ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 2, pp. 1-9
DOI:10.33168/JLISS.2024.0201

Understanding Public Opinions of Government Measures Against COVID-19 Through Twitter Sentiment Analysis

Gilang Gumelar, Abba Suganda Girsang

Computer Science Department, BINUS Graduate Program –Master of Computer Science, Bina Nusantara University, Jakarta, 11380, Indonesia.

gilang.gumelar@binus.ac.id, agirsang@binus.edu

Abstract. This study performs sentiment analysis on Twitter data to classify user opinions regarding Indonesian government policies during the COVID-19 pandemic. Prior research applying deep learning techniques like long short-term memory (LSTM) networks for sentiment analysis of social media data is limited. A dataset of 1000 tweets was collected using keywords related to government policies on vaccination, public activity restrictions, health protocols, and online learning. The tweets were preprocessed and word embeddings were generated using Word2Vec. An LSTM model was developed for sentiment classification and policy categorization. The results indicate that the combination of skip-gram Word2Vec and LSTM achieves an accuracy of 88% for sentiment analysis, outperforming other methods. Health protocols garnered the most positive sentiments while distance learning garnered the most negative sentiments. The study demonstrates the effectiveness of LSTM networks for sentiment classification of social media data to gain insights into public opinions.

Keywords: Covid-19, Long Short-Term Memory, Natural Language Processing, Reccurent Neural Network, Sentiment Analysis, Word2vec.

1. Introduction

The COVID-19 pandemic struck the world in early 2020, infecting almost every region globally. On March 2, 2020, the Indonesian government declared it a national disaster (Almuttaqi, 2020). By September 15, 2020, Indonesia had recorded 225,000 fatalities, with 161,000 recoveries and 8,965 deaths. The government implemented various measures to combat the pandemic, including large-scale social restrictions (PSBB), penalties for violating the new normal rules, online learning, and restrictions on public activities. However, some individuals still neglected preventive measures like wearing masks and maintaining physical distancing. The introduction of vaccines offered hope, but faced challenges such as social acceptance and vaccine hesitancy.

To gauge public opinions towards the vaccines, researchers aimed to conduct sentiment analysis on Twitter user opinions. Previous studies have used various methods, Ruales (2014) also conducted research comparing the LSTM method with several other methods for sentiment classification of movie reviews. The error rate of the LSTM method was found to be 0.134, which was lower than the error rate of the Recurrent Neural Network (RNN) without LSTM, which was 0.248. Araque (2017) The LSTM method was utilized for conducting sentiment analysis on Spanish tweets. The researcher employed two distinct types of features, namely word embeddings and sentiment lexicon values. The findings demonstrated that the amalgamation of these two features enhanced the performance of sentiment analysis.

Similarly, Kristiyanti (2018) employed SVM and Naïve Bayes for sentiment analysis during West Java's regional elections. The study collected campaign jargon from Twitter accounts associated with the candidates, yielding a 94% accuracy for Naïve Bayes in the campaign of Deddy Mizwar – Dedi Mulyadi. Other studies by Windasari (2017), Rosdiana (2019), and Afshoh (2017) Naïve Bayes to analyze opinions on Gojek, Makassar's government policies, and the cigarette price increase, respectively In this paper, we aim to use long-short term memory (LSTM) for sentiment analysis to categorize opinions into positive and negative sentiments regarding the government's COVID-19 policies. The analysis will encompass vaccination processes, the implementation of public activity restrictions (PPKM), health protocols, online learning, and non-categorized topics. The research endeavors to gain valuable insights into public sentiments and their responses to government measures during the COVID-19 pandemic.

2. Literature Review

Sentiment analysis is a contemporary method employed to analyze user data and gain insights for making informed business decisions. The studies mentioned below have served as the foundation for our research.:

Wang (2016) conducted a study to determine the aspect-level sentiment of sentences. Sentiment polarity of a sentence is not only based on its content but also closely related to the sentence's context. For example, in the sentence "The appetizers are great, but the service is slow," the sentiment polarity in terms of the taste aspect is positive, but it turns negative when referring to the service aspect. The researcher created a Deep Learning model utilizing Long Short Term Memory (LSTM) to capture and understand such subtleties.

Ruales (2014) also conducted research comparing the LSTM method with several other methods for sentiment classification of movie reviews. The error rate of the LSTM method was found to be 0.134, which was lower than the error rate of the Recurrent Neural Network (RNN) without LSTM, which was 0.248. Araque (2017) The LSTM method was utilized for conducting sentiment analysis on Spanish tweets. The researcher employed two distinct types of features, namely word embeddings and sentiment lexicon values. The findings demonstrated that the amalgamation of these two features enhanced the performance of sentiment analysis.

Furthermore, sentiment analysis can be applied to understand opinions regarding the services of online marketplaces, as demonstrated by Kusumawati (2019) in their research. They compared two algorithms, namely Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM), for sentiment analysis on Twitter data related to the services of the online marketplace Tokopedia. The analysis focused on classifying opinions as positive or negative. NBC and SVM are algorithms commonly used for data classification. They differ in their classification approaches, with NBC utilizing text frequency to calculate probabilities in each class, while SVM is more complex, employing hyperplane equations to effectively separate data into multiple classes. The research aimed to compare these two algorithms and found that SVM achieved higher accuracy (83.34%) compared to NBC (75%). The positive opinions frequently mentioned product quality and discounts/sale programs, while negative opinions mainly addressed delivery and payment procedures.

These studies highlight the effectiveness of LSTM-based models in aspect-level sentiment analysis and sentiment classification tasks. They also emphasize the applicability of sentiment analysis in various domains, such as social media and online marketplaces, to gain insights into public opinions and improve decision-making processes.

3. Proposed Method

The proposed system model for this paper can be divided into various stages that can be better understood through phases.

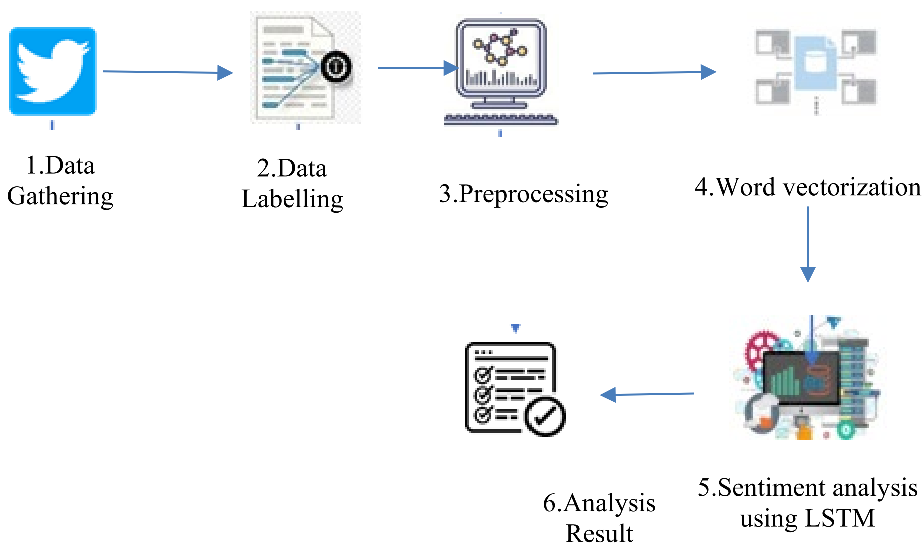

Fig. 1: Proposed Method.

3.1. Data Gathering

the data was collected by scraping Twitter, a popular social media platform. The data consisted of tweets, with each tweet limited to 140 characters. Tweets were obtained through hashtag or keyword searches. The data collection technique utilized the Twitter API provided by Twitter itself, combined with publicly available data from Kaggle, specifically the "Indonesian Tweets COVID-19 Handling (2020)" dataset. The dataset consisted of 500 scraped tweets and 500 publicly available tweets from Kaggle.

a. Tweet – User Tweet Text.

b. Sentiment – “Positive” and “Negative.

c. Category – category in form of “vaksinasi”, “pendidikan daring”, “protokol kesehatan”, “PPKM” and “uncategorized”

The data retrieval process involves extracting information using a set of predetermined keywords in table 1. Each keyword is used to gather approximately 100 tweets, adhering to the specified limit.

This methodology ensures that relevant and contextual data are obtained for further analysis in our research study. By employing these keywords and limiting the tweet count, we aim to capture a diverse range of perspectives and opinions from social media, which will contribute to a comprehensive understanding of the subject under investigation.

Table. 1: Data Gathering Keyword And Hastag.

Keyword and hastag
#lawancovid19
Sekolah tatap muka
#ppkm
#ayovaksin
#dirumahaja
#protokolkesehatan
#semuawajibpakaimasker

3.2. Data Labelling

The second stage is the process where the data that has been scrapped is then labeled so that the model can understand how the contents of the writing are whether the sentiment is positive or negative and also makes a model to find out the category of the writing.

The labeling process was conducted by engaging five expert respondents, who participated in assigning labels, and their consensus was sought for the labels provided.

3.3. Preprocessing

Data Preprocessing: The collected data underwent preprocessing to ensure its quality and usability. This involved tasks such as removing irrelevant symbols or characters, handling missing values, and normalizing text.

1) Case Folding – Convert all text to lowercase.
2) Tokenizing – Cutting documents into small pieces which can be chapters, sub-chapters, paragraphs, sentences, and words (tokens).
3) Stop Word Removal – removing sentences that have no meaning
4) Stemming – reducing a word to its base or root form

3.4. Word Vectorization

Word2Vec Embedding: The Word2Vec algorithm was employed to generate word embeddings from the preprocessed text data. Word2Vec captures semantic relationships between words and represents them in a vector space

3.5. Sentiment Analysis Using LSTM

Sentiment analysis phase: The dataset will undergo division into two parts: training data and testing data. The training data will make up 90% of the dataset, while the remaining 10% will be designated for testing.. The RNN model will be utilized to obtain sentiment analysis results, classifying tweets as either positive or negative, and to perform tweet categorization.The tweet categorization process will be followed by the sentiment classification step using the RNN model. RNN, specifically Long Short-Term Memory (LSTM), will be employed to predict sentiments based on the tweets. LSTM consists of four main components: input, recurrent connections, forget gate, and output. By employing this approach, we aim to utilize RNN with LSTM to perform sentiment analysis and categorization of tweets. This will enable us to determine the sentiment of tweets (positive or negative) and categorize them based on specific aspects related to government policies.

Table. 2: Hyperparameter Table.

Code	Value	Description
model.add(LSTM(100)	100	neuron
model.add(Dense(10, activation='softmax'))	softmax	Activation function
model.compile(loss='categorical_crossentropy')	Categorical_crossentropy	Loss function
model.compile(optimizer='adam')	adam	optimizer
model.fit(epochs=50)	50	epoch
model.fit(batch_size=100)	100	Batch size

3.6. Analysis Result

Analysis Result: This phase involves assessing the performance of all implemented models and obtaining the results for our analysis. For more details, refer to section 4.

4. Analysis Results

4.1. Vectorization

In the process of implementing Word2Vec, the Gensim library provided by Python utilized. This research aimed to compare the performance of two methods of word2vec such as Continous bag of words and skip-gram, by evaluating their ability to check word similarity. To compare the accuracy of similarity, the trained models were called upon to measure the similarity accuracy between skip-gram and CBOW. The comparison focused on the words "vaksin" (vaccine) and "prokes" (health protocols), which are expected to have a high similarity.

Based on the results obtained, it was found that skip-gram outperformed CBOW in terms of accuracy. Despite the small training dataset size, skip-gram achieved a higher accuracy of 0.98, while CBOW scored 0.51. Overall, the results indicated that skip-gram performed better than CBOW in this study, particularly when working with a limited training dataset.

4.2. Classification and Categorization

To automate sentiment prediction and classify it as positive or negative, LSTM modeling is necessary. In this research, sentiment classification will be performed using the LSTM, which is suitable for binary and non classification tasks. The input layer will take the preprocessed and vectorized data obtained through either the Skip-gram or CBOW method. In this study, the sigmoid activation function and the

ADAM optimizer (Adaptive Moment Estimation) were utilized. The ADAM optimizer optimizes the network weights and minimizes the loss during the training process, while the sigmoid activation function generates output within the range of 0 and 1.

To evaluate the performance of the constructed model, 100 test data points were randomly selected from Twitter based on the keywords used to collect the training data. The capabilities of each vectorization method combined with LSTM were compared in accomplishing the sentiment classification task. Based on the processed data from the entire test dataset, it was found that 76% of Twitter users' positive sentiments towards the measures implemented by the Indonesian government to combat the COVID-19 pandemic, while 24% expressed negative sentiments towards those policies. These findings can serve as a benchmark for the government in formulating policies if a future pandemic were to occur.

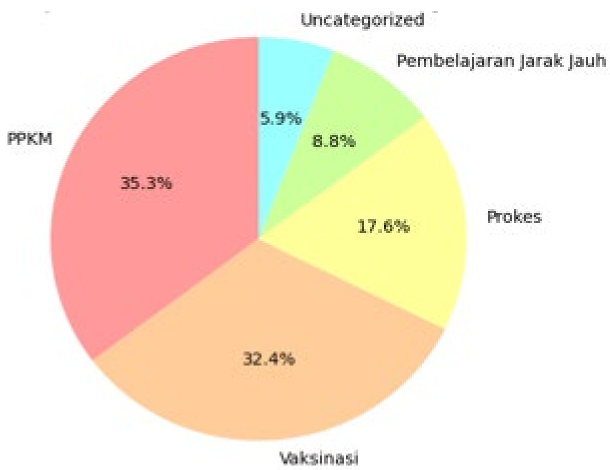

Fig. 2: Chart From Positive Sentiment

4.3. Evaluation and Result Discussion

The The accuracy rates for various factors, namely PPKM, Vaccination, Prokes (Health Protocols), Distance Learning, and Uncategorized, were calculated. PPKM achieved an accuracy of 86.67%, Vaccination had 76.32% accuracy, Prokes achieved 88% accuracy, Distance Learning had 85% accuracy, and Uncategorized had 60% accuracy. Precision values were also calculated, with PPKM achieving 85.71% precision, Vaccination at 73.33%, Prokes at 85.71%, Distance Learning at 60%, and Uncategorized at 40%. Recall or sensitivity values were determined as well, with PPKM, Prokes, and Distance Learning achieving 85.71% recall, Vaccination at 68.75%, and Uncategorized at 66.67%. The F1 scores were also calculated, with PPKM, Prokes, and Distance Learning achieving an F1 score of 85.71%, Vaccination at 70.95%, and Uncategorized at 50%. These results provide a comprehensive evaluation.

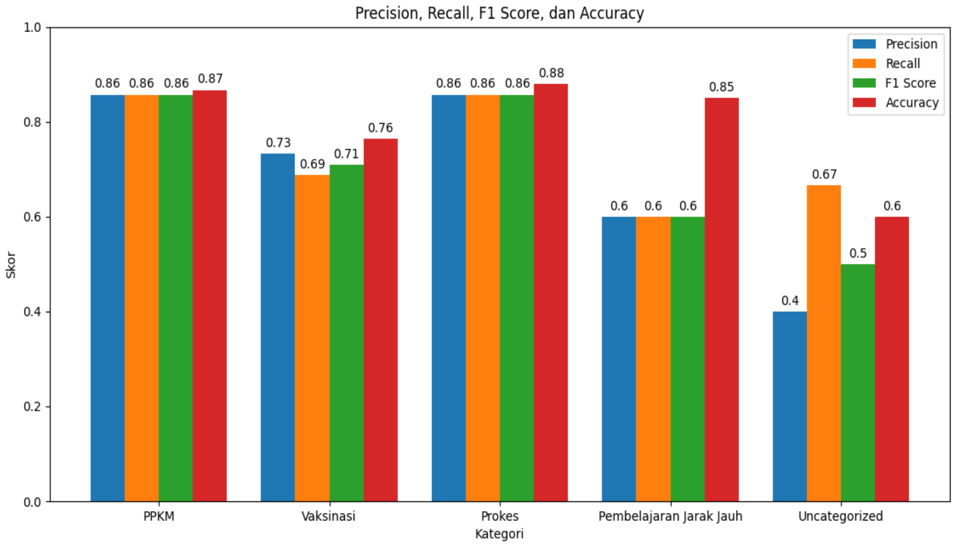

Fig. 3: Evaluation Result.

Practicality This research utilized word2vec embedding and LSTM models for sentiment analysis and categorization tasks. The findings provide valuable insights into Drawing from the outcomes, this refers to the sentiments expressed by Twitter users regarding government policies amid the COVID-19 pandemic, several recommendations for further research and areas of improvement can be suggested. Firstly, in future research, the possibility of investigating the application of advanced embedding techniques like BERT (Bidirectional Encoder Representations from Transformers) could be considered. BERT has shown remarkable performance in various natural language processing tasks and could potentially enhance the accuracy and contextual understanding of the sentiment analysis.

Moreover, considering the rapid advancements in deep learning, researchers can investigate other deep learning architectures beyond LSTM. Models such as Transformer-based architectures (e.g., GPT, BERT) or even hybrid models combining LSTM with attention mechanisms may yield more accurate results. Additionally, expanding the training dataset could lead to improved accuracy.

Increasing the volume and diversity of the training data can help the model learn more comprehensive patterns and generalize better to real-world sentiments. This could be achieved by collecting data from various sources and incorporating more recent data to capture the evolving sentiment trends. Furthermore, it is crucial to consider the limitations and biases in the data collection process. Twitter data is inherently noisy and subject to bias due to user demographics and the nature of the platform. Researchers should address these limitations by incorporating data from multiple platforms, incorporating user demographics, and employing techniques to mitigate bias.

Overall, this research provides a foundation for sentiment analysis and categorization of Twitter data related to government policies during the pandemic. By incorporating the suggested recommendations, future studies can further enhance the accuracy, robustness, and applicability of sentiment analysis models, providing more nuanced insights into public sentiments and aiding policymakers in making informed decisions.

5. Conclusion

Based on this research, several conclusions can be drawn: Firstly, this study presents the use of word2vec embedding and LSTM for sentiment analysis and categorization tasks. The combination of these techniques proved to be effective in analyzing and categorizing text data. Secondly, the results indicate that the skip-gram method outperformed CBOW in measuring word similarity or contextual meaning between words. The skip-gram method achieved a similarity accuracy of 0.98, while CBOW scored 0.51. Furthermore, in terms of sentiment analysis with binary output, the skip-gram embedding method yielded better results compared to CBOW. Additionally, the analysis of Twitter users' sentiments revealed that the majority of users expressed negative sentiments towards distance learning policies, while the process of vaccination received the highest number of positive sentiments. Lastly, the data obtained from this research can provide valuable insights into Twitter users' comments on various government policies during the COVID-19 pandemic. This data can serve as a benchmark for the government in improving their policies to better align with public sentiments. In conclusion, the utilization of word2vec embedding and LSTM, along with the analysis of sentiment and categorization, provides a comprehensive understanding of user opinions and can help inform policy-making decisions during times of crisis such as the COVID-19 pandemic.

By incorporating the suggested recommendations, future studies can further enhance the accuracy, robustness, and applicability of sentiment analysis models, providing more nuanced insights into public sentiments and aiding policymakers in making informed decisions

Based on the findings of this research, several recommendations for further research development can be outlined:

1) Utilization of more advanced embedding methods, such as BERT, could be considered.
2) Exploration of alternative deep learning approaches is advisable.
3) Increasing the size of the training dataset can contribute to enhancing accuracy.

These suggestions have the potential to advance the field and refine the outcomes of subsequent research endeavors.

References

Afshof, F. (2017). *Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter*. 1–17.

Almuttaqi, A. I. (2020). Kekacauan Respons Terhadap COVID-19 di Indonesia. *Thc Insigjts*, *1*(13), 1–7.

Fahmi, A., Ramadhan, I., Studi, P., Informasi, S., & Komputer, F. I. (2020). *Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi Covid-19*. *1*(1), 608– 617.

Irfan, M. R., Fauzi, M. A., Tibyani, T., & Mentari, N. D. (2018). Twitter Sentiment Analysis on 2013 Curriculum Using Ensemble Features and K-Nearest Neighbor. *InternationalJournal of Electrical and Computer Engineering (IJECE)*, *8*(6), 5409. https://doi.org/10.11591/ijece.v8i6.pp5409-5414

Kristiyanti, D. A., Umam, A. H., Wahyudi, M., Amin, R., & Marlinda, L. (2019). Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter. *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, *Citsm*, 1–6. https://doi.org/10.1109/CITSM.2018.8674352

Kusumawati, R., D‟Arofah, A., & Pramana, P. A. (2019). Comparison Performance of Naive Bayes Classifier and SupportVector Machine Algorithm for Twitter‟s Classification of Tokopedia Services. *Journal of Physics: Conference Series*, *1320*(1). https://doi.org/10.1088/1742-6596/1320/1/012016

Pratama, M. O., Satyawan, W., Jannati, R., Pamungkas, B., Raspiani, Syahputra, M. E., & Neforawati, I. (2019). The sentiment analysis of Indonesia commuter line using machine learningbased on twitter data. *Journal of Physics: Conference Series*,

Romadhan, A. (2018). Word2Vec. [online] Tersedia di:

https://medium.com/@arifromadhan19/word2vec-95c5df46e045 [Diakses 19 Juni 2023].

Rosdiana, R., Eddy, T., Zawiyah, S., & Muhammad, N. Y. U. (2019).*Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar*. 87–93.

Rossi, A., Lestari, T., Setya Perdana, R., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. *Jurnal Pengembangan TeknologiInformasi Dan Ilmu Komputer*, *1*(12), 1718–1724. http://j- ptiik.ub.ac.id

Ruales, J. (2017). Recurrent Neural Networks for Sentiment Analysis. *In SemEval*, *2014*,573–580. http://arxiv.org/abs/1704.06125

Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik RumahTangga. *Creative Information Technology Journal*, *2*(3), 207– 217.

Skymind. (2017). *A Beginner's Guide to Recurrent Networks and LSTMs*. Diambil kembali dari Deep Learning for Java: https://deeplearning4j.org/lstm.html

Sutskever, I. (2013). *Training Recurrent Neural Networks.* Toronto.

Tanulia, Y., & Girsang, A. S. (2019). Sentiment analysis on twitterfor predicting stock exchange movement. *Advances in Science, Technology and Engineering Systems*, *4*(3), 244–250. https://doi.org/10.25046/aj040332

Wang, Chen-Kai, Singh, O., Tang, Z.-L., & Dai, H.-J. (2017). Usinga Recurrent Neural Network Model for Classification of Tweets Conveyed Influenza-related Information. *Afnlp*, *2017*, 33–38. https://nlp.stanford.edu/projects/glove/

Windasari, I. P., Uzzi, F. N., & Satoto, K. I. (2017). Sentimentanalysis on Twitter posts: An analysis of positive or negative opinion on GoJek. *2018-January*, 266–269. https://doi.org/10.1109/ICITACEE.2017.8257715