American Journal of Open Research, 2(7), July, 2023, 1-7
E-ISSN: 2834-7668

SENTIMENT ANALYSIS AND TOPIC MODELING OF CITIZEN SATISFACTION WITH THE INDONESIAN GOVERNMENT IN HANDLING A PANDEMIC

Siti Nurmalasari[1], Achmad Nizar Hidayanto[2], Labibah Alya Huwaida[3], Hapsari Wulandari[4]*
[1,2,3,4]Universitas Indonesia, Jakarta, Indonesia
*siti.nurmalasari@ui.ac.id**

*Corresponding Author

Abstract: *When the COVID hit the world, many countries issued policies to stop the spread of the virus. In Indonesia, various opinions and surveys have emerged regarding citizen satisfaction with the government's performance in handling the pandemic. But the survey method still has many weaknesses, for example: bias from the researcher, lack of confidentiality, hello effect, and so on. The purpose of this study is to classify this phenomenon into positive and negative sentiments taken from Twitter data. Every record is pre-processed to clean the data. Data labelling using a lexicon consisting of positive or negative polarities. Sentiment classification using Support Vector Machine (SVM). Each positive and negative sentiment will be processed using Latent Dirichlet Allocation (LDA) method to find out the interpretation of the main topics that are often discussed, then made into a visualization using a word cloud. The best model obtained was the model with TF-IDF feature extraction with a precision value of 0.87, a recall of 0.95, an accuracy of 0.89, and an F1-measure of 0.93. Our findings indicate that people are more likely to be satisfied with the performance of the government's fight against COVID than with the policies they introduce. People are also satisfied because they can feel mudik (go back to hometown) again after two years of the pandemic. Dissatisfaction comes from people who think that there is a business game in vaccine policy as well as the government's lack of transparency regarding the number of COVID cases.*

Keywords: *Sentiment analysis, Topic modeling, COVID, Citizen satisfaction, Government performance, Support vector machine, Latent dirichlet allocation, Twitter.*

INTRODUCTION

The COVID pandemic that has occurred over the past three years has made a lot of changes in the way of life of the people in Indonesia and the world. Since the emergence of the first case with two positive people of COVID in Indonesia (Ihsanuddin, 2020), until finally designated by WHO (World Health Organization) as a global pandemic on March 9, 2020, (Satgas COVID-19, 2020) positive patient cases continue to increase every day. Various steps and policies have been taken by the government to deal with the impact of this virus, although there are various pro and contra opinions from the citizen regarding these policies. The policies including Turn Wisma Atlit into COVID Emergency Hospital that can accommodate approximately 3000 patients (Kurniawan, 2020). Provide incentives for health workers of IDR 5-15 million/month (Kemenkes RI, 2020). Ministry/Agency budget reallocation of IDR 62 trillion (Putri, 2020). Importing Health equipment such as PPE (Personal Protective Equipment), Masks, and Rapid Tests (Wareza, 2020). Providing BLT (Direct Cash Assistance) to people directly affected by the pandemic (Kompas.com, 2020). Make rules for social restrictions (social distancing) known as PSBB (Large-Scale Social Restrictions) to PPKM (Enforcement of Restrictions on Community Activities) (Nurita, 2020; Ramadhan, 2021). Requiring the public to be vaccinated against COVID as an effort to protect the body from the attack of the COVID virus (Kompas.com, 2021).

Currently, after three years of fighting against COVID, Indonesia's situation is starting to recover. Based on data from Johns Hopkins University CSSE (Center for Systems Science and Engineering) COVID on September 12, 2021, Indonesia is considered one of the best in the world in handling COVID cases. This is because Indonesia has succeeded in reducing COVID cases by 58% in just two weeks. The news was also confirmed by the spokesperson for the COVID Ministry of Health, dr. Siti Nadia Tarmizi (Azizah, 2021).

Citizen satisfaction with the government in dealing with COVID has become one of the hot topics among Indonesian people. This is very important because it can be taken into consideration for the government in making better policies in the future. Various opinions and surveys have emerged regarding public satisfaction regarding the government's performance in handling the pandemic in Indonesia, both in print, electronic, and social media, one of which is Twitter.

DOI:
https://opsearch.us/index.php/us/index

Twitter is what is happening and what people are talking about right now (Twitter.com, n.d.). Twitter is a social media that provides facilities for users to communicate with each other, exchange information, and express their thoughts through short messages written in a tweet with a limited size, which is a maximum of 280 characters. In Indonesia, Twitter became the fifth most popular social media site in 2020 with the percentage of users reaching 56% of the 160 million active users of social media (Jayani, 2020). Information that is shared on Twitter is called a tweet. In making a tweet, users usually also include links, videos, and so on.

In recent years, Twitter has exerted a great deal of influence in generating information resources. Because survey methods often have several weaknesses including: bias from researchers, hello effect, depending on people's responses, samples do not represent the entire population, lack of confidentiality, allows for errors and subjectivity (Hayati, 2022).

Sentiment analysis is a computational study in recognizing and expressing opinions, sentiments, evaluations, attitudes, emotions, subjectivity, judgments, or views contained in a text. Sentiment analysis can also be interpreted as finding the polarity of the text and classifying it into positive or negative which can help in evaluating a product or performance, as well as in decision making (Eldha Oktaviana & Arum Sari, 2022).

There are many researchs discuss about sentiment analysis in term of COVID pandemic. (Sulistyono et al., 2021) performed sentiment analysis on Twitter to classify Twitter user responses about COVID vaccine into positive, negative, and neutral responses using Support Vector Machine (SVM) as classification method and evaluation using confusion matrix and cross validation. The result of the study obtained positive sentiment of 43.5%, negative of 19.1%, and neutral of 37.4%. The results of classification evaluation are less than optimal because there are still Indonesian regional languanges that are also process and the result of labelling classification process have different result that affect the algorithm accuracy. (Sahir et al., 2021) conducted the study to analyze public opinion for online learning during COVID pandemic using Twitter data in October 2020. The research carried out document-based text mining and feelings on Twitter data analyzed using the Naïve Bayes method. Analysis in that period found 25% positive sentiment, 74% negative sentiment, and 1% neutral sentiment. Several tweets showed that the words ‘stress’ and ‘covid’ were the most commonly spoken in the month. This study suggested to use of various algorithms to provide more reliable performance in the future experiments. (Bayu Nugroho et al., 2021) performed a text analysis on Twitter to depict the Indonesian's public impression of new normal conditions during COVID pandemic in July 2020. The result then described in a word cloud map and sorted in a flipped bar chart. This study also performed a network bi-gram analysis to identify word correlations then identify sentiments from Twitter text. The results reveal that Indonesian people understand the concept of "new normal" in accordance with the expectations of Indonesian's government. (Sujiwo et al., 2021) presented a sentiment analysis of Indonesian's government policies in overcoming COVID through Twitter data using several classification methods, namely SVM, Naive Bayes, and LSTM. The study found that Indonesian people on Twitter gave negative sentiments.

As previously stated, sentiment analysis has been applied in various research relating to COVID. However, no similar study on citizen satisfaction with the government in handling a pandemic has been conducted using Twitter data. Before, Chen et al. have researched based on an open-source study of 14 nations—the United States, Canada, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden, United Kingdom, Australia, Japan, and South Korea—factors drive the satisfaction of citizens with government reactions to COVID. Number of confirmed cases per million people, number of fatalities per million people, governments' health and containment programs, stringency policies, and economic assistance policies are the contributing variables. The approval of any national reaction to COVID is influenced by economic support and health policy. The findings make some recommendations to governments when they start implementing programs that balance economic support, livelihoods, and public health. Therefore, this study will measure sentiment analysis of citizen satisfaction with the government in handling a pandemic in Indonesia using Twitter data, also the text mining approach.

According to (Annur, 2022) Twitter is a social media platform that is most widely used by its users to find news, including to provide opinions on government policies. While sentiment analysis and topic modeling can assist to understand sentiments and topics being discussed on a related platform

to be used in designing strategies to improve specific services or products (Pathak et al., 2021). There are various typical approaches that are utilized for sentiment analysis and topic modeling. Machine learning is a popular sentiment analysis approach because of its simple methodology and excellent classification accuracy (Balakrishnan et al., 2022)(Wankhade et al., 2022)(Birjali et al., 2021). Moreover, Latent Dirichlet Allocation (LDA) is the most often used approach for modeling topics since it is well suited to modeling general topics with a variety of data (Kherwa & Bansal, 2018) (Vayansky & Kumar, 2020). Due of these characteristics, this study will employ various methodologies to address research questions.

Several theoretical explanations that are relevant to this research as supporting theories will be explained in this section.

Citizen Satisfaction

Citizens are, in fact, the true owner of the governments. Therefore, it is logical that governments should be citizen-centric (Ghodousi et al., 2019) . With the introduction of a citizen-centric approach over the last two decades, the public sector has changed dramatically. The primary changes are decentralization and citizen participation in decision-making. In this regard, citizen satisfaction surveys have become an unavoidable tool for measuring government performance. According to studies, performance measurements fail to achieve their goal if citizens' needs are not considered. As a result, the process for governmental decision-making will become less effective (Ghodousi et al., 2019) (Ho, 2003).

The Expectancy Disconfirmation Model (EDM) is a theoretical framework that explains how individuals form and revise their expectations about a product or service, and how these expectations influence their satisfaction or dissatisfaction with their actual experience of using or consuming the product or service. This model has long been used in the private sector to investigate consumer satisfaction, but Van Ryzin (Van Ryzin, 2004) has recently linked it to citizen satisfaction with local government services.

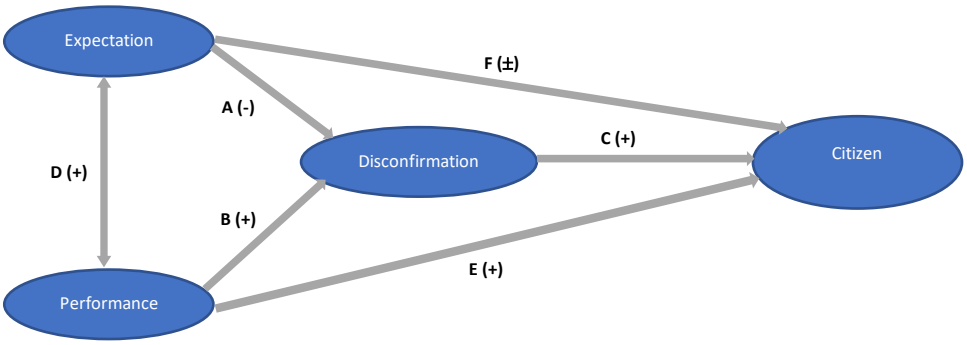

Figure 1. The Expectancy Disconfirmation Model (EDM) (Chatterjee & Suy, 2019)

(Chatterjee & Suy, 2019) identified the relationship between the different variables as shown in Figure 1. This provides insight into the important hypotheses associated with this model and its implications for the analysis of citizen satisfaction in public sectors.

Table 1. The Expectancy Disconfirmation Model (EDM) (Chatterjee & Suy, 2019)

Link	Variable
A	It is hypothesized that high expectations give rise to more negative disconfirmation
B	It represents that high performance results in more positive disconfirmation
C	Disconfirmation, in turn, is positively influenced to frame satisfaction judgment
D	Expectations and performance are also believed to be positively interlinked
E	It shows performance positively connected with satisfaction. It indicates that performance is a preliminary element which can influence the satisfaction beyond varying expectation level of individual

Link	Variable
F	Expectations can influence satisfaction when people are unable to evaluate the service performance of local government directly then the evaluation is done on the basis of previously hold expectation for reasons of dissonance reduction or ego defensiveness

Sentiment Analysis

Sentiment analysis (SA) is a study that analyzes people's opinions, sentiments, judgments, attitudes, and emotions toward entities such as products, services, organizations, individuals, events, issues, or topics which written as either positive, neutral, or negative, and can be applied to a variety of industries, including consumer goods, healthcare, tourism, hospitality, financial services, social events, and political elections (Liu, 2020). SA can be used by researchers, business organizations, and governments to gain business insights and make better decisions by analyzing public emotions and views (Birjali et al., 2021). Machine learning, lexicon-based approaches, and hybrid approaches are three commonly used approaches for sentiment analysis (Birjali et al., 2021; Wankhade et al., 2022) (Ben-Hur & Weston, 2010). This research used Support Vector Machine (SVM) as primary method due to its simple algorithm and high accuracy results (Birjali et al., 2021), (Wankhade et al., 2022). SVM is a non-probabilistic classifier that can divide data linearly or non-linearly and deal with discrete and continuous variables. This classifier analyzes data and determines the best hyperplane to divide it into different classes. The hyperplane with the greatest margin to the closest training point of the two classes demonstrates effective separation (Birjali et al., 2021). The hyperplane separation is set to be as large as possible (Tripathy et al., 2016). SVM falls under the category of kernel methods, which are algorithms that rely on data only through the dot-product or can be replaced by a kernel function (Birjali et al., 2021).

Topic Modeling

Topic modeling is a method for discovering and annotating large collections of documents. It is a statistical method for analyzing the words in a text to determine the topics contained within it, as well as how they are related and change over time. This algorithm can be used to find patterns in a variety of data types, including genetic data, images, and social networks. This algorithm requires no prior labeling because topics emerge from an analysis of the original text, allowing us to manage electronic records on a scale that manual labeling would not allow (Blei, 2012).

The simplest and most widely used method for finding topics in text documents is Latent Dirichlet Allocation (LDA). The basic idea behind LDA is that documents are represented as a random mix of hidden topics, with each topic defined by a word distribution. LDA assumes that topics are generated before documents. As a result, the word for each document is generated in two stages. First, I chose a topic distribution at random. Second, for each word in the document, choose a topic at random from the distribution of topics in step 1, followed by a word at random from the appropriate vocabulary distribution (Blei, 2012).

The research questions that will be addressed by this study are as follows: 1) What is the sentiment towards citizen satisfaction with the government in handling a pandemic? 2) What recommendations may be offered to enhance citizen satisfaction based on the topics derived from negative sentiments? This paper is organized as follows: section 1 provides the introduction and research background; section 2 discusses the relevant literature of this study. The proposed method for this study is presented in section 3. Section 4 discusses analysis and experimental result. Conclusions, implications, and future work are presented in section 5.

MATERIALS AND METHODS

The purpose of this study is to classify positive and negative sentiments taken from Twitter data about citizen satisfaction with the government in dealing with pandemic and then modelling topics from the results. In data collection to classification, machine learning methods are used, including the Support Vector Machine (SVM) and topic modelling using the Latent Dirichlet Allocation (LDA). The flow diagram of the proposed method shown in Fig. 2. This study has 7 stages, namely crawling Twitter data, data pre-processing, feature extraction, sentiment classification, model evaluation, topic modelling, and topic visualization.

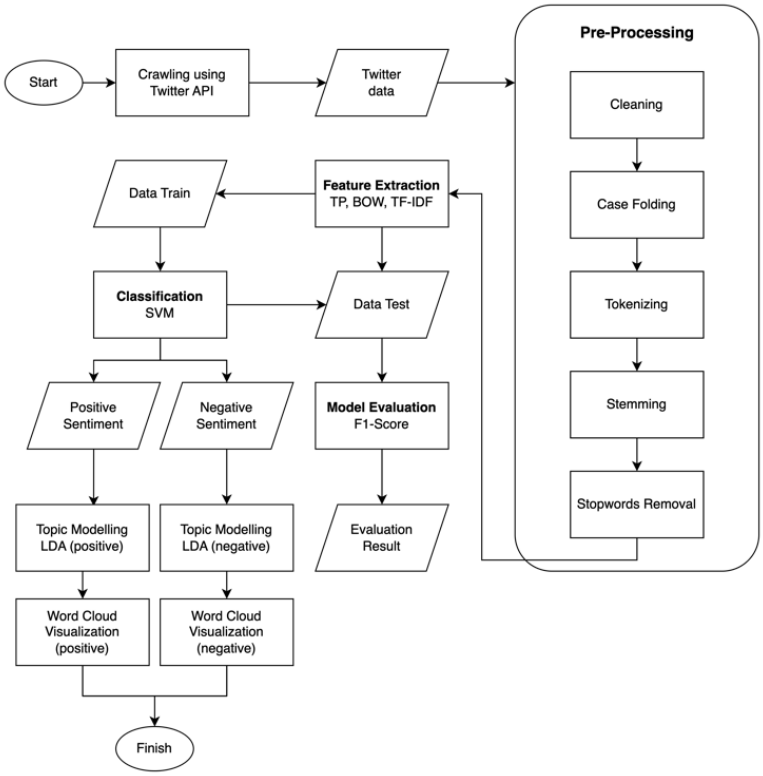

Figure 2. Research method

A. Data Collecting

The authors collect Twitter data from April-May 2022 containing keywords related to government policies in dealing with COVID pandemic in Indonesia. An example of successfully collected tweet data is shown in Table 2. The authors made observation using data from Twitter API. Twitter API is used to collect tweet data and stored it in a database. Twint is a Python library that is used to access the Twitter API. Twint has a feature to retrieve a set of data using certain keywords and a predefined time limit. While searching for tweets, query matching with the tweets includes the geospatial information, user, etc., got down the exact data required to perform the evaluation. Acquiring the data leads to find the information which may or may not be helpful, and when the data collected is not applicable cleans the required data. Following is the list of key-value pairs similar to Python "dictionary" data type: user id, username, original tweet, hashtag, location, timestamp.

B. Data Preprocessing

The second stage is text cleaning which aims to prevent the dataset from containing noise so that it can affect sentiment classification modeling. The initial data processing steps in this study follow the primary stages:

1. Cleansing: removes non-ASCII characters, URL addresses, hashtags, punctuation, numbers, new lines, and extra spaces.
2. Case folding: converts all characters in the review data to lowercase.
3. Stopping removing stopwords from the text so we can focus more on essential words, in this study, using Sastrawi and NLTK libraries in Python.
4. Stemming: changing words into forms without affixes, in this study, using Sastrawi library in Python.
5. Tokenization: dividing the existing text into smaller and meaningful elements, in this study, splitting sentences into words.

C. Sentiment Analysis

The sentiment classification model was trained using Supervised Machine Learning, which includes several popular classifiers, one of which is SVM. The classifier model will be evaluated using accuracy, precision, recall, and F-score measurements, as well as K-fold cross-validation with 5-fold.

At this point, the best classifier model obtained will be used to identify sentiments in the entire dataset.

D. Topic Modeling

At this stage, the review data from the previous stage will be used for topic modeling based on positive and negative sentiments. The goal is to identify the main topics of discussion for each sentiment. Positive sentiment reviews can provide information about topics that users like, while negative sentiment reviews can provide information about topics that users dislike. For topic modeling, the LDA algorithm is used in this study.

RESULTS AND DISCUSSION

The following are the results and discussion of the study that has been conducted based on the methodology described in section 3:

A. Data Collection

The authors collect Twitter data from April-May 2022 containing keywords related to government policies in dealing with COVID pandemic in Indonesia and obtained 95,952 tweets. An example of successfully collected tweet data is shown in Table 2.

Table 2. Example of Tweet

Class	Tweet
Positive	Alhamdulillah kepuasan publik semaksimal meningkat nih gaes terhadap kinerja Pak Jokowi selama ini, apalagi didorong dengan keberhasilan penanganan pandemi COVID dan lancarnya penyelenggaraan mudik tahun ini #IndonesiaBisaLebihMaju
Negative	Siapa sih investor vaksin nusantara? apa mereka ketakutan jika vaksin nya gagal ngasilin cuan? #Kompascom #Indonesia #Vaksin #vaksinnusantara #idi

B. Data Preprocessing

The second stage is text cleaning which aims to prevent the dataset from containing noise so that it can affect sentiment classification modeling. The cleaning process is carried out by deleting urls, hashtags, mentions, casefolding, removing stopwords, stemming, and tokenization. Based on this text cleaning process, 5,990 tweets were obtained, which were used in the analysis.

Table 3. Example of Data Cleaning

Step	Before	After
Cleansing	Pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi COVID. Seperti apa peraturan baru tersebut? Simak selengkapnya. https://t.co/N0xXUilj0b	Pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi COVID Seperti apa peraturan baru tersebut Simak selengkapnya
Case folding	Pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi COVID Seperti apa peraturan baru tersebut Simak selengkapnya	pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi covid seperti apa peraturan baru tersebut simak selengkapnya
Stopwords Removal	pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi covid seperti apa peraturan baru tersebut simak selengkapnya	pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi covid seperti apa peraturan baru tersebut simak selengkapnya
Stemming	pemerintah mengeluarkan aturan baru soal sekolah tatap muka di situasi masa pandemi covid seperti apa peraturan baru tersebut simak selengkapnya	perintah keluar atur baru soal sekolah tatap muka di situasi masa pandemi covid seperti apa atur baru sebut simak lengkap
Tokenization	perintah keluar atur baru soal sekolah tatap muka di	[‘perintah’, ‘keluar’, ‘atur’, ‘baru’, ‘soal’, ‘sekolah’,

	situasi masa pandemi covid seperti apa atur baru sebut simak lengkap	'tatap', 'muka', 'di', 'situasi', 'masa', 'pandemi', 'covid', 'seperti', 'apa', 'atur', 'baru', 'sebut', 'simak', 'lengkap']

C. Sentiment Analysis

Based on this text cleaning process, 5,990 tweets were obtained, which were used in the analysis. The next step is to label each sentence using a lexicon consisting of positive and negative polarities. The result shown in Table 4.

Table 4. Classification of the tweets

Total tweets	Positive tweets	Negative tweets
5,990	4,179	1,811
100%	75%	25%

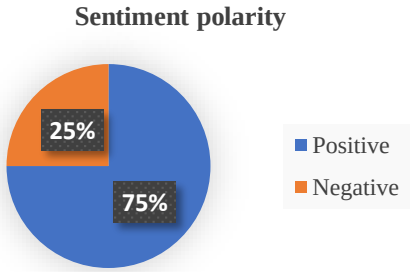

Figure 3. Sentiment analysis of citizen satisfaction with the Indonesian goverment in handling a pandemic

The cleaned and transformed data is divided into 2, namely 70% train data to train the model (3,920 tweets) and 30% test data to evaluate the model's performance (1,670 tweets). The model training was carried out using a data train containing 75% positive labels (2,958 tweets) and 25% negative labels (968 tweets) where the proportions of the two labels were not balanced (imbalanced data). Table 5 compares modeling using Term Presence (TP), Bag of Words (BoW), and Term Frequency-Inverted Document Frequency (TF-IDF) feature extraction.

Table 5. Evaluation between models

	Term Presence	***Bag of Words***	***TF-IDF***
Precision	0.75	0.78	0.84
Recall	0.79	0.78	0.61
Accuracy	0.89	0.9	0.88
F1-Measure	0.77	0.78	0.71

Based on the comparison of scores above, the three models have an average F1-Measure of 0.75 so that the model is improved by performing SMOTE to overcome imbalanced data and model validation using the K-fold cross validation method with a value of k=5.

K-Fold Cross Validation used to determine the test results and evaluate the maximum algorithm performance. With this method, the test used is by folding k data and doing k iterations. The following table compares modeling using term presence, Bag of Words, and TF-IDF feature extraction using SMOTE and K-fold cross validation methods.

Table 6. Evaluation between models Using K-Fold Cross Validation

	Term Presence	***Bag of Words***	***TF-IDF***
Average Precision	0.70	0.72	0.87
Average Recall	0.78	0.79	0.95
Average Accuracy	0.86	0.75	0.89
F1-Measure	0.74	0.83	0.93

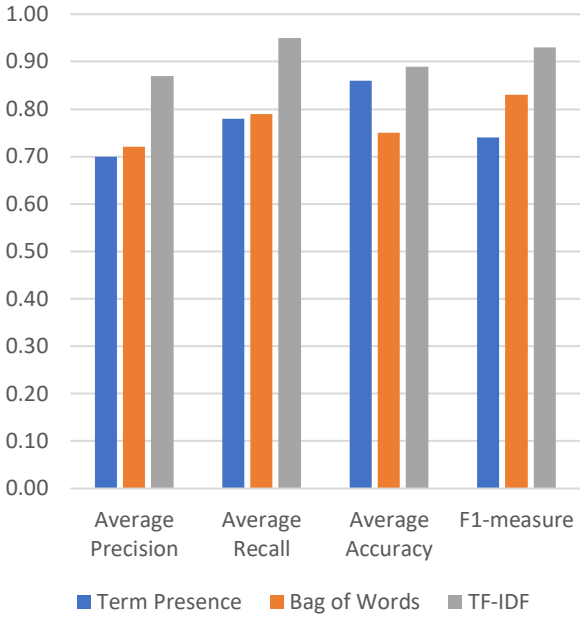

Figure 4. Evaluation between models Using K-Fold Cross Validation

Based on the comparison of scores above, the best model is the model with TF-IDF feature extraction to predict sentiment in the test data.

D. Topic Modeling

After conducting sentiment analysis, topic modeling was carried out on 4,179 positive sentiments (75%) 1,811 negative sentiments (25%). This study uses the LDA algorithm for topic modeling. The results are represented in Table 7.

Table 7. Modeling the Topic of positive sentiment

Topic	Word Distribution
Presidential performance	kerja, puas, jokowi, masyarakat, presiden
Vaccine policy	prokes, vaksinasi, kunci, kendali, lebaran
Recovery rate	sembuh, update, orang, indonesia, hasil
Handling of the covid virus	covid, virus, tugas, corona, tangan
Covid task force	prokes, covid, lebaran, patuh, satgas

The results of the extraction of negative sentiment topics with Word Distribution are represented in Table 8.

Table 8. Modeling the topic of negative sentiment

Topic	Word Distribution
PPKM policy	gagal, ppkm, covid, miris, level
Against vaccine	vaksin, covid, corona, tangan, hasil
Vaccine business	covid, corona, atur, perintah, vaksin

Discussion

This study employs sentiment analysis and topic modeling to assess citizen satisfaction with the government's response to the pandemic. This study uses Twitter data to look at public opinion regarding government policies in dealing with pandemics. According to the findings of this study, TF-IDF model with gives better performance than TP and BoW as a whole. This model produces accuracy 0.89, precision 0.87, recall 0.95, and F-score 0.93.

This study also produces sentiment classifications with positive and negative sentiments for citizen satisfaction to government regarding policies in dealing with pandemic as seen in Table 4. Positive and negative sentiments have a significant difference, namely 75% positive sentiment with 4,179 tweets and 25% negative sentiment with 1,811 tweets. In other words, citizen satisfaction with the goverment regalding policies in dealing with pandemics still dominate by positive sentiment, although there are still those who give negative sentiment.

The results of the topic modelling are then visualized in the form of a word cloud in Figure 5. Words that appear most frequently are visualized as more significant in size and darker in color. Blue words represent positive sentiments and red words represent negative sentiments. Word cloud visualization shows that tweets with positive sentiment relate to citizen satisfaction with the government's performance in handling the pandemic, especially the mandatory vaccine regulations and the formation of the COVID task force. Moreover, tweets with negative sentiment relate to people who disagree with vaccines because they are considered to be just a business field and also there are PPKM policies in various regions.

Figure 5. Word Cloud Visualization

Table 9 shows the results of the author's attempt to map the results of sentiments with one of the sources of tweets that represent these sentiments.

Table 9. Mapping tweets and its sentiment

Tweets for positive sentiment	Tweets for negative sentiment
Kepuasan masyarakat terhadap kinerja Presiden Jokowi meningkat karena dinilai berhasil tangani Pandemi COVID (Twitter.com, 2022f).	*Proyek vaksinasi di seluruh dunia itu gagal total. Vaksin COVID tidak mencegah penularan, sakit dan mati, dibuktikan oleh fakta dan data. Jadi, pemerintah Indonesia juga harus jujur, tidak buang2 APBN dan hentikan vaksinasi. Bahaya vaksin COVID itu fatal @KemenkesRI* (Twitter.com, 2022c)
Hore tahun ini bisa mudik lagi. Vaksin 2 dan booster bebas tes PCR & ANTIGEN. Vaksin 2 dosis tapi belum booster wajib tes ANTIGEN. Vaksin 1 dosis wajib tes PCR. Bagi masyarakat yang ingin melakukan mudik lebaran dipersilahkan, dengan syarat sudah 2 kali vaksin dan 1 kali booster (Twitter.com, 2022a).	*Epidemiolog dari Griffith University Australia Dicky Budiman mengatakan, kebijakan pemerintah melonggarkan penggunaan masker masih kurang tepat. Pasalnya, situasi pandemi virus corona di Indonesia belum cukup aman* (Twitter.com, 2022e).
Ketua satgas penanganan COVID mendorong agar masyarakat segera melakukan vaksin booster. Hal ini harus dibarengi dengan disiplin prokes yang ketat. Karena vaksin booster dan prokes saling melengkapi (Twitter.com, 2022d).	*PPKM Diperpanjang lagi, Sidoarjo dan Gresik Gagal ke Level 1* (Twitter.com, 2022b).

Sentiment classification results are further analyzed with topic modeling out to find out how public opinion is related to government policies in dealing with pandemics using LDA, 5 topics for positive sentiments and 3 topics for negative sentiments was obtained.

For topics with positive sentiment, discuss the performance of the president, vaccine policy, COVID recovery rates, COVID handling, and the COVID task force. The public sees the government's performance as being very serious in dealing with COVID, starting with the formation of the COVID Task Force and strict health protocols that resulted in a high recovery rate. The public can also go mudik or return to homecoming, every Eid al-Fitr with mandatory vaccine regulations. Based on the concept of EDM (Chatterjee & Suy, 2019), especially Link E, "It shows performance positively connected with satisfaction. It indicates that performance is a preliminary element which can influence the satisfaction beyond varying expectation level of individual". For topics with negative sentiment, discuss the PPKM policy, against vaccine, and vaccine business. Public feel that the PPKM policy makes it difficult for them to earn a living and meet their daily needs. The public also believes that the vaccine policy is the work of unscrupulous individuals who use it as a business. To overcome this problem, the government must make PPKM policies that are not detrimental to both employees and business owners, for example work-from-home work rules, and assistance for business owners affected by Covid. The government also needs to be more transparent in vaccine policy so that people don't think that this vaccine is a business game.

Our findings have some implications on how the government might implement future COVID

prevention measures. The government should provide easy access for the citizenry to health facilities such as vaccines, antigens, and PCR. The government must also collaborate with social influencers to provide socialization regarding the importance of vaccines, explaining that this vaccine is important to prevent the transmission of COVID, and not for business purposes. The government should also make PPKM policies that have less impact on both business owners and employees.

The theoretical implication of this research is to conduct a sentiment analysis on citizen satisfaction with government policies in handling COVID in Indonesia using data mining with the Support Vector Machine and SMOTE methods using 5-fold cross validation, which has never been done in previous research. The topic modeling in this study uses the Latent Dirichlet Allocation (LDA) method for positive and negative sentiments. The results of the topic modeling are then visualized in the form of word distribution and word clouds, making it easier to see the issues that are most frequently discussed by the public through Twitter. This analysis was conducted to answer problems in the survey method that still have weaknesses, such as: bias from researchers, the hello effect, depending on people's responses, samples do not represent the entire population, lack of confidentiality, allowing for errors, and subjectivity (Hayati, 2022).

Suggestions for further research include modeling with a more specific scope, for example in certain areas so that the results can also be more specific.

CONCLUSION

The COVID pandemic has presented significant difficulties for governments worldwide, particularly in Indonesia. The imminent threat to public health at the onset of the health crisis led the government to impose quick and tough policies, but there will be always pros and cons among the public regarding these policies. Thus, this study uses sentiment analysis and topic modeling to find out how satisfied citizens are with their government's COVID response policies. According to the findings of this study, TF-IDF model with gives better performance than TP and BoW as a whole. This model produces accuracy 0.89, precision 0.87, recall 0.95, and F-score 0.93. Positive and negative sentiments have a significant difference, namely 75% positive sentiment with 4,179 tweets and 25% negative sentiment with 1,811 tweets. In other words, citizen satisfaction with the goverment regalding policies in dealing with pandemics still dominate by positive sentiment, although there are still those who give negative sentiment. Our findings indicate that people are more likely to be satisfied with the performance of the government's fight against COVID than with the policies they introduce. People are also satisfied because they can feel mudik again after two years of the pandemic. Dissatisfaction comes from people who feel that there is a business game in vaccine policy as well as the government's lack of transparency regarding the number of COVID cases. Suggestions for further research include modeling with a more specific scope, for example in certain areas so that the results can also be more specific.

REFERENCES

Ihsanuddin, “Fakta Lengkap Kasus Pertama Virus Corona di Indonesia,” Mar. 03, 2020. https://nasional.kompas.com/read/2020/03/03/06314981/fakta-lengkap-kasus-pertama-virus-corona-di-indonesia?page=all (accessed May 01, 2022).

Satgas COVID, “Tanya Jawab,” 2020. https://covid19.go.id/tanya-jawab?search=Apa%20yang%20dimaksud%20dengan%20pandemi (accessed May 01, 2022).

D. Kurniawan, “HEADLINE: Wisma Atlet Kemayoran Jadi RS Darurat Penanganan Corona COVID, Bagaimana dengan Daerah?” Mar. 25, 2020. https://www.liputan6.com/news/read/4209977/headline-wisma-atlet-kemayoran-jadi-rs-darurat-penanganan-corona-COVID-bagaimana-dengan-daerah (accessed Jun. 10, 2022).

Kemenkes RI, “Tenaga Kesehatan Tangani COVID Dapat Insentif dan Santunan Kematian,” Apr. 29, 2020. https://www.kemkes.go.id/article/view/20043000001/tenaga-kesehatan-tangani-COVID-dapat-insentif-dan-santunan-kematian.html (accessed Jun. 10, 2022).

A. C. Putri, “Sri Mulyani Alihkan Uang APBN Rp 62 T untuk Perangi Corona,” Mar. 20, 2020. https://www.cnbcindonesia.com/news/20200320194857-4-146602/sri-mulyani-alihkan-uang-apbn-rp-62-t-untuk-perangi-corona (accessed Jun. 10, 2022).

M. Wareza, "RI Impor Ribuan Alat Tes Corona China, 15 Menit Hasil Keluar!" Mar. 18, 2020. https://www.cnbcindonesia.com/news/20200318135438-4-145808/ri-impor-ribuan-alat-tes-corona-china-15-menit-hasil-keluar (accessed Jun. 10, 2022).

Kompas.com, "Ada 7 Bantuan Pemerintah Selama Pandemi COVID, Berikut Rinciannya," Aug. 26, 2020. https://nasional.kompas.com/read/2020/08/26/09222471/ada-7-bantuan-pemerintah-selama-pandemi-COVID-berikut-rinciannya?page=all (accessed Jun. 10, 2022).

D. Nurita, "Istana Jelaskan 3 Pertimbangan Jokowi Terapkan PSBB," Apr. 01, 2020. https://nasional.tempo.co/read/1326530/istana-jelaskan-3-pertimbangan-jokowi-terapkan-psbb (accessed Jun. 10, 2022).

F. M. Ramadhan, "Pemerintah Terapkan PPKM pada 11 Januari 2021 Demi Meredam Kasus Aktif COVID," Jan. 08, 2021. https://grafis.tempo.co/read/2465/pemerintah-terapkan-ppkm-pada-11-januari-2021-demi-meredam-kasus-aktif-COVID (accessed Jun. 10, 2022).

Kompas.com, "Pemerintah Tegaskan Masyarakat Wajib Divaksinasi COVID," Jan. 08, 2021. https://nasional.kompas.com/read/2021/01/08/13411591/pemerintah-tegaskan-masyarakat-wajib-divaksinasi-COVID?page=all (accessed Jun. 10, 2022).

K. N. Azizah, "Diapresiasi Dunia, RI Disebut Jadi Salah Satu Negara Terbaik Tangani COVID," Sep. 19, 2021. https://health.detik.com/berita-detikhealth/d-5730142/diapresiasi-dunia-ri-disebut-jadi-salah-satu-negara-terbaik-tangani-COVID (accessed May 01, 2022).

Twitter.com, "About." https://about.twitter.com/en (accessed May 01, 2022).

D. H. Jayani, "10 Media Sosial yang Paling Sering Digunakan di Indonesia," Feb. 06, 2020. https://databoks.katadata.co.id/datapublish/2020/02/26/10-media-sosial-yang-paling-sering-digunakan-di-indonesia (accessed May 01, 2022).

Hayati, "14 Kelebihan dan Kekurangan Penelitian Deskriptif," Mar. 21, 2022. https://penelitianilmiah.com/kelebihan-dan-kekurangan-penelitian-deskriptif/ (accessed Jun. 10, 2022).

N. Eldha Oktaviana and Y. Arum Sari, "ANALISIS SENTIMEN TERHADAP KEBIJAKAN KULIAH DARING SELAMA PANDEMI MENGGUNAKAN PENDEKATAN LEXICON BASED FEATURES DAN SUPPORT VECTOR MACHINE," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 357–362, 2022, doi: 10.25126/jtiik.202295625.

A. Sulistyono, S. Mulyani, E. H. Yossy, and R. Khalida, "Sentiment Analysis on Social Media (Twitter) about Vaccine-19 Using Support Vector Machine Algorithm," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 62–67. doi: 10.1109/ISRITI54043.2021.9702775.

S. H. Sahir, R. S. Ayu Ramadhana, M. F. Romadhon Marpaung, S. R. Munthe, and R. Watrianthos, "Online learning sentiment analysis during the COVID Indonesia pandemic using twitter data," *IOP Conf Ser Mater Sci Eng*, vol. 1156, no. 1, p. 012011, Jun. 2021, doi: 10.1088/1757-899x/1156/1/012011.

D. Bayu Nugroho, M. Ar Rochmah, F. Khasanah, F. Fatmawati, and A. R. Sena, "The New Normal in Indonesia: A Twitter Based Social Media Analysis," *BIO Web Conf*, vol. 41, p. 04004, 2021, doi: 10.1051/bioconf/20214104004.

B. Sujiwo, A. Wibowo, and D. R. S. Saputro, "Sentiment Analysis of Indonesian Government Policies in Handling Covid 19 Through Twitter Data," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 453–457. doi: 10.1109/ISRITI54043.2021.9702784.

C. M. Annur, "Survei Reuters: Twitter Banyak Digunakan untuk Mencari Berita," Mar. 28, 2022. https://databoks.katadata.co.id/datapublish/2022/03/28/survei-reuters-twitter-banyak-digunakan-untuk-mencari-berita (accessed Mar. 01, 2023).

A. R. Pathak, M. Pandey, and S. Rautaray, "Topic-level sentiment analysis of social media data using deep learning," *Appl Soft Comput*, vol. 108, p. 107440, Sep. 2021, doi: 10.1016/j.asoc.2021.107440.

V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, "A deep learning approach in predicting products' sentiment ratings: a comparative analysis," *J Supercomput*, vol. 78, no. 5, pp. 7206–7226, Apr. 2022, doi: 10.1007/s11227-021-04169-6.

M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl Based Syst*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.

P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review," *ICST Transactions on Scalable Information Systems*, vol. 0, no. 0, p. 159623, Jul. 2018, doi: 10.4108/eai.13-7-2018.159623.

I. Vayansky and S. A. P. Kumar, "A review of topic modeling methods," *Inf Syst*, vol. 94, p. 101582, Dec. 2020, doi: 10.1016/j.is.2020.101582.

M. Ghodousi, A. A. Alesheikh, B. Saeidian, B. Pradhan, and C. W. Lee, "Evaluating citizen satisfaction and prioritizing their needs based on citizens' complaint data," *Sustainability (Switzerland)*, vol. 11, no. 17, Sep. 2019, doi: 10.3390/su11174595.

A. Ho, "Perceptions of Performance Measurement and the Practice of Performance Reporting by Small Cities," *State and Local Government Review*, vol. 35, no. 3, pp. 161–173, Sep. 2003, doi: 10.1177/0160323X0303500302.

G. G. Van Ryzin, "Expectations, performance, and citizen satisfaction with urban services," *Journal of Policy Analysis and Management*, vol. 23, no. 3, pp. 433–448, 2004, doi: 10.1002/pam.20020.

R. Chatterjee and R. Suy, "An Overview of Citizen Satisfaction with Public Service: Based on the Model of Expectancy Disconfirmation," *Open J Soc Sci*, vol. 07, no. 04, pp. 243–258, 2019, doi: 10.4236/jss.2019.74019.

B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 2nd edition. Cambridge University Press, 2020. doi: 10.1017/9781108639286.

A. Ben-Hur and J. Weston, "A User's Guide to Support Vector Machines," 2010, pp. 223–239. doi: 10.1007/978-1-60327-241-4_13.

A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Syst Appl*, vol. 57, pp. 117–126, Sep. 2016, doi: 10.1016/j.eswa.2016.03.028.

D. M. Blei, "Probabilistic topic models," *Commun ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012, doi: 10.1145/2133806.2133826.

Twitter.com, "Kepuasan masyarakat terhadap kinerja Presiden Jokowi meningkat karena dinilai berhasil tangani pandemi COVID," May 29, 2022. https://twitter.com/TwittKita31/status/1530731709586874368?s=20&t=7R1PvjqjG0rj7uq6_8KK2w (accessed Oct. 01, 2022).

Twitter.com, "Proyek vaksinasi di seluruh dunia itu gagal total," Apr. 28, 2022. https://twitter.com/AbdulAzizlatte/status/1519502417662218241 (accessed Oct. 01, 2022).

Twitter.com, "Hore tahun ini bisa mudik lagi," Apr. 04, 2022. https://twitter.com/CirebonPolresta/status/1510898701623246848?s=20&t=7R1PvjqjG0rj7uq6_8KK2w (accessed Oct. 01, 2022).

Twitter.com, "Epidemiolog dari Griffith University Australia Dicky Budiman," May 18, 2022. https://twitter.com/tribunnewswiki/status/1526741556262162432 (accessed Oct. 01, 2022).

Twitter.com, "Ketua satgas penanganan COVID mendorong agar masyarakat segera melakukan vaksin booster," May 11, 2022. https://twitter.com/Anggita_lung/status/1524191290753122305?s=20&t=7R1PvjqjG0rj7uq6_8KK2w (accessed Oct. 01, 2022).

Twitter.com, "PPKM Diperpanjang Lagi, Sidoarjo dan Gresik Gagal ke Level 1," Apr. 05, 2022. https://twitter.com/jawapos/status/1511345065758797833 (accessed Oct. 01, 2022).