

PAPER • OPEN ACCESS

Naives Bayes Algorithm for Twitter Sentiment Analysis

To cite this article: Samsir *et al* 2021 *J. Phys.: Conf. Ser.* **1933** 012019

View the [article online](#) for updates and enhancements.

You may also like

- [Merit System in The Placement of Civil Servants and its Effect toward Performance of Sub-district Office in Labuhanbatu Region Nort Sumatera](#)
Z M E Siregar and J Lubis
- [Fish biodiversity in false gharial habitat \(*Tomistoma schlegelii* Müller, 1838\) in Labuhan Batu district](#)
I Hariyadi, R Machrizal, R H Dimenta et al.
- [Growth Response of Labuhanbatu Local Rice With Gamma Ray Irradiation in M1 Generation](#)
S H Y Saragih, K D Sitanggang, W Lestari et al.



245th ECS Meeting
San Francisco, CA
May 26–30, 2024

PRiME 2024
Honolulu, Hawaii
October 6–11, 2024

Bringing together industry, researchers, and government across 50 symposia in electrochemistry and solid state science and technology

Learn more about ECS Meetings at
<http://www.electrochem.org/upcoming-meetings>



Save the Dates for future ECS Meetings!

Naives Bayes Algorithm for Twitter Sentiment Analysis

**Samsir¹, Deci Irmayani⁴, Firman Edi⁵, Junaidi Mustapa Harahap², Jupriaman⁶
Rizki Kurniawan Rangkuti⁷, Basyarul Ulya², Ronal Watrianthos^{1*}**

¹Departement of Informatics Engineering, Universitas Al Washliyah Labuhanbatu, Indonesia

²Department of Management, Universitas Al Washliyah Labuhanbatu, Indonesia

⁴Department of Informatics Management, Universitas Labuhanbatu, Indonesia

⁵Department of Engineering Management, Institut Teknologi Batam, Indonesia

⁶Department of English Education, Universitas Al Washliyah Labuhanbatu, Indonesia

⁷Department of Mathematics Education, Universitas Al Washliyah Labuhanbatu, Indonesia

*ronal.watrianthos@gmail.com

Abstract. On 2 March 2020, the Indonesian government, through President Joko 'Jokowi' Widodo, announced the first two cases of COVID-19 in Indonesia. This is the first case of COVID-19 officially confirmed in that country. Several cases have continued to increase since then. President Jokowi began issuing policies on the spread of this virus. This is different from other countries, such as Malaysia and Singapore, which responded from the previous month when the Indonesian government still stated that coronavirus does not exist in Indonesia. Our case study is to find a public opinion through social network analysis of Indonesian public policy during the beginning of the Indonesian COVID-19 pandemic in March 2020. This research implements text mining and document-based sentiments on Twitter data that is reprocessed through machine learning techniques using the Naïve Bayes method. We found negative opinions in the period more dominant by 46%, while that was 35% positive sentiment and 20% neutral. This research shows that anticipation, sadness, and anger are very dominant in the emotional analysis.

1. Introduction

The World Health Organization (WHO) has declared the coronavirus (COVID-19) a global pandemic on March 11, 2020[1]. This has a significant impact on the global economy in the short term[2]. This pandemic began in the city of Wuhan, Hubei province, China, at the end of December 2019, and spread rapidly in various countries. In the second week of April 2020, the virus has caused 1.5 million people infected with 88 thousand deaths in 209 countries. Although the vaccine has not been found, in general, COVID-19 is a disease that can heal itself but can also be deadly with a mortality rate of 5%[3][4].

Since the outbreak of the pandemic, several countries, including Indonesia, have issued policy plans to deal with it. Following the announcement by President Joko 'Jokowi' Widodo of the first case of COVID-19 on 2 March 2020[5], the Indonesian government began issuing policies on the spread of this virus. This is considered too late because several other countries, such as Malaysia and Singapore, have responded since the previous month[6][7]. In February, the Indonesian government even declared that there was no coronavirus in Indonesia[8]. The Indonesian government also considers that the opening of positive patient data for COVID-19 is not transparent. A survey conducted by the Indonesian Institute of Sciences (LIPI) in collaboration with the National Disaster Management Agency (BNPB) to find out



public perceptions about information transparency concerning COVID-19 showed 97 percent of 15,101 respondents agreed that the recent travel history of positive patients must be opened transparently by the government[9].

This study analyzes the opinion of the Indonesian people through a social network analysis of the government policy towards COVID-19 during the early pandemic period of March 2020. We used sentiment analysis by crawling Twitter with the keywords Jokowi and Coronavirus. The keyword 'jokowi' is the representation of President Jokowi as the government and the coronavirus is the alias name of COVID-19. This process is done by drawing up Twitter opinions on an event or issue so that positive or negative opinions can be concluded[10]. The classification process of the Naïve Bayes method uses simple statistical probability values that assume strong independence from Bayesian rules[11].

2. Methodology

2.1 Social Network Analysis

Social Network Analysis (SNA) examines user interactions and relationships. The main concept is to study human relationships, such as friendship and family relationships, or the same interests. This relationship is then visualized in such a way that it is easily analyzed[12][13][14]. Social networks have an important role to play in the development of social computing. Some interesting topics, such as semantic analysis, content mining, or sentiment analysis, are widely used[12].

2.2 Sentiment Analysis

Sentiment analysis is the process of determining sentiments and grouping the polarity of texts in documents or sentences so that they can be categorized as positive, negative, or neutral sentiments [15][16]. Currently, sentiment analysis is widely used by researchers as one branch of research in computer science along with the explosion of information on the internet. Social media like Twitter are usually used in sentiment analysis to determine people's perceptions[17]. Sentiment analysis can also be likened to opinion mining because it focuses on opinions that state positive or negative [18].

2.3 Text Mining

Text mining aims to produce information from the collection of documents. Text mining is a solution for the retrieval of information by processing, grouping and analyzing large amounts of unstructured data[19]. The process of retrieving information in the mining text may result in a sentiment analysis that identifies the emotional nature of a statement, whether positive or negative. Text mining is currently widely used in a variety of research areas, such as software development, online media, marketing, academics, and politics. Like data mining, text pre-processing is necessary for the analysis of text mining through several stages, namely: case folding, tokenizing, filtering, stemming, tagging, and analysis[20].

2.4 Naïve Bayes

The Naïve Bayes method is a classification method for text mining sentiment analysis. This method can be classified and has the capability of data accuracy and computation[21]. Naïve Bayes has recently been widely used in classification techniques, particularly in social media networks such as Twitter, using several methods, including Unigram Naïve Bayes, Multinomial Naïve Bayes and Maximum Entropy Classification[22]. Calculation of the probability categories in Naïve Bayes uses the Bayes algorithm approach using the following equations:

$$P(c|z) = \frac{p(z|c) \times p(c)}{p(z)} \quad (1)$$

Equation (1) shows that c is the class label, z is the attribute applied, while $P(c)$ and $P(z | c)$ are the previous probabilities of the class. In the Naïve Bayes classification, the data set that is processed will

be annotated in three classes, positive, negative, and neutral[23]. The main feature of the Naïve Bayes classification is getting strong assumptions from each condition or event [24].

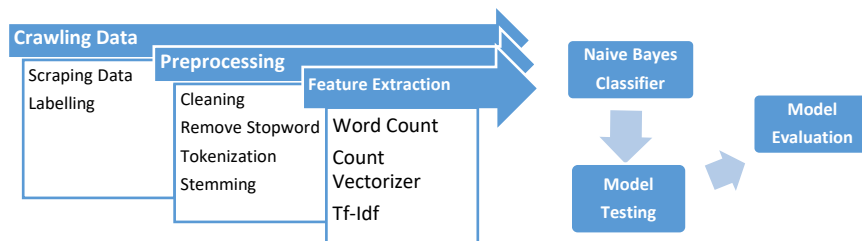


Figure 1. Research Method

Figure 1 above shows the research method used in this study. Naïve Bayes is used in opinion-mining as a method. Crawling data uses Twitter data by providing keywords within a given timeframe. The data collected are labeled to determine feelings and are transformed into structured data. The cleaning process is carried out at this stage to reduce noise. The last step is to remove the word 'stop' and delete meaningless words like me and, or[25]. The next process is the identification of words in the text by breaking up sentences into terms based on spaces and punctuation. This process is called tokenization, which ends with the process of stemming changing the affix words into basic words[19][26].

The process of creating a feature is carried out in the extraction feature to facilitate the classification of Naïve Bayes. The final stage is the stage at which the classification process begins based on the sentiments expressed in the document. This stage produces a model and is used to show the accuracy of the results of the classification. The level of accuracy at this stage is measured by the value of the performance of the classification by performing model tests and model assessments.

3. Result and Discussion

Research data was drawn from Twitter using the keyword 'Jokowi' at the beginning of the period of the COVID-19 pandemic in Indonesia. We used the public streaming of the Twitter API Drone Emprit to search[27] and collected data from 1 March to 31 March 2020. Crawler data successfully collected 306,284 tweets from the sample in Table 1.

Table.1 Crawling Data

<i>@PermanaRikie</i>	Selamat pagi. Perkenalkan saya pasien 10 di Cirebon. Dinyatakan positif Covid-19 14 Maret lalu. Semoga surat ini bisa sampai ke Pak Presiden @jokowi & Pak Menkes @KemenkesRI https://t.co/fp8Gt2RmvA
<i>@zenrs</i>	Pak @jokowi, pejabatmu dusta soal pasien Cianjur? Anak istri pasien tsb kini positif. Brp lg yg terkena [or hrs ditracking] krn pejabat bilang alm negatif? Stop menyimpan [apalagi jk berdusta] soal data/fakta. Kalian berbahaya betul main2 dg nasib kami https://t.co/cKPotOH6aZ
<i>@sandalista1789</i>	Penulis twit yang tak direspons admin akun @jokowi ini telah meninggal dunia sebagai pasien PDP corona R.I.P.
<i>@anandabadudu</i>	Pak Menkes Terawan dan pemerintahan @jokowi baiknya menunjuk seseorang/tim yang kompeten di dunia medis untuk menjadi juru bicara publik dlm menghadapi krisis #CoronaVirusIndonesia. Sudah statement2 denial, perbanyak info mitigasi
<i>@jansen_jsp</i>	Sejak awal kami sudah menunda politik pak @jokowi. Tapi anda malah menghadapi wabah corona ini dgn pendekatan politik. Darurat sipil ini kebijakan "cuci tangan" dan power oriented. Anda ingin mengendalikan publik tapi tidak mau menanggung hidup mereka.

Preprocessing steps were carried out to remove noise, clarify features, tokenize, and convert data[28]. Data cleaning needs to be done at this stage to determine the class and to delete unneeded attributes such as URL, mention, username, RT, hashtag, and punctuation, as shown in Table 2.

Table.2 Preprocessing Data

Positive	saya pasien 10 di Cirebon dinyatakan positif Covid-19 14 Maret lalu. Semoga surat ini bisa sampai ke Pak Presiden
Negative	Pak Jokowi pejabatmu dusta soal pasien Cianjur Anak istri pasien tsb kini positif. Brp lg yg terkena Stop menyimpan data/fakta.
Negative	Penulis twit yang tak direspons admin akun Jokowi ini telah meninggal dunia sebagai pasien PDP
Positive	Pak Menkes Terawan dan pemerintahan Jokowi baiknya menunjuk seseorang/tim yang kompeten di dunia medis untuk menjadi juru
Negative	Sejak awal kami sudah menunda politik Pak Jokowi. Tapi anda malah menghadapi wabah corona ini dgn pendekatan politik

The next stage calculates the frequency of words that often appears in both categories to produce positive or negative and visualize them using word cloud as shown in Figure 2 and Figure 3.



Figure.2 Wordcloud Positive

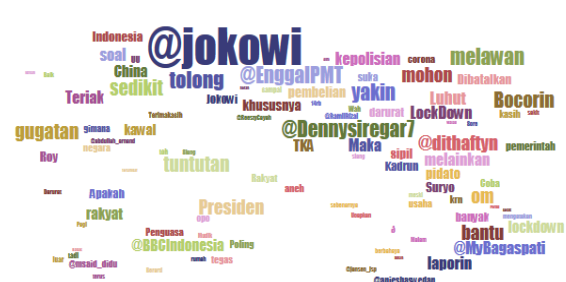


Figure.3 Wordcloud Negative

The results of the word cloud were analyzed for emotions using Plutchik's Wheel of Emotions model. This model divides emotions into eight categories: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation[29]. Figure 4 shows the dominant emotions in aggregate in March 2020 are anticipation, sadness, and anger. The most dominant emotion is anticipation, indicating the public hopes for readiness and anticipation in the future[30].

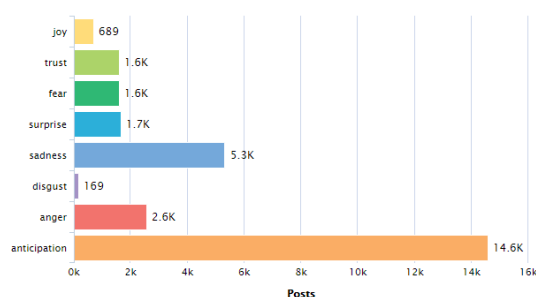


Figure 4. Emotion Analysis

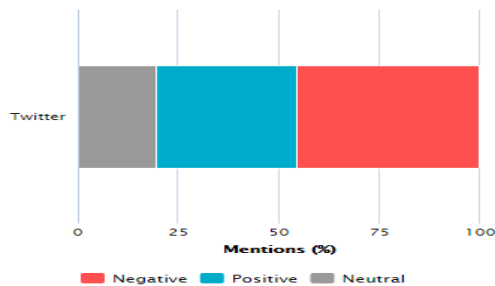


Figure 5. Analysis Sentiment

Sentiment analysis generates 35% positive, 46% negative, and 20% neutral as shown in Figure 5. Negative sentiment is stronger because it produces more negative mentions than positive mentions.

4. Conclusion

This study applies the Naïve Bayes algorithm for sentiment classification. This sentiment model can detect the value of polarity and emotion. The research data was drawn from Twitter using the keyword 'Jokowi' at the beginning of the COVID-19 pandemic period in Indonesia in March 2020. This keyword considered a representation of President Jokowi as the government. Research in that period showed 35% positive sentiment, 46% negative sentiment, and 20% neutral. The negative opinion generated allegedly

due to public dissatisfaction with President Jokowi's policies in the early days of the start of the COVID-19 pandemic in Indonesia.

In that period, we found dominant emotions in anticipation, sadness, and anger in the emotional analysis. The most dominant emotion is anticipation, which indicates public hopes for readiness and anticipation in the future. Further research is expected to use different algorithms to obtain more accurate results in the analysis of opinions.

References

- [1] J. Ducharme, "World Health Organization Declares COVID-19 a 'Pandemic.' Here's What That Means," *Time Magazine*, 2020. .
- [2] W. McKibben and R. Fernando, "Centre for Applied Macroeconomic Analysis The Global Macroeconomic Impacts of COVID-19: Seven Scenarios," 2020.
- [3] Worldometers, "COVID-19 Coronavirus Pandemic," *worldometers.info*, 2020. .
- [4] R. Verity *et al.*, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *Lancet Infect. Dis.*, Mar. 2020.
- [5] M. I. Gorbiano, "BREAKING: Jokowi announces Indonesia's first two confirmed COVID-19 cases," *The Jakarta Post*, 2020. .
- [6] NewsDesk, "Malaysia reports 19th COVID-19 case," <http://outbreaknewstoday.com>, 2020. .
- [7] N. S. Times, "Singapore Covid-19 cases surge to 72," www.nst.com.my, 2020. .
- [8] R. C. Paddock and D. M. Sijabat, "Indonesia Has No Reported Coronavirus Cases. Is That the Whole Picture?," *the New York Times*, 2020. .
- [9] J. Post, "Majority of Indonesians want transparency about COVID-19 patients," *thejakartapost.com*, 2020. .
- [10] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, Sep. 2019.
- [11] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, Sep. 2019.
- [12] A. Alamsyah, F. Putri, and O. O. Sharif, "Social Network Modeling Approach for Brand Awareness," in *2nd International Conference on Information and Communication Technology (ICoICT)*, 2014, pp. 448–453.
- [13] I. Yuliana, P. I. Santosa, and N. A. Setiawan, "Analisis Jejaring Media Sosial untuk Pemetaan pada Komunitas Online," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 2015, pp. 14–17.
- [14] R. H. S. Siburian, R. Karolina, P. T. Nguyen, E. L. Lydia, and K. Shankar, "Leaf disease classification using advanced SVM algorithm," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special Issue, pp. 712–718, Aug. 2019.
- [15] D. A. Ramadhan and E. B. Setiawan, "ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE," in *e-Proceeding of Engineering*, 2019, pp. 9376–9743.
- [16] A. M. Zuhdi, E. Utami, and S. Raharjo, "ANALISIS SENTIMENT TWITTER TERHADAP CAPRES INDONESIA 2019 DENGAN METODE K-NN," *J. Inf. Politek. Indonusa Surakarta*, vol. 5, no. 2, p. 7, 2019.
- [17] R. Ferryawan, Kusriani, and F. W. Wibowo, "ANALISIS SENTIMEN WISATA JAWA TENGAH MENGGUNAKAN NAÏVE BAYES," *J. Inf. Politek. Indonusa Surakarta*, vol. 5, no. 3, pp. 55–60, 2019.
- [18] C. B. Saputra, A. Muzakir, and D. Udariansyah, "ANALISIS SENTIMEN MASYARAKAT TERHADAP #2019GANTIPRESIDEN BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," in *Bina Darma Conference on Computer Science*, 2019, pp. 403–413.
- [19] A. Sholihin, Haviluddin, N. Puspitasari, M. Wati, and Islamiyah, "Analisis Penyakit Difteri

- Berbasis Twitter Menggunakan Algoritma Naïve Bayes,” *SAKTI – Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 1, no. 1, pp. 7–15, 2019.
- [20] A. Nugraha, Y. H. Chrisnanto, and R. Yuniarti, “Prediksi Sentimen Pada Sosial Media Twitter Mengenai Produk Smartphone Menggunakan Algoritma K-NN Classification,” in *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 2019, pp. 251–258.
- [21] J. M and V. H, “Opinion Mining For Sentiment Data Classification,” *Int. J. Res. Inf. Technol.*, vol. 3, no. 1, pp. 1–13, 2014.
- [22] R. P and M. M, “Sentiment Analysis of User Generated Twitter Updates using Various Classification,” 2009.
- [23] A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter Sentiment Analysis: A Case Study for Apparel Brands,” in *IOP Conference Series: Journal of Physic 1176 022015*, 2018.
- [24] N. Rochmawati and S. C. Wibawa, “Opinion Analysis on Rohingya using Twitter Data,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018.
- [25] R. Watrianthos, S. Suryadi, D. Irmayani, M. Nasution, and E. F. S. Simanjorang, “Sentiment Analysis Of Traveloka App Using Naïve Bayes Classifier Method,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 07, pp. 786–788, 2019.
- [26] D. Setian and I. Seprina, “ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET LAZADA MENGGUNAKAN TEXT MINING DAN ALGORITMA NAIVE BAYES CLASSIFIER,” in *Bina Darma Conference on Computer Science*, 2020, pp. 998–1004.
- [27] I. Fahmi, “Drone Emprit: Konsep dan Teknologi,” *IT Camp on Big data and Data Mining*. Jakarta, 2017.
- [28] A. D. P. Wardhani, I. Slamet, and Sugiyanto, “ANALISIS SENTIMEN PENGGUNAAN KOTAK SUARA KARDUS MENGGUNAKAN SVM LINEARLY SEPARABLE DATA,” in *SEMINAR NASIONAL PENELITIAN PENDIDIKAN MATEMATIKA (SNP2M) 2019*, 2019, pp. 167–174.
- [29] D. Chafale and A. Pimpalkar, “Review on Developing Corpora for Sentiment Analysis Using Plutchik’s Wheel of Emotions with Fuzzy Logic,” *Int. J. Comput. Sci. Eng.*, vol. 2, no. 10, pp. 14–18, 2014.
- [30] I. Fahmi, “Analisis Persepsi dan Emosi Publik Terhadap Penanganan #COVID19 Oleh Pemerintah (7-13 Maret 2020),” 2020.