ITB Graduate School Conference (IGSC) ISSN: 2963-718X
December 21, 2021

Sentiment Analysis against the Community Activities Restrictions Enforcement (PPKM) Data using ANN (Artificial Neural Networks)

Andri Fauzan Adziima

Bandung, Indonesia
Email: 20920302@mahasiswa.itb.ac.id

Abstract. Enforcement of Community Activities Restrictions Enforcement (PPKM in Indonesian) was announced and enforced in July 2021 and has had a significant impact on many communities in Jakarta, with social networking media is becoming popular as a powerful tool for influencing people and sharing their views with the general public, to understand the emotions of the public in social media sentiment analysis is carried out using deep learning (ANN) with tweets dataset related to the enforcement of community activity restrictions (PPKM) in Indonesia. The results of this study show that while the model is being trained, the model tends to overfit after achieving the highest accuracy at a particular epoch. Comparing the three experiments, using features with TFIDF values above the median makes a small contribution to improving the model, while using ngram doubles the number of important vocabularies and other vocabularies. An important vocabulary for producing more influential features that can be combined with, which can significantly improve the accuracy of the model.

Keywords: *ppkm; sentiment analysis; tfidf; ann (artificial neural networks); deep learning.*

1 Introduction

Enforcement of Community Activities Restrictions Enforcement (PPKM in Indonesian) was announced and enforced in July 2021 and has had a significant impact on many communities in Jakarta. Many offices are reintroducing telecommuting for their employees, and many shops and businesses are temporarily closed. Because of this, many people were at home. With PSBB, the use of social media continues to grow, and Internet networks are now being used by a wider community to freely communicate their needs and opinions, including the use of social media [1].

Today, social networking media is becoming popular as a powerful tool for influencing people and sharing their views with the general public. For example, Twitter is one of the social media channels that Indonesian politicians often use for political campaigns [2]. Of the several social networking sites, Twitter is one of the most popular tools used by politicians [3]. Twitter has been used by many Indonesian politicians to evoke public sympathy and increase popularity before the election [4]. Politicians use Twitter to post their opinions, thoughts, and activities to attract and influence those who vote in parliamentary elections. In addition, Twitter was used as a data source to analyze public sentiment in elections in some countries such as Indonesia, India, Germany, the United States, the United Kingdom and Bulgaria [1].

Recently, deep learning has become one of the most powerful learning methods to quickly learn about the characteristics of data and provide up-to-date prediction results [5].

The main concept of the deep lean algorithm is to automatically extract information from the data. Deep learning algorithms use large amounts of unstructured data to automatically extract information from complex unstructured data. These deep learning algorithms are primarily naturally inspired and have the common goal of mimicking the activity of the human brain. The human brain is a great processor. How it works exactly is still a mystery. The most basic element of the human brain is the neuron. Deep learning algorithms form an abstract representation. This is because many abstract expressions have been constructed that accumulate non-abstract expressions. The main advantage of abstract representations is that they can be invariant to local changes in the input data. Learning these invariant features is the main goal of pattern recognition. Deep learning algorithms are actually deep architectures that consist of many consecutive layers. In this case, a non-linear transformation is applied to the input to provide a representation of the output.

In this study, sentiment analysis will be carried out using deep learning (ANN) with tweets dataset related to the enforcement of community activity restrictions (PPKM) in Indonesia.

2 Related Works

2.1 Sentiment analysis based on dictionary approach

Bhagwat et.al [6] conducted a sentiment analysis survey using a dictionary approach in the context of online reviews. Whenever a consumer makes an online purchase, they make decisions based on the reviews of other consumers. Every rating (good or bad) gives the consumer an impression of the product and helps them decide whether to buy or learn. Therefore, this type of review-tweet is preprocessed into more structured information. Tokenization is used to identify words in the text. Text pre-processing is followed by tweet analysis. This analysis step is considered to be the core of text mining. This is because it extracts some useful and important knowledge from the text. Emotion classification performance is accurately assessed with this SVM (hybrid method).

2.2 Sentiment analysis on social media

Chopra et al. [7] conducted sentiment analysis on social media. This paper proposes a system that uses Facebook as a social media site to collect public opinion through Facebook posts. 1000 Facebook posts from La7 and Rail news broadcasts from Italy will be collected. KMean's clustering algorithm is used to cluster news items from La7 and Rail. Includes semantic and linguistic approaches to improve results. Analyze each sentence, assign the appropriate meaning to the context, and get accurate results. Then preprocess the post by removing the various parts of speech. nouns, verbs, adverbs and adjectives to reduce complexity. Sentiment analysis is performed by polar mining and syntactic trees. The Bayesian learning method is used for the classification process. The recall and precision are used as metrics to calculate the performance of the proposed system.

2.3 Public opinion sentiment on the effects of PSBB

Syarifuddin et al. [8] conducted a survey related to public opinion analysis of the impact of PSBB on Twitter using Decision Tree, KNN, and Naïve Bayes Algorithm. This is a collection of public opinions' posts or comments from Twitter users regarding the impact of PSBB. Data on the impact of PSBB received up to 170 opinions and was processed using a data mining technique with a text mining process, tokenization, transformation, classification, and truncation. It is then converted into three different algorithms to compare. The algorithms used are decision trees, k-nearest neighbors (KNN), and naive Bayes classifiers, with the goal of finding the highest accuracy of these classifiers. The best result of this study is a decision tree algorithm with an accuracy value of 83.3%, an accuracy of 79%, and a recall of 87.17%.

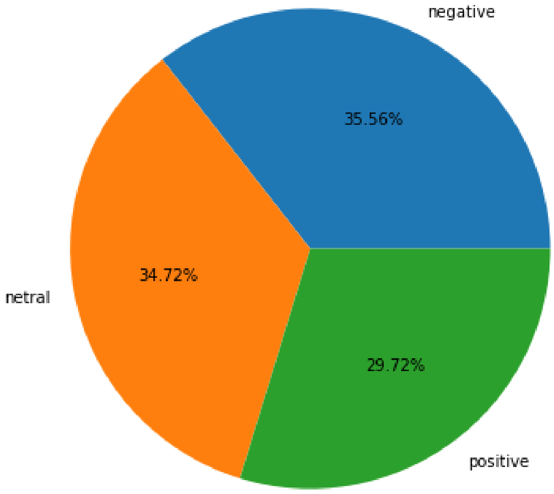
3 Methods

3.1 Dataset

In this study we use secondary data obtained from twitter API, the tweets contain the word "PPKM" are retrieved and saved in csv format. 740 tweets are obtained in October of 2021 and after manually check the relevance of the tweet and the context only 360 tweets are selected.

3.2 Labelling Dataset

At this stage, the data is labeled automatically using VADER (Valence Aware Dictionary for Sentiment Reasoning) which is a model used for text sentiment analysis that is sensitive to each polarity.



VADER Sentiment Analysis of Tweets Dataset The result of this labeling (Fig. 1) generates data with negative sentiment 35.5%, neutral sentiment 34.7%, and positive sentiment 29.7%.

Table 1 VADER Sentiment Analysis Example

No	word_ind	word_en	Polarity
1	ppkm dan vaksinasi dapat menurunkan kasus covid	PPKM and vaccination can reduce covid cases	Positive
2	polres lakukan edukasi warga terkait ppkm level satu	Police conduct public education regarding PPKM level one	Netral
3	ppkm bikin mood ancur bikin frustrasi gimana cara dapet uang gimana cara biar bisa bayar ini bayar itu	PPKM makes the mood break, it makes you frustrated, how do you get money, how do you pay for this, pay for that	Negative

As we can see from the example of the sentence in Table I. to predict the polarity of the sentences VADER needs input in English, therefore the data was translated from Indonesia to English using TextBlob which is one of the libraries in python that is able to automatically translate bulk of sentences. Also based on the sentiment analysis generated by VADER we can conclude that the result is quite good and represent the actual sentiment, because VADER is a dictionary-based method that use bag of words to define the polarity of a sentence.

3.3 Tweets Preprocessing

Data cleaning or pre-processing is the process of finding and cleansing the data by replacing, modifying, or deleting the dirty or coarse data [7]. The data cleaning steps involved as follows:

- Digits removal
- Emojis removal
- Links removal
- Symbols removal
- Stop-words removal
- Prepositions removal
- Punctuations removal
- Retweet removal
- Clean hash tags
- Removal of screen name
- Convert tweets to lower case
- Stemming and language normalization
- Tokenization

3.4 TF-IDF

Term Frequency-Inverse Document Frequency (TFIDF) is a method of calculating the weight of each extracted word. This technique is commonly used to count common words in information retrieval. The TFIDF weighting model is a method of integrating the term frequency (tf) model and the inverse document frequency (idf) model. term frequency (tf) is a way to count the frequency of occurrence of terms in a document, and inverse document frequency (idf) is used to calculate terms that appear in various documents (comments) which are considered general terms, and are considered not important [9].

The step of weighting with TF-IDF are:

1. Count term frequency (tf)
2. Count weighting term frequency (W_{tf})

$$Wtf_{t,d} = \begin{cases} 1 + log10tf_{t,d}, \ if \ tf_{t,d} > 0 \\ 0 \end{cases} \quad (1)$$

3. Count document frequency (df)
4. Count the weight of inverse document frequency (idf)

$$idf_t = Log\frac{N}{dft} \quad (2)$$

5. Count the weight of TF-IDF

$$W_{t,d} = Wtf_{t,d} \times idf_t \quad (3)$$

Notes:
$tf_{t,d}$ = term frequency
$Wtf_{t,d}$ = weight of term frequency
df = the number of times the document contains a term
N = the total number of documents
$W_{t,d}$ = weight of TF-IDF

3.5 ANN (Artificial Neural Networks)

Once the automatic labeling is done using VADER (Dictionary Based) approach, sentiment analysis is performed using ANN. The tweet stored in the CSV file are retrieved, then the train and test data are split in the ratio 80:20. The configuration of the ANN as follows:

- Criterion: NLLLoss.
- Optimizer: AdamW.
- Dropout: 0.2.
- 2 hidden layers of 64 nodes each layer.
- Activation Relu and Logsoftmax.

4 Result and Discussion

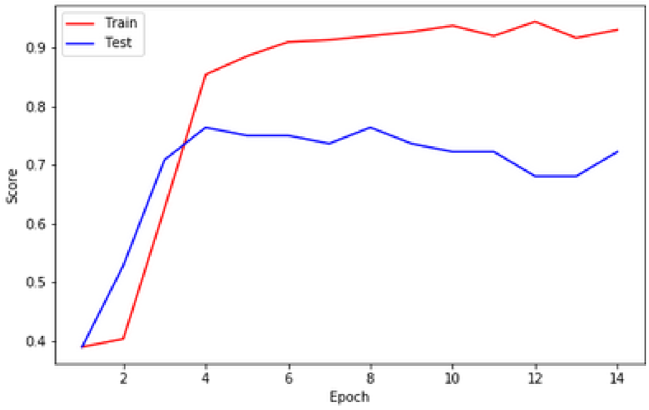
This section presents the results and analysis for the experiments conducted in this research. Three experiments that have been done to find the best model are as follows:

1. Sentiment model using all features for all TF-IDF score and n-gram range = (1,1)

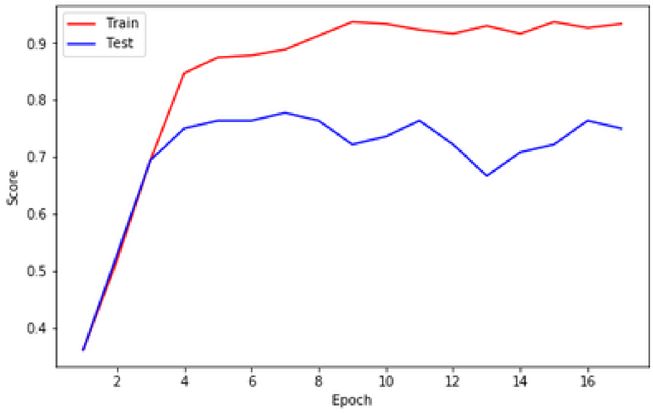
2. Sentiment model using features with TF-IDF score above median and n-gram range = (1,1)

3. Sentiment model using features with TF-IDF score above median and n-gram range = (1,2)

Sentiment Model using All Features for All TF-IDF Score and N-Gram Range = (1,1)



Sentiment Model using Features with TF-IDF Score Above Median and N-Gram Range = (1,1)



Sentiment Model using Features with TF-IDF Score Above Median and N-Gram Range = (1,2)

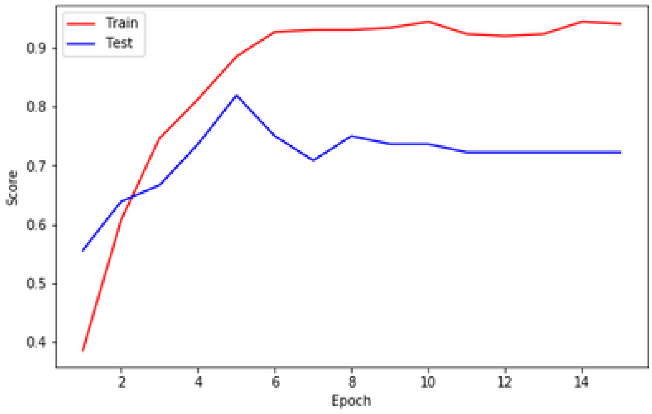


Figure 1 Train-Test Accuracy Score for Sentiment Model using Different Features and N-Grams

Figure 2. shows that the test accuracy for each experiment reaches the best accuracy at epoch 4, 7, and 5 then keep decreasing after that, while the training accuracy keeps increasing which means after reaching the best accuracy the model tends to overfit.

Table 2 Comparison of Each Best Model Result

	all features for all TF-IDF score and n-gram range = (1,1)	features with TF-IDF score above median and n-gram range = (1,1)	features with TF-IDF score above median and n-gram range = (1,2)
Train Loss	0.7446	0.2715	0.4846
Test Loss	0.7255	0.7787	0.6692
Train Accuracy	0.8542	0.8889	0.8854
Test Accuracy	0.7639	0.7778	0.8194

As we can see from Table II. the result of experiments 1 and 2 shows that the use of features with TF-IDF score above the median and removing unnecessary features only slightly improve the model test accuracy even though more than half of the features were removed in this process. Therefore, the features with TF-IDF score below the median does not have many contributions towards the model result.

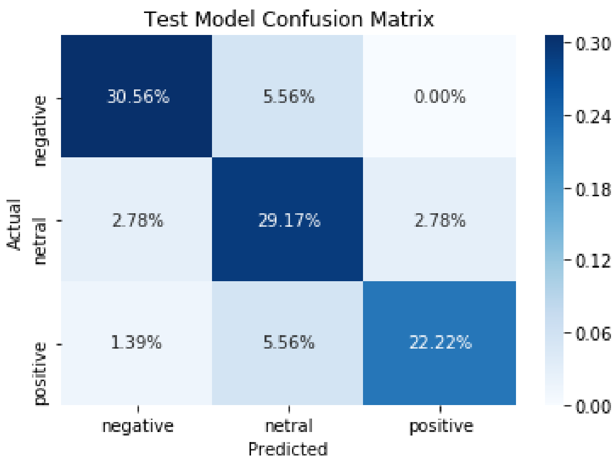


Fig 3. Test Prediction Model Confusion Matrix

Based on the confusion matrix result in Fig. 3, we can see that the model creates error mostly when it is related to neutral sentiment, 11.12% of the prediction failed to predict the negative and positive sentiment and instead predicts it as neutral sentiment and 5.56% of the prediction failed to predict the neutral sentiment, only 1.39% failed to predict positive sentiment and instead predicts it as negative sentiment. The bias in predicting the neutral sentiment is probably caused by the lack of data used in the train model because there is only 288 tweets used as the train and 72 tweets as the test, therefore when a sentence does not have strong positive or strong negative polarity then the model will put the sentence to neutral polarity, this concludes that in this ANN model the neutral polarity has wider polarity range compared with the negative and positive polarity.

On the other hand, the comparison between experiments 2 and 3 shows that the use of the n-gram range can give better improvement to the model accuracy, because the n-gram range is able to double the number of important features and combine it with other important words, so that the model has more important vocabulary which can influence the model result.

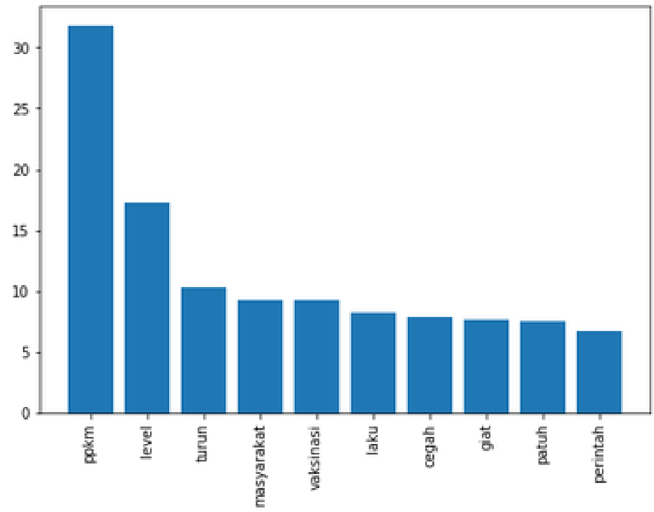


Fig 4. Top 10 Most Important Vocabulary with N-Gram Range = (1,1) in Experiment 2

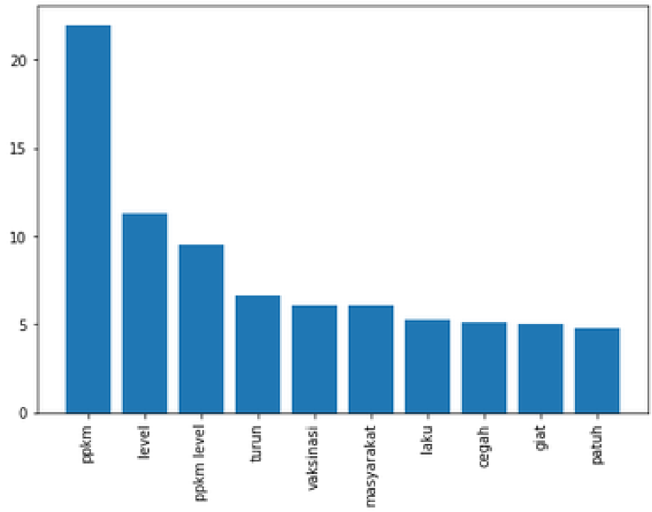


Fig 5. Top 10 Most Important Vocabulary with N-Gram Range = (1,2) in Experiment 3

The comparison between Fig 4. and Fig 5. shows the importance of the n-gram range, in experiment 2 and 3 the most important vocabulary is "ppkm" which mean this word could be the most influential feature in the model and as we can see in Fig 5. by using n-gram range = (1,2) a new feature "ppkm level" is emerging, with the using of n-gram range it is possible to improve the model accuracy because there is more important vocabulary which can influence the model result.

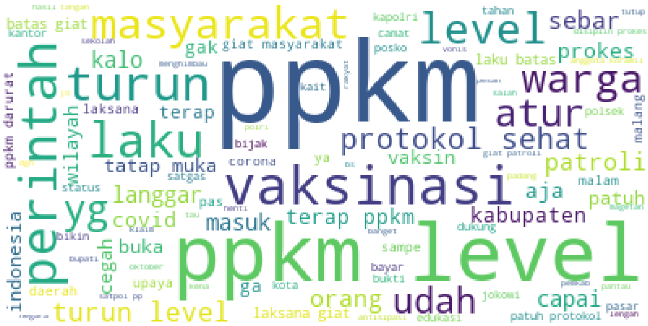


Fig 6. Word Cloud of the Tweets Data

We can conclude that the result of this study shows that the model using N-Gram Range = (1,2) has better results compared with the other methods. In Fig.5 and Fig.6 we can see which vocab has more influence on the model, it makes sense if the word 'PPKM' has more influence compared to others, because that word appears in every data. As we use ANN as the method, we can not know which words are associated with the positive, negative, or neutral sentiment as ANN is a black box machine learning. Therefore, to improve this study we need to add the correlation between each word to the sentiment so that we can analyze the result of the model comprehensively.

5 References

Within the text, references should be cited by giving the last name of the author(s) and numbered consecutively starting with [1], i.e:

"Some results from the experiment were given by Wijaya and Riyanto in [1], Wijaya, *et.al* in [2], Majerski and Przybylo in [3], Nurdin, *et al.* in [4] and [5]."

Note that in the case of three or more authors, only the last name of the first author is cited and the others are denoted by *et al*. **The same rule is also held for the header title on even pages (see Header in top of Page 2).**

Within the Reference chapter, use the same typeface as the body of the text for the references, or just find *Reference* in **Styles Windows**. In References chapter you should write based on the order of appearances, not alphabetically. Example of References:

References

[1] L. Septiani and Y. Sibaroni, "Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia," J. Linguist. Komputasional, 2019, doi: 10.26418/jlk.v2i2.23.
[2] Alashri S, Parriott E, Kandala S S, Awazu Y, Bajaj V and Desouza K C 2018 The 2016 US Presidential Election on Facebook: An Exploratory Analysis of Sentiments Proc. of the 51st Hawaii Int. Conf. on System Sciences pp 1771–80
[3] Bansal B and Srivastava S 2018 On predicting elections with hybrid topic-based sentiment analysis of tweets Procedia Computer Sci. 135 pp 346–53
[4] Hidayatullah A F and Sn A 2014 Analisis sentimen dan klasifikasi kategori terhadap tokoh public pada Twitter Seminar Nasional Informatika 2014 (semnasIF 2014) pp 115–22
[5] Daniel Jurafsky and James H. Martin ―Naive Bayes and Sentiment Classification‖ Published on August 7,2017.
[6] ApekshaPande, Reshma Bhonde, Binita Bhagwat ,et.al, Sentiment Analysis Based on Dictionary Approach, International Journal of

Emerging Engineering Research and Technology Volume 3, Issue 1, January 2015, PP 51-55.

[7] Fatehjeet Kaur Chopra and Rekha Bhatia, Sentiment Analyzing by Dictionary based Approach, International Journal of Computer Applications (0975-8887).

[8] M. Syarifuddin, “Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Naïve Bayes,” vol.15, pp.87–94,2020, doi:0.33480/inti.v15i1.1433

[9] C. D. Manning, P. Raghavan, and H. Schutze, Introduction to Information Retrieval. 2008.