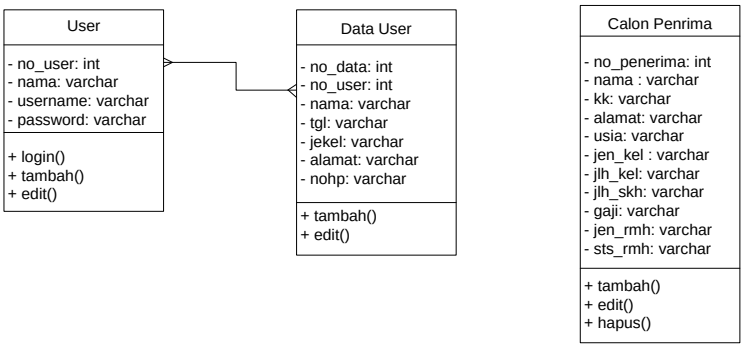

Figure 5. Class Diagram

3.2. Implementation

After the information system is designed, the next stage is implementing the design into a system that can be used. The following are the views of the information system created.

Recipient Candidate Data Page

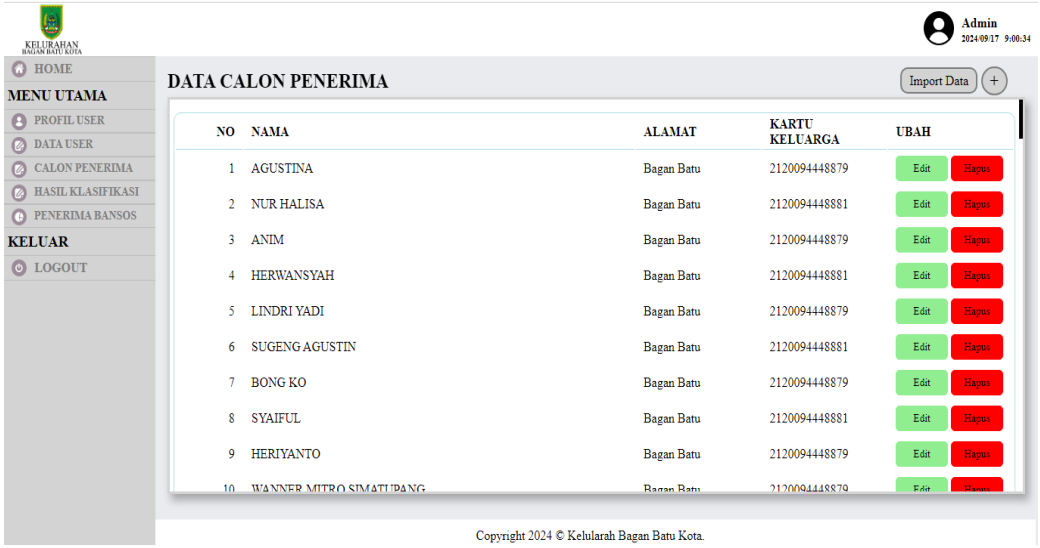

Figure 6. Prospective Recipient Data Page

This page will display the classification results of residents registered to receive social assistance. The admin can also access this page to register other prospective recipients, change existing prospective recipient data, or delete existing data.

Candidate Registration Page

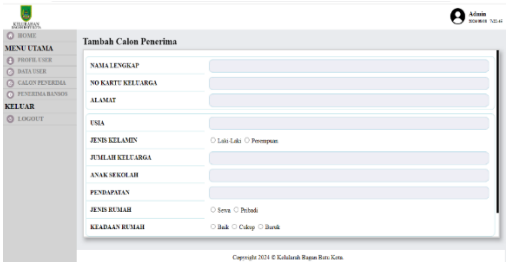

Figure 7. Prospective Recipient Registration Page

The Prospective Recipient Registration page can be accessed using the “+" button on the Prospective Recipient Data page; on this page, the admin can register residents who apply as recipients of social assistance to be determined by the system whether the citizen's application is accepted.

Receiver Review Page

Figure 8. Recipient Review Page

The Recipient Review page is a page that field officers can access; on this page, the officer will check the submitted data of the residents received and then match the registered data. Furthermore, if the data registered does not match, the status of the citizen's application will be cancelled. In contrast, if the data registered matches, the system will determine the assistance that suits the citizen's situation.

Social Assistance Recipient Page

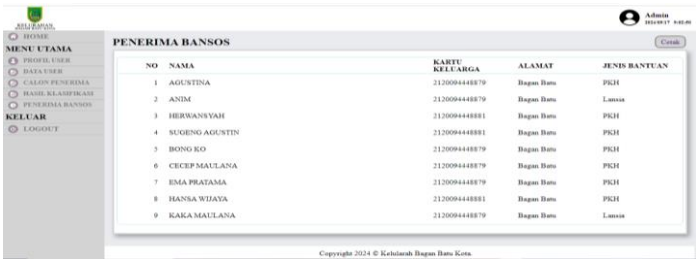

Figure 9. Social assistance recipient page

The Social Assistance Recipient page is accessible to all users. It displays the

recipients' names and the type of assistance they receive.

3.1 SYSTEM EVALUATION RESULTS

From the resulting system, a system evaluation is carried out using previous data to review whether the resulting system has provided good classification results. The assessment is done by comparing the actual data classification results with the classification results produced by the system, then getting the following results, which are shown in **Figure 10**.

Figure 10. System Evaluation Results

Here are the data that got the wrong classification result, which can be presented in **Table 1**.

Table 1. Misclassification Results

No.	Name	Actual Results	Classification Results
1	LUKI FARHAN	Receive	Not Accepted
2	MIRA MARDIANA	Receive	Not Accepted
3	NANA PRIHANTO	Receive	Not Accepted
4	ONA PROMO	Not Accepted	Receive
5	PUTRA HARTINI	Not Accepted	Receive
6	QISMA KUSUMA	Not Accepted	Receive
7	RAHMA MAULANI	Not Accepted	Receive
8	SRI CIPRIANI	Not Accepted	Receive
9	TARA WIJAYA	Receive	Not Accepted
10	ULI SUSAN	Receive	Not Accepted
11	VITA AGUSTIN	Receive	Not Accepted
12	WINA SURYONO	Receive	Not Accepted
13	YULI WIDJAJA	Not Accepted	Receive
14	ZARA HARMONY	Not Accepted	Receive
15	AGUS WIBOWO	Receive	Not Accepted
16	BUDI SANTOSO	Receive	Not Accepted
17	CICI GUNAWAN	Receive	Not Accepted
18	DEWI HIDAYAT	Receive	Not Accepted
19	ERIK SUPARI	Not Accepted	Receive
20	FANI PANGESTU	Not Accepted	Receive
21	GITA HERMAWAN	Not Accepted	Receive
22	HADI SURYANI	Receive	Not Accepted
23	INTAN MAULANA	Receive	Not Accepted
24	JOKO NURHALIZA	Receive	Not Accepted
25	KIKI PREMADE	Receive	Not Accepted
26	LINA SUTOMO	Not Accepted	Receive
27	MAYA WULANDARI	Not Accepted	Receive
28	NINA WIJAYA	Receive	Not Accepted
29	OPIK SEKAR	Not Accepted	Receive
30	PANDU PUTRA	Not Accepted	Receive
31	LUKI FARHAN	Receive	Not Accepted
32	MIRA MARDIANA	Receive	Not Accepted
33	NANA PRIHANTO	Receive	Not Accepted

The evaluation of the classification results was carried out using 1100 sample data. There is an error of 3% error or 33 data that has a different value from the actual data. The error consists of 20 "Receive" results classified as "Not Receive" and 13 "Not Receive" results classified as "Not Receive". The cause of the error is often due to the amount of monthly income that falls between "Receive" or "Not Receive". Nevertheless, based on the evaluation results, it can be seen that the system still has a high level of classification accuracy.

4. Conclusion

Based on the research process that has been carried out, it is known that this system is successful in classifying recipients of social assistance funds. Out of 1100 social assistance applicants, the number of recipients is 303. Those who do not receive 797 models resulting from random forest classification have a high accuracy value of 97%, which with such a high level of accuracy, is expected to be able to produce a classification that is indeed accurate and can be used as a reference in determining the classification of potential recipients of social assistance. And by applying the model to a system will help the kelurahan bagan batu kota determine residents who are entitled to social assistance and what type of social assistance is more suitable for them.

References

[1] O. Saputra and I. Ismail, "Classification on Book Reading Literacy by Students Using Random Forest Algorithm in Lampung University," *J. Ilmudata.org*, vol. 2, no. 11, pp. 1-15, 2022.

[2] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Using K-Means Clustering Method to Determine Single Tuition Amount," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80-89, 2020, doi: 10.52158/jacost.v1i2.102.

[3] B. F. Supit and E. Lumingkewas, "Implementation of Cash Social Assistance Policy in Talikuran Utara Village, Kawangkoan Utara Sub-district, Minahasa," *Acad. Educ. J.*, vol. 14, no. 2, pp. 1059-1068, 2023, doi: 10.47200/aoej.v14i2.1981.

[4] D. Librado, T. Prabawa, and H. A. Triyanto, "Clustering Social Assistance Recipients Using the Simple Additive Weighting Method," *JIKO (Journal of Inform. and Computers)*, vol. 7, no. 1, p. 30, 2023, doi: 10.26798/jiko.v7i1.677.

[5] A. Aldiyansyah, A. Irma Purnamasari, and I. Ali, "Comparison of Accuracy Levels of Decision Tree and Random Forest Algorithms in Classifying Bpnt Social Assistance Recipients in Slangit Village," *JATI (Journal of Mhs. Tech. Inform.*, vol. 8, no. 1, pp. 127-132, 2024, doi: 10.36040/jati.v8i1.8290.

[6] Ardi Ramdani, Christian Dwi Sofyan, Fauzi Ramdani, Muhamad Fauzi Arya Tama, and Muhammad Angga Rachmatsyah, "Data Mining Classification Algorithm to Predict Communities in Receiving Social Assistance," *J. Ilm. Sist. Inf.*, vol. 1, no. 2, pp. 39-47, 2022, doi: 10.51903/juisi.v1i2.363.

[7] M. Fardi, Inka Nusamuda Pratama, Darmansyah, and Amil, "Transparency of Data

Collection for the Cash Social Assistance Program in Poverty Reduction in Bebidas Neighborhood, Pegesangan Village, Mataram City," *J. Soc. Policy Issues*, vol. 3, pp. 150-155, 2023, doi: 10.58835/jspi.v3i3.209.

[8] S. Lestari and S. Agustiansyah, "Implementation of Data Mining Clustering of Poor Population Data Using the C4.5 Algorithm Method to Recommend Social Assistance in RT. 05/01 Jati Mekar Village, Jati Asih District, Bekasi City," *J. Tek. Electro and Computing*, vol. 5, no. 1, pp. 95-104, 2023.

[9] M. Yazid, D. A. Sani, and N. M. Anggadimas, "Recommendation System for Recipients of APBD Social Assistance Using the Support Vector Machine Method," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 6, no. 1, pp. 54-61, 2024, doi: 10.28926/ilkomnika.v6i1.619.

[10] S. Abdy and M. Alda, "Designing a Pharmacy Information System Using the CodeIgniter Framework and MySQL Database," *J. Inf. Comput. Log.*, vol. 1, no. 4, 2020.

[11] A. N. Riswanto and D. J. Lubis, "Scientific Journal of Informatics and Computers Volume. 1, Number. 1, June 2024, pp 23-29 Keywords-One-Time Password (OTP), Random Forest, Classification, Throwaway Prototyping," *Informatech J. Ilm. Inform. and Comput.*, vol. 1, no. 1, 2024.

[12] D. P. Sinambela, H. Naparin, M. Zulfadhilah, and N. Hidayah, "Implementation of Decision Tree and Random Forest Algorithms in Postpartum Hemorrhage Prediction," *J. Inf. and Technol.*, vol. 5, no. 3, pp. 58-64, 2023, doi: 10.60083/jidt.v5i3.393.

[13] M. I. Baihaqi, A. Syaripudin, and F. A. Nugroho, "Implementation of the Random Forest Algorithm in Stock Price Predictions Based on Historical Data," *Jubitek J. BIG DATA AND TECHNOLOGY. Inf.*, vol. 1, pp. 42-51, 2023.

[14] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Comparative Analysis Classification of Optimization Algorithms in Random Forest for Bank Marketing Data Classification," *J. RESTI (Engineering Systems and Technol. Information)*, vol. 5, no. 1, pp. 187-192, 2021, doi: 10.29207/resti.v5i1.2813.

[15] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, "Classification of Mental Health Data in the Technology Industry Using Random Forest Algorithm," *IJESPG J.*, vol. 1, no. 3, pp. 237-253, 2023.

[16] A. Rustamana, P. Wahyuningsih, M. F. Azka, and P. Wahyu, "Quantitative Method Research," *Sindoro Cendikia Pendidik.*, vol. 5, no. 6, pp. 1-10, 2024.

[17] B. Setiawan and T. Triase, "Implementation of Ui/Ux Design of Ourticle Application into Android-Based Application," *SIBATIK J. J. Ilm. Bid. Sos. Econ. Culture, Technol. and Educ.*, vol. 2, no. 3, pp. 805-818, 2023, doi: 10.54443/sibatik.v2i3.665.

[18] S. Surendra, T. Triase, and S. Afzalena, "Implementation of Job Order Costing Method in Web-Based Production Information System," *JS (School Journal)*, vol. 4, no. 2, pp. 97-106, 2020.

[19] M. Prabowo, *METHODOLOGY OF INFORMATION SYSTEM DEVELOPMENT*. LP2M IAIN Salatiga, 2020.