Artificial Intelligence Approach For BAZNAS Website Using K-Nearest Neighbor (KNN)

Yuslena Sari
Information Technology
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
yuzlena@ulm.ac.id

Mutia Maulida
Information Technology
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
mutia.maulida@ulm.ac.id

Endi Gunawan
Informatics Engineering
STMIK Indonesia Banjarmasin
Banjarmasin, Indonesia
endy.stmiki@gmail.com

Johan Wahyudi
Informatics Engineering
STMIK Indonesia Banjarmasin
Banjarmasin, Indonesia
johan77@stmik.id

Abstract—**Amil Zakat National Agency (BAZNAS) is a national institution for the distribution of zakat. As one of the main foundations in Islam, zakat is, obviously, very important to be fulfilled. However, it is very often that the data of the recipient became unclear that it caused problems in terms of a fair distribution of zakat. This research tried to offer a solution by doing a classification of the recipient of zakat on the BAZNAS websites into two categories: indigent and poor, using K-Nearest Neighbor method. This research concluded that the accuracy of KNN method by using classification report, confusion matrix, and ROC-AUC respectively resulted in accuracy of 97%, 96.7%, and 97.7%**

Keywords—Zakat, Islam, K-Nearest Neighbor, BAZNAS

I. INTRODUCTION

Zakat is one type of worship that must be carried out by Muslims, in the form of giving a small portion of the property owned to other people in need, namely people in need. Along with the progress of the times, society is becoming more and more complex. If in the past people could distribute their zakat independently, now it is a bit difficult to do. Especially with the emergence of types of gated community housing or elite housing complexes that are far from the environment of underprivileged people [1].

As a country with largest population of Muslim, Indonesia certainly sees the distribution of zakat as very important. Therefore, the problem of distributing zakat is trying to be overcome by establishing National Amil Zakat Agency which a non-structural official organization build by government for distributing zakat, infaq, and alms. Along with the Government, BAZNAS are responsible for controling zakat management based on Islamic rule, trust, benefit and integration [2].

However, management by BAZNAS also often experiences problems, especially in determining the recipients of zakat due to the large amount of data available. This is certainly a problem for BAZNAS in carrying out the principles in accordance with Islamic law, trustworthiness, and justice because if the available data is not appropriate or experiencing problems, it is not impossible for zakat to be distributed with the wrong target and demands to give zakat to people who really need it are not met . Thus, data management of zakat recipients becomes important.

Data on zakat recipients obtained from the social service must be classified according to certain criteria in accordance with the provisions of Islamic law into groups of zakat recipients, namely the indigent and poor. In this case, the use of information technology can be relied upon, namely by applying artificial intelligence to the classification process. As research conducted by [3] those who use KNN as a method for classifying zakat recipients, a good accuracy value is 85%. This result is supported by another study by [4] comparing several classification methods to classify students' success factors on the thesis exam. The KNN method has the highest accuracy value compared to the Naïve Bayes and Decision Tree methods, which is 80.39%.

From those research results, it can be seen how the performance of KNN shows high accuracy as a classification method. Contribution of this study is to classify zakat recipients using the K-Nearest Neighbor classification method based on indigent and poor categories. The data used in this study are zakat recipients and prospective zakat recipients in 2017 - 2018 which were obtained from the National Amil Zakat Agency of Banjar Regency. Rest of this journal will be written into several parts which the second part describes related research and theories of classification methods. Research method used will be explained in the third section. As for research result and conclusion will be described in sections 4 and 5.

II. LITERATURE REVIEW

In carrying out a research, appropriate literature is needed in order to produce good quality research. Kind of literature review carried out in this study is finding theories related to zakat, data mining processes and the KNN algorithm. Other thing to do is reviewing related studies that have been carried out by several researchers about data classification.

A. Zakat

Zakat is one of the five pillars of Islam in the form of giving a certain part of the total assets that have met the requirements to people in need, called mustahik. In terms of Islamic jurisprudence (fiqh), the pillars are something that must exist or must be done [5]. The pillars of Islam thus mean a set of things that must exist or must be done by Muslims. So because zakat is included in the pillars of Islam, it means that all Muslims are obligated to pay zakat.

There are two types of zakat, namely zakat on property or zakat mal, namely zakat which is imposed on various types of assets whose acquisition does not conflict with religious provisions, and zakat fitrah which is carried out once a year, usually paid towards the end of the month of Ramadan [6]. As for the provisions of zakat, including the distribution method, the amount to be issued, the type of goods to be

given, the implementation time, the criteria that must be issued and those who are entitled to receive it, etc. ijtihad.

B. Data Mining

Data mining is the process of finding patterns in large data sets by involving many classifications algorithm. Many researchers who work in information technology area uses this technique to find different patterns from data sets. One of data mining technique that mostly use is classification which needed to classify an instance of testing dataset into associated class of training datasets [7]. There are many type of classification technique that can be used in data mining such as Naïve Bayes, Neural Networks, Apriori, K-Nearest Neighbor,etc. [8].

There are many approaches that can be used to expand different classification definition. Artificial neural networks, statistics-based and machine learning are often used to complete classification process in data mining. Application of classification technique mostly use for computer vision, geostatistics, speech recognition and biological classification [9]–[11].

C. K-Nearest Neighbor

A classification algorithm that works by classifying a class or group of an individual as part of the same class as the majority of its nearest neighbors. Thus, the basic algorithm is to classify individuals into certain classes, first we must find the k-nearest neighbors, then see which class is the majority of these neighbors. In the training phase, the method only reserves the vectors of the features and classifications of the training sample data. [12]

Data testing are used to calculate same features in classification stage. The sum of the nearest K values is taken from the distance of the new vector to the entire vector in the training data set. The highest classification value of these points is the classification point value of prediction. The Euclidean formula show in the following equation (1) used to measure the distance between two points between training and testing data.

$$D(x - y) = \frac{\sqrt{\Sigma k - 1\, n\, (xk - yk)^2}}{1} \tag{1}$$

Where D is the distance between the points in the training data x and the test data points y to be classified, where x = x1, x2, ..., xi and y = y1, y2, ..., yi and i represent the values attribute and n is the attribute dimension [13].

D. Related Research

Setyadi et al. conducted a study to classify the potential distribution of zakat at the Lazismu branch in Yogyakarta. This study uses the K-Nearest Neighbor algorithm to classify 14 branches by using the Super Independent, Independent, Self-Independent and Less Independent classes. The results showed that the K-NN method has an accuracy value of 85% with the ratio of test data and testing data is 80:20 [3].

Another study was also conducted by Solichin in 2019 related to the use of data mining algorithms. Classification is done to examine factors that affect students thesis graduation at Budi Luhur University. To get the best classification results, this study compares three classification methods, namely Naïve Bayes,k-Nearest Neighbors (kNN) and

Decision Tree. Based on result comparasion, it is known that KNN algorithm has best accuracy of 78.20% for classifying student thesis graduation [4].

In another study, Faisal et al. using the KNN algorithm as a classification method to analyze customer satisfaction with services at Cikarang Camera Rental. The variables used in this study consisted of prices, facilities, and loyalty services with satisfactory and unsatisfactory results. As for the results of this study, it is known that the K-NN algorithm used achieves an accuracy value of 98% and a precision value of 100%. Based on these results, it is used as a reference in developing an application to find out information related to customer satisfaction [14].

In 2021, the use of KNN as a classification method is also used by Hilal and Hasan in their research. This research aims to detect the SARS-CoV-2 virus that causes COVID-19 disease against other types of corona viruses. This detection process uses a classification method with K-NN using the complete genome sequence of the human coronavirus recorded in the 2019 Novel Coronavirus Resource. The results show that the use of K-NN as a classification method has an accuracy value of 98.4%[15].

III. RESEARCH METODOLOGY

Aiming to determine the accuracy of K-NN method in classifying zakat recipients on BAZNAS, the appropriate stages of research methods are needed. The research method used in this research are literature studies review, research data collection, research data analysis and performance evaluation.

A. Data Collection

This research used secondary data of zakat recipient from Badan Zakat Amil Nasional of Banjar district. The data will include different group of zakat recipient which are indigent and poor based on BAZNAS criteria. Resource of data are from surveys and district report that have been collected from 2017 to 2018.

B. Data Analysis

There are several stages that must be carried out in the data analysis process using the K-NN method. As described in Figure 1, the analysis phase begins with data pre-processing where variables that are not relevant to the classification requirements are removed. This stages is done to get a better result for classification accuracy. Some examples of zakat recipient data that have been processed according to the required variables are shown in Table 1.

To simplify the classification process using K-NN, zakat recipient data is converted from words to number. In this research, the numbers used to represent the data are 1 and 2. For the age variable with old value will be convert to 1 and middle age is 2, while for other variable with value Yes will be convert to 1 and No to 2. After going through the conversion process, the results of Table 1 conversion are shown in Table 2.

TABLE I. PRE-PROCESSING DATA FOR ZAKAT RECIPIENTS

Age	Residential status	Job Status	Disability	Chronic Diseases	Class
Old	no	no	no	yes	Indigent
Middle	alone	no	no	no	Poor

age					
Middle age	alone	yes	no	no	Poor
Middle age	alone	yes	no	no	Poor
Old	no	no	no	yes	Indigent
Old	no	no	no	no	Indigent
Middle age	alone	yes	no	no	Poor
Old	no	no	no	yes	Indigent
Middle age	alone	yes	no	no	Poor
Middle age	no	no	no	yes	Indigent
Middle age	alone	no	no	no	Poor
Middle age	alone	yes	no	no	Poor
Middle age	alone	yes	no	no	Poor
Old	no	no	no	yes	Indigent

2	1	1	2	2	Poor
1	2	2	2	1	Indigent
2	1	1	2	2	Poor
2	2	2	2	1	Indigent
2	1	2	2	2	Poor
2	1	1	2	2	Poor
2	1	1	2	2	Poor
1	2	2	2	1	Indigent

Data analysis was carried out to obtain results and performance using KNN classification method. This method will be implemented into a system using Python programming. The phyton package used is scikit-learn which facilitate the use of the KNN method. The value of k is defined as 1, because there are only two class.
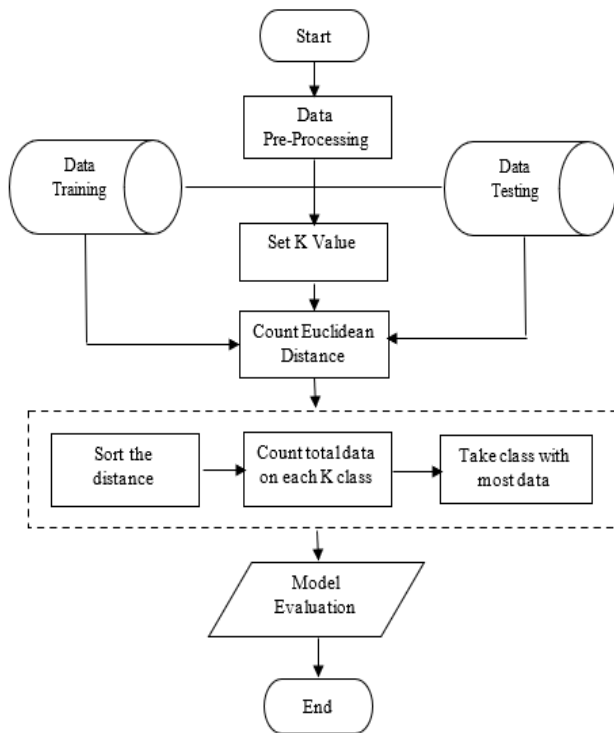

Fig. 1. Flowchart of Classification Using K-NN

TABLE II. ZAKAT RECIPIENTS DATA AFTER CONVERSION

X1	X2	X3	X4	X5	Class
1	2	2	2	1	Indigent
2	1	2	2	2	Poor
2	1	1	2	2	Poor
2	1	1	2	2	Poor
1	2	2	2	1	Indigent
1	2	2	2	2	Indigent

C. Performance Evaluation

Performance evaluation is the stage of researchers to measure method accuracy that have been applied to the system. There are three evaluation method are used in this research such as classification report, confusion matrix and ROC-AUC score. Classification report is one of method that can be used to measure performance value of classification method.This method will calculate precision, recall and F1-Score of each class in classification [16]. Formula of each measurement are shown respectively in equation 2-4 below.

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

$$F1-Score = \frac{Precision*Recall}{Precision+Recall} \qquad (4)$$

Another evaluation method used in this research is confusion matrix which used to know the quality of the classifier of different classes. Confusion matrix shown as table with size n x n, where n is number of different classes which shows predicted and actual classification value. Based on data from confusion matrix, it could calculate prediction accuracy [17].

IV. RESULT AND DISCUSSION

The number of data used in this research for both class is 300, with each of 152 data with the class "Indigent" and 148 for the label "Poor". With a random data collection and ratio of 70:30 between training data and test data 70:30 performance test is obtained as follows.

Performance measurement using the KNN algorithm is done with a classification report using Python programming language. Classification report is a method that comes from one of the packages available with the Python programming language. The results of the classification report can be seen in Table 3.

TABLE III. CLASSIFICATION REPORT

Class	Precision	Recall	F1-score
Indigent	98%	96%	97%
Poor	96%	98%	97%
Accuracy			97%

By using precision, recall, and F1-score, the micro accuracy results obtained from each class are 97%. Meanwhile, in confusion matrix, the amount of test data used is 90 and 210 for training data for the determination of "Indigent " and "Poor" class. The data is taken randomly and the confusion matrix obtained as shown in Table 4.

TABLE IV.   CONFUSION MATRIX

	Positive	Negative
Positive	45	1
Negative	2	42

Based on data from confusion matrix, it used to calculate precision, recall and accuracy of K-NN as classification method. Precision and recall value of K-NN are 97,8% and 95,7%. As for accuracy value of K-NN to clasify data of recipient zakat is 96,7%. In addition to the confusion matrix, the ROC-AUC method is also used in measuring performance in this study. By using this method, the results obtained are 97.7%.

V. CONCLUSION

Based on this study it shows that K-Nearest Neighbor are promising classification method that can be used to classify zakat recipient based on right criteria. From testing results of accuracy value with classification report, confusion matrix and ROC-AUC score are 97%, 96,7% and 97,7% respectively. Limitation of this research are data sources used are only limited to Banjar districts which affected to less amount of training data sets.

Further research related to classification for both zakat recipient data and other objects can be carried out using other classification methods. Using other methods such as Naïve Bayes, Decision Tree, Support Vector Machine may be able to provide a better accuracy value. Determination of variables used to perform calculations and experiments on using different K values can provide better accuracy results for the K-NN method.

REFERENCES

[1] N. Mohamed Salah and H. M. Ayad, "Why people choose gated communities: A case study of Alexandria metropolitan area," *Alexandria Eng. J.*, 2018, doi: 10.1016/j.aej.2017.10.008.

[2] B. A. Z. Nasional, "Tentang BAZNAS," 2019. https://baznas.go.id/profil (accessed May 21, 2021).

[3] "KLASIFIKASI POTENSI ZAKAT DI LAZISMU DIY MENGGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN) BERBASIS WEB FRAMEWORK," *JSTIE (Jurnal Sarj. Tek. Inform.*, 2019, doi: 10.12928/jstie.v5i3.12370.

[4] A. Solichin, "Comparison of decision tree, Naïve Bayes and K-nearest neighbors for predicting thesis graduation," 2019, doi: 10.23919/EECSI48112.2019.8977081.

[5] D. Djuwaini, "Pengantar Fiqih Muamalah," *Yogyakarta: Pustaka Pelajar*, 2008.

[6] B. A. Z. Nasional, *Fikih Zakat Kontekstual Indonesia*. Jakarta: Badan Amil Zakat Nasional, 2018.

[7] "The Cambridge handbook of artificial intelligence," *Choice Rev. Online*, 2015, doi: 10.5860/choice.187061.

[8] N. Satyanarayana, C. Ramalingaswamy, and Y. Ramadevi, "Survey of Classification Techniques in Data Mining," *IJISET -International J. Innov. Sci. Eng. Technol.*, 2014.

[9] A. S. Fitrani and W. Novarika, "Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection in Jemirahan Village , Jabon District," *Int. J. Informatics Comput. Sci.*, vol. 3, no. 2, pp. 66–79, 2019, doi: 10.30865/ijics.v3i2.1391.

[10] C. Lin and F. Yan, "The Study on Classification and Prediction for Data Mining," 2015, doi: 10.1109/ICMTMA.2015.318.

[11] W. Lefebvre-Ulrikson, F. Vurpillot, and X. Sauvage, *Atom Probe Tomography: Put Theory Into Practice*. 2016.

[12] A. R. Lubis, M. Lubis, and Al-Khowarizmi, "Optimization of distance formula in k-nearest neighbor method," *Bull. Electr. Eng. Informatics*, 2020, doi: 10.11591/eei.v9i1.1464.

[13] A. M. Siregar and A. Pusphabuana, *Data Mining: Pengolahan Informasi Data dengan Rapid Miner*. Surakarta: Kekata Grup, 2017.

[14] S. F. Sutan Faisal and N. Sutan Faisal, "Implementation of K-Nearest Neighbor Algorithm for Customer Satisfaction," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, 2020, doi: 10.36805/bit-cs.v1i2.886.

[15] H. Arslan and H. Arslan, "A new COVID-19 detection method from human genome sequences using CpG island features and KNN classifier," *Eng. Sci. Technol. an Int. J.*, 2021, doi: 10.1016/j.jestch.2020.12.026.

[16] Y. Sari, M. Alkaff, and M. Maulida, "Classification of Rice Leaf using Fuzzy Logic and Hue Saturation Value (HSV) to Determine Fertilizer Dosage," 2020, doi: 10.1109/ICIC50835.2020.9288585.

[17] N. Nofriani, "Comparisons of Supervised Machine Learning Techniques in Predicting the Classification of the Household's Welfare Status," *J. Pekommas*, 2019, doi: 10.30818/jpkm.2019.2040105.

ICIC 2021

IEEE INDONESIA SECTION

APTIKOM
ASOSIASI PENDIDIKAN TINGGI INFORMATIKA DAN KOMPUTER

UNIVERSITAS NUSA MANDIRI

Certificate of Participation

This is to certify that

Yuslena Sari

has presented for the Paper titled

Artificial Intellegence Approach For BAZNAS Website Using K-Nearest Neighbor (K-NN)

at the 6th International Conference on Informatics and Computing (ICIC2021)
held between 3-4 November 2021
at Virtual Conference of ICIC 2021

General Chair of APTIKOM,

Zainal A. Hasibuan

APTIKOM
ASOSIASI PENDIDIKAN TINGGI INFORMATIKA DAN KOMPUTER

Organising Chair of ICIC 2021

Yusuf Durachman

ICIC 2021

Sponshorship :

LEMBAGA SERTIFIKASI PROFESI
INFORMATIKA

Co-Host :

BINA INSANI UNIVERSITY

MDP

UNIVERSITAS BUDI LUHUR

UNIVERSITAS MIKROSKIL

BINUS UNIVERSITY

Telkom University

STMIK LIKMI
School of Business & IT

STIKI

UNIVERSITAS AMIKOM YOGYAKARTA