

Sentiment Analysis to Identify Public Opinion for Zakat Implementation in Indonesia Using Machine Learning Algorithms

SRI REDJEKI*, SETYAWAN WIDYARTO**

*Informatics Department, Indonesia Digital Technology University, Karang Jambe, 55198, Yogyakarta, INDONESIA

** Faculty of Communication Visual Art and Computing, Universiti Selangor, Shah Alam, Selangor, MALAYSIA

Abstract: - This research aims to explore and identify public opinion related to Zakat in Indonesia by utilizing big data technology through sentiment analysis. The source of Twitter social media public opinion data is used in the research. Proper keywords need to be determined before crawling Twitter data collections. The Twitter data collected in the research was 1060 Twit at the beginning of the year 2020. Indonesian community opinion identification, a part of the machine learning method, was a method used in training and testing data requirements. Opinions are classified into 3 classes i.e. positive, neutral and negative. The identification shows that the Random Forest Classifier method and the Naïve Bayes method provide an accuracy value of 86% and 81%. While the method that produces the lowest accuracy is the Support Vector Machine method of 53%. The accuracy results show that the sentiment of analysis can be used as an "early warning system" for the decision-makers of Zakat in Indonesia.

Key-Words: - Machine Learning, Opinion, Sentiment Analysis, Twitter, Zakat

Received: June 18, 2021. Revised: March 25, 2022. Accepted: April 27, 2022. Published: July 2, 2022.

1 Introduction

According to We are Social data, there are 175.4 million internet users in Indonesia in early 2020. Compared to the previous year, there was an increase of 17% or 25 million users. The data shows that 64% of Indonesia's population is connected to the internet and social media. The size of the population, the rapid growth of internet and telephone users are potential for the national digital economy. One source of data on big data is social media that has high quality and can be directly consumed by the public. Many researchers have given more attention to social media. Data analysis using social media is not easy because of its incompleteness and dynamic nature [1,2].

The analysis of social media data that is most widely used to identify public opinion is sentiment analysis. This method is widely used for the prediction or early identification of an organization in capturing the opinions of social media [3,4]. Social media attract computer scientists and statisticians as research opportunities because social media provides an impressive amount of data about users and their interactions. The most discussed and interesting thing is predicting future events and developments from social media data [5,6].

Sentiment Analysis or Opinion Mining is a way to compare people's emotions, attitudes, and opinions about a particular entity. Entities can represent individuals, events, or topics [7,8]. The main focus of Opinion Mining is to extract and analyze people's opinions about an entity while sentiment analysis identifies the sentiments expressed in the text and then analyzes them.

This study aims to model the analysis of public opinion related to the management of zakat by zakat institutions in Indonesia in the form of sentiment analysis. Zakat, seen from the perspective of economics, has a positive correlation with consumption figures that will drive the economy. The macro consumption model is determined by basic consumption and consumption that comes from income. If seen from the *mustahik* side, then zakat will increase the aggregate of basic consumption, namely the accumulation of basic consumption [9].

Management and distribution of zakat that is less than the maximum or the emergence of inefficiency is a concern for zakat managers in Indonesia [10]. Good zakat management will have an impact on people's welfare [11]. The standardization of zakat management in Indonesia is managed by the government zakat institution, the National Zakat

Board (BAZNAS). The potential and realization of zakat receipts in Indonesia increased during the 2011-2015 period. However, there is a very large gap between the potential value and the amount of realization. During 2011-2015 the realization of zakat receipts was less than 1%. One possibility is that the majority of Indonesians prefer to distribute their zakat directly to muzakki, so it is not recorded [12]. There are increasing problems at Indonesian Zakat Institutions (IZI), namely socialization costs, the number of volunteers, the amount of zakat collected, and the distribution of consumptive zakat, which cumulatively reduces the intermediation function of zakat institutions. IZI must resolve this problem to improve the efficiency of the zakat institution [13].

Zakat institutions are institutions that are responsible for managing the collection and distribution of zakat funds. The Qur'an explains the appointment of officials to collect and distribute zakat, which must be in the form of state departments or public funds that are fully managed by public bodies [14]. Zakat has become an important source of funds for the development of Islam since the entry of Islam in Indonesia. During the Dutch colonial period, the Colonial Government-issued *Bijblad* Number 1892 on August 4, 1893, which contained the Colonial Government's policy regarding zakat. The purpose of the issuance of this regulation is to prevent the occurrence of zakat financial misappropriation by the magicians. The *naib* worked to carry out the administration of the power of the Dutch Colonial Government without getting a salary to finance their lives [9].

Some critical opinions in all involved parties would be categorized into several groups. Opinions from a group who received zakat (*mustahik*), who distribute the zakat (*amil*), and who pay the zakat (*muzakki*). However, there is a significant percentage of parties that are not in those groups. They don't either pay or receive the zakat and they are also not the distributor. Indeed, based on Islamic guidelines there are very clear requirements for every group. However, in real life, it's not easy to identify each group as no sufficient data, no law enforcement, relative term definition, and other human being related characters.

From the *mustahik* groups, the consumptive patterns and productive patterns would be important issues in utilizing Zakat aid. It's believed *mustahik* would utilize zakat in productive spending and empower them in terms of poverty alleviation [15]. It is also supported by [16].

World Bank reported in the last two decades the majority of the poor and vulnerable in Indonesia have

uplifted from poverty into the aspiring middle class even though determinant of those groups tends to measure by ending up in how much money of the spending. Based on [18] there are approximately 115 million people who belong in the aspiring middle-class category.

On the other hand, based on [18] the potential total of Zakat in Indonesia right now is Rp 233,84 trillion. Furthermore, the reinforcement of zakat management through financial technology systems has been studied [20] but it is difficult to find an existing work of social media involvement in Zakat matters in Indonesia. The realization of BAZNAS fund distribution for the Covid-19 handling program has reached more than Rp 16 billion [19].

Report dissemination through social media would help in transparency and trust from Zakat stakeholders. One of social media is Twitter and in analyzing comments from Twitter is sentiment analysis. However, sentiment classification will result in sentimental polarities. It is common knowledge that most universe phenomena have been in pairing. In our sentiment analysis, we also put word labeling in two labels i.e. positive and negative but we are also aware of a position in the middle or neutral. Thus, sentiment classification is a central issue in sentiment analysis [20]. Positive could be interpreted as good and negative as bad.

Machine learning creates a model based on regularities found in vast volumes of data, which is then used to generate predictions and categorize raw data. There are a number of different machine learning algorithms which all have different use cases and may result in different accuracy with the same case. This paper research applied seven algorithms in machine learning i.e. Support Vector Machine (SVM), Logistics Regression, K-Nearest Neighbor (KNN), Decision Tree, Gaussian Naïve Bayes, Boost Classifier, and Random Forest Classifier.

2 Methodology

In conducting research, identification of public opinion related to zakat management uses a sentiment analysis approach from Twitter social media. Data on public opinion was obtained from tweets during the March period with a total of 1060 data tweets that have been cleared. The process of crawling twitter data in this study is illustrated in Figure 1. Before performing the data crawling process, first, identify the source of the research object, reference sources are obtained from several journals and books related to zakat, sentiment analysis, and machine learning. All available sources are used to determine the right keywords at the time of crawling the data.

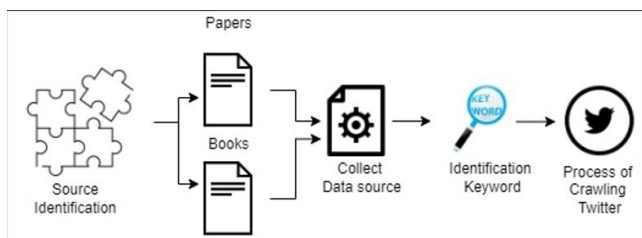


Fig. 1 Block Diagram of Crawling Process

The stages of the research carried out are shown in Figure 2. The first step taken is to define data requirements. The data used is social media data on Twitter, so it needs to determine the right keywords in crawling data. The main objective of this study is to obtain input from the public on the management of zakat in Indonesia by the National Alms Board (BAZNAS). The second step in the form of data collection in the form of text from Twitter was taken in the period March 2020 because this month the government began implementing the period of large-scale social restrictions (PSBB) and the community began to be affected by the economic downturn and how BAZNAS assisted the community with the impact of Covid-19 on the people's economy. After the data collection is done the data cleaning aims to clean the tweet data so that the text data will be easily processed. Data cleaning steps can be seen in Figure 3.

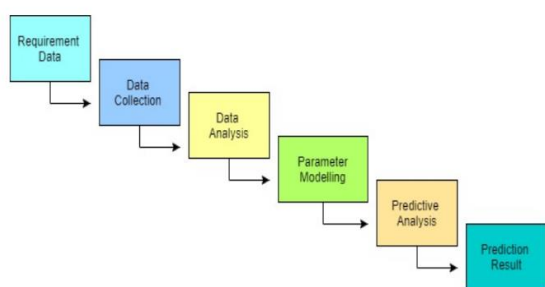


Fig. 2 Stages of Research Methodology

The third step is to analyze the data by determining the model for opinion classification in the form of sentiment analysis. The next step is to determine the parameters of the model used to classify. The dataset will be divided into 2 namely training data and testing data. The model parameters used are to conduct training and data testing is used to determine the accuracy of the existing models. The fifth and final step is to test predictions of new opinion data and display prediction results based on the model with the best accuracy.

This study uses 7 machine learning algorithm models to classify the support vector machine (SVM), Logistics Regression, KNN, Decision Tree, Gaussian Naïve Bayes, Boost Classifier, and Random Forest Classifier. The software used for data processing uses the sklearn library in Python that is run using Jupyter.

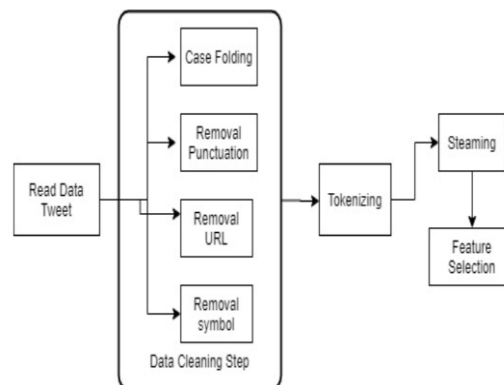


Fig. 3 Data Cleaning Process

The data cleaning process shown in Figure 3 includes:

- a. Case Folding is the first pre-processing step we will take is to turn our tweets into smaller cases.
- b. Removing Punctuations because when handling text data, it does not add any additional details. Removing all instances of this would also help us reduce the size of the training results.
- c. Removal URL is deleting the URL in the Twitter text
- d. Removal Symbol is removing all symbols that have no meaning
- e. Tokenization refers to breaking the text into a series of words or phrases.
- f. Stemming is the process of changing a word into its root word by removing affixes to words in documents or changing a verb to a noun.

The stages of data cleaning carried out in this study used modules in python. One of the stages of data cleaning tokenization is shown in the coding snippet in Figure 4.

The frequency of occurrence of words from the three opinion groups is displayed in the form of a histogram in Figures 9-11.

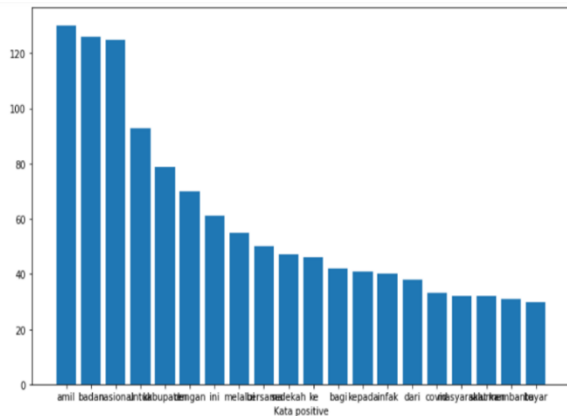


Fig. 9 Frequency Positive Opinion

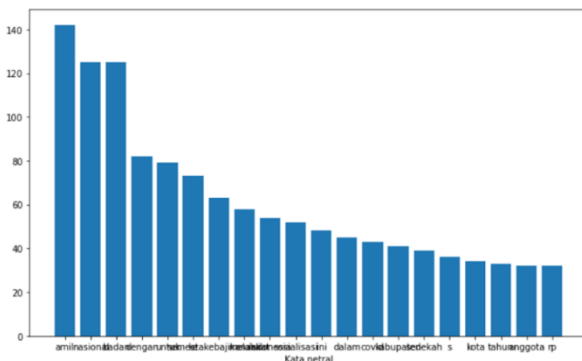


Fig. 10 Frequency Neutral Opinion

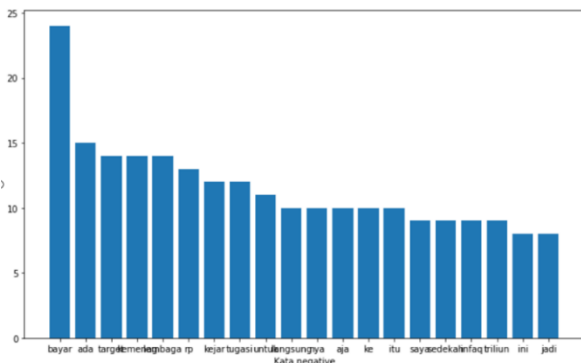


Fig. 11 Frequency Negative Opinion

The dataset used in this study is divided into 2, namely for training 80% or 1038 data and for testing 20% or 260 data. The distribution of the dataset for training and testing is shown in Figure 12.

```
X_train, X_test, y_train, y_test = train_test_split(df.new_text, df.label.astype(float),
                                                test_size=0.2, random_state=42)

print("Text train shape : ", X_train.shape)
print("Label train shape : ", y_train.shape)
print("Text test shape : ", X_test.shape)
print("Label test shape : ", y_test.shape)

Text train shape : (1038,)
Label train shape : (1038,)
Text test shape : (260,)
Label test shape : (260,)
```

Fig. 12 Dataset of Research

Figures 9 and 10 show that the words of the national amil body have almost the same frequency, meaning that the public still sees that one of the national amil bodies has pretty good performance because the number of neutral opinions is still more than positive opinions. While the frequency of words that appear on negative opinions relates to the payment of zakat and there are still people who have not fully entrusted the payment of zakat to the national zakat body. The public also sees a negative thing related to the BAZNAS target in collecting zakat.

The model used to test the class identification of public opinion shows varying accuracy from some of the models used. The accuracy of each model is shown in Figures 13 and Table 1.

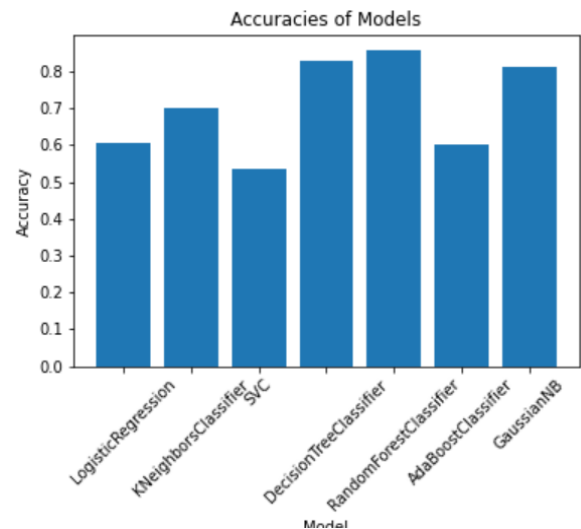


Fig. 13 Histogram of Model Accuracy

Table 1 Machine Learning Algorithm Classification Accuracy

Machine Learning Algorithms	Accuracy
Logistic Regression	0,6069
KNeighbors Classifier	0,7013
SVM	0,5346
Decision Tree	0,8270
Random Forest Classifier	0,8553
AdaBoost Classifier	0,6006

Gaussian

0,8113

The 7 classification models used in this study obtained an accuracy of above 80% is the 86% Random Forest Classifier, 83% Decision Tree Classifier, and 81% Naïve Bayes Classifier. While the model with the smallest accuracy is in the Support Vector Machine model of 53%.

4 Conclusion

In this study, the results obtained that the identification of community opinion about the management of zakat by the Indonesian BAZNAS by using social media data-based sentiment classification approach can be used as additional knowledge by BAZNAS in decision making. The big data approach used to identify opinions in this study is able to provide an accurate value of 86% for the Random Forest Classifier method, while other methods have an accuracy value below that value. Data on the classification of positive public opinion towards BAZNAS was 42%, more than 50% showed a neutral attitude and less than 10% showed negative opinions. Twitter sampling as a dataset is strongly influenced by several parameters including the keywords used when crawling data, the retrieval period, and the amount of data taken. This parameter will affect the results of the final identification of public opinion on existing research objects.

References:

- [1] P. Wang, B. W. Xu, Y. R. Wu, and X. Y. Zhou, 2014. Link prediction in social networks: the state-of-the-art, *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–38. Doi: 10.1007/s11432-014-5237-y.
- [2] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein. 2015. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLoS Comput. Biol.*, vol. 11, no. 10, pp. 1–15. Doi: 10.1371/journal.pcbi.1004513.
- [3] M. V. Mäntylä, D. Graziotin, and M. Kuuttila. 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, vol. 27. Elsevier Ireland Ltd, pp. 16–32. Doi: 10.1016/j.cosrev.2017.10.002.
- [4] J. L. O. Hui, G. K. Hoon, and W. M. N. W. Zainon. 2017. Effects of Word Class and Text Position in Sentiment-based News Classification. *Procedia Comput. Sci.*, vol. 124, pp. 77–85. Doi: 10.1016/j.procs.2017.12.132.
- [5] H. Schoen, D. Gayo-Avello, P. Takis Metaxas, E. Mustafaraj, M. Strohmaier, and P. Gloor. 2013. The power of prediction with social media. *Internet Res.*, vol. 23, no. 5, pp. 528–543, Oct. Doi: 10.1108/IntR-06-2013-0115.
- [6] M. S. Neethu and R. Rajasree. 2013. Sentiment analysis in twitter using machine learning techniques. 4th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT. Doi: 10.1109/ICCCNT.2013.6726818.
- [7] M. Ridzwan Yaakub, M. Iqbal Abu Latiffi, and L. Safra Zaabar. 2019. A Review on Sentiment Analysis Techniques and Applications. *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 551, no. 1. Doi: 10.1088/1757-899X/551/1/012070.
- [8] C. Whitelaw, N. Garg, and S. Argamon. 2005. Using appraisal groups for sentiment analysis. *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 625–631. Doi: 10.1145/1099554.1099714.
- [9] Muhammad Aziz I. 2014. *Regulasi Zakat Di Indonesia; Upaya Menuju Pengelolaan Zakat Yang Profesional*. vol. IV, no. 01.
- [10] S. Al Parisi. 2017. Tingkat Efisiensi dan Produktivitas Lembaga Zakat di Indonesia. *Esensi*, vol. 7, no. 1. Doi: 10.15408/ess.v7i1.3687.
- [11] C. Canggih, K. Fikriyah, and A. Yasin. 2017. Potensi Dan Realisasi Dana Zakat Indonesia. *al-Uqud J. Islam. Econ.*, vol. 1, no. 1, p. 14. Doi: 10.26740/al-uqud.v1n1.p14-26.
- [12] S. Al-Ayubi, A. Ascarya, and B. T. Possumah, 2018. Examining the efficiency of zakat management: Indonesian zakat institutions experiences. *Int. J. Zakat*, vol. 3, no. 1, pp. 37–55. Doi: 10.37706/IJAZ.V3I1.66.
- [13] I. M. Alonso. 2015. Crowdfunding in Islamic finance and microfinance: A case study of Egypt, vol. 1.
- [14] S. Sumai, A. N. Mutmainnah, Nurhamdah, and M. Arsyad. 2019. Role of zakat in poverty reduction and food security. *IOP Conf. Ser. Earth Environ. Sci.*, vol. 343, no. 1. Doi: 10.1088/1755-1315/343/1/012254.
- [15] M. Y. Qamaruddin, S. M. Anwar, M. Surullah, Azaluddin, and Jusni. 2019. Zakat expenditure patterns and its relationships with the improvement of prosperity and environment. *IOP Conf. Ser. Earth Environ. Sci.*, vol. 343, no. 1. Doi: 10.1088/1755-1315/343/1/012147.
- [16] Aspiring Indones. 2019. *Aspiring Indonesia—Expanding the Middle Class*. Aspiring Indones. Middle Cl. Doi: 10.1596/33237

- [17] P. K. Strategis, B. Amil, and Z. Nasional, OUTLOOK. 2020.
- [18] M. Aulia Rachman and A. Nur Salam. 2018. The Reinforcement of Zakat Management through Financial Technology Systems. Int. J. Zakat, vol. 3, no. 1, pp. 57–69.
- [19] BAZNAS, “LAPORAN BAZNAS DALAM PENANGANAN PANDEMI COVID-19,” p. 400, 2020.
- [20] Y. Xu, L. Li, H. Gao, L. Hei, R. Li, and Y. Wang. 2020. Sentiment classification with adversarial learning and attention mechanism. Comput. Intell., no. December 2019, pp. 1–25. Doi: 10.1111/coin.12329.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

**Sri Redjeki carried out the data crawling and the programming.
Setyawan Widyarto has analysed the Algorithms used.**

[Follow: www.wseas.org/multimedia/contributor-role-instruction.pdf](http://www.wseas.org/multimedia/contributor-role-instruction.pdf)

Sources of funding for research presented in a scientific article or scientific article itself

This research is fully funded by Indonesia Digital Technology University (UTDI) and was done in Universiti Selangor, Malaysia.