Accreditec Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 1 (2023) 138 - 145 ISSN Media Elektronik: 2580-0760

Sentiment Analysis Against Political Figure's Billboard During Pandemic Using Naïve Bayes Algorithm

Ade Bastian[1], Ardi Mardiana[2], Dinda Sri Wulansari[3]
[1,2,3]Informatic, Faculty of Engineering, Universitas Majalengka
[1]adebastian@unma.ac.id, [2]aim@unma.ac.id, [3]dindasriwulansari29@gmail.com

Abstract

In the midst of the Covid-19 Pandemic, many Indonesians have reacted negatively to the placement of political individuals' billboards with very huge sizes on the streets. The early political campaign that was run was thought to be contentious. On social media like Twitter, the majority of people freely share their thoughts. The purpose of this study is to investigate how the general public reacted to the placement of billboards advertising political figures during the epidemic and to categorize those responses. It is envisaged that it would also provide advice for connected parties that may be used when making judgments regarding the policy of constructing billboards for political figures during a pandemic based on the results of data analysis. Twitter users tend to be more expressive because of the character limits, which means they have sentimental or emotional values. Using the Nave Bayes Algorithm, it is possible to do sentiment analysis on the sentiment data by categorizing user comments into positive, negative, and neutral attitudes. Regarding the sentiments expressed on billboards showing political leaders during the pandemic, tweets were sorted into three categories: liked, unfavorable, and neutral. The accuracy rate from Naive Bayes categorization of political personalities during the pandemic on social media Twitter was 83.3% with a precision value of 89%, recall 83%, and f-1 score of 82%.

Keywords: *covid-19 pandemic; twitter; sentiment analysis; naïve bayes clalssification algorithm*

1. Introduction

The internet and other technical advances have made knowledge easily accessible to Indonesians. Out of Indonesia's 266.911 million people, 196.7 million utilize the internet, and 170 million of those use social media regularly [1]. The internet and social media have a very close relationship and have combined to form one cohesive system. The internet, especially social media, is crucial for knowledge dissemination [2]. Social media can be used to access news and public opinion that is not broadcast on national television or in newspapers. Social media can be an effective communication tool in urgent situations, like the Covid-19 pandemic [3].

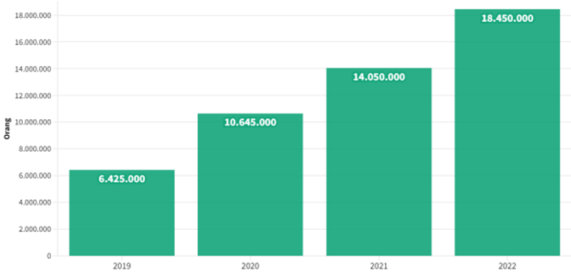
Spain is experiencing economic instability, loneliness, and dread as a result of the Covid-19 pandemic's uncertainty over the length of the quarantine [4]. When the closure was implemented during the first pandemic wave in Canada, it sparked criticism of the government [5]. The use of social media by candidates for office in 2022 is driven by Filipinos' reliance on it, but the local government has not been able to oversee and control its usage, particularly in relation to fake news about the general election in 2022 [6]. During the Covid-19 emergency, political officials in India have been communicating with the people on Twitter, including advising them to keep a physical distance [7]. Even the fluids that a person exhales when they cough or sneeze can be infected with the Corona virus, making it an extremely contagious illness. Because to the application of social distancing in Indonesia, the country has seen a decrease in the number of instances of Covid-19 [8]. Article 58 paragraph (1) of KPU Regulation 13/2020 stipulates that political parties or coalitions of political parties, candidate pairs, campaign teams, or other parties prioritize the campaign method of limited meetings, face-to-face meetings. This was prompted by the existence of the Covid-19 outbreak, which caused the KPU (Komisi Pemilihan Umum) to issue a new regulation regarding campaigning. face-to-face interaction as well as conversation through social media and other forms of online media [9]. Installing campaign billboards is a common kind of advertising that is utilized when campaigning for various political offices. The visual text package on political billboards takes the shape of party logos, candidate photographs, and color schemes, while the written text takes the form

Accepted: 03-11-2022 | Received in revised: 25-01-2023 | Published: 03-02-2023

of candidate slogans, serial numbers, names, and election districts [10]. Billboards have the unintended consequence of giving cities an unattractive appearance, which further contributes to the disarray of public spaces from an aesthetic standpoint [11].

Figure 1. Examples of Political Figures Billboard

One of the methods that can be utilized to familiarize voters with the presidential candidates is to place billboards in public areas, as seen in the illustration to the right (Figure 1). The majority of people believe that it is unethical for political officials to have billboards promoting themselves up during the COVID-19 outbreak. In addition to the length of the general election, which is expected to continue for some time, this does not offer any significant benefits because its sole purpose is to portray political people [12]. In 2021-2022, many political billboards will be installed in the context of the presidential election campaign for the 2024-2029 period [13]. The public believes that the construction of billboards with the slogan "Flap the Wings of Diversity" and an image of Puan Maharani Nakshatra Kusyala Devi, Chairman of the DPR RI (Dewan Perwakilan Rakyat Republik Indonesia), is a presidential candidate campaign for the general election in 2024 [14]. The advertisements featuring Puan Maharani were put up on billboards in a number of different provinces in Indonesia. In the midst of the Covid-19 pandemic, billboards of political figures that were installed on the streets with very large sizes received many responses from the public, and the campaigns that were carried out a long time ago were considered controversial, so they invited many people to give their opinions and opinions. Because of the epidemic caused by the Covid-19 virus, it is judged improper for billboards to appear in 2024 [15].



Fugure 2. Twitter User Data in Indonesia 2019-2022

According to the data presented in Figure 2 above, the number of Twitter users in Indonesia will continue to rise between the years 2019 and 2022. Twitter users in Indonesia are expected to reach 18.45 million in 2022 [16]. On social media, the vast majority of users regularly and openly share their thoughts, beliefs, and perspectives. Twitter is one of the social media platforms that is well-known to the general public as a place to devote oneself completely. Twitter is a social media platform that offers an Application Programming Interface (API), which may be utilized for the purpose of processing data collected from user tweets that contain hashtag references [17]. Twitter is a social network based on a web service that belongs to the category of microblogging and is limited to 280 characters. Twitter has four patterns of communication, including broker patterns, communication patterns, and tweet or twitter communication patterns in the sense of loading an article commonly referred to as status. patterning of communication that is good and interactive [18]. Because each tweet is limited to a certain number of characters, users are typically more expressive on Twitter than they are on other social networking platforms. In order to give the content an emotional or sentimental significance and to position it as a useful resource for carrying out analysis of sentiments. Using sentiment analysis, data from written text can be turned into information contained in a single opinion sentence [19]. Using sentiment analysis, many components of an entity, such as views, evaluations, attitudes, emotions, sentiments, and judgements, may be investigated and researched. Entities include individuals, organizations, events, and subjects, among other things [20]. The primary focus of both sentiment analysis and topic modeling is on the activity of dataset exploration [21]. The fields of marketing, healthcare, banking, and politics are just some of the areas that are frequently investigated for their potential applications of sentiment analysis of social media [22]. The Naïve Bayes classifier is an algorithm that is commonly used for classification techniques. This algorithm makes use of probability theory in order to determine the most appropriate possibility [23]. The Naïve Bayes algorithm is highly effective when used to datasets obtained from social media platforms for the purpose of giving sentiment categorization [24]. In spite of the fact that it contains a great number of distinct permutations, the Naïve Bayes model is consistently regarded as having a very high degree of accuracy [25]. Several earlier studies and sentiment analyses explored the general difficulties associated with the practice of installing billboards for political personalities. These difficulties are related to the study of political science. During the Covid-19 Pandemic in Indonesia, various events took place that were documented and caught the attention of the public. These events included the success of implementing limitations, social distance, and the administration of vaccinations, amongst many others.

During the COVID-19 epidemic, the goal of this study is to use the Naïve Bayes algorithm to evaluate and categorize the sentiment of public reactions to the construction of billboards for political personalities.

2. Research Methods

The data collection process was initiated by taking use of Twitter's Application Programming Interface (API) while using the.csv file extension. The unprocessed data goes through many steps of processing, including cleaning, case folding, tokenizing, removing stopwords, and stemming. After that, the word weight is created by utilizing the term frequency-inverse document frequency, and then Naive Bayes is used to determine the term frequency before moving on to the classification step. Figure 3 provides a visual representation of the study framework, which may be seen below.

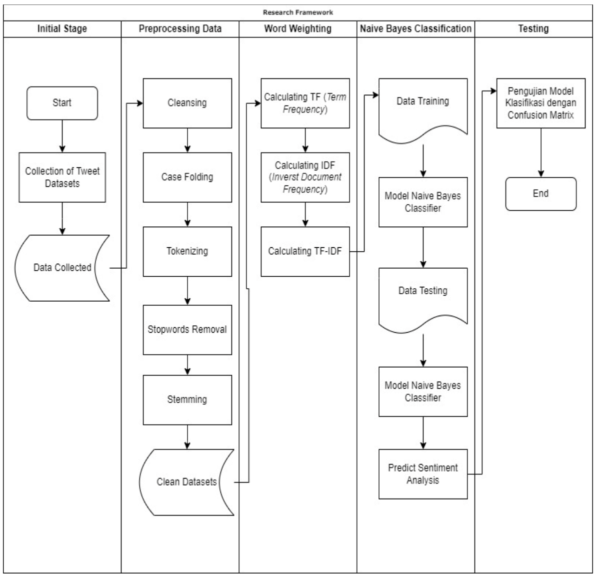


Figure 3. Research Framework

2.1 Data collection

This research publication makes use of secondary data, more specifically data from Twitter. This research makes use of the Sentiment Analysis Dataset of Political Personalities Billboards During the Epidemic, which was gathered by crawling the Twitter API. This dataset was used to analyze the sentiments expressed by billboards displaying political figures during the pandemic. The crawling of datasets produces files in the.csv format, which may then be used for additional data preparation.

2.2 Preprocessing

At this step of the classification process, the preprocessing is carried out with the purpose of reducing qualities that have less of an effect on the overall process. The data that is input at this step is still in the form of raw data, which means that it is still unclean; therefore, the end result of this procedure is a quality document that is supposed to make the categorization process easier. The preprocessing step involves a few different processes, including cleaning, case folding, tokenization, removing stopwords, and stemming. The outcomes of these procedures are outlined in the following tables, beginning with table 1 and continuing through table 5.

In the process of cleaning up the sentences that were collected, there is typically still noise, which may be described as errors at random or variations in the variables that were assessed. Therefore, whatever it is that causes the noise needs to be eliminated. HTML characters, keywords, emotion symbols, hashtags (#), usernames (@username), urls (http://website.com), and email addresses (name@website.com) are not included in the list of omitted terms.

Table 1. Cleansing Process

Input Process	Output Process
Banyak Baliho Tokoh Politik Tak Berizin, Kesbangpol DIY Akan Panggil Petinggi Parpol https://t.co/irzy0snhpx https://t.co/z8ZY5GvQwy	Banyak Baliho Tokoh Politik Tak Berizin, Kesbangpol DIY Akan Panggil Petinggi Parpol

The process of shifting the size of the letters in each word from upper case to lower case or lowercase is referred to as case folding. The objective is to get rid of redundant data that consists of the same information only with lowercase characters.

Table 2. Case Folding Process

Input Process	Output Process
Banyak Baliho Tokoh Politik Tak Berizin, Kesbangpol DIY Akan Panggil Petinggi Parpol	banyak baliho tokoh politik tak berizin, kesbangpol diy akan panggil petinggi parpol

Tokenizing is the process of segmenting tweets into individual chunks by analyzing each individual word that makes up each tweet. The text in question consists solely of words that are all punctuated by spaces between them. Therefore, the conclusion of this procedure is a single word, which is then inserted into the database for the purpose of weighing.

Table 3. Tokenizing Process

Input Process	Output Process
banyak baliho tokoh politik tak berizin, kesbangpol diy akan panggil petinggi parpol	'banyak' 'baliho' 'tokoh' 'politik' 'tak' 'berizin' 'kesbangpol' 'diy' 'akan' 'panggil' 'petinggi' 'parpol'

The step known as "Stopwords Removal" refers to the process of deleting words that contribute little to the overall meaning of the document. Words like "which" and "if" are examples of less significant words, but there are plenty of others. Even the presence of words does not compromise the reliability of the assessment of the document's emotional state. The terms that are intended to be eliminated are compiled in a database known as stopwords. If there is a tweet in the document that matches the word in the stopword, the word will be deleted and replaced with a space character. If there is not a tweet that matches the word in the stopword, the word will not be erased.

Table 4. Stopword Removal Process

Input Process	Output Process
banyak baliho tokoh politik tak berizin, kesbangpol diy akan panggil petinggi parpol	banyak baliho tokoh politik tak berizin, kesbangpol panggil petinggi parpol

The process of stemming converts any word that has an affix into the stem (base word) of the word that is produced when a stopword is applied to it.

Table 5. Stemming Process

Input Process	Output Process
banyak baliho tokoh politik tak berizin, kesbangpol panggil petinggi parpol	banyak baliho tokoh politik tak izin kesbangpol panggil tinggi parpol

2.3 Weighting

At this point, the word weight computation (term) will be carried out using the available data utilizing the TF-IDF technique (Term Frequency-Inverse Document Frequency), which stands for Term Frequency-Inverse Document Frequency. The frequency with which a phrase appears in a given text is referred to as its level of frequency, which is referred to as the term's frequency. In the meanwhile, the number of papers in which a phrase appears is referred to as its document frequency. The tweets that have been preprocessed may be seen in table 6 further down.

Tabel 6. Preprocessed Tweets

	Tweet	Class
D1	banyak baliho tokoh politik tak izin kesbangpol panggil tinggi parpol	Negative
D2	baca gerak tokoh lain gencar pasang baliho bisa tingkat elektabilitas popular	Neutral
D3	lomba panjat baliho tokoh politik seru	Positive

Combine all of the papers in order to search for the number of times each term appears in each of the documents. The results of the computation of the Term Frequency (TF) are presented in Table 7 below:

Table 7. Term Frequency Calculation Results

Term	TF			Term	TF		
	D1	D2	D3		D1	D2	D3
baca	0	1	0	panggil	1	0	0
baliho	1	1	1	panjat	0	0	1
banyak	1	0	0	parpol	1	0	0
bisa	0	1	0	pasang	0	1	0
elektabilitas	0	1	0	politik	1	0	1
gencar	0	1	0	popular	0	1	0
gerak	0	1	0	seru	0	0	1
izin	1	0	0	tidak	1	0	0
kerbangpol	1	0	0	tinggi	1	0	0
lain	0	1	0	tingkat	0	1	0
lomba	0	0	1	tokoh	1	1	1

2.4 Classification

The complete dataset is broken down into many steps, including training data and test data, as shown in Figure 4. This classification method is referred to as Naive Bayes. In the beginning phases of the dataset, preparations were done, beginning with preprocessing and continuing all the way through the weighting process utilizing the TFIDF technique. For the purpose of determining the category probability value and the probability of each word in each term for each class using the training data that has been obtained with positive count = 3, neutral count = 12, and negative count = 7, each word whose TF-IDF weight value has been determined will serve as a reference for the other words. Computing the likelihood of each category (prior), which includes positive, neutral, and negative, and then calculating the probability of each word over all documents, where the number of documents is determined by the outcomes of the preprocessing of the data. During the initial phase of the training process, the implementation of the Naive Bayes classifier is carried out in order to construct probabilities based on the training data. After that, input the data that will be tested. After that, use the Nave Bayes method to compute the values of each characteristic that has been uncovered. Additional analysis of the test data using the classification model that was developed using the training data. At this point in the process, the data is sorted into one of three groups based on its tenor: positive, neutral, or negative. When the process of classification has been finished, the information obtained from the outcomes of the classification will be saved in preparation for the assessment process.

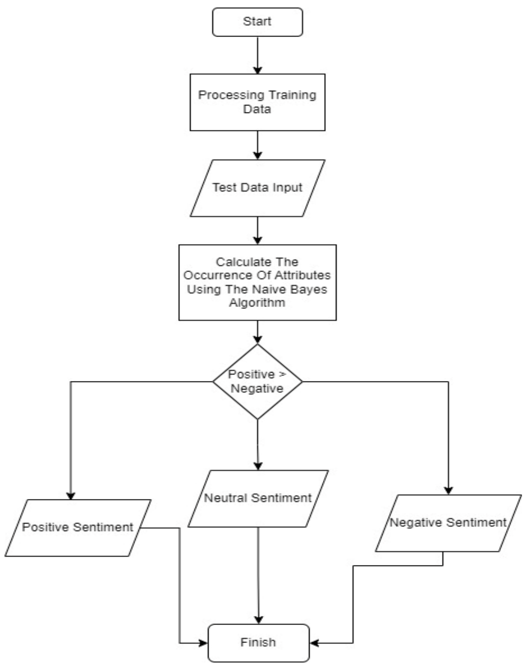


Figure 4. Naïve Bayes Classifier Flowchart

DOI: https://doi.org/10.29207/resti.v7i1.4643

The equation (1) is the formula that will be used to calculate each of the terms in the text.

$$P(w|pos/neg/net) = \frac{(nk \mid pos/neg/net)+1}{(n|\, pos/neg/net)+|kos|} \quad (1)$$

Where P (w| pos / neg / net) is the probability of occurrence of a word in a category, w is a word that appears in a category, (n, pos / neg/ net)+1 is the number of times the word appears in the category, n is the occurrence of every word. words in categories and vocabulary are all words from all categories.

Based on equation (1) above, the processed probabilities include ‘baca’, ‘baliho’, ‘banyak’, ‘bisa’, ‘elektabilitas’, ‘gencar’, ‘gerak’, ‘izin’, ‘kesbangpol’, ‘lain’, ‘lomba’, ‘panggil’, ‘panjat’, ‘parpol’, ‘pasang’, ‘politik’, ‘popular’, ‘seru’, ‘tidak’, ‘tinggi’, ‘tingkat’, and ‘tokoh’.

2.5 Testing

At this point in the process, an assessment is carried out to evaluate the findings of the classification. This is accomplished by assessing the performance value of the system that has been developed up to this point. Accuracy, precision, recall, and f1-Score are the test parameters that are utilized in the assessment process. The computation for these scores is derived from the confusion matrix table. The procedure of testing is depicted in figure 5, which is a diagram.

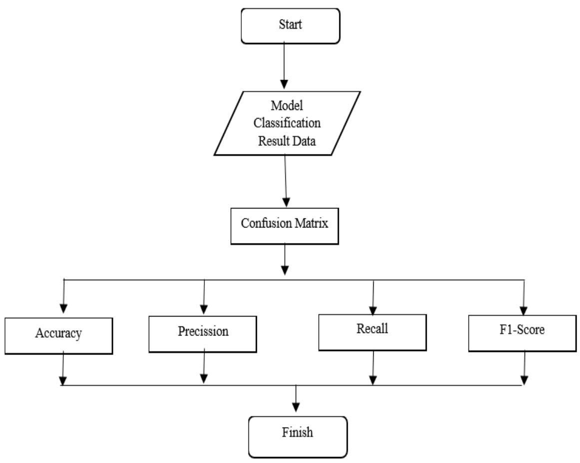


Figure 5. Testing Process Diagram

The system will automatically use the confusion matrix table to illustrate the outcomes of the classification process carried out by the model. Following this, the values for accuracy, precision, recall, and f1-score may be derived from each model that has been created by utilizing the confusion matrix. It is possible to form a judgment about which model receives the greatest categorization value based on this collection of values.

3. Result and Discussion

The preprocessed dataset will be split up into two pieces, namely training data and testing data. These will be the two sections that will be used moving forward. Following the preprocessing, there are a total of 170 tweets in the datasets, of which 76 tweets have a negative emotion, 52 tweets have a positive attitude, and 424 tweets have a neutral sentiment. Figure 6 provides a visual representation of the amount of sentiment information that may be gleaned from tweets.

```
dataset.label.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe043fa36d0>
```

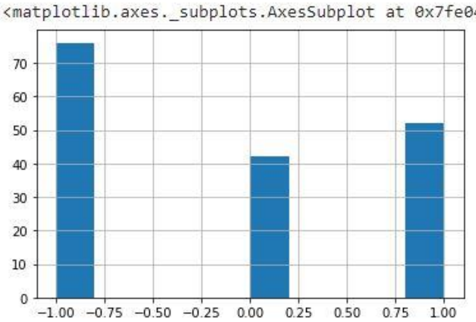


Figure 6. Volume Sentimen

It is required to conduct an even distribution of the volume of each sentiment that is utilized; that is, each sentiment has a total of 20 tweets in order to have a high level of accuracy. This distribution must be done in order to have a good level of accuracy. The procedure of determining word weights with TFIDF (Term Frequency-Inverse Document Frequency) will then be carried out once the data have been separated in the manner described above. TFIDF is a method that is used to determine the word weights (terms) based on the data that is already there. At this point, the TfidfVectorizer function included in the sklearn package is called upon to perform the necessary work. In addition, sentiment classification will be done based on NBC by computing the likelihood of testing data documents by referring to the word probabilities in the training data. This will be done by calculating the chance of testing data documents. The MultinomialNB function found in the sklearn package is utilized in order to complete this NBC procedure.

An evaluation of the model that had been developed was carried out so that the performance of the Naive Bayes Algorithm could be determined. Confusion matrices will be used to represent the outcomes of the classification when they are displayed visually. The confusion matrix is a tool that evaluates the performance of a classification technique. It provides data that compares the results of a classification that was performed by the system with the results of a classification that should have been performed. A visualization of the confusion matrix was created with seaborn, which is an open-source visualization package

DOI: https://doi.org/10.29207/resti.v7i1.4643

built on top of the matplotlib library. This visualization may be found in Figure 7.

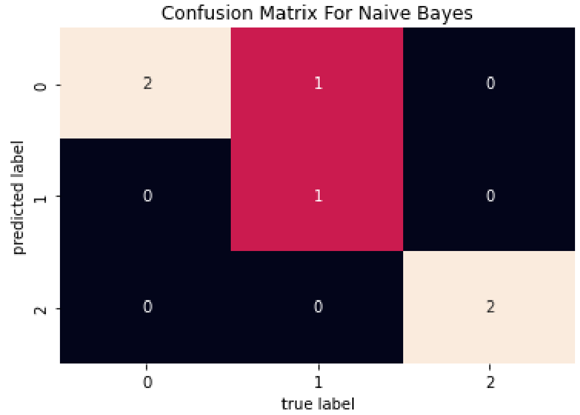


Figure 7. Confusion Matrix

It is clear that the model properly identifies two of the data as being positive, one data as being neutral, and two data as being negative. In addition, the model's prediction of one data point as belonging to a positive class when it should be neutral is incorrect (False Positive). The outcome of the confusion matrix is presented in table 8. The outcomes of the Confusion Matrix are summarized in the table that can be found below.

Table 8. Confusion Matrix Results

		Predict Class		
Actual Class		Positive	Neutral	Negative
	Positive	2	1	0
	Neutral	0	1	0
	Negative	0	0	2

According to the confusion matrix table 8, the model properly identifies two pieces of information as being positive, one piece of information as being neutral, and two pieces of information as being negative. In addition, the model's prediction of one data point as belonging to a positive class when it should be neutral is incorrect (False Positive). Equation (2) is used in the manual computation of the correctness of the matrix calculation described above:

$$\text{Accuracy} = \frac{True\ Positif + True\ Netral + True\ Negatif}{Total\ Data\ yang\ di\ Uji} x100\% \quad (2)$$

$$= \frac{5}{6} \text{ x } 100\ \% = 0.8333333333 \text{ or } 83.33\ \%$$

Accuracy is a measure of how well the data can be classified based on how well the model that was created did its job. Calculating the proportion of correct predictions to the whole data set is how one arrives at an accurate estimate of the model's performance. The level of the system's ability to find the accuracy between the information desired by the user and the answers given by the system can be stated as a function of the value that accuracy plays in the overall performance of the system, which can be determined by knowing the value that accuracy has on the overall performance of the system. In this particular research endeavor, the system had a success rate of 83.33 percent when it came to discovering information.

In addition, the value of accuracy, recall, and f1 score in each classification class may be used to determine the performance of the classification of each class. This can be done so in order to observe how well each classification class performs. The amount of accuracy that exists between the data that was requested and the findings that were produced by the model is referred to as precision. The computation of the ratio of correct predictions to the total positive projected outcomes is how precision is determined. This ratio is compared to the overall positive expected results. The ability of the model to successfully retrieve the information that was entered into the test is referred to as its recall. The results of the calculation of the ratio of true positive predictions to the total data that is true positive are used to derive recall. Recall is produced from these findings. The F1-Score is a single metric that measures the success of retrieval by combining recall and precision into a single score. The following are the findings from the performance of the sentiment analysis, which was produced with a precision value of 89%, recall of 83%, and an f-1 score of 82%, as can be seen in the figure that follows (Figure 8).

	precision	recall	f1-score	support
-1	0.67	1.00	0.80	2
0	1.00	0.50	0.67	2
1	1.00	1.00	1.00	2
accuracy			0.83	6
macro avg	0.89	0.83	0.82	6
weighted avg	0.89	0.83	0.82	6

Figure 8. Precision, Recall and fl-score NBC Model

The findings of the sentiment analysis may be retrieved once the data cleaning and classification processes have been carried out. A wordcloud is a type of data visualization that describes a collection of terms that are widely available in a text analysis. Wordclouds are created by analyzing large amounts of text. Figure 9 provides a visual representation of the messages sent by political billboards around the country.

Figure 9. Wordcloud Political Billboard

4. Conclusion

The public's reaction to the installation of billboards for political figures during the pandemic on Twitter social media can be seen from the number of tweets expressing negative sentiments, which number as many as 76, while the number of tweets expressing positive sentiments, which number as many as 52, and the number of tweets expressing neutral sentiments, which

DOI: https://doi.org/10.29207/resti.v7i1.4643

number as many as 42. The Naive Bayes categorization of billboards of political personalities on Twitter social media during the pandemic yielded an accuracy level of 83.3% with data points. This result was derived via the analysis of data. Based on these findings, it appears that the Naive Bayes algorithm is capable of doing a good job of classifying [26]. Tweets were classified into favorable, negative, and neutral categories about thoughts on billboards depicting political leaders during the epidemic. It is recommended that further research be done to increase the number of datasets so that the accuracy level can be more accurately determined. Additionally, it is recommended that further research be done to apply the five categories of the Naive Bayes classification model that are applied, specifically positive, very positive, negative, very negative, and neutral, so that more categories can be used as evaluation parameters.

Acknowledgments

We would like to express our gratitude to Twitter, APJII, KPU Indonesia, dataindonesia.id, and the Faculty of Engineering at Universitas Majalengka for the use of their respective research facilities.

References

[1] APJII, "APJII Internet Survey Report 2019-2020 (Q2)," *https://apjii.or.id/content/read/39/521/LaporanSurvei-Internet-APJII-2019-2020-Q2*, 2020.

[2] S. Supadiyanto and S. Murti, "Social Media Portraits and YouTubers during the COVID-19 Pandemic in Indonesia," *Technium Education and Humanities*, vol. 2, no. 1, pp. 35–47, Jun. 2022, doi: 10.47577/teh.v2i1.6815.

[3] A. N. Vyas *et al.*, "The Girl Rising 'We Dream, We Rise' Social Media Campaign in India: Reach, Engagement and Impact," *Journal of Creative Communications*, vol. 15, no. 1, pp. 106–124, Mar. 2020, doi: 10.1177/0973258619878354.

[4] C. de las Heras-Pedrosa, P. Sánchez-Núñez, and J. I. Peláez, "Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and Its Impact on Digital Ecosystems," *Int J Environ Res Public Health*, vol. 17, no. 15, 2020, doi: 10.3390/ijerph17155542.

[5] R. Chandrasekaran, V. Mehta, T. Valkunde, and E. Moustakas, "Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study," *J Med Internet Res*, vol. 22, no. 10, p. e22624, Oct. 2020, doi: 10.2196/22624.

[6] A. A. ARUGAY, "Stronger Social Media Influence In The 2022 Philippine Elections," *https://fulcrum.sg/stronger-social-media-influence-in-the-2022-philippine-elections/*, Jun. 22, 2022.

[7] M. Kaur, R. Verma, and S. Ranjan, "Political Leaders' Communication: A Twitter Sentiment Analysis during Covid-19 Pandemic," *Jurnal The Messenger*, vol. 13, no. 1, p. 45, Sep. 2021, doi: 10.26623/themessenger.v13i1.2585.

[8] G. A. M. Kukuh Jaluwana, Gusti Made Arya Sasmita, and I Made Agus Dwi Suarjaya, "Analysis of Public Sentiment Towards Goverment Efforts to Break the Chain of Covid-19 Transmission in Indonesia Using CNN and Bidirectional LSTM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 511–520, Aug. 2022, doi: 10.29207/resti.v6i4.4055.

[9] Komisi Pemilihan Umum, Peraturan Komisi Pemilihan Umum Republik Indonesia Nomor 13 Tahun 2020 Tentang Perubahan Kedua Atas Peraturan Komisi Pemilihan Umum Nomor 6 Tahun 2020 Tentang Pelaksanaan Pemilihan Gubernur Dan Wakil Gubernur, Bupati Dan Wakil Bupati, Dan/Atau Wali Kota Dan Wakil Wali Kota Serentak Lanjutan Dalam Kondisi Bencana Nonalam Corona Virus Disease 2019 (Covid-19). Jakarta: KPU, 2020.

[10] A. Priana, "Efektivitas Iklan Baliho Grace Natalie Dalam Meningkatkan Citra Diri Sebagai Ketua Umum Partai Solidaritas Indonesia Tahun 2019" *Commed : Jurnal Komunikasi dan Media*, vol. 4, no. 2, pp. 108–122, Feb. 2020, doi: 10.33884/commed.v4i2.1652.

[11] I. A. Sutarini, "Pengukuran Efektivitas Penggunaan Media Baliho Pada Pemilihan Umum 2019 Terhadap Generasi Milenial," *SENADA (Seminar Nasional Manajemen, Desain dan Aplikasi Bisnis Teknologi)*, vol. 2, pp. 237–241, Feb. 2019, [Online]. Available: https://eprosiding.idbbali.ac.id/index.php/senada/article/view/217

[12] detikcom, "Survei Litbang Kompas: Baliho Politik Dianggap Nggak Ngaruh!," *https://news.detik.com/berita/d-5692372/survei-litbang-kompas-baliho-politik-dianggap-nggak-ngaruh*, Aug. 23, 2021.

[13] R. A. Yanuartha, R. R. Siahainenia, P. Hergianasari, and K. Netanyahu, "Strategi Komunikasi Politik Penggunaan Baliho Versus Media Sosial Oleh Partai Demokrasi Indonesia Perjuangan Tahun 2021," *Administraus*, vol. 6, no. 1, pp. 21–40, Jan. 2022, doi: 10.56662/administraus.v6i1.137.

[14] R. Farasonalia, "Makna 'Kepak Sayap Kebhinekaan' di Baliho Bergambar Puan Menurut Politikus PDI-P," *https://regional.kompas.com/read/2021/08/07/064402878/makna-kepak-sayap-kebhinekaan-di-baliho-bergambar-puan-menurut-politikus?page=all*, Aug. 07, 2021.

[15] M. I. Khatami, "Eksistensi 'Baliho 2024' Dalam Pertarungan Elektabilitas: Analisis Pencitraan Tokoh Politik," *Jurnal Ilmu Sosial Dan Ilmu Politik Universitas Jambi*, vol. 5, no. 2, Oct. 2021.

[16] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," *https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022*, Aug. 10, 2022.

[17] R. P. Dakhi and A. M. B. Aji, " Sentiment Analysis Of Public Opinions On The Effectiveness Of Online Learning Using Naïve Bayes Algorithm," *JISICOM (Journal of Information System, Informatics and Computing)*, vol. 6, no. 1, pp. 273–279, 2022, doi: 10.52362/jisicom.v6i1.822.

[18] N. M. G. P. dan Mochammad Fauzi dan Indriati Indriati dan Liana Dewi, " Identifikasi Tweet Cyberbullying pada Aplikasi Twitter menggunakan Metode Support Vector Machine (SVM) dan Information Gain (IG) sebagai Seleksi Fitur," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 11, pp. 5236–5332, 2018, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3276

[19] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER: Journal of Information Technology*, vol. 2, no. 1, Mar. 2017, doi: 10.31284/j.integer.2017.v2i1.95.

[20] Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, and A. Zhou, "Towards online anti-opinion spam: Spotting fake reviews from the review sequence," in *2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014)*, 2014, pp. 261–264.

[21] Arminditya Fajri Akbar, Harry Budi Santoso, P. O. H. Putra, and S. B. Yudhoatmojo, "User Perception Analysis of Online Learning Platform 'Zenius' During the Coronavirus Pandemic Using Text Mining Techniques," *Jurnal Sistem Informasi*, vol. 17, no. 2, pp. 33–47, Oct. 2021, doi: 10.21609/jsi.v17i2.1065.

[22] C. Iglesias and A. Moreno, "Sentiment Analysis for Social Media," *Applied Sciences*, vol. 9, p. 5037, Nov. 2019, doi: 10.3390/app9235037.

DOI: https://doi.org/10.29207/resti.v7i1.4643

[23] M. Bramer, "Principles of Data Mining. Undergraduate Topics in Computer Science," *Ch. 12: Estimating the Predictive Accuracy of a Classifier*, Nov. 2013.

[24] R. Risnantoyo, A. Nugroho, and K. Mandara, "Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm,"*Journal Of Informatics And Telecommunication Engineering*, vol. 4, no. 1, pp. 86–96, Jul. 2020, doi: 10.31289/jite.v4i1.3798.

[25] N. Umar and M. Adnan Nur, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 585–590, Aug. 2022, doi: 10.29207/resti.v6i4.4179.

[26] Putu, Ni Luh Merawati, Ahmad Zuli Amrullah, Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naïve Bayes dan Latent Dirchlet Allocation", *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5 no. 1 (2021), pp. 123-131. DOI: https://doi.org/10.29207/resti.v5i1.2587

DOI: https://doi.org/10.29207/resti.v7i1.4643