Perception analysis of the Indonesian society on twitter social media on the increase in BPJS kesehatan contribution in the Covid 19 pandemic era

F C Permana\textsuperscript{1*}, Z M Wicaksono\textsuperscript{2}, C Kurniawan\textsuperscript{2}, A S Abdullah\textsuperscript{3} and B N Ruchjana\textsuperscript{4}

\textsuperscript{1}Study Program of Multimedia Education, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No.229 Bandung, 40154 Indonesia.
\textsuperscript{2}Student of Master of Management Program, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia.
\textsuperscript{3}Department of Computer Science, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia.
\textsuperscript{4}Department of Mathematics, Universitas Padjadjaran, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia.

*fahmi.candrap@upi.edu

Abstract. In the conditions of the Covid-19 Pandemic, the president of Indonesia, Ir Joko Widodo, on May 13, 2020, announced an increase in BPJS Kesehatan contribution. This announcement made everyone in the community boisterous, including Indonesians on social media twitter. Currently, Indonesia is ranked fifth in the world using social media twitter, so data mining on Twitter is a good opportunity to see the public’s response to a developing issue. This paper will discuss perceptions analysis of the Indonesian people on social media regarding the issue of increasing the contribution of BPJS Kesehatan. The issue of increasing the contribution of BPJS Kesehatan has been a topic of conversation on Twitter for a long time. After 30 days of data crawling, 145,359 tweets were obtained. This amount of data proves that the Indonesian people are very active in issuing opinions regarding the issue of increasing contribution BPJS Kesehatan. Various kinds of differences of opinion in each conversation are classified into three types of opinion using the Naive Bayes method. The three types of opinions are classified into positive opinions, negative opinions, and neutral opinions. The opinion classification results obtained were 73% containing negative opinions, 18% positive opinions, and 9% neutral opinions. This can serve as an early warning for the government to see the public’s response in every policy taken. So that each policy can be evaluated for the better.

1. Introduction
Indonesia officially announced the first case of Covid 19 on March 2, 2020 (https://nasional.kompas.com/read/2020/03/03/06314981/fakta-lengkap-kasus-pertama-virus-coronadi-indonesia?page=all), this marked the start of the Covid 19 Pandemic conditions announced by the government of the Republic of Indonesia. This condition has an impact on various aspects, one of which is the economic and health aspects (https://katadata.co.id/timdatajournalism/analisisdata/5f12658719721/simalakama-mitigasi-covid-19-kesehatan-atau-ekonomi). In the declining economic condition due to the effects of the COVID-19 pandemic, on May 13, 2020, through Presidential Decree Number 64 of 2020, Indonesian President Ir. Joko Widodo announced an increase in BPJS Kesehatan contributions.
The announcement of the increase in BPJS contributions has become a polemic for the Indonesian people in general, including netizens on social media in Indonesia (https://finet.detik.com/cyberlife/d-5012988/bpjs-kesehatan-naik-lagi-warga-twitterland-meradang). Even on social media twitter hashtags related to BPJS Kesehatan such as #Iuran BPJS, # BPJS Naik, #Iuran BPJS Naik, #BPJS Kesehatan immediately became a trending topic for several days after the announcement of the increase in BPJS Kesehatan contributions by the President on May 13, 2020 (https://economy.okezone.com/read/2020/05/13/320/2213632/iuran-bpjs-kesehatan-naik-jadi-trending-topic-begini-reaksi-warganet). The discussion on the topic of BPJS Kesehatan in social media Twitter which has become a trending topic worldwide is one of the implications of the number of Twitter users in Indonesia.

Currently, Indonesia is in the fifth rank as the country with the highest number of Twitter users in the world, with 19.5 million active users, as of 2020 (https://kominfo.go.id/content/detail/23666/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media). With this very large number, every issue that develops in social media can be an alternative for the government in seeing the public's response to every government policy. Peter Vermeij in his article in Journalism Practice (Vol. 6, 2012) entitled "Twitter Links Between Politicians and Journalists" states that there are two perspectives of social media. First as a disseminator of information and second as a relationship builder. Facebook is a social media with a second perspective, while Twitter is a social media with a first perspective (https://tirto.id/bagaimana-twitter-memengaruhi-opini-publik-dan-preferensi-politik-cGre). Meanwhile, John Parmelee, author of Politics and the Twitter Revolution (2011), told The Guardian that Twitter can set the agenda about what the journalistic world wants to do (https://tirto.id/bagaimana-twitter-memengaruhi-opini-publik-dan-preferensi-politik-cGre). One of the reasons why Twitter is considered a spreader of information is the fact that half of Twitter trending topics become CNN headlines. Because Twitter has the power to be the mouthpiece of conventional media news (https://tirto.id/bagaimana-twitter-memengaruhi-opini-publik-dan-preferensi-politik-cGre). Therefore, Twitter social media can be used as an alternative media to evaluate any policies that will be carried out by the government, including those related to the government's decision to increase BPJS Kesehatan contributions during the Covid 19 pandemic.

This paper discusses the analysis of public perceptions on social media twitter after the President of the Republic of Indonesia Ir. Joko Widodo announced an increase in BPJS Kesehatan contributions. By analyzing public perceptions using opinion classifications based on sentiment analysis and classifying some of the words contained in the keyword "BPJS Contribution" in social media twitter. This paper is written by taking several references related to opinion classification using the naïve Bayes method and sentiment analysis in social media twitter. One of them is a study entitled "Naive Bayes as an opinion classifier to evaluate student satisfaction based on student sentiment in Twitter Social Media" which discusses public reactions in social media regarding Ebola Virus Information [3].

2. Method
This research was conducted using a two-stage method. The first method is to use the Text Mining Method as the first step in data collection and cleaning, then after that the Social Network Analysis process is carried out to analyse each related word in the text that has been taken and visualize it in various visualization categories. The two processes we carry out can be illustrated in Figure 1.
2.1. **Text Mining**

The first stage is the Text Mining stage, which is a method that can be used to search, visualize, and evaluate various collections of knowledge from text documents. In this study, the text document analysed was text-based on social media. Information that is analysed in text mining usually involves the process of adjusting text input, then providing a pattern into more structured data then evaluating the output of the information source (https://katadata.co.id/timdatajournalism/analisisdata/5f12658719721/simalakama-mitigasi-covid-19-kesehatan-atau-ekonomi). The following are several stages in the Text Mining method, which can be seen in Figure 2.

![Figure 1. The process carried out in the method](https://katadata.co.id/timdatajournalism/analisisdata/5f12658719721/simalakama-mitigasi-covid-19-kesehatan-atau-ekonomi)

Based on Figure 2 Some of the steps taken are as follows:

2.1.1. **Data Acquisition**

The process that occurs in data acquisition is data collection, usually cleaning irrelevant data is usually carried out (https://katadata.co.id/timdatajournalism/analisisdata/5f12658719721/simalakama-mitigasi-covid-19-kesehatan-atau-ekonomi). In this stage, the data source is taken from social media twitter. To get access to this data retrieval, we must be registered as an account on the https://developer.twitter.com/ as a developer account on twitter to get an API (Application Programming Interface) Key. Then after that, the data crawling process is carried out using the Pyton programming language with the API Key that has been obtained as a token for access to Twitter's API periodically every 7 days, for 30 days from 1 - 31 May 2020. In this data acquisition process, 145,359 data tweets were taken.

2.1.2. **Text pre-processing**

The next process after data acquisition or data retrieval is the storage of text data that has been collected, this process is called the text pre-processing process. Before doing this process, we need to clean the data up by removing the noisy, irrelevant topic and duplicated tweet from the raw data with the search
term “iuranBPJS”. This cleaning process gives a significant decreasing amount of data tweets from 145,359 to 11,732 data tweets. This process needs to be provided since in most of the grab Twitter data there is so much noise and duplicated data [4].

After the data is cleaned, the next process is the real text pre-processing, also known as the per word process, the documents to be analysed will go through a word order process and delete words that are considered unimportant, and meaningless. These words are used as stop words. The stop word process is the process of eliminating conjunctions that have no significant meaning, for example in, to, from, etc. [5].

2.1.3. Modelling
The next process is data modelling, each data in the form of text will go through an evaluation and validation process, usually through a feedback loop process. This modelling process is usually better known as a machine learning process where each text will go through a classification and clustering process. Machine Learning itself is a collection of algorithms that take data input and study the characteristics of the data so that they can find patterns, limitations, patterns, and data processing features. Techniques commonly used in Machine Learning have two processes, the classification, and clustering process [5]. In this stage, each opinion and tweet data are divided into three categories, namely positive opinion, negative opinion, and neutral opinion using the Naïve Bayes method as an opinion classification method and the Waikato Environment for Knowledge Analysis (WEKA) as an application in classifying opinions. WEKA was developed by the University of Waikato. WEKA is an open-source machine learning software. WEKA has a lot of data pre-processing and modelling techniques to support several standard data mining tasks, in particular, such as pre-processing data, clustering, classification, regression, visualization, and feature selection [8].

2.2. Social Network Analysis (SNA)
After finishing with the text mining method, the next step is to carry out the Social Network Analysis method. This stage is carried out to analyses and visualize each word contained after the data has been properly collected and processed at the text mining stage. SNA is a mapping of topics, conversations, locations, and relationships between users. In SNA, there are network nodes or points that describe users, locations, and other information. The lines connecting each node describe the relationships between the nodes. SNA is very well done for analysing sentiment and social media conversation topics [8].

![Figure 3. Social Network Relation](7)

SNA can be illustrated to determine the centre point of a node seen from the number of connections. The blue node is the highest central point with 8 connections, while the lowest is the red dot with 4 connections and is located in the outermost network of the social network [9]. The social network analysis process uses the existing functions on the twitter platform, retweet, and mention functions. A retweet shows agreement while mentioning indicates a discussion of the issue [9]. In this stage, we use software to visualize the results of this SNA, including Tableu as the Analytic Visualization data and also using Gephi as the SNA Job Creator.
3. Result and Discussion

The result of this paper is an analysis of public perceptions based on the results of data visualization related to BPJS Kesehatan Contribution Increase in Twitter Social Media with the keyword taken “iuranBPJS”. This keyword was chosen because it is directly related to the public's response to the topic of increasing or adjusting BPJS Kesehatan Contribution on social media twitter. To see the results of this analysis, this paper will show some of the results of data analysis based on several aspects and characteristics of the user, the type of tweet, and also some words related to the keyword “iuranBPJS”.

3.1. Total Mention and data trend social media (Twitter Vs Google Trend)
The issue of increasing BPJS Kesehatan contributions began after the Constitutional Court decision on May 1, 2020. This decision requires the government to cancel Presidential Regulation 75 of 2019 in the article on adjusting the National Health Insurance premium for Participants who do not receive Wages, on the grounds that there is no adequate legal umbrella (https://money.kompas.com/read/2020/04/21/15113322/pembatalan-kenaikan-iuran-bpjs-kesehatan-mulai-berlaku-per-april-2020). But on May 13, 2020, the government issued Perpres 64 in 2020 concerning National Health Insurance, this Perpres was issued to improve the previous Presidential Decree, which in essence the government again announced an increase in BPJS Kesehatan contributions [5,6]. This has made the issue of increasing BPJS contributions to be discussed again, in this paper we see quite fluctuating trends in google trends based on information search sites, and in twitter based on social media[7]. Figure 4 shows a similar graphic, on the 13th May 2020 got the same peak increase.

![Timeline trend Twitter and Google](image)

**Figure 4. Timeline trend Twitter and Google**

Figure 4 also shows that there are so many tweets and searches related to BPJS fees on the google site, it can be seen that almost within 24 hours on May 13, 2020, Indonesian netizens are busy discussing and looking for information related to the increase in BPJS Kesehatan contributions on the google site. This shows that the majority of Indonesian netizens are looking for the truth and validation of BPJS Kesehatan dues not only through social media twitter but also check the validity of the news again on the google site. With the amount of data taken from the Twitter social media platform in the range 1-31 May 2020, there were 145,359 tweets. Meanwhile, the highest amount of data occurred on May 13, 2020, with 33,078 tweet data.
3.2. Conversation characteristics
Apart from the number of tweet data and trends in each day from 1 to 31 May 2020, we analyzed the location of the user in tweeting related to this “iuranBPJS”. The distribution of user locations related to this “iuranBPJS” can be seen in Figure 5.

![Figure 5. Location of the user with the tweet "iuranBPJS"](image)

In Figure 5, it can be seen that the location of the user with the tweet related to “iuranBPJS” does not only come from Indonesia, but there are also some tweets from several ASEAN countries, Australia, and even from the United States. Then in Figure 6, it can be seen that the majority of this tweet data comes from Twitter for Android application, which is 109,941 users and the second most tweet data is through the direct Twitter website as many as 17,488 users.

![Figure 6. Related words and Tweet data sources](image)

Figure 6 shows the words most contained in the keyword “iuranBPJS”, where the words “Jokowi” and “naik” (go up) are among the words most used by users, with a total frequency of 31,459 and 21,095 words. This shows that when BPJS dues do indeed “go up”, at that time the Indonesian president Jokowi became a topic of conversation among netizens. Apart from that, we also analyzed to see which accounts were mentioned the most about this “iuranBPJS” conversation. Figure 7 shows that the @BPJSKesehatan account as the official BPJSK Kesehatan account is one of the most mentioned. Besides, the @CNNIndonesia account has also been mentioned a lot, this can happen because CNN as
a news portal previously made an article related to an announcement from President Jokowi regarding the increase in BPJS Kesehatan contributions.

One of the characteristics of Twitter is the existence of trending topics using hashtags or so-called #hashtag. This is done so that each issue can be linked to a campaign or a moment that occurs at that time. In Figure 8, it can be shown that the Top Hashtag is dominated by various kinds of sentiments such as #BPJS recorded 3,159 tweets, #Hensat recorded around 819 tweets, this hashtag was pioneered by a politician and several hashtags indicated to be top hashtags with indications of linking BPJS fees towards 375 tweets of #MerakyatTapiBohong, 284 tweets of #PuraPuraBagia, 284 tweets of #RezimParanoid and many others that indirectly criticize all policies taken by the Indonesian government.

After analyzing the top hashtags, we also analyzed the age and number of followers of accounts that tweeted with the keyword “iuranBPJS”. This was done to see whether these tweets were natural or not, the majority of boot accounts were when they were created in 2020, and detected as boot/spam then at that time the account will be deleted by twitter.
Based on Figure 9, it can be seen that the dominance of users who join this conversation is 36,000 users who have less than 1 year of age, the range of users who have 0-100 followers is 34,648 users, the verified accounts are 1,623 users, and the users who have the most 0 followings with the number of 463 users. From these characteristics, it can be seen that this trend is quite rational, normal, and can represent the real Indonesian netizens. Due to a large number of followers and verified accounts and a long account age, this can prove that the majority of accounts with “iuranBPJS” tweet come from real accounts, or not from buzzers with fake accounts.

3.3. Opinion Analysis Using WEKA

From the data of 145,359 Tweets, we processed the 11,732 data tweet that has been cleaned to make 700 training data, and 300 testing data were taken that had been cleaned through preprocessing data. Each data is manually assigned a sentiment label positive with +1 score, neutral with 0 scores, and negative with -1 score based on a Linguist as expert judgment to give a label (positive, neutral, negative) every training data, and validation of testing data. The proportionality between training data and testing data has an ideal ratio with 70% data training and 30% data testing [10]. For example, if we have a sentence “bpjs naikin iuran 1 januari 2020 tp duitnya ga dibalikin yg januari maret” in subjectively we also have known that this sentence has a negative sentiment. Therefore, this sentence will have a negative label with -1 score from the Linguist. Then the training data that already has a label is tested with WEKA using the multilayer perceptron method and gives results as shown in Figure 10.
Figure 10. Opinion Classification Process manually and conversion in .arff files

WEKA's training data classification uses a multilayer perceptron. The training data is a file in .arff format which contains the values of the specified attributes according to the parameters used (number of positive words, number of neutral words, number of negative words, number of positive emoticons, number of neutral emoticons, number of negative emoticons). The classification result is the content of sentiment (positive, neutral, negative) on each data in the training data [11]. After the opinion classification process is done manually and a .arff file is formed, the results of this WEKA processing can be seen in Figure 11.

Figure 11. Classification output results using WEKA.

After the .arff file is formed, the next step is to carry out the training process and data testing using WEKA. Figure 1 shows the results of the training and data testing process using WEKA, where
Previously we have weighted the value of each opinion and the results have been stored in WEKA. After the results of weighing the value of each opinion in the testing data, the next step is to validate any data that has been issued by WEKA; this is done to test how high the accuracy and quality of weighting by WEKA is. After previously we carried out the process of training data on machine learning that was processed by WEKA. The data validation process can be seen in Figure 12.

![Image](image.png)

**Figure 12.** Validation results for each weighting opinion testing data from WEKA.

From the testing data that was processed using WEKA, we got a value of 92% of the results of testing using WEKA which can be seen in Figure 12. This shows that the results of the classification of opinions using the naïve Bayes method using WEKA are quite accurate. After validating the final process, we calculated the presentation of this testing data for each category which we previously labeled 1 for positive opinions, 0 for neutral opinions, and -1 for negative opinions. The results showed that 73% of the public gave negative opinions, 18% gave positive opinions, and 9% gave a neutral opinion on the issue of increasing BPJS Kesehatan contributions.

4. **Conclusion**

Based on the results of the analysis and data visualization carried out, the topic of increasing BPJS Kesehatan contributions has subsided and received a positive response from the public on social media, namely when the Constitutional Court cancelled the increase in BPJS Kesehatan premiums on May 1, 2020. Social media was being discussed again when the Government through Perpres 64 in 2020 announced an increase in BPJS Kesehatan Contributions on May 13, 2020. After 30 days of data crawling, 145,359 tweets were obtained. This amount of data proves that the Indonesian people are very active in issuing opinions regarding the issue of increasing BPJS contributions. Various kinds of differences of opinion in each conversation are classified into three types of opinion using Text Mining and Naive Bayes as classifier method. The three types of opinions are classified into positive opinions, negative opinions, and neutral opinions. The opinion classification results obtained were 73% containing negative opinions, 18% positive opinions, and 9% neutral opinions. This can serve as an early warning for the government to see the public’s response in every policy taken. So that each policy can be evaluated for the better.

**Acknowledgments**

The authors thank to Rector Universitas Padjadjaran who provided financial assistance for the dissemination of research through Academic Leadership Grant with contract number: 1427/UN6.3.1/LT/2020. Also thanks to the reviewers for the valuable review of this paper.
References

[1] Permana F C, Rosmansyah Y, Abdullah A S 2017 Naive bayes as opinion classifier to evaluate students satisfaction based on student sentiment in twitter sosial media Journal of Physics: Conference Series 893 012051.

[2] Vermeij P 2012 Twitter Links Between Politicians and Journalists Journalism Practice 6.

[3] Tran T, Lee K 2016 Understanding Citizen Reactions and Ebola-Related Information Propagation on Social Media IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[4] Bouazizi M, Otsuki T 2016 Pattern-Based Approach for Sarcasm Detection on Twitter IEEE Access 4 106.

[5] Purbo O W 2017 Text Mining : Analisa MedSos, Kekuatan Brand & Intelejen di Internet (Yogyakarta: Penerbit Andi).

[6] Rasyada I, Setiowati Y, Barakbah A, and Al Islami M T F 2020 Sentiment Analysis of BPJS Kesehatan’s Services Based on Affective Models 2020 International Electronics Symposium (IES) pp. 549-556 doi: 10.1109/IES50839.2020.9231940.

[7] Laagu, Asnoer M, Arifin A S 2020 Analysis the Issue of Increasing National Health Insurance (BPJS Kesehatan) Rates through Community Perspectives on Social Media: A Case Study of drone Emprit International Conference on Smart Technology and Applications (ICoSTA). IEEE.

[8] Bouckaer R R, Frank E, Hall M, Kirkby R, Reutemann P, Seewald A, et al 2015 WEKA Manual (New Zealand: University of Waikato).

[9] Andrianti S D 2018 Sentiment Analysis On Twitter Account Using Naive Bayes Classifier Algorithm Case Study: Indonesia Healthcare and Social Security Agency (BPJS Kesehatan) Journal of Information System, Applied, Management, Accounting, and Research 2 (2) ISSN: 2598-8719.

[10] James G, et al. 2013 An Introduction to Statistical Learning: with Applications in R (New York: Springer-Verlag).

[11] Habibi R R, Setyo Hardi D B, Ernawati 2016 Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation”, Tesis Magister Teknologi Informasi, Universitas Atmajaya Yogyakarta.