Review of Accounting, Finance and Governance

Journal homepage: https://akuntansi.pnp.ac.id/rafgo

Sentiment Analysis of Indonesian Government Policy in the Environmental Sector Using Machine Learning Method

Amy Fontanella[1], Hendrick[2], Hidayatul Ihsan[3], Zhi-Hao Wang[4] , Igra [5]

[1]Accounting Department, Padang State Polytechnic, Indonesia
[2]Electronic Department, Padang State Polytechnic, Indonesia
[3]Accounting Department, Padang State Polytechnic, Indonesia
[4]Department of Information Management, Southern Taiwan University of Science and Technology, Taiwan
[5]Accounting Department, Padang State Polytechnic, Indonesia

ARTICLE INFO

Keywords:
Coal waste,
FABA waste,
Toxic waste, Sentiment Analysis

Received : 12 May 2021
Accepted : 25 May 2021
Published : 4 June 2021

ABSTRACT

The study aims to analyse government policies related to the issuance of PP no. 22/2021 based on sentiment analysis on social media, especially Twitter. Data collection is using three keywords such as coal waste, FABA waste and toxic waste through Twitter API. 236 tweets are obtained and labelled into positive and negative sentiments. The dataset is grouped into training and testing data. Training data consists of 50 tweets and testing data consists of 186 tweets. The cleansing process is carried out by tokenizing, transform cases, stop word filters and comparison of classification models. The results showed that the public opinions tended to be negative sentiments with the accuracy rate is 77.40%.

Introduction

One of the hot topics that are currently being discussed on social media is when President Joko Widodo excluded coal waste from the category of hazardous and toxic waste (B3). This is stated in Government Regulation (PP) Number 22/2021 concerning Implementation of Environmental Protection and Management in attachment 14 of PP Number 22/2021 which states that the types of coal waste removed from the B3 waste category are fly ash and bottom ash (FABA), provided that the two wastes are sourced from the coal combustion process at a steam power plant (PLTU) or from other activities using technology other than stocker boilers and/or industrial furnaces. Whereas from the previous regulation in article 54 paragraph 1 letter a of PP Number 101/2014 concerning Hazardous Waste Management, it is stated that fly ash from burning coal in PLTU activities belongs to the category of B3 waste, as quoted from the Kompas.com page on March 11, 2021.

The emergence of PP No. 22/2021 has sparked a fierce debate from experts in the field of environment and society. Experts are concerned that the release of FABA from the hazardous waste category will make it increasingly difficult to control the level of environmental pollution, which will result in decreased air quality that can harm living things around it. The loud public response to PP No. 22/2021 gives a signal that the government lacks communication when issuing a regulation related to the public interest. For this reason, it is necessary to conduct research to analyse government policies related to the issuance of PP no. 22/2021. Usually to collect public opinion on certain topic/issues, researchers do the interviews or distributing the questionnaires and then analysing the responses. Researchers today can find out public response based on sentiment analysis on social media, especially Twitter.

This study uses the Natural Language Toolkit (NLTK) sentiment analysis method. Natural language is a part of computer science, which is a more specific subsection of artificial intelligence, which allows computers to understand, process

and manipulate the language used by humans. The computer's ability to understand language is not only in the form of audio but is also able to translate text or writing (Shayaa et al., 2018). All responses in the form of text or writing earlier, will be analyzed using the sentiment analysis method. Sentiment analysis will group all the text into positive, negative and neutral sentiments. Based on this, we can find out how social media users respond to the hazardous waste regulations.

From this research, it is hoped that it will be known how the community actually responds to government policies related to the release of FABA from the hazardous waste category. The results of this sentiment analysis can be used as input for the government to re-analyze public policies that are pros and cons in society. This will improve the image of the government in the eyes of the public as a government that is responsive to the dynamics that occur in society.

Literature Review

Machine Learning

Machine learning is a method that consists of stages of data collection, pre-processing, feature extraction and training. In general, machine learning is divided into three groups in the training process, namely supervised learning, unsupervised learning and reinforcement learning. Machine learning has evolved to be able to perform classifications such as writing, and voice to be specific to text mining. Text mining, also known as text data mining, is a process of converting unstructured text into structured text to recognize concepts, patterns, topics and attributes in the data (F. R. Lucini et al., 2017). One method of text mining is sentiment analysis which has the ability to analyze a problem or product based on text.

Natural Language Toolkit

Natural language toolkit or commonly abbreviated as NLTK is a python-based platform developed to process NLTK text data equipped with more than 50 corpora and lexical resources. In addition, NLTK also provides libraries for text mining ranging from classification, tokenization, stemming, tagging, parsing. One of the features possessed by NLTK is sentiment analysis.

Sentiment Analysis

Sentiment analysis is a method used to measure and identify feelings related to certain events (Saura, Herraez, &Reyes-Menendez, 2019). Sentiment analysis is a part of data mining whose main purpose is to capture and categorize the feelings expressed in words which are then grouped into positive or negative sentiments.

Currently, many people share their ideas, thoughts or opinions on various social media about various things, ranging from product reviews, political talks, films, music, sports and others. By conducting sentiment analysis, producers can find out how consumers respond to their products, political parties can find out how the image of their cadres in the eyes of the community, the government can find out public opinion about government policies, and so on.

The results of sentiment analysis are believed to be more accurate and describe the actual situation than the survey methods that have been carried out so far, because this sentiment analysis is based on public opinion that develops in real time and is natural without being engineered, so that decision making based on sentiment analysis is expected to be more logical, rational. and has a relatively small risk of rejection (Prastyo, Sumi, Dian, &Permanasari, 2020).

Twitter

Twitter is an online social networking and microblogging service that allows users to send and read text-based messages of up to 280 characters known as tweets (Elbagir &Yang, 2019). Twitter was founded in March 2006 by Jack Dorsey and in July, the social networking site was launched. Twitter Inc. is based in San Francisco, with additional servers and offices in New York City, Boston, and San Antonio.

Twitter is an open platform managed by people who are competent in the field of business, program developers, and marketing. This team will later assist the customer/advertiser in analyzing the advertised product. The results of this analysis can be used for decision making for advertisers.

Twitter has a feature that uses several symbols, including the @ symbol which is used to call someone's username on Twitter or to send a message to the person concerned. Another popular symbol is the # symbol (hashtag) which is used to mark a particular topic or issue on Twitter. The hashtag symbol also makes it easier for us to find news related to the topic used. For example #prayforNTT which is used to mark news about the current tropical cyclone disaster (6/4/2021). With these characteristics, Twitter is an ideal source for providing data sources for use in text mining research and sentiment analysis (Hasan, Maliha, &Arifuzzaman, 2019).

Research methods

The research process will be carried out in several stages, namely data collection, data pre-processing and data processing as shown in Figure 1 below.

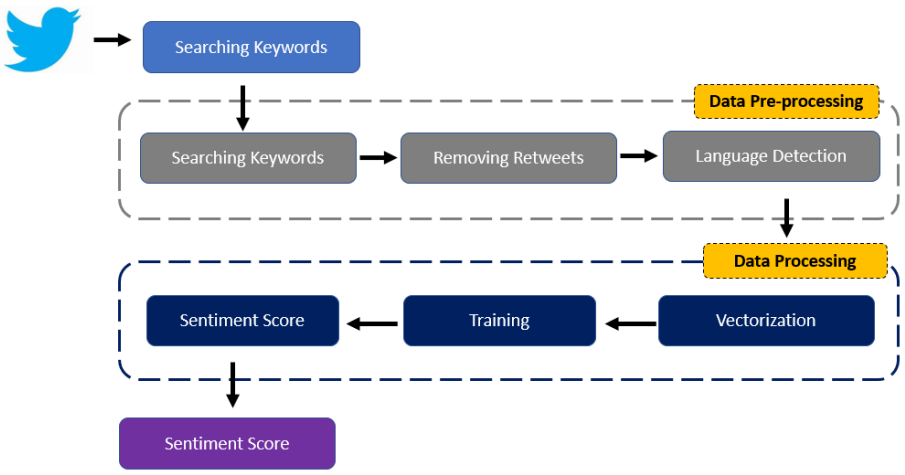

Figure 1. Research Block Diagram

Data Collection

Data collection is carried out through several stages:

1. Twitter API registration. The difference between the Twitter API and regular twitter lies in the access it has. The Twitter API account can be used through the program so that the text collection process can be done easily.
2. Searching keywords using search engines in the NLTK library. The keywords that are searched use Indonesian language settings and try to use words that are easily recognizable on Twitter. This research uses keywords such as : coal waste and FABA waste.

Pre-Processing Data

Data Pre-processing is the initial process carried out to obtain information about Twitter users' responses to the keywords mentioned before. All Twitter user responses will be saved in a file of type csv. Furthermore this file will be referred to as a dataset. The next pre-processing stage is to remove retweet from Twitter users. This is used to eliminate responses that sometimes have nothing to do with the topic being discussed on Twitter. The last stage in pre-processing is language detection. This process serves to narrow down the comments obtained on Twitter so that comments that will be taken are only in Indonesian language.

Data Processing

This stage is a decisive stage in analysing all text or words in the dataset. The datasets obtained from the data collection and pre-processing process will be labelled according to the desired targets such as positive sentiment, negative sentiment and neutral sentiment. To carry out training, a machine learning process will create a model according to the target. The final process of training will produce a model that is able to give sentiment to words that appear on Twitter.

Results

The collected tweets responses is separated into two groups which are dataset and data test. Dataset is used to create a model in training process consists of 50 tweets, while the testing data consists of 186 tweets. All 236 tweets were labelled into positive and negative sentiment. The labelling process is done manually, because the Rapid miner cannot detect text in Indonesian language.. The cleansing process is carried out by tokenizing, transform cases, stop word filters and comparison of classification models so that the design model process is obtained as shown in Figure 2.

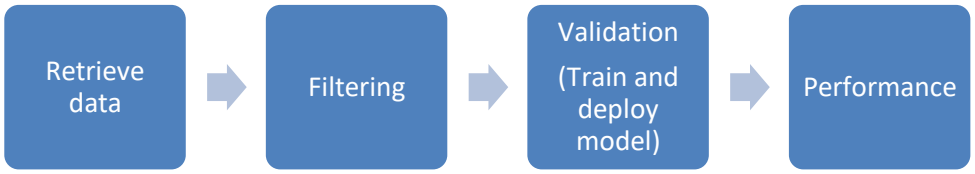

Figure 2. Design Model Process

The results showed that the public opinions on the environmental policies tended to be negative sentiments and has average as shown in Table 1 below:

Table 1. Performance Measurement

	True Negative	True Positive	Class Precision
Prediction Negative	89	20	81.65%
Prediction Positive	22	55	71.43%
Class Recall	80.18%	73.33%	
Accuracy: 77.40%			

From a total of 111 negative sentiments, 89 sentiments showed true negative results, meaning that 89 negative sentiments were suitable with the prior prediction, namely having negative values, only 22 sentiments turned out to be positive. Furthermore, from 75 tweets that had positive sentiment, 55 tweets showed true positive values, meaning that the tweets were in accordance with predictions, and there were 20 tweets that were initially predicted to have positive sentiments but turned out to be true negative results. So, we can calculate the accuracy rate using the following formula :

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}\ x\ 100\%$$

$$Accuracy = \frac{55 + 89}{55 + 20 + 89 + 22}\ x\ 100\%$$

$$Accuracy = \frac{144}{186}\ \text{x}\ 100\%$$

$$Accuracy = 0.774\ x\ 100\%$$
$$Accuracy = 77.4\%$$

With an accuracy rate of 77.4%, it means that the ability of this model to predict positive and negative sentiments is 77.4%. This value is obtained by using a total of 236 data with the number of positive sentiment as many as 75 tweets and negative sentiment as many as 111 tweets.

Discussion and Conclusions

Higher negative sentiment from the public comes from those who care about environmental issues. By removing coal ash waste from the category of toxic and hazardous materials, it is considered a backward step in efforts to reduce environmental damage in Indonesia. In addition, it is feared that the impact is also very dangerous for public health. There has been a lot of news that reveals the emergence of public health problems around coal waste landfills, such as the appearance of skin diseases and respiratory disorders (news.detik.com, 10 Oct 2018). In addition, this negative sentiment is also related to the alleged economic motive behind the issuance of coal ash waste (FABA) from the waste list. People think that the government is more concerned with economic factors than the health of its citizens.

On the other hand, those who support the removal of FABA from the list of toxic wastes refer to other countries that have no longer included FABA in the hazardous waste category, such as Japan, Europe, Australia and the United States. In addition, from the results of laboratory tests, FABA from PLTU does not contain elements that harm the environment, so there will be an opportunity to use it as construction raw materials such as cornblocks, building materials to cement.

The large number of negative opinions that develop in the community shows the government's lack of effort to disseminate information about this new regulation. The public needs to be informed that there are two types of coal ash waste, namely those from industrial furnaces and those from PLTU. Ash waste from industrial boiler is still categorized as hazardous waste, while waste from PLTU does not contain hazardous materials so that it can be used for infrastructure material.

In the future, the government should invite environmental activists, researchers and academics in the environmental field as well as other parties affected by these regulations. When the government has listened to the aspirations of various parties before issuing a regulation, it is hoped that the community will be able to support the regulation.

References

Elbagir, S., &Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and Vader sentiment. *Lecture Notes in Engineering and Computer Science*, *2239*, 12–16

F. R. Lucini et al. (2017). Text mining approach to predict hospital admissions using early medical records from the emergency department. *Int. J. Med. Inf*, *100*(April), 1–8

Hasan, M. R., Maliha, M., &Arifuzzaman, M. (2019). Sentiment Analysis with NLP on Twitter Data. *5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2019*, (December).

https://doi.org/10.1109/IC4ME247184.2019.9036670

Prastyo, P. H., Sumi, A. S., Dian, A. W., &Permanasari, A. E. (2020). Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel. *Journal of Information Systems Engineering and Business Intelligence*, *6*(2), 112. https://doi.org/10.20473/jisebi.6.2.112-122

Saura, J. R., Herraez, B. R., &Reyes-Menendez, A. (2019). Comparing a traditional approach for financial brand communication analysis with a big data analytics technique. *IEEE Access*, *7*, 37100–37108. https://doi.org/10.1109/ACCESS.2019.2905301

Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Seuk Wai, P., Wai Chung, Y., …Al-Garadi, M. A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, *6*, 37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311