TIERS Information Technology Journal
Vol.5, No.1, June 2024, pp. 34~40
ISSN: 2723-4533 / E-ISSN: 2723-4541
DOI: 10.38043/tiers.v5i1.5378

Implementation Of Decision Tree Algorithm For Classification Of Eligibility In Social Assistance Fund Distribution

Hartana[1], Raden Deasy Mandasari[2]
hartana.han@bsi.ac.id[1], deasy.rde@bsi.ac.id[2]
Electrical Engineering Study Program, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Jakarta, Indonesia

ABSTRACT

This study examined the use of decision tree algorithms, specifically C4.5, in classifying the eligibility for social assistance fund distribution in Kelurahan Bangka Belitung Laut, Kecamatan Pontianak Tenggara. Inaccuracies in distribution were caused by recipient selection based on village officials' recommendations, which only considered the type of occupation. To address this issue, this study developed a system based on the C4.5 algorithm to process economic census data. This research used a quantitative method with a descriptive analytical approach, collecting data through literature studies and economic census, and processing data using the Rapid Miner application. The classification model generated was evaluated using 10-fold cross-validation to ensure high accuracy. The results showed that the C4.5 algorithm achieved 100% accuracy, precision, and recall. The decision tree model indicated that the main attributes determining eligibility were occupation and income. Some rules derived from this model, such as those who are unemployed with an income below Rp29,500 being eligible for assistance, provided clear guidelines for policymakers. The implementation of this algorithm is expected to improve fairness and effectiveness in the distribution of social assistance funds in Kelurahan Bangka Belitung Laut, reduce public dissatisfaction, and prevent potential social conflicts. This study recommended adopting the model in other areas with adjustments to local data to enhance broader fairness in aid distribution.

Keywords: C4.5 Algorithm; Cross-Validation; Decision Tree; Eligibility Classification; Social Assistance Distribution.

Article Info

Accepted : 02-06-2024
Revised : 11-04-2024
Published Online : 25-06-2024

Correspondence Author:

Raden Deasy Mandasari
Electrical Engineering Study Program, Faculty of Engineering and Informatics,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450.
Email: deasy.rde@bsi.ac.id

1. INTRODUCTION

In the distribution of social assistance in Kelurahan Bangka Belitung Laut, Kecamatan Pontianak Tenggara, there are inaccuracies in the distribution of aid funds. This is due to the selection of recipients based on recommendations from village officials, especially RT, which only considers the type of occupation of the residents. As a result, residents who are classified as underprivileged often do not receive the assistance they should. These inaccuracies not only cause dissatisfaction among the community but also have the potential to cause broader social problems if not addressed immediately.

The rationale for this research is the importance of a more accurate and objective system in determining the eligibility of social assistance recipients. Previous research has shown that traditional methods based on personal recommendations are often inaccurate and unfair. According to research by [1], [2],

classification algorithms like Decision Trees can provide higher accuracy in various predictive applications [3], [4].

The urgency of this research lies in the urgent need to improve the social assistance distribution system in Kelurahan Bangka Belitung. Research by [5] shows that the application of data mining algorithms, including Decision Trees, can increase the effectiveness and efficiency in determining aid recipients [6], [7]. Therefore, the implementation of the C4.5 algorithm, which is a variant of the Decision Tree, was chosen as a potential solution to address this problem.

The main problem in this research is how to improve the accuracy of determining social assistance recipients in Kelurahan Bangka Belitung. Alternative solutions considered include various data mining classification techniques such as Logistic Regression, Naive Bayes, and Support Vector Machines [8]–[12]. However, based on studies by [13], the Decision Tree algorithm, especially C4.5, shows superior performance in terms of interpretability and accuracy on similar datasets.

The solution chosen in this study is the implementation of the C4.5 algorithm for the classification of social assistance recipients [14]–[18]. This algorithm is not only capable of generating accurate predictive models but also easy to interpret, so it can be used by local governments and related stakeholders to make more precise decisions [19], [20].

Referring to these references, this study aims to develop a system based on the C4.5 algorithm that can process economic census data of residents in more detail and accurately in determining the eligibility of social assistance recipients. The results of this research are expected to make a significant contribution to improving the fairness and effectiveness of the distribution of social assistance funds in Kelurahan Bangka Belitung.

2. RESEARCH METHOD

2.1. Type of Research, Time, and Place

This research is a quantitative study with a descriptive analytical approach aimed at developing a classification model for social assistance fund recipients using the Decision Tree algorithm, specifically C4.5. The study was conducted from January to June 2024 in Kelurahan Bangka Belitung Laut, Kecamatan Pontianak Tenggara, West Kalimantan.

2.2. Research Target and Subjects

The target of this research is all residents of Kelurahan Bangka Belitung Laut registered in the 2021 economic census data. The research subjects include resident data consisting of various economic and social variables, including occupation, income, and social assistance recipient status. The sampling technique used is simple random sampling, where from a population of 5,167 residents, a sample of 98 people was taken. The sample selection used the Slovin formula with an error rate of 10% [21]–[25].

2.3. Research Procedures, Instruments, and Data Analysis Techniques

The research procedure began with data collection through literature studies and economic census of residents. Literature studies were conducted to collect secondary data from scientific journals, articles, and literature related to the use of Decision Tree algorithms in data classification. Meanwhile, economic census data of residents were collected from Kelurahan Bangka Belitung Laut. Instruments used in this research included questionnaires with Likert scales to measure residents' social and economic variables and the Rapid Miner application for data processing and implementation of the C4.5 algorithm.

After the data was collected, data preprocessing was performed to clean and prepare the data before analysis. The data was then processed using the C4.5 algorithm with the Rapid Miner application, where the classification model was generated through the 10-fold cross-validation method to ensure high accuracy. Data analysis techniques used include descriptive analysis to process and interpret the collected data. The analysis results are presented in the form of tables, graphs, and diagrams to facilitate interpretation and decision-making. The decision tree model generated is used to determine the eligibility of recipients based on relevant variables such as occupation and income.

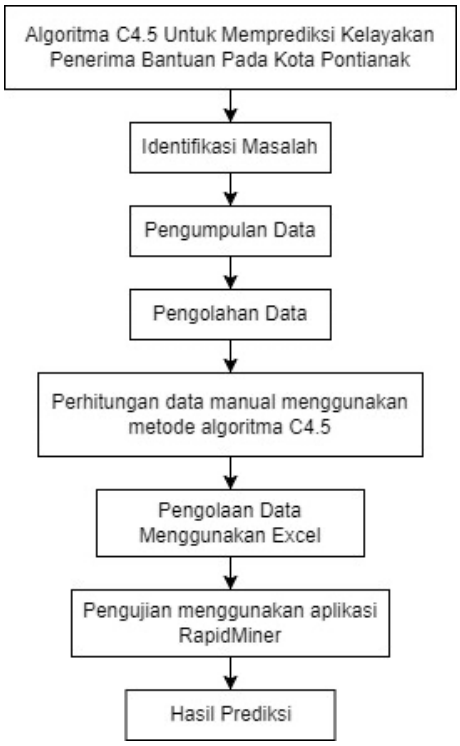

Figure 1. Research Stages

3. RESULTS AND DISCUSSION

3.1. Population Data Analysis

Population data was analyzed based on various categories of occupation and assistance recipient status. Table 1 below shows the distribution of residents based on occupation type and assistance recipient status.

Table 1. Population Data of Kelurahan Bangka Belitung Laut According to Occupation and Assistance Recipient Status

No	Occupation	Total	Receiving Assistance	Not Receiving Assistance
1	Unemployed	4.375	2.875	1.500
2	Government Officials	884	120	764
3	Educators	129	45	84
4	Entrepreneurs	3,873	1.400	2.473
5	Agriculture and Livestock	29	15	14
6	Fishermen	11	7	4
7	Students	2.242	1.120	1.122
8	Healthcare Workers	30	5	25
9	Retirees	167	55	112
10	Other Jobs	2.563	1.200	1.363

3.2. Data Processing with C4.5 Algorithm

The collected data was then processed using the C4.5 algorithm. Data preprocessing was carried out to clean and prepare the data. The figure below shows the data preprocessing process.

.

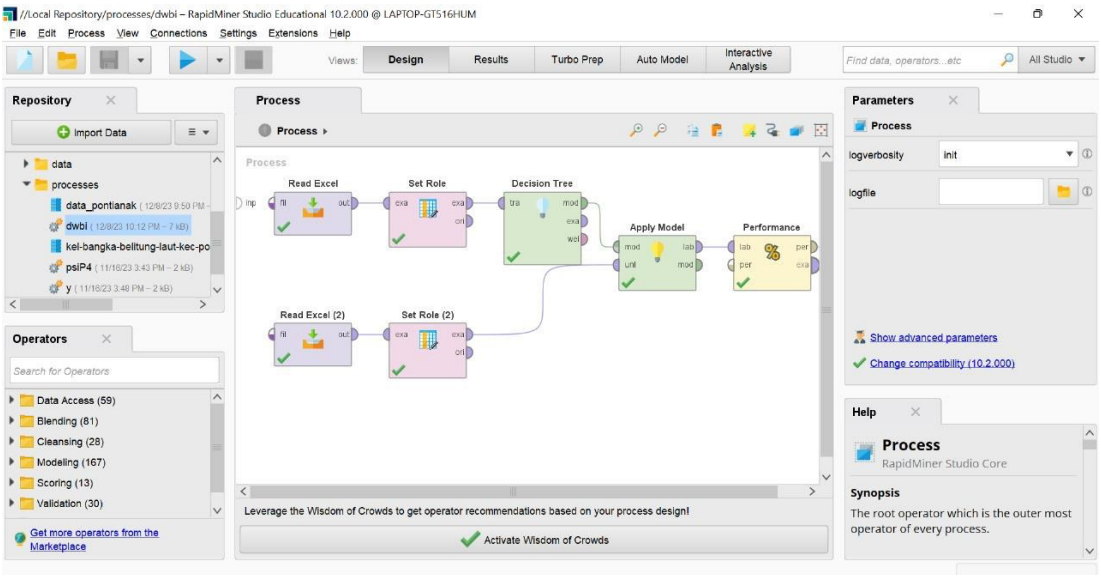

Figure 2. Preprocessing

Based on the data processing results through the Rapid Miner application, a Confusion Matrix was obtained to measure the accuracy of the C4.5 algorithm. The evaluation results showed that this algorithm achieved 100% accuracy, precision, and recall.

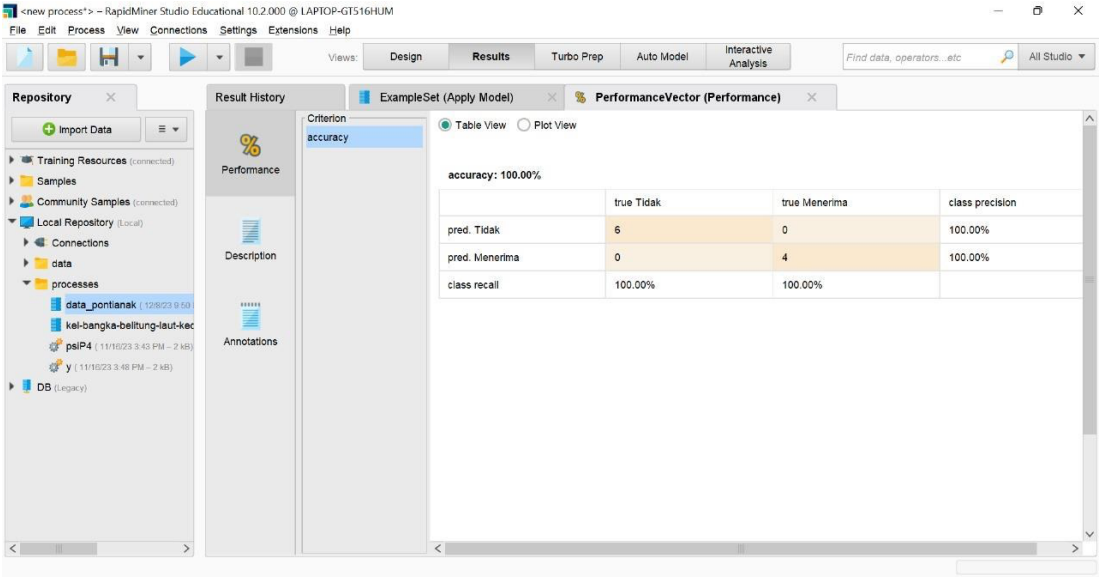

Figure 3. Accuracy

3.3. Decision Tree Model

The decision tree model generated from the C4.5 algorithm is shown in Figure 4 below. This model helps determine the eligibility of recipients based on relevant variables such as occupation and income.

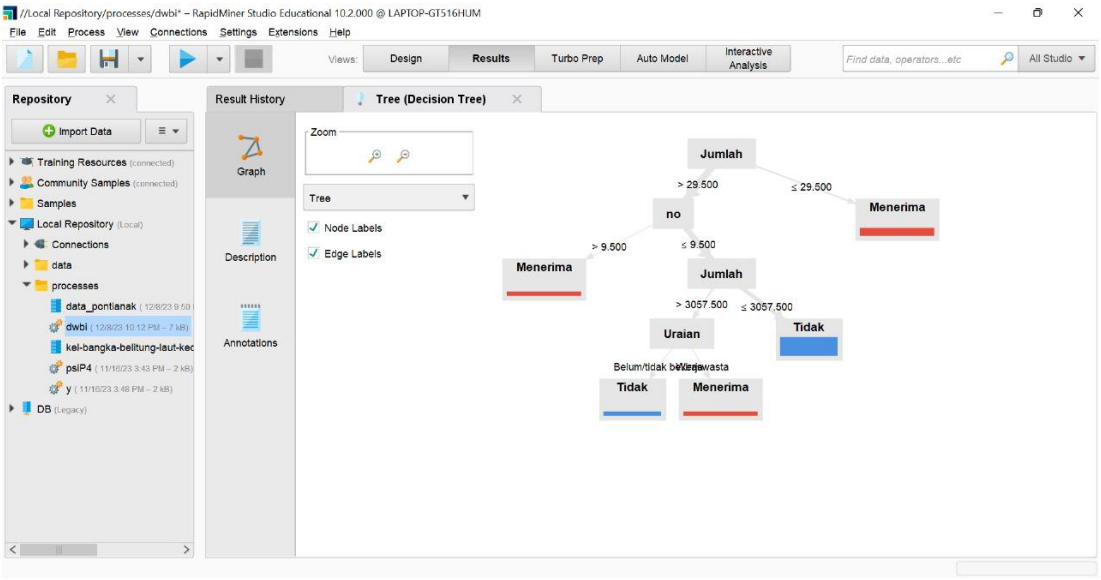

Figure 4. Decision Tree Model

Explanation of the decision tree model generated: the main attribute that becomes the root is "Occupation Type", followed by branches that show income and other variables affecting the decision to receive assistance.

3.4. Rules from the Decision Tree

Based on the decision tree model, several rules generated are:

1. If the occupation is "Unemployed" and income < Rp29,500, then the recipient is eligible for assistance.
2. If the occupation is "Entrepreneur" and income > Rp29,500, then the recipient is not eligible for assistance.

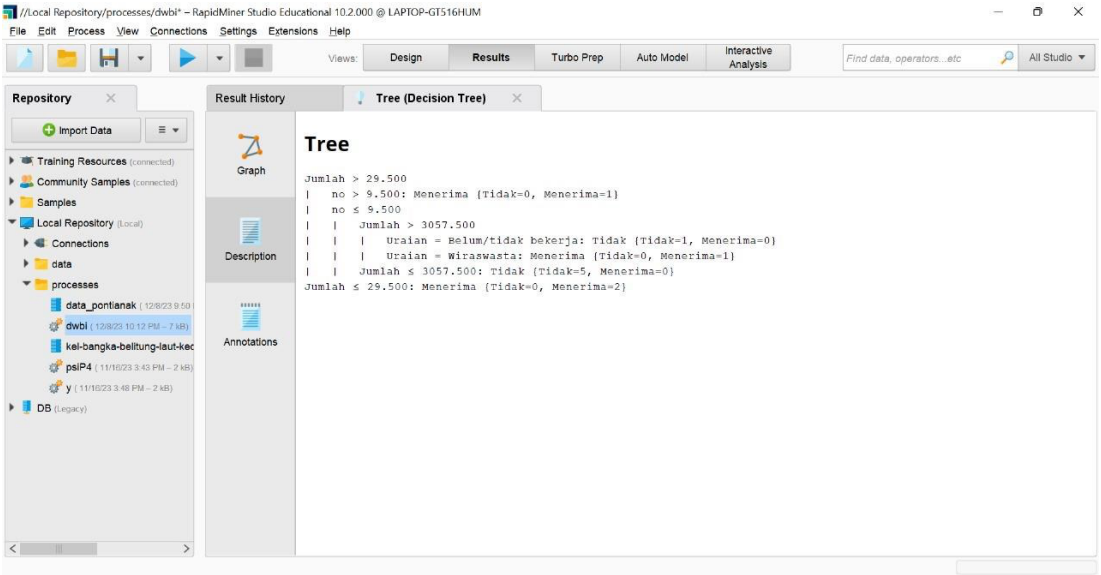

Figure 5. Description Tree

4. CONCLUSION

Based on the research results, the implementation of the C4.5 algorithm in classifying the eligibility of social assistance recipients shows excellent results with accuracy, precision, and recall reaching 100%. This indicates that the algorithm is effective in determining eligible recipients based on economic census data. The decision tree model generated can be easily interpreted and provides clear rules in determining eligibility. The main attributes used in this model are occupation and income, which are proven relevant in eligibility

.

classification. The high validity of the model indicates that the C4.5 algorithm can handle datasets well and produce accurate predictions. There are no misclassifications in the Confusion Matrix, which confirms the reliability of this model.

This study successfully identifies and implements a fairer and more accurate solution for determining social assistance recipients. By using more detailed economic census data, this model can address the inaccuracies in the previous aid distribution system based solely on personal recommendations. Based on the findings of this study, several recommendations for the next steps include adopting and implementing this model in other areas with adjustments to local data. This can help local governments improve the fairness of social assistance distribution more broadly. It is recommended to use more detailed and up-to-date data to increase the accuracy of the classification model. Regular data collection can help adjust the model to changes in social and economic conditions. It is important to provide training and socialization to village officials and local governments on the use of this model. A good understanding of the model interpretation will help in making more accurate decisions. Developing a web-based system that integrates the C4.5 algorithm to facilitate access and use by various stakeholders can be an effective tool in determining eligibility for assistance.

REFERENCES

[1] A. Abdulhafedh, "Comparison between Common Statistical Modeling Techniques Used in Research, Including: Discriminant Analysis vs Logistic Regression, Ridge Regression vs LASSO, and Decision Tree vs Random Forest," *OALib*, vol. 09, no. 02, pp. 1–19, 2022.

[2] Sugiono, A. Taufik, and R. Faizal Amir, "Penerapan Penerapan Teknik Pso Over Sampling Dan Adaboost J48 Untuk Memprediksi Cacat Software," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 198–203, 2020.

[3] Y. Suhanda, L. Nurlaela, I. Kurniati, A. Dharmalau, and I. Rosita, "Predictive Analysis of Customer Retention Using the Random Forest Algorithm," *TIERS Inf. Technol. J.*, vol. 3, no. 1, pp. 35–47, 2022.

[4] S. Annisa and Y. Rahkmawati, "Resampling Techniques in Rainfall Classification of Banjarbaru using Decision Tree Method," *TIERS Inf. Technol. J.*, vol. 4, no. 2, pp. 122–128, 2023.

[5] P. Ladianto, "Implementasi Data Mining Untuk Bantuan Dana Bansos Dengan Menggunakan Algoritma C4. 5 Di Pemerintahan Kabupaten Empat Lawang," *JUPITER (Jurnal Penelit. Ilmu dan Tek. ...*, no. 13, pp. 287–297, 2022.

[6] N. Aulia, N. Suarna, and W. Prihartono, "Klasifikasi Penentuan Penerima Program Indnesia Pintar Di Krwilbidikcam Greged Menggunakan Algoritma C4.5," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3913–3919, 2024.

[7] A. Nur, A. Rohim, A. I. Purnamasari, and I. Ali, "Komparasi Efektifitas Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Manfaat Program Keluarga Harapan (Studi Kasus : Kecamatan Cicalengka Kabupaten Bandung)," *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2355–2362, 2024.

[8] V. H. Nhu *et al.*, "Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020.

[9] P. Golpour *et al.*, "Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, pp. 1–9, 2020.

[10] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *Int. J. Inf. Technol.*, vol. 13, no. 4, pp. 1503–1511, 2021.

[11] V. H. Nhu *et al.*, "Comparison of support vector machine, bayesian logistic regression, and alternating decision tree algorithms for shallow landslide susceptibility mapping along a mountainous road in the west of Iran," *Appl. Sci.*, vol. 10, no. 15, 2020.

[12] A. Tariq *et al.*, "Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data," *Heliyon*, vol. 9, no. 2, p. e13212, 2023.

[13] I. Ba'abbad, T. Althubiti, A. Alharbi, K. Alfarsi, and S. Rasheed, "A Short Review of Classification Algorithms Accuracy for Data Prediction in Data Mining Applications," *J. Data Anal. Inf. Process.*, vol. 09, no. 03, pp. 162–174, 2021.

[14] M. I. Sa'ad, D. Bryan, Kusrini, and Supriatin, "Decision Support System for Covid19 Affected Family Cash Aid Recipients Using the Naïve Bayes Algorithm and the Weight Product Method," *2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020*, pp. 120–125, 2020.

[15] R. Rusdiansyah, H. Supendar, and T. Tuslaela, "Data Mining using K-means method for feasibility selection of Non-cash food Assistance recipients in the Era of Covid-19," *SinkrOn*, vol. 6, no. 1, pp. 25–33, 2021.

[16] M. Mardison, S. Defit, and S. Alturky, "Prediction of Scholarship Recipients Using Hybrid Data Mining Method with Combination of K-Means and C4.5 Algorithms," *Int. J. Artif. Intell. Res.*, vol. 5, no. 2, pp. 168–179, 2021.
[17] P. Nabillah, I. Permana, M. Afdal, F. Muttakin, and A. Marsal, "A Comparative Study of the Performance of KNN , NBC , C4 . 5 , and Random Forest Algorithms in Classifying Beneficiaries of the Kartu Indonesia Sehat Program," *JUSIFO (Jurnal Sist. Informasi)*, vol. 10, no. 1, pp. 17–26, 2024.
[18] F. Riandari and S. Defit, "The Application of C4.5 Algorithm for Selecting Scholarship Recipients," *ComTech Comput. Math. Eng. Appl.*, vol. 13, no. 1, pp. 11–21, 2022.
[19] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018.
[20] F. Pirmansyah and T. Wahyudi, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Evaluasi Anggota Satuan Pengamanan Studi Kasus Pt. Yimm Pulogadung," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1566–1580, 2023.
[21] Y. Septiani, E. Arribe, and R. Diansyah, "MENGGUNAKAN METODE SEVQUAL (Studi Kasus : Mahasiswa Universitas Abdurrab Pekanbaru)," *J. Teknol. dan Open Source*, vol. 3, no. 1, pp. 131–143, 2020.
[22] K. Purba and K. Sudibjo, "The Effects Analysis of Transformational Leadership, Work Motivation and Compensation on Employee Performance in PT. Sago Nauli," *Budapest Int. Res. Critics Inst. Humanit. Soc. Sci.*, vol. 3, no. 3, pp. 1606–1617, 2020.
[23] M. Meilinda, L. Louis, J. Jason, and H. Nazmi, "The Influence of Discipline, Selection, and Organizational Culture on Employee Performance," *Almana J. Manaj. dan Bisnis*, vol. 5, no. 1, pp. 20–28, 2021.
[24] T. D. Anugraheni, L. Izzah, and M. S. Hadi, "Increasing the Students' Speaking Ability through Role-Playing with Slovin's Formula Sample Size," *J. Stud. Guru dan Pembelajaran*, vol. 6, no. 3, pp. 262–272, 2023.
[25] Maurizio Zanardi and Stephen Martin, "THE EFFECT OF SOCIAL SECURITY ON EMPLOYEE PERFORMANCE WITH WORK MOTIVATION AS AN INTERVENING VARIABLE (Case Study on Employees of United Kongdom Medical Dcotor and Nurses)," *MEDALION J. Med. Res. Nursing, Heal. Midwife Particip.*, vol. 1, no. 3, pp. 90–95, 2020.

.