Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, Number 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

The Comparison of Decision Tree and K-Nearest Neighbor Performance for Determining *Mustahik*

Noor Amelia[1], Winda Aprianti[2*]
[1,2]Politeknik Negeri Tanah Laut, Indonesia
[1]noor.amelia@politala.ac.id, [2]winda@politala.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 14-10-2023
Accepted: 15-10-2023
Published: 23-10-2023

Keywords:
AUC; Confusion Matrix; Decision Tree; KNN; zakat

ABSTRACT

The problem of poverty is one of the fundamental issues of concern to the Indonesian government. One of the methods used by Islam to alleviate poverty is through zakat from Badan Amil Zakat Nasional (BAZNAS). Currently, the distribution of zakat is divided into two, namely in the form of consumptive zakat and productive zakat. Productive zakat is aimed at people who need business capital. To assist zakat managers in managing their funds, a mechanism is needed that can process *mustahik* data so that it can be selected more quickly and precisely using data mining. In this research, the data mining methods that will be used are K-nearest neighbor (KNN) and Decision Tree. The dataset used in this research is data obtained from BAZNAS and has been preprocessed to obtain a dataset with 7 attributes and 144 records. Decision trees, KNN Manhattan, and KNN Euclidean are used to predict *mustahik* candidates who are worthy of receiving zakat. The performance of the third method was tested using AUC and confusion matrix namely Accuracy, Precision, Recall, and F1 in each dataset split scenario of 70%:30%, 75%:25%, and 80%:20%. Based on the number of false positive and false negative results, the best performance obtained is KNN Euclidean with a dataset division scenario of 80%:20%.

INTRODUCTION

The problem of poverty is one of the fundamental issues of concern to the Indonesian government. Based on data from the Central Statistics Agency (BPS) in September 2022, the number of poor people in Indonesia reached 26.32 million (Badan Pusat Statistik, 2023). Islam with all the noble teachings contained in it has a projection that is far ahead with the aim of maintaining the interests and benefits of mankind. One of the ways that Islam uses to alleviate poverty is in the form of zakat. Badan Amil Zakat Nasional (BAZNAS) is one of the amil zakat institutions that targets zakat collection of IDR 26 trillion in 2022 so that the enormous potential can be relied upon as a real mechanism and solution in overcoming poverty in Indonesia. The distribution of BAZNAS RI throughout 2022 includes five main aspects, namely education (45,814 beneficiaries), health (371,500 beneficiaries), social humanity (1,602,898 beneficiaries), da'wah (76,391 beneficiaries), and economy (19,335 beneficiaries) (Badan Amil Zakat Nasional, 2023). Based on research by (Amelian, Machfiroh, & Fitriyani, 2020) shows that there is economic growth of MSMEs through capital assistance from BAZNAS, which has an impact on increasing their income and profits .

Along with the times, the distribution of zakat has been modified, not only distributed in the form of consumptive zakat, but also in the form of productive zakat which means the distribution of zakat to the poor to be used as business capital, as a form of their livelihood, with this business it is hoped that they will be able to meet their own needs (Safradji, 2018). In giving zakat, determining whether someone can be categorized as receiving zakat is very important. Determination of the criteria for zakat recipients (mustahik) must be clearly delineated so that there are no errors in the allocation and distribution which ultimately have an impact on the failure of the mustahik empowerment. However, the limited funds of each zakat management organization in Indonesia result in only a small portion of the underprivileged community who can enjoy zakat. Therefore, an accurate mechanism for determining zakat recipients is needed through the use of data mining.

Data mining is a process of finding meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques (Han, Pei, & Tong, 2022). Some researchers have conducted research in the field of poverty alleviation and the determination of zakat recipients by utilizing data mining. One of the methods studied by (Setiadi, Parnolo, & Prayitno, 2017) by implementing the K-Nerest Neighbor (K-NN) algorithm in web framework-based data classification. (Hasanah & Mahdiyah, 2021) also conducted research using the C4.5 algorithm in Baznas Padang City to determine recipients of zakat funds. Meanwhile, (Firdaus, 2021) used the Algorithm (K-NN) to classify the determination of mustahik from poor people in Surade District. However, this research is only in the form of calculating and classifying whether

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, Number 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

someone is capable or not, so a system is needed that is interconnected between analysis to sort mustahik and information systems that can manage mustahik data so that it is easier for users to understand and zakat giving is more targeted. (Ahmad, Saharuna, & Raharjo, 2020) applied data mining in determining recommendations for mustahiks using the classification method with artificial neural network algorithms with accurate results.

Based on the thesis written by (Kamila, 2020) states that the application of data mining for BAZNAS zakat recipients through various condition attributes using the C.45 decision tree algorithm as attribute selection is also applied in this study. However, this method has not been equipped with the SPK method to obtain the results of mustahik selection decisions based on eligibility results that are specific to productive zakat recipients.The proposed research will use data mining methods to classify prospective recipients of productive zakat into eligible or ineligible. Then the system will rank the eligible data to determine the priority of productive zakat recipients.

LITERATURE REVIEW

The following is previous research related to this research.

1. A comparison of Decision Trees and KNN was made by (Hassonah, Rodan, Al-Tamimi, & Alsakran, 2019) to predict churn using accuracy, precision, recall, F-Measure (F1), Area Under Receiver Operating Characteristic Curve (AUROC) or (AUC), and lift measure. Overall, the Decision Tree is better than KNN. However, the results of the confusion matrix show KNN predicts positive values better than Decision Tree.
2. The application of KNN to predict zakat recipients has been carried out by (Sari, Maulida, Gunawan, & Wahyudi, 2021) and the result of it shows accuracy results above 90%.
3. Prediction of determining credit eligibility using classification algorithms Decision Tree C4.5, Naïve Bayes, K-NN, and Random Forest was made by (Oktafriani, Firmansyah, Tjahjono, & Widodo, 2023). The result of performance testing using Accuracy, Precision, and Recall shows that KNN is the best method for determining credit eligibility.
4. Classification of zakat potential in Lazismu DIY using K-Nearest Neighbor (K-NN) method based web framework made by (Setiadi, Parnolo, & Prayitno, 2017). The results of the confusion matrix with a ratio of 80:20 of test data and data testing data resulted in an accuracy value of 85% and an error rate of 25%. Results accuracy of more than 85% is said to be good in classification, proving that the factors selected are close to significant values. The selected attribute value factors are close to the significant value.
5. Information System For Determining Recipients Of Zakat Funds Using C4.5 Algorithm In Baznas Padang Panjang City was made by (Hasanah & Mahdiyah, 2021). The results of calculations used 1084 training data to get 954 rules. The application of data mining using the C4.5 algorithm on zakat fund data at the Baznas Padang Panjang resulted in 83.7638%, a precision value of 85.2273%, and a recall value of 7.8261%.
6. Determination of poor people receiving zakat using The K-Nearest Neighbor Algorithm was made (Firdaus, 2021), One of the efforts to help overcome poverty in Sukabumi District, especially Surade District, is to offer zakat to the poor, which is a classification method using the neighboring k-nearest algorithm, aiming to classify the poor in Surade District so that zakat recipients are really poor and entitled to receive zakat according to the conditions required to receive zakat.
7. A study on Decision Tree and KNN Algorithm for Intrusion Detection Systems was made by (Pathak & Pathak, 2020) and the result of it has found that the Decision tree algorithm gives a better result with an accuracy of 99,15%. Also, they found that the time taken to build the Decision tree algorithm was far less then the KKN algorithm. Therefore, the results show that Decision Tree algorithm gives an overall better result as compared to KKN.
8. Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer was made by (Rajaguru & R, 2019) and the result show the comparative analysis results indicate that the KNN classifier outperforms the result of the decision-tree classifier in the breast cancer classification.
9. KNN is outperform than Decision Tree for student performance prediction in research by (Yulianto, Triayudi, & Sholihati, 2020).

In this research, Decision Tree, KNN Manhattan, and KNN Euclidean will be implemented to predict *mustahik* and then their performance will be compared based on AUC, Accuracy, Precision, Recall, and F1.

METHOD

The research stages began with data collection at BAZNAS. Data from BAZNAS Banjarmasin City and BAZNAS Tanah Laut Regency. This data then becomes a Dataset in implementing data mining. The implementation of data mining is presented in Figure1. to determine the classification of data that is or is not entitled to receive Zakat (*Mustahik*).

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 3, Number 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

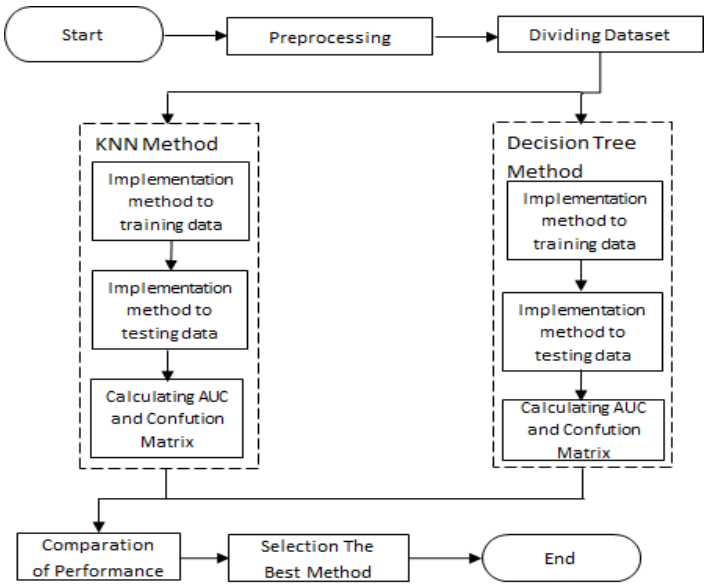

Figure 1. The Implementation of Method

RESULT

The data obtained from BAZNAS Banjarmasin City and BAZNAS Tanah Laut Regency amounted to 178 data. Data from BAZNAS Banjarmasin City includes the attributes name, address, assistance program, number of dependents, husband's job, wife's job, total income, total expenditure, assistance program, water source, and survey officer recommendations. Data obtained from BAZNAS Tanah Laut Regency includes name, address, number of dependents, husband's occupation, wife's occupation, total income, savings, types of savings, debt, and total expenses, condition of the house, the existence of toilets, ownership of rice fields, vehicle ownership, water sources, electrical power, livestock ownership, personality, and survey officer recommendations.

The dataset still has different attributes for data from the Banjarmasin City Baznas and Tanah Laut Regency Baznas, redundant data, missing values, and data inconsistencies. For redundant data, one of the object records will be deleted so that 144 data are obtained. Meanwhile, attributes that have more than 50% missing values will be removed and the attributes from the two Baznas data will be standardized so that 6 attributes will be obtained, namely number of dependents, husband's job, wife's job, total income, total expenses, and survey officer recommendations. Seeing the importance of financial conditions every month, a new attribute was created, namely the difference between income and expenses, making the dataset have 7 attributes. Meanwhile, for example, there are inconsistencies in the husband's or wife's job attributes, filling in non-working data has different variants, such as not working and "-". So it needs to be changed to Not Working. The next stage is to divide the dataset into training data and testing data with several division scenarios, namely 70%:30%, 75%:25%, and 80%:20%. Next, the Decision Tree, KNN with Manhattan, and KNN with Euclidean are applied to the dataset with various dataset division scenarios. The results of Decision Tree performance testing are presented in Table 1 and Figure 2.

Table 1. The Result of Decision Tree Performance Testing

Testing Techniques	70%:30%	75%:25%	80%:20%
AUC	54.10%	43.70%	41.90%
Accuracy	75.90%	73.60%	74.50%
F1	75.90%	73.33%	74.50%
Precision	75.90%	73.00%	74.50%
Recall	75.90%	73.60%	74.50%

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

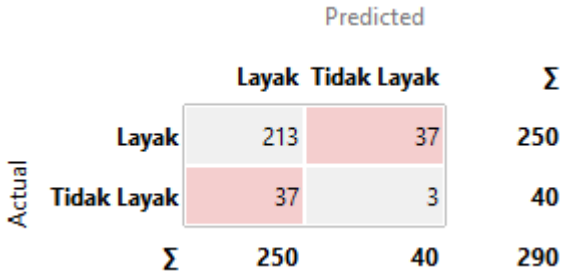

Figure 2. The Result of Confusion Matrix for Decision Tree

The performance test results of KNN with Manhattan on the Dataset are presented in Table 2 and Figure 3.

Table 2. The Result of KNN with Manhattan Performance Testing

Testing Techniques	70%:30%	75%:25%	80%:20%
AUC	56.00%	58.10%	55.50%
Accuracy	80.70%	79.70%	79.30%
F1	75.80%	75.80%	76.70%
Precision	72.20%	72.40%	74.30%
Recall	80.70%	79.70%	79.30%

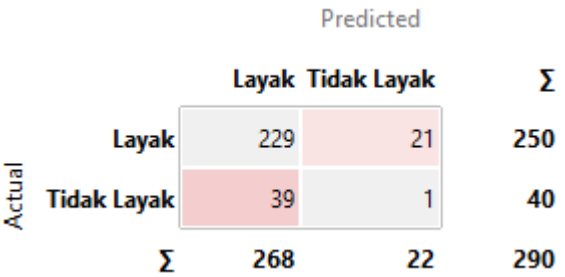

Figure 3. The Result of Confusion Matrix for KNN Manhattan

The results of testing KNN performance with Euclidean on the Dataset are presented in Table 3 and Figure 4.

Table 3. The Result of KNN with EuclideanPerformance Testing

Testing Techniques	70%:30%	75%:25%	80%:20%
AUC	55.60%	56.40%	53.30%
Accuracy	80.90%	79.20%	90.70%
F1	75.90%	75.50%	77.40%
Precision	72.30%	72.30%	74.60%
Recall	80.90%	79.20%	80.70%

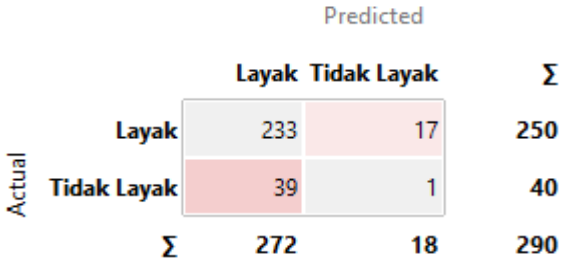

Figure 4. The Result of Confusion Matrix for KNN Euclidean

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 3, Number 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

DISCUSSION

Based on the results in Table 1 to Table 3 and Fig. 2 to Fig. 4, the following are obtained.

1. The comparison of the number of False Positives and False Negatives presented in Fig. 2 shows the same number, then the testing technique that is being considered is accuracy. The best accuracy value for the Decision Tree lies in the dataset with a 70:30 dataset division scenario.
2. A comparison of the number of False Positives and False Negatives presented in Fig. 3 shows that the numbers are not close, so the testing techniques that are considered are F1 and Precision. The best F1 value and precision for KNN Manhattan lies in the dataset with the 80:20 dataset division scenario.
3. A comparison of the number of False Positives and False Negatives presented in Fig. 4 shows that the numbers are not close, so the testing techniques that are considered are F1 and Precision. The best F1 and Precision values for Euclidean KNN lie in the dataset with the 80:20 dataset division scenario.

Based on the results, the performance of the three methods is the most stable for dividing the dataset by 80%:20%. The summarized results in Table 1, Table 2, and Table 3 are presented in Figure 5.

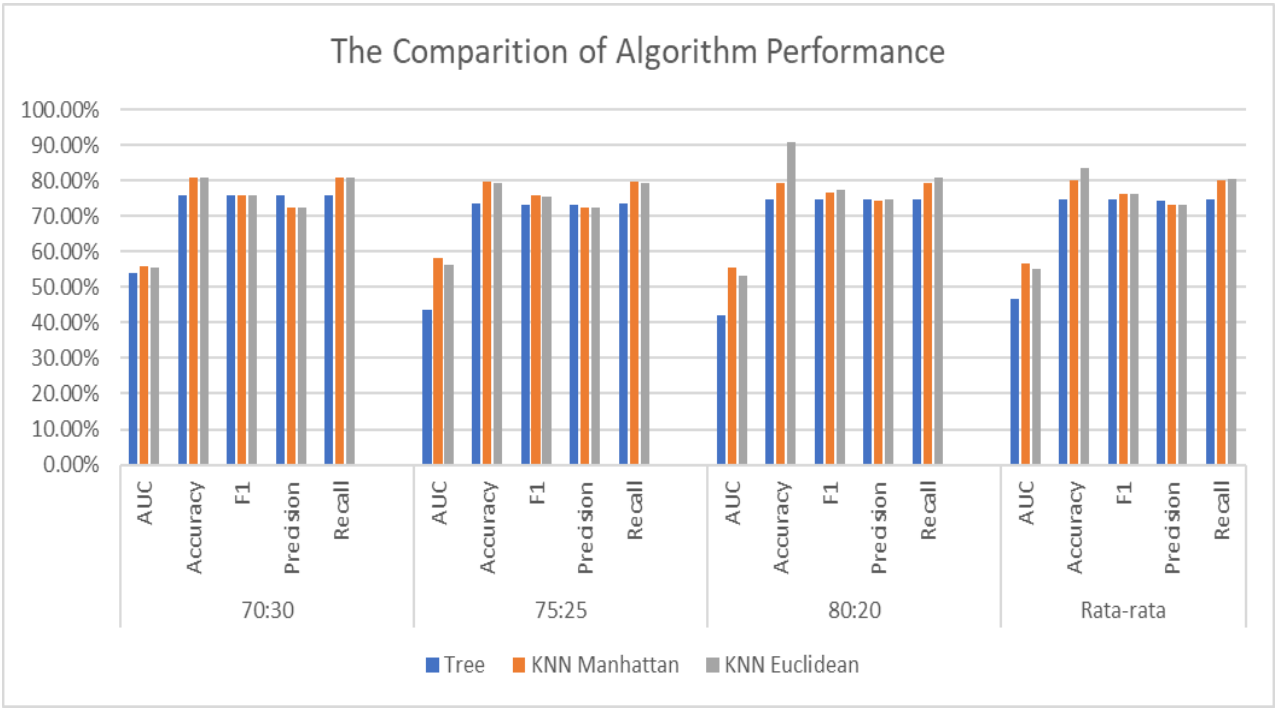

Figure 5. Performance Results of the Three Methods on the Dataset

Based on Fig. 5, it can be seen that the best method is KNN with distance measurement using Euclidean. KNN method is better than Decision Tree for predicting, especially positive values in line with research Hassonah et al (2019) and Oktafriani (2023). But, if the result of this research is compared with research by (Sari et al, 2021), the performance of KNN and Decision Tree is lower. This is due to the limited number of datasets and the imbalance in target class data.

CONCLUSION

The dataset used in this research is data obtained from BAZNAS and has been preprocessed to obtain a dataset with 7 attributes and 144 records. Decision trees, KNN Manhattan, and KNN Euclidean are used to predict *mustahik* candidates who are worthy of receiving zakat. The performance of the third method was tested using AUC, Accuracy, Precision, Recall, and F1 in each dataset split scenario of 70%:30%, 75%:25%, and 80%:20%. Based on the number of false positive and false negative results, the best performance obtained is KNN Euclidean with a dataset division scenario of 80%:20%.

ACKNOWLEDGMENT

Thank you to the Direktorat Akademik Pendidikan Tinggi Vokasi Direktorat Jenderal Pendidikan Vokasi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi for providing grants through the Beginner Lecturer Research Grant.

E-ISSN : 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.2953

REFERENCES

(2023, April). Retrieved from Badan Amil Zakat Nasional: http://baznas.go.id

(2023, Januari 16). Retrieved from Badan Pusat Statistik: http://www.bps.go.id

Ahmad, A. A., Saharuna, Z., & Raharjo, M. F. (2020). Pemanfaatan Data Mining dalam Penentuan Rekomendasi Mustahik (Penerima Zakat). *Elektron Jurnal Ilmiah, 12*(1), 67-73. doi:https://doi.org/10.30630/eji.12.2.182

Amelian, N., Machfiroh, I. S., & Fitriyani, Y. (2020). Analisis Pengaruh Penyaluran Dana Zakat terhadap Perkembangan Usaha Kecil dan Menengah (UKM) Mustahik. *Jurnal Akuntansi, Ekonomi dan Manajemen Bisnis, 8*(1), 45-51. doi:https://doi.org/10.30871/jaemb.v8i1.1707

Firdaus, B. A. (2021). Penentuan Masyarakat Miskin Penerima Zakat Menggunakan Algoritma K-Nearest Neighbor. *JASISFO (Jurnal SIstem Informasi), 2*(2), 191-204.

Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques, Fourth Edition.* Cambridge: Morgan Kaufmann.

Hasanah, M., & Mahdiyah, E. (2021). *Sistem Informasi Penentuan Penerima Dana Zakat Menggunakan Algoritma C4.5 di BAZNAS Kota Padang Panjang.* Riau: Universitas Riau.

Hassonah, M. A., Rodan, A., Al-Tamimi, A.-K., & Alsakran, J. (2019). Churn Prediction: A Comparative Study Using KNN and Decision Trees. *2019 Sixth HCT Information Technology Trends (ITT)* (pp. 182-186). Ras Al Khaimah, United Arab Emirates: IEEE. doi:doi: 10.1109/ITT48889.2019.9075077

Kamila, I. (2020). *Penerapan Data Mining dan Smarter Model untuk Klasifikasi dan Pendukung Keputusan Penerima Zakat Baznas.* Riau: Universitas Islam Negeri Sultan Sayarif Kasim.

Oktafriani, Y., Firmansyah, G., Tjahjono, B., & Widodo, A. M. (2023). Analysis of Data Mining Applications for Determining Credit Eligibility Using Classification Algorithms C4.5, Naïve Bayes, K-NN, and Random Forest. *Asian Journal of Social and Humanities, 1*(12), 1139-1158. doi:https://doi.org/10.59888/ajosh.v1i12.119

Pathak, A., & Pathak, S. (2020). Study on Decision Tree and KNN Algorithm for Intrusion Detection System. *International Journal of Engineering Research & Technology (IJERT), 9*(5), 376-381. doi:10.17577/IJERTV9IS050303

Rajaguru, H., & R, S. C. (2019). Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer. *Asian Pacific Journal of Cancer Prevention, 20*(12), 3777-3781. doi:10.31557/APJCP.2019.20.12.3777

Safradji. (2018). Zakat Konsumtif dan Zakat Produktif. *Tafhim Al-'Ilmi Jurnal Pendidikan dan Pemikiran Islam*, 59-66.

Sari, Y., Maulida, M., Gunawan, E., & Wahyudi, J. (2021). Artificial Intelligence Approach For BAZNAS Website Using K-Nearest Neighbor (KNN). *2021 Sixth International Conference on Informatics and Computing (ICIC)* (pp. 1-4). Jakarta, Indonesia: IEEE. doi:doi: 10.1109/ICIC54025.2021.9632954

Setiadi, T., Parnolo, A., & Prayitno. (2017). Klasifikasi Potensi Zakat di Lazismu DIY Menggunakan Metode K-Nearest Neighbor (K-NN) Berbasis Web Framework. *Jurnal Sarjana Teknik Informatika, 5*(3), 1-9. doi:http://dx.doi.org/10.12928/jstie.v5i3.12370

Yulianto, L. D., Triayudi, A., & Sholihati, I. D. (2020). Implementation Educational Data Mining For Analysis of Student. *Jurnal Mantik, 4*(1), 441-451.