E3S Web of Conferences 317, 05013 (2021) https://doi.org/10.1051/e3sconf/202131705013
ICENIS 2021

Prediction of The Level of Public Trust in Government Policies in the 1st Quarter of The Covid 19 Pandemic using Sentiment Analysis

Zahra Nur Aziza[1*], *Daniel Yeri Kristiyanto*[2]

[1]Department of Information Systems, Institut Teknologi Telkom Purwokerto, Central Java, Indonesia
[2]Departement of Information Systems, Institut Teknologi Telkom Purwokerto, Central Java, Indonesia

Abstract. The covid-19 pandemic has made changes in society, including Government policy. The policy changes led to mixing responses from the public, namely netizens. Netizen shares their opinion in social media, including Twitter. Their opinion can represent the public's trust in the Government. Sentiment analysis analyses others' opinions and categorises them into positive opinions, negative opinions, or neutral opinions. Sentiment analysis can analyze large numbers of opinions so that public opinion can be analyzed quickly. This paper explains how to analyze public trust using sentiment analysis and to use Naïve Bayes classification method to analyze sentiment. The data research was taken from Twitter in the first quarter of the Covid-19 pandemic, with around 3000 tweets. The tweets were related to Covid-19 and the Government from several countries such as the United States, Australia, Ireland, Switzerland, Italy, Philippines, Sri Lanka, Canada, Netherlands, United Kingdom, Germany, and Lebanon. This study aims to determine the level of public trust in the Government in the first quarter of the Covid-19 pandemic. The research result is expected to be used as a reference for the public policy stakeholders to determine future policies.

1 Introduction

Covid-19 has become the most popular topic, often discussed worldwide, since the World Health Organization (WHO) set Covid-19 into a global pandemic. Covid-19 has made several changes in social life, including policies set by the government. Many countries have set lockdowns or prohibited their citizens from leaving their homes to prevent the spread of the coronavirus. The policies that the government has set have caused various reactions from the public. Many people respond to government policies through social media such as Twitter.

The use of social media is currently proliferating. When the research was conducted, social media users reached 3.8 million users [1]. Many people use social media to share their opinions and thoughts. One of the social media platforms that people widely use to share

* Corresponding author: 1913109@ittelkom-pwt.ac.id.

their opinions today is Twitter. Through Twitter, many people disseminate various information. Based on these data, it can be concluded that many Twitter users are voicing their opinions regarding current issues, including information about the Covid-19 pandemic. Opinions in the form of tweets can represent public confidence in the government.

Public trust is an essential variable in building good governance. Public trust generates legitimacy that can create social assets for the government, used to gain political support and government activities [2]. In this case, sentiment analysis needs to be done. Sentiment analysis can automatically categorize opinions with pre-existing data and learning data. Sentiment analysis can categorize responses from the public into positive, negative, and neutral sentiments [2]. The better way to analyze public trust in the government uses sentiment analysis because this method can classify opinions in large numbers to shorten the time and energy required. However, an analysis of public trust in government policies during a pandemic is fundamental. Public trust analysis is needed because public trust is the most crucial element of the general administration's legitimacy [3].

Naïve Bayes is used because this algorithm has higher accuracy than other algorithms such as decision tree and random forest. This is approved by the research conducted by Fitri, Andreswari, and Hasibuan on the research: LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm [4]. Their research resulted in data accuracy with Naïve Bayes is 86.43% and with Decision Tree and Random Forest is 82.91%. Another reason for using Naïve Bayes in this study is that Naïve Bayes has a good performance in sentiment analysis. According to the research conducted by Mubarok et al. [5], it is stated that the result of their research resulted in Naïve Bayes being suitable in his research by producing the best F1-Measurement of sentiment analysis is 78.12%.

The purpose of this study was to analyze public trust in the government during the Covid-19 pandemic. The main objective of this study is to determine the degree of public trust in Government policies during a pandemic through sentiment analysis on Twitter.

2 Research Method

Sentiment analysis can be done by several methods, one of which is using Naïve Bayes. Naïve Bayes is a method in analysis that uses Bayesian probability. This method has several processes: data collecting, pre-processing data, and applying the Naïve Bayes classifier.

2.1 Data Collection

This study takes data from Kaggle.com [6] and recalculates using a different method. The original data was analysed using the Natural Language Processing (NLP) method, and this study uses the Naïve Bayes classification method. The original data amounted to 3000 tweets, but the data used were 100 tweets—the final number of tweets obtained by using a simple random sampling technique. The sample size uses the following formula:

$$n = \frac{N}{1+Ne^2} \quad (1)$$

Where:
n = a sample size
N = Population size
e = the desired margin of error

2.2 Pre-processing Data

This process is necessary because the data from Twitter still contains content other than sentiment such as emoticons, website links, hashtags, white space, etc. This process removes website links, special symbols, usernames (start with @), hashtags, and website, and also transforms emoticons into word equivalents. Pre-processing is carried out so that the sentiment analysis more accurate [7].

2.3 Naïve Bayes Classification

The primary methodology of research on sentiment analysis on tweets, especially using the Naïve Bayes Classifier method, is to classify tweets into positive, negative, or neutral tweets.

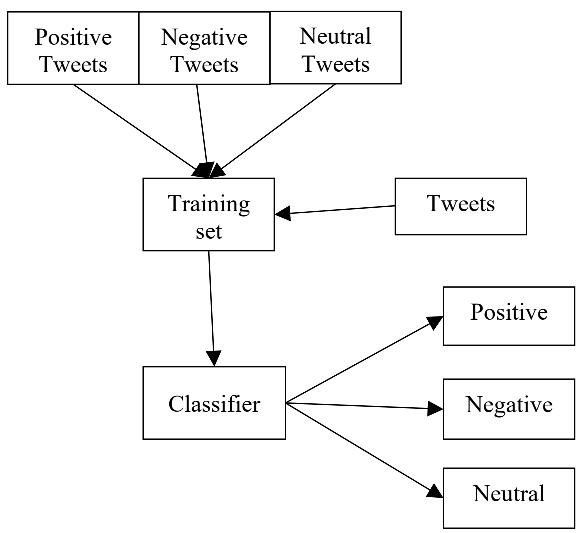

Fig. 1. Main Methodology of Naïve Bayes Classifier.

Naïve Bayes is a statistical analysis algorithm that processes data on numerical data using Bayesian probability. Naïve Bayes Classification classifies texts based on keywords probability in comparing training documents (past knowledge) and test documents (observed data) [7]. This study uses the Naïve Bayes classification because it is an algorithm that fits the purpose of this study. Naïve Bayes classification assumes that a particular feature in a class is not related to the existence of other rule [8]. Here is the formula:

$$P(c|Z) = \frac{P(c)P(Z|c)}{P(Z)} \quad (2)$$

Where,
P(c) = class prior probability
P(Z) = predictor prior probability
P(c|Z) = posterior probability
P(Z|c) = predictor prior probability

2.4 Evaluation

The last stage after the data is classified using the Naïve Bayes algorithm is the evaluation stage. This stage aims to determine the accuracy of the algorithm that has been carried out. The result of the evaluation can be used to find conclusions. In the classification process, there is a False Statement [4]. This means that in the classification process, errors can occur in grouping sentences. For example, a sentence is classified into a positive sentence even though the sentence is a negative sentence or vice versa. This can be avoided by conducting a performance evaluation. Accuracy can be measured using this formula:

$$Accuracy = \frac{TP+T+TN}{TP+FN+FP+T+TN+FN} \quad (3)$$

Where,
TP = True positive
T = True negative
FN = False negative
FP = False positive
TN =True neutral
FN = False neutral

3 Result

3.1 Data Collection

The original data or the population size (N) = 3000 and e = 10 %, so the calculation becomes:

$$n = \frac{3000}{1 + 3000(0.1)^2} = 96.774 \approx 100$$

Table 1. The Several Raw Data used in This Study.

Username	Screen Name	Location	Tweet At	Original Tweet	Sentiment
3877	48829	Cornwall, England	16-03-2020	@TheJoshuaTurner @Loreign83 @peanut_astro @my_amigouk @afneil @BorisJohnson @patel4witham This is both disgusting and disgraceful charging over inflated prices for items for stopping the spread of COVID-19, the government really needs to do something about	Negative
3892	48844	Leicester, England	16-03-2020	Pretty sure within a week or two, supermarket supply chains will dry up as more countries are effected by Covid-19 (and possibly go into lockdown). If so, would the government introduce a form of rationing so that the people can eat? Somehow	Positive
3943	48895	United Kingdom	16-03-2020	When the government says to start social	Positive

				distancing, but you work retail so you can't just not talk to customers in the store lol fml I'm 100% going to catch covid-19	
4013	48965	Montreal, QC	16-03-2020	#FashionA's #Coronavirus Collapse: #Travel, #stockmarkets, #economic outlooks Â- they are all slipping as #governments and #businesses try to get a handle on #COVID-19. https://t.co/v#n6Jrz7mq #retail #apparel #socialdistancing #store #closures #revenues	Negative
4031	48983	Toronto, Ontario	16-03-2020	Between US air strikes and the #COVID_19 crisis, the Iraqi government may find itself dealing with its greatest governance challenge since the 2014 ISIS invasions. @MiddleEastEye https://t.co/DVthW3RTRj	Negative
4063	49015	Lahore, Pakistan	16-03-2020	Due to fear of #COVID2019 people have started hoarding items of daily use. This will impact the regular level of supply in markets, if these commodities will disappear from markets it will create shortage and prices will go up. Then people will start cursing government.	Negative

3.2 Pre-processing Data

This process removes elements other than sentiment such as website links, special symbols, usernames (start with @), hashtags, white space, and transforming emoticon into word equivalents. Here is the example:

Tweet before pre-processing data:

#Coronavirus is "an exposure of all the holes in the social safety net" says NELP Government Affairs Director Judy Conti

#UI #Unemployment #PaidLeaveForAll
https://t.co/BrCY9IJWSv

Tweet after pre-processing data:
is an exposure of all the holes in the social safety net, says NELP Government Affairs Director Judy Continu.

Table 2. The Example Tweets before Pre-processing.

User Name	Screen Name	Location	Tweet At	Original Tweet	Sentiment
30	44982	United States	07-03-2020	#Coronavirus is "an exposure of all the holes in the social safety net," says NELP Government Affairs Director Judy Conti #UI #Unemployement #PaidLeaveForAll https://t.co/BRCY9IJWSv	Positive
51	45003	Australia	09-03-2020	@7newsSydney @ScottMorrisonMP Some factors have brought on the recession 1) Ignoring per-capita recession Â', 2) wage growth stagnant 3) Consumer debt levels are through the roof 4) casualized work! These changes in the economy are driven by government not #co	Negative
56	45008	United States	09-03-2020	@realDonaldTrump Separate Table of US Corona Virus Statistics. Very high stats of corona virus. US Government's inability to control corona virus. High prices Disinfectant gel in USA High price of mask in USA. Food prices are raising day by day	Negative

Table 3. The Example Tweets After Preprocessing.

UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment
30	44982	United States	07-03-2020	is an exposureof all the holes in the social safety net, says NELP Government Affairs Director Judy Conti	Positive
51	45003	Australia	09-03-2020	Some factors have brought on the recession 1) Ignoring per-capita recession Â', 2) wage growth stagnant 3) Consumer debt levels are through the roof 4) casualized work! These changes in the economy are driven by government not	Negative
56	45008	United States	09-03-2020	Separate Table of US Corona Virus Statistics. Very high stats of corona virus. US Government's inability to control corona virus. High prices Disinfectant gel in USA High price of mask in USA. Food prices are raising day by day	Negative

3.3 Naïve Bayes Classification

This process is to classify the tweets with Bayesian Classifier. Before classifying tweets, group the training data into positive, negative, and neutral tweets.

After the data training is group, the next step is to calculate the probability of each word in data training based on the positive or negative or neutral tweet. The data test can be calculated with formula number 2. That step provided all the probability of each word in each category (positive, negative, or neutral). To classify each sentence in the test data, use the probability of each word based on the result provided by the before step. The result of each category needs to be compared to know the sentiment of the tweets; the highest result will decide the category of that sentence.

Table 4. Result of Naïve Bayes Classification.

Sentiment	True Positive	True Negative	True Neutral
Prediction Positive	18	26	3
Prediction Negative	23	26	3
Prediction Neutral	0	0	0

Based on the result, from the 100 data, obtained 18 tweets declared true positive, 26 tweets declared true negative, 3 tweets declared true neutral. From the result the accuracy of this research is 44.44%.

4 Conclusion

The Naïve Bayes successfully applies to the analysis of tweet sentiment during the first quarter pandemic Covid-19. The analysis result of this study shows that the public is not satisfied with the government's policies during the first season of the Covid-19 pandemic. Based on the analysis, it is hoped that the government can use the result of this study as a reference for making new policies to make better decisions. Besides that, it can also be used to help people choose the next government.

References

1. "Digital 2020: 3.8 billion people use social media - We Are Social." https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media (accessed May 30, 2021)
2. "Twitter and News: How people use Twitter to get news." https://www.americanpressinstitute.org/publications/reports/survey-research/how-people-use-twitter-news/ (accessed May 30, 2021)
3. F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Teknol. Inf. ITSmart*, vol. **2**, no. 2, p. 35, doi: 10.20961/its.v2i2.630 (2016)
4. V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "ScienceDirect ScienceDirect Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm ScienceDirect Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm," doi: 10.1016/j.procs.2019.11.181 (2019)
5. M. S. Mubarok, A. Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," *AIP Conf. Proc.*, vol. **1867**, doi: 10.1063/1.4994463 (2017)
6. "Coronavirus tweets NLP - Text Classification | Kaggle." https://www.kaggle.com/datatattle/covid-19-nlp-text-classification (accessed May 30, 2021)
7. H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," in *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016*, Apr. 2017, pp. 416–419, doi: 10.1109/ICATCCT.2016.7912034 (2017)
8. A. Goel, J. Gautam, and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," in *Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies, NGCT 2016*, Mar. 2017, pp. 257–261, doi: 10.1109/NGCT.2016.7877424 (2017)