Journal of Advances in Information Systems and Technology **5**(1) ISSN 2714-9714
April 2023, 91-105

Analysis of Public Opinion on the Impact of the Implementation of Community Activity Restrictions (PPKM) During the Covid-19 Pandemic Using Long Short Term Memory and Latent Dirichlet Allocation

Gebyar Bintang Taufikurohman [1]*, Alamsyah [1]

[a] Departement of Computer Science, Faculty of Mathematics and natural Sciences, Semarang State University, Semarang, Indonesia
*Corresponding author: Gtaufik27@students.unnes.ac.id

ARTICLE INFO

Article history
Received 26 March 2023
Revised 7 April 2023
Accepted 17 April 2023

Keywords
Text Mining
Sentiment Analysis
Twitter
Classification

ABSTRACT

Technology social is the fastest and *most up-to-date* source of information. A model that can provide mapping will help in sorting out information more precisely and quickly. Public opinion in the mass media always develops quickly to talk about an issue in just a few days or even hours, so we do not know what the opinions of the people in the mass media are on the issue. In this study, the author applied topic modeling to the results of sentiment analysis on PPKM. The source of data in this study was obtained from twitter using *SNScrape*. The collected data was analyzed sentiment using the *Long Short-term Memory* (LSTM) method, so that public opinion was obtained with positive, negative, and neutral sentiments. The classification obtained from the results of the sentiment analysis process is continued with the topic modeling process using the *Latent Dirichlet Allocation* (LDA) method and visualized in the form of *a wordcloud* to find out the relationship between one topic and another. The sentiment analysis process produces a model with an accuracy rate of 90.8% and the topic modeling process successfully presents topics that are easy to interpret so that conclusions can be known about an issue.

1 Introduction

The COVID-19 pandemic or *Coronavirus disease 2019* has spread to all countries in the world, including Indonesia. In dealing with this pandemic, the Indonesian government has set various policies. This policy aims to reduce the spread of the COVID-19 virus and maintain the stability of various aspects of people's lives affected by the COVID-19 pandemic. Among the fields affected by the COVID-19 pandemic include trade, economy, tourism education, investment and health (Gössling *et al.*, 2020; Taofik, 2021). The confirmed case of COVID-19 reaches 138.416.498 people, 2.975.875 death cases, and 192 countries. Reducing the rate of virus growth is one of the efforts that are continuously being made. The virus with rapid transmission also needs quick overcome. The most used methods against the virus spread are diagnosis, testing, quarantine, and isolation. The faster treatment can be applied. It can be one of the best efforts to reduce the rate of growth (Alamsyah et al., 2021).

One of the policies of the Indonesian government is to suppress the spread of the COVID-19 virus by establishing the implementation of restrictions on community activities (PPKM). The implementation of PPKM is carried out by implementing various multilevel or *level* systems based on the situation and conditions of the region (Mihardja & Harja, 2021).

One of the well-known social media platforms in Indonesian society is Twitter. Twitter has an interaction concept that has a focal point on writing between users that is limited to 140 characters. Within a day, Twitter was able to generate 500 million short messages called tweets (Karami *et al.* , 2018) . Total 500 million tweets generated by its users, there is a variety of information that can be unearthed. One of them is the opinion of the people consisting of various *hatsc*. For example, *hatc* personal life, social, business and evloen covers politics.

Sentiment analysis is an analysis process obtained from various social media *platforms* and the internet. The goal is to find out how people feel about an issue, *capc*, product or policy implementation. One of the *most common capc* sentiment analysis is to focus on the degree of polarity of opinion. This type of sentiment analysis will group responses or opinions into several categories such as positive, neutral and negative (Liu, 2010).

Specific approaches can be looked at and analyzed beforehand, this would be useful for saving resources and a more targeted approach. Predictions that can be made using text mining are basically formed by machine learning (Muslim et al., 2021).

To determine the polarity of such opinions requires further analysis such as the opinions of people, types of sentiments, evaluations, judgments, attitudes, and emotions towards entities of products, services, organizations, individuals, problems, events or a *hatc*. Therefore, we need *machine learning* that aims to detect and classify sentiment. In principle, sentiment analysis is part of *natural language processing (NLP)* and *machine learning* how it works is by classifying positive, negative and neutral words (Astari *et al.*, 2021).

Time series analysis has several purposes, namely forecasting, modeling, and control. Forecasting problems are related to the formation of models and methods that can be used to produce accurate estimates. The difference between modeling and forecasting is that predicting is more likely to be a "black-box" model to obtain forecasts, while modeling tends to be a model that can be interpreted to explain what happens regarding the relationship between variables in time series data. Since then, publications related to time series analysis have grown rapidly. Until 1980, most of the research was focused on linear time series models, especially linear models of the Autoregressive Integrated Moving Average class. developed a complete procedure for the ARIMA modeling methodology which until now has been used as a standard procedure in building linear time series models (Alamsyah & Walid., 2017).

LSTM itself is one of the *deep learning* which is a development of *the Recurrent Neural Network* (RNN). The RNN method uses the calculation of weights repeatedly, so that the accuracy value obtained is better than that of simple artificial neural networks (Naury *et al.* , 2021) . Meanwhile, LSTM is an improvement of *RNN* that can store memory so that it can be selected by adding an *attention* mechanism so that each word fits the context.

After the sentiment analysis classification process, topic modeling was then carried out using *latent dirichlet allocation* (LDA). LDA works by taking into account the relationship of words to words in the set of documents being analyzed (Al-Sultany & Aleqabie, 2018). From this process we can find out what topics are being discussed by the community and how the community tends towards these topics. LDA is one of the topic *modeling* methods that is able to map *the topic* of a document, besides that it also requires labeling to map which *topics* are *positive*, *negative* and *Neutral*. Therefore, this study proposes a classification method using a *long short term memory* (LSTM) algorithm to overcome this. That way it will improve the performance of the clarification results that will be calculated to get the maximum possible results. Based on the above problems, this research will focus on the LDA and LSTM methods, with what works as *a machine* le*arning* is LSTM with the help of the *topic modeling* process with LDA to achieve maximum results.

2 The Proposed Algorithm

2.1 Text Mining

Text mining is one of the inventions in document text by using text mining which aims to simplify the process of searching or extracting data that has important and hidden properties from a collection of text. (Ramanathan & Meyyappan, 2019) Dor text hidden from a document can then be known to be often referred to as implicit text. The knowledge of this implicit text is not written clearly and clearly in a text and it is necessary to have certain human comprehension techniques and computer algorithms to know it. (Thu & New, 2017)

2.2 Twitter

Twitter was launched in July 2006 as a *microblogging-based* social media service that encourages users to express their thoughts and activities in *real time*. As a social media, *Twitter* has a role in the process of exchanging information between its users. As well as making its users interconnected (Bucher, 2013; Azeharie & Kusuma, 2014) .

Twitter users can express their opinions or comments and thoughts in the form of text with a limit of 280 characters. Any expression sent in twitter media is often also *a* tweet or tweet. These tweets can be attached by different types such as images, videos as well as links. *Twitter* is also categorized as a Web 2.0-based application that can provide various online facilities, such as social networks, *online* communities, information production and sharing co-production, content production, and consumption from users

2.3 Implementation of Restrictions on Community Activities

The Covid-19 virus pandemic that spread throughout the world overwhelmed all countries in anticipation of the rapid spread of the Covid-19 virus (Haridison *et al.* , 2021) . The implementation of Community Activity Restrictions (PPKM) is one of the government's efforts to deal with the spread of the Covid-19 virus in Indonesia. The PPKM policy was set by the government to control community mobility to suppress Covid-19. PPKM began to be implemented since 2021 and is focused on deployment points in Java and Bali (Yunia *et al.* , 2021) .

2.4 Data Collection

Data collection is a data collection process used to conduct a study. Research data can be obtained through social media(Corner & Deshpande, 2019) *twitter*. The method of data retrieval or commonly called *crawling* data originating from twitter in *real time* is using the *twitter streaming API* which provides access to data transfer on demand with certain conditions (Gabarron *et al.*, 2019). *Twitter* also has many APIs that have their own uses such as *public streaming* APIs and *REST APIs* that make the process of retrieving data easy. In this article, it uses the twitter API from the *tweepy* program package which is an additional program package in the python language specifically to complete *twitter* data processing. The program package from *tweepy* also supports *twitter* access through *basic authentication* and the latest is the *OAuth method*(Singh *et al.*, 2016).

2.5 Tokenizing

Tokenization is one of the processes at the *text preprocessing* stage that aims to break a document of a text into words, terms, symbols, or some other meaningful element called a token. *Tokenization* has advantages especially in linguistics as well as in computer science, where this process is part of lexical analysis. Generally, the *tokenization* process occurs at the word level. *Toctenization* is also often interpreted as the task of breaking down the sequence of characters so that they become several (words/phrases) called tokens. The *tokenization* stage also allows at the same time to discard certain unimportant characters such as punctuation marks (Rai & Borah, 2021).

2.6 Lemmatization

In research (Ravi *et al.*, 2019) it is stated that an idea that contains non-standard words is often used in communicating and interacting with other people. Words that have non-standard properties are formed a lot as a result of human interaction itself and have a level of degree that is far from the actual use of the word , namely the dictionary of the original language. In a study with the object *of sentiment* analysis, words that have non-standard properties are very influential on the results obtained, there is data analysis. To be able to improve the sentiment quality results from data that has non-standard words, it is necessary to improve or convert non-standard words into standard words. *Lemmatization* is the process of converting non-standard words into the original language. An example of *lemmatization* is from the sentence "Aqu sdng not in the office today", from the sentence found 3 non-standard words, namely "Aqu", "sdng" and "tdk". In the process of *lemmatization* non-standard words found earlier will be changed to "I", "medium" and "no".

2.7 Remove Stopwords

In research there are (Rahutomo *et al.*, 2019) ideas that have textoccuments in which there are words that are not very useful such as prepositions, conjunctions, adjectives, *slank* words, pronouns and much more. These words are very often encountered with the same main word which can make the word have no unique value and no meaning certain special. Words that have a less contributing nature in the analysis text are commonly referred to as *stopwords* or *stoplists*. *Stopword* has no potential to be used as an index document. *Stoplists* have a unique property, namely thateach language has its own *stoplist*. There are methods that can improve the quality of *sentiment analysis* by removing *stopwords* in the *dataset*.

2.8 Stemming

Stemming is the process of removing additional attributes on a word such as the removal of "me-" and -kan" from "create" to "create". (Julian *et al.*, 2019)It could also be in its simple sense that *stemming* has the meaning of making the word berimbuhan into its basic word. *Stemming* is often used for the implementation of a system that has the purpose of retrieving information such as search engines and other analytical texts. *Stemming* has algorithms that are based on certain rules and also heuristic processes to cut or remove extra characters at the end of a word. The process that has the purpose of removing a character can be done with the initial stage of seeing a word after which it will be verified and if it meets the criteria for the algorithm additional words then the word will be deleted. *Stemming* itself has two problems that often arise in its implementation. The problem is over-stemming and *under-stemming*. *Over-stemming* can occur due to the deletion of words that can be judged too much or excessively while under-stemming can occur because when the word deletion process has inappropriate values that have an impact on changes in the meaning in a word. In this study, the use of *the stemming* process was optimized with the help of a special program package from Sastrawi to be able to help analysis in Indonesian.

2.9 Data Analysing

Data analysing is the core process of research, where data is analyzed using one or several specific methods with the aim of achieving a predetermined final result. In this study,(Youth & Maktoubian, 2019) *data analysing* was to carry out the process of *topic modeling* and *sentiment analysis*. *Topic modeling* using *latent dirichlet allocation* and *sentiment analysis* using *the long short term memory* method. A combination of the two methods above is used to meet the initial purpose of the study.

2.10 Embedding

Embedding is one of the techniques of converting words that allows words with the same meaning to have similar values or understandings into numerical (Brownlee, 2017). After going through the *text processing* process, *embedding* will be applied to be converted into a number sequence. *Embedding* generally has several characteristics. Namely, among them have high dimensions, there is *noise* in the data, and there is a bad number structure. In this study, data from *Twitter* was used. Data coming from *Twitter* is quite complex. This is because the characteristics of *Twitter* are the use of language that does not match the standard language and the many spelling errors in *writing tweets* (Zhou *et al*., 2014).

2.11 Machine Learning

Machine learning or machine learning as it progresses prioritizes generalization as a search through the possibilities of a hypothesis. Machine learning techniques are divided into two including Regression and Classification (Witten *et al.* , 2016) . Some artificial network-based machine learning is often used in various ways, one of which is as a forecasting or forecasting function. Artificial networks such as *Artificial Neural Network* (ANN) have often been used but in the ANN type, some weaknesses are found such as relatively long time use, for which several artificial network models have begun to be developed to overcome these problems (Martellotta *et al*., 2017; Ahmad *et al*., 2014). Artificial Neural Network (ANN) is a part of artificial intelligence that adopts the work of the human nervous system. Artificial Neural Network techniques can be used to conduct training in the classification process. One technique in Artificial Neural Networks, namely the perceptron multi-layer network. This network consists of several layers of neurons (computing units) that are connected hierarchically in a feed forward. Multilayer perceptron uses several learning techniques, the most popular being back-propagation. The backpropagation neural network architecture is a hierarchical design that contains nodes at each layer that are fully interconnected (Prasetiyo et al., 2020).

2.12 Long Short Term Memory

Long *short term memory* (LSTM) is a method resulting from the development of *the algorithm recurrent neural network*. This model was developed and introduced by Hochreiter and Schmidhuber, based on the problems discovered by Bengio. Bengio argues that the RNN model cannot correlate information that has been stored for a long time, sothat the stored information expires and is updated again. The LSTM algorithm can extractk informasi from data for a long time (Naury *et al*., 2021; Idris *et al*., 2019).

The LSTM model can solve problems that the RNN cannot handle with memory management capabilities. This memory setting capability uses memory cells and gate units. To deal with the problem of *disappearing gradients*, LSTM introduced a unit called *the Constant Erro Carousel* (CEC)(Seo *et al*., 2020). The researchers then developed LSTM to be applied in areas such as, *classification*, *forecasting*, and *speech recognition* (Li & Qian, 2016).

2.13 Latent Dirichlet Allocation

LDA is a probabilistic topic model in which each document is represented randomly or a mixture of more than one set of latent topics and each topic is represented as a distribution of vocabulary (sahria Y *et al* ., 2005). In this *topic modeling*, LDA can detect a collection of similarity sentences, languages and topics that have not been labeled before, analisis LDA hasnyai hidden characteristics of some sentences or topics hidden in some mixed documents (Setijohatmo *et al.* , 2020) . The main point of the LDA for topic modeling is carried out for each dataset with a positive label, a negatively labeled dataset, and a neutral labeled dataset, so that it is obtained what topics are discussed by the *online* mass media related to a news issue, be it a topic derived from a headline with positive, negative, and neutral sentiments. The discussion can underlie a modeling that leads to labeled data to obtain visualizations. With this visualization we will make it easier to get topics in online media talks so that we can get links and slices between topics (Naury *et al.* , 2021) .

3 Method

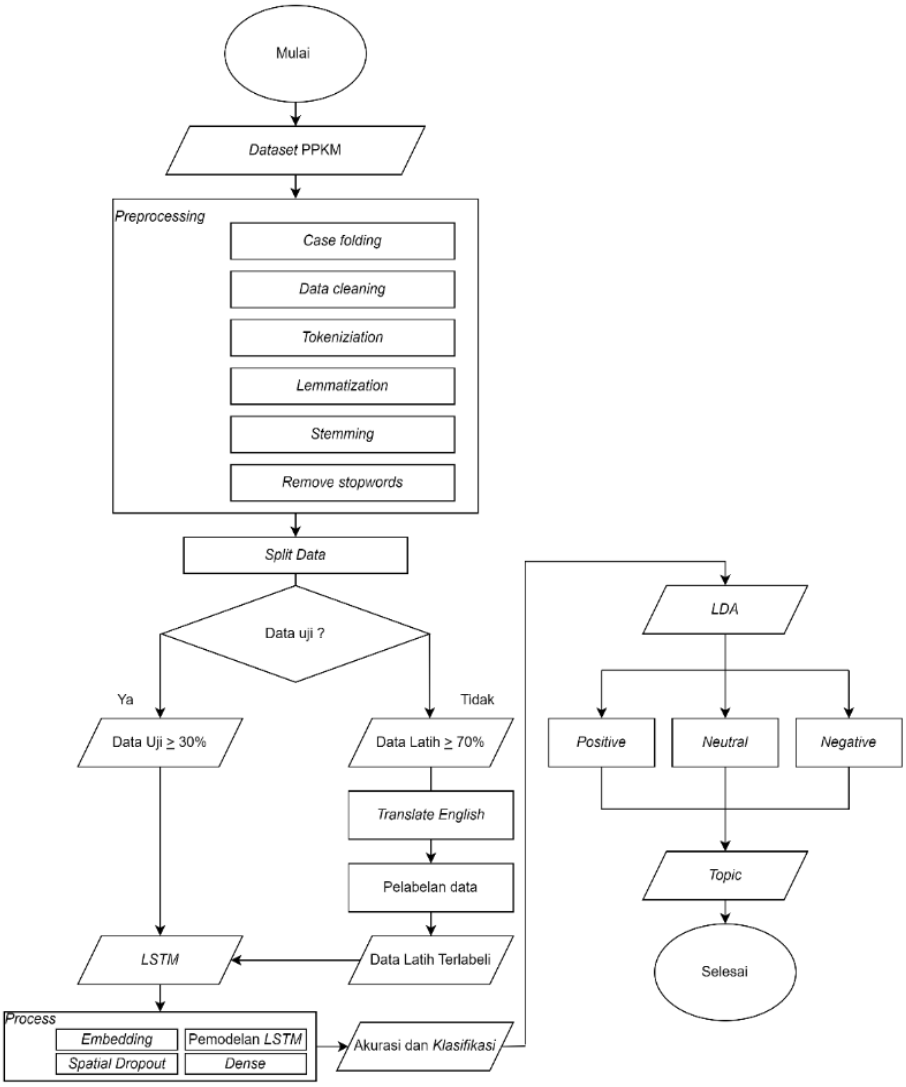

Figure 1. *Flowchart* Research Design

At stage Figure 1. where the stages of this research are composed of the first stage of *text processing* with6 stages in it with the aim of normalizing data.the second stage is data separated into training data Test data with a data ratio of 70% : 30%. The training data will be processed first using *textblob* to do *sentiment* labeling with 3 categories *of positive, neutral* and *negative*. The next process of training data is entered into the *lstm architecture* and processed using 3 *hidden layers* and *dense (softmax)* the results obtained are LSTM *models* and accuracy. The LSTM *model* is used to test data to perform sentiment labeling. The test data is labeled as entered by LDA to perform a topic search. *The topics* obtained were 5 *wordclouds* in each sentiment category. The following are some of the stages of research results including Case Folding, Data Cleaning, Lemmanitization, Tokenization, Stemming and Remove Stopwords.

3.1 Case Folding Results

At this stage, it is the result of the processing of the case folding method. The right column is the end result of the case folding process. It can be clearly seen in the right column that all typefaces have been changed to the same without any different typeface and can be seen in the Table 1.

Table 1. Case Folding

full_text	***Case folding***
PPKM = Clever You're looking for Mesjid ðŸ¤§,	ppkm = clever you find mesjid ðŸ¤§,
@ksjincha ppkm yes :(,	@ksjincha ppkm yes :(,
…	…
Opo nek bar ppkm ki viruse ilang? Lakok do postponed all belehe.	opo nek bar ppkm ki viruse ilang? Lakok do postponed all belehe. ...
trubus rahardiansyah: IN THE FUTURE, PPKM MUST BE TIGHTENED ""	Trubus Rahardiansyah: In the future, PPKM, must be tightened ""
@tvOneNews @RadenPardede:after looking at the data it looks like PPKM is likely to be extended ...	@tvonenews @radenpardede:after looking at the data it looks like ppkm is likely to be extended ...
1 word for KPC-PEN DOG!! https://t.co/KPnyd00cr7"@SatgasCovid19	1 word for kpc-pen anjijng!! https://t.co/KPnyd00cr7"@satgascovid19

3.2 Data Cleaning

At this stage is data cleaning, from the left column containing text data from *case folding* which is still mixed with many characters or parts of the data that are not used in the research process. The right-hand column contains the cleanup text data at the *data cleaning* stage. The cleanup results are clean text without any url, "@", inappropriate characters, *username*, *hashtag* and can be seen in Table 2. Data Cleaning

Table 2. Data Cleaning

full_*text*	***Data Cleaning***
ppkm = clever you find mesjid ðŸ¤§,	PPKM is smart you are looking for mesjid
@ksjincha ppkm yes :(,	ppkm yes
opo nek bar ppkm ki viruse ilang? Lakok do postponed all belehe. ...	opo nek bar ppkm ki viruse ilang Lakok do postponed kabeh belehe
…	…
Trubus Rahardiansyah: In the future, PPKM, must be tightened ""	
@tvonenews @radenpardede:after looking at the data it looks like ppkm is likely to be extended ...	Trubus Rahardiansyah In the future PPKM must be tightened after seeing data it seems that PPKM is likely to be extended the word for KPC Pen Dog
1 word for kpc-pen anjijng!! https://t.co/KPnyd00cr7"@satgascovid19	
@alergisefood is ppkm sksk	kan ppkm sksk

3.3 Lemmanitization

At this stage is Lemmanitization, the left column shows the data before going through the *Lemmatization* stage. The right-hand column shows the data has gone through the *Lemmatization* stage with word normalization with designed ones and can be seen in Table 3.

Table 3. Lemmatization

Tokenization	***Lemmatization***
pasti,ppkm	definitely ppkm
iyes,sih,definitely,di,batas,also,which,help,his, now,where,ppkm,bange	Yes, it must be at the limit also that helps him now where PPKM is very
ppkm,clever,clever,kaulah,find,mesjid	ppkm clever you find mesjid
opo,nek,bar,ppkm,ki,viruse,ilang Lakok,do,postponed,kabeh,belehe	what if the bar ppkm ki viruse is gone, how come do you postpone all qurban
trubus,rahardiansyah,future,ppkm,must,tightened,after,see,data,like,ppkm,likely,will be extended,word,create,kpc ,pen,dog	Trubus Rahardiansyah In the future PPKM must be tightened after seeing data it seems that PPKM is likely to be extended the word for KPC Pen Dog
kan,ppkm,sksk	kan ppkm sksk
makanya,plunge,lead,direct,ppkm,pak,biar,tahu,condition,field,useless,lip,service,continue,reversed,palace	That's why I jumped in to lead PPKM directly, sir, so that you know the condition of the field for free lip service continues behind the palace
mirip,malaysia,dozens,hotel,and,restaurant,di, garut,pasang,flag,white,because,ppkm	Similar to Malaysia dozens of hotels and restaurants in Garut put up white flags because of PPKM

3.4 Tokenization

At this stage is Tokenization, the left column shows the data before going through the *tokenization* stage. The right-hand column shows that the data has gone through the *tokenization* stage with non-standard normalization of words and can be seen in Table 4. Tokenization.

Table 4. Tokenization

Data Cleaning	***Tokenization***
Yes, it must be at the limit also that helps now where PPKM is really	i yes,sih,definitely,di,batas,also,which, help,his,now,where,ppkm,bange
PPKM is smart you are looking for mesjid	p pkm,clever,clever,kaulah,find,mesjid
opo nek bar ppkm ki viruse ilang Lakok do postponed kabeh belehe	o po,nek,bar,ppkm,ki,viruse,ilang Lakok,do,postponed,kabeh,belehe
Trubus Rahardiansyah In the future PPKM must be tightened after seeing data it seems that PPKM is likely to be extended the word for KPC Pen Dog	t rubus,rahardiansyah,future,ppkm,must ,tightened,after,see,data,like,ppkm,lik ely,will be extended,word,create,kpc ,pen,dog

3.5 Stemming

At this stage is Stemming, the left column shows text data that still shows affixes while the left column shows text data that has been cleared at the stage *stemming* and can be seen in Table 5. Stemming.

Table 5. Mood

Tokenization	*Stemming*
Yes, it must be at the limit also that helps him now where PPKM is very	Yes, it must be at the limit also that helps him now where PPKM is very
…	…
what if the bar ppkm ki viruse is gone, how come do you postpone all qurban	what if the bar ppkm ki viruse is gone anyway do retreat all qurban
Trubus Rahardiansyah In the future PPKM must be tightened after seeing data it seems that PPKM is likely to be extended the word for KPC Pen Dog	Trubus Rahardiansyah Depan PPKM Must Be Strict After Seeing Data It Seems That PPKM Is Likely to Be Long Words for KPC Pen Dog
That's why I jumped in to lead PPKM directly, sir, so that you know the condition of the field for free lip service continues behind the palace	That's why I jumped in to lead PPKM directly, sir, let you know the condition of the field for free lip service and continue to return to the palace

3.6 Remove Stopwords

At this stage is Remove Stopwords, th left column shows text data that is still filled with stopwords and the right column shows data that has been cleared from *stopwords* and can be seen in Table 6. Remove Stopwords.

Table 6. Remove Stopwords

Stemming	*Remove Stopwords*
ppkm fren	PPKM
definitely ppkm	
Yes, it must be at the limit also that helps him now where PPKM is very	it's definitely at the limit as well which helps where ppkm is very
ppkm clever you find mesjid	ppkm clever you find mesjid
…	…
what if the bar ppkm ki viruse is gone anyway do retreat all qurban	What if PPKM Viruse is gone, how come all qurban is postponed

4 Results and Discussion

This section is divided into two parts, results and discussion. Results are a picture of data and findings obtained using the methods and procedures described in the data collection method. The discussion is a presentation of results that answer research questions more comprehensively.

```
Process - Predicting Label
Epoch 1/10
53/53 [==============================] - 5s 52ms/step - loss: 0.9444 - accuracy: 0.5433
Epoch 2/10
53/53 [==============================] - 3s 50ms/step - loss: 0.7920 - accuracy: 0.6776
…………………………………………………………………………………………………...
…………………………………………………………………………………………………...
Epoch 10/10
53/53 [==============================] - 3s 48ms/step - loss: 0.2678 - accuracy: 0.9086
Accuracy; 0.9086
```

4.1 LSTM Performance

At this stage, the results of the LSTM architecture will be displayed, namely the results of calculating accuracy, sentiment labeling, system strength in *labeling* and data classification results outside PPKM.

4.1.1 Accuracy Results

The accuracy result obtained from Figure 4.10 is 0.9086 with repeated calculations 10 times passing 3 *hidden layers*, 1 *dense* and *Adam* as *optimizer*. Per *epoch* able to read 53 *iterations* in which there were 32 *batches.* The result of 0.9086 was obtained 5 times the loop and the best results were taken from the 5 experiments conducted and can be seen in Figure 2. Accuracy Results

Figure 2. Accuracy Results

4.1.2 Labelling Results

The result of the *labelling* process using the LSTM algorithm. The results of the LSTM algorithm for the classification of sentiment analysis resulted in three groups, namely: positive, negative, and neutral. The results of the analysis are a process of the LSTM *model* studied from latih data and can be seen in Table 1. Labelling Process Results

Table 1. Labelling Process Results

Remove Stopwords	***Predicted***
it's definitely at the limit as well which helps where ppkm is very	*Positive*
What if PPKM Viruse is gone, how come all qurban is postponed	*Negative*
Trubus Rahardiansyah Depan PPKM Must Strictly See PPKM Data Possible Word Length for KPC Pen Dog	*Neutral*
That's why I jumped in to lead PPKM directly, sir, let you know the condition of the field for free lip service and continue to return to the palace	*Positive*
Similar to Malaysia dozens of hotels and restaurants garut put up white flags because of PPKM	*Neutral*
…….	……

4.2 Topic Modelling Results

Before entering the *topic, modeling* stage using the LDA algorithm, the dataset is processed using the LSTM algorithm as an algorithm in the *labelling* process according to the explanation above. At this stage, it is divided into three *topic modeling* based on labelling results, namely positive, negative, and neutral sentiment topics. For discussion of *topics* that already have labels such as sample data from the *labelling data* stage in Table 2. Topic Positive Results, Table 3. Topic Neutral Results and Table 4. Negative Topic Results

Table 2. Topic Positive Results

Topic Positive									
Wordcloud 1		*Wordcloud 2*		*Wordcloud 3*		*Wordcloud 4*		*Wordcloud 5*	
Text	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight
PPKM	0,048	PPKM	0,055	china	0,026	PPKM	0,050	PPKM	0,047
do not	0,020	create	0,016	corona	0,015	emergency	0,025	house	0,013
People	0,019	community	0,011	come	0,010	adha	0,023	Day	0,012
So	0,018	do not	0,011	PPKM	0,010	Eid al-Fitr	0,023	citizen	0,011
more	0,012	look for	0,011	put on	0,008	kalua	0,013	Of	0,011
result	0,012	July	0,010	ken	0,007	Executable	0,012	Prayer	0,010
face	0,010	can	0,010	Visit	0,007	covid	0,011	pandemic	0,009
president	0,010	What	0,009	Wed	0,006	day	0,010	Want	0,009
Karna	0,009	old	0,009	of	0,006	Prayer	0,010	Mosque	0,009
jokowi	0,009	Person	0,008	appear	0,006	Same	0,010	forbid	0,009

Table 3. Topic Neutral Results

Topic Neutral									
Wordcloud 1		*Wordcloud 2*		*Wordcloud 3*		*Wordcloud 4*		*Wordcloud 5*	
Text	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight
PPKM	0,093	PPKM	0,078	PPKM	0,094	PPKM	0,084	PPKM	0,078
So	0,018	emergency	0,025	day	0,015	Plan	0,026	Eid al-Fitr	0,022
People	0,016	house	0,022	recovered	0,013	emergency	0,019	adha	0,020
Indonesian	0,014	saleable	0,015	Indonesian	0,013	long	0,014	of	0,019
emergency	0,013	covid	0,012	emergency	0,012	giat	0,013	Satpol	0,012
Not	0,012	do not	0,011	end	0,012	police	0,011	want	0,011
Again	0,012	Spread	0,011	Order	0,011	orderly	0,009	emergency	0,009
recovered	0.012	time	0,010	two	0,009	i	0,008	day	0,009
road	0,010	border	0,010	teng	0,007	my	0,006	Indonesian	0,009
javanese	0,010	prokes	0,010	citizen	0,007	just	0,006	recovered	0,008
….	…	….	….	…..	…..	……	…..	……	….

Table 4. Negative Topic Results

Topic Negative									
Wordcloud 1		*Wordcloud 2*		*Wordcloud 3*		*Wordcloud 4*		*Wordcloud 5*	
Text	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight	*Text*	Weight
road	0,039	PPKM	0,032	PPKM	0,062	PPKM	0,078	PPKM	0,058
PPKM	0,038	saleable	0,026	emergency	0,053	long	0,047	Order	0,036
impact	0,025	Trillion	0,018	covid	0,022	So	0,028	same	0,030
city	0,020	Rp	0,018	long	0,021	day	0,020	social assistance	0,029
lonely	0,019	We	0,016	impact	0,018	emergenc y	0,016	salur	0,028
photograph	0,016	Not	0,016	city	0,017	If	0,014	long	0,019
edar	0,016	beritabiasar sindonews	0,016	recovered	0,017	just	0,014	So	0,011

preweddin g	0,016	dana	0,016	community	0,011	end	0,012	saleable	0,011
Voice	0,016	All	0,015	citizen	0,011	Motion	0,011	impact	0,011
saw	0,016	add	0,015	case	0,011	head	0,011	will	0,011
….	….	….	….	….	….	….	….	….	….

4.3 Analysis of Public Opinion

At this stage, we will discuss the label results produced using the LSTM algorithm and also the topic *modeling* results using the LDA algorithm. The use of these two algorithms produces an analysis of public opinion based on the results of labels created in the form of positive, negative, and neutral. For discussion topics that have positive labels such as sample data from the data stage of data *labelling* in Table 5. Furthermore, for negative topics can be seen in Table 6 and the last topics that have neutral labeling of data are seen in Table 7.

Table 5. Public Opinion Analysis Results from *Topic Modelling* in *Positive* Labels

Sentence	*Labelling* Results	Topic
Polsek Jetis Salur Bansos PPKM Darurat Complete News Click Polresjogja Polsekjetisjogja BersamalawanCovid TetapDisciplinary Covid PPKMDARURAT	*Positive*	PPKM, Covid, Emergency
New name I guess emergency so ppkm real transition ppkm leveling	*Positive*	DArurat, PPKM
Kudus about worship there are opponents indeed mo PPKM PSBB lock down et al still gas to hold a village mosque is indeed a solution as long as brenti total positive numbers rise 12ethics so green zone not medical personnel many ready army chuaaksz	*Positive*	PPKM, Village, Mosque

Table 6. Public Opinion Analysis Results from *Topic Modelling* in *Negative* Labels

Sentence	*Labelling* Results	Topic
gagal ppkm national way of failing central government failed president ri bkk jokowi keslhn ftl ppkm adi citizen sdr limit mobility his economy smntr ngr order full ignore the main needs of food ftl due to	*Negative*	ppkm
cm nentuin ppkm length anything confused what regime if pandemic so commander adl scientists field healthy not opung taste cost-effective life saving destroyed dech all destroyed tang incompetent people country win	*Negative*	ppkm, orang
PPKM Emergency Long Perindo Remembers About The Prosperous Economy Of The People Sindonews Not Ordinary News	*Negative*	PPKM, Emergency

Table 7. Public Opinion Analysis Results from *Topic Modelling* in Neutral Labels

Sentence	*Labelling* Results	Topic
menteri is waiting for the results of the discussion mr. jokowi jokowi jokowi decided to ask the president of ppkm so what not	*Neutral*	PPKM, President, Jokowi
masjid around the house in the city of the former schola ied adha Bangka ppkm emergency smug religious wrap	*Neutral*	ied, ppkm, emergency, home, adha
PPKM clever you find mosque	*Neutral*	PPKM

5 Discussion

This study applies *Long Short Term Memory (LSTM*) for labelling data which tests the level of accuracy of *sentiment analysis* related to the Implementation of Community Activity Restrictions (PPKM) spread on public opinion on social media forums twitter. After that, it is continued by applying *Latent Dirichlet Allocation (LDA)* as a *topic modeling* mapping. Between the two algorithms, namely LDA and LSTM, it was also used for previous research to test the level of accuracy and *topic modeling*, but the use of different datasets. In this study, a comparison of the results was carried out between each use of the algorithm applied to the *proposed method* with the previous study and also followed by different datasets between each study and better accuracy results could be obtained.

After the implementation of the program code and the process of implementing the stages in the Community Opinion Analysis system on the Impact of the Implementation of Community Activity Restrictions (PPKM) During the COVID-19 Pandemic Using *Long Short Term Memory* and *Latent Dirichlet Allocation*. Modeling results were obtained for testing the train dataset and testing on *sentiment analysis* in the form of different accuracy, namely 0.9002 for the lowest value and 0.9 086 for the highest.

Based on the accuracy results obtained, the best results that will be taken from the *proposed method* and compared to previous studies using LDA and LSTM algorithms as *modeling* in different dataset cases can be seen in Tabel 7. Comparison of Other Research Methods and Topics

Writer	Dataset	Topic	Method	Accuracy
Huang *et al* (2021)	Sina-Weibo, Wechat, QQ groups	*Cryptocurrency*	LSTM	87%
Widi Widayat (2021)	Andrew *et al*	*Movie Review*	LSTM – Word2Vec	85,86%
Priyantina, Sarno (2019)	Traveloka	*Hotel Review*	LDA – Semantic Similarity – LSTM	93%.
Reza Amalia Priyantina (2019)	Traveloka	*Hotel Review*	Word Embedding - LSTM	93%
Proposed Method	Twitter	PPKM	LSTM – LDA	90,86%

In this study, the authors combined the use of LSTM as *labelling* and LDA as *topic modeling* which resulted in an accuracy rate of 90.86%. When compared with the accuracy of previous studies according to Tabel 7 this study provides an accuracy value that can compete among others. In the next study, this model can be used as a basic basis for the sentiment analysis test process with other *twitter datasets*. For this reason, this model is able to provide test results that have a fairly high level of accuracy compared tothe research beforenya.

6. Conclusion

In the process of labeling and topic search. It requires data and the data we want to take is PPKM. I took the data from *Twitter* with a time span starting from July 03, 2021 – July 20, 2021, a total of 3000 *texts*, the data will be normalized first using *preprocessing.* The data has been normalized and will be divided into 70% training data and 30% test data. First the train data will be processed first using *textblob. Textblob* is a library of *python* the main function of the *library* to label with 3 categories of *positive, neutral* and *negative* before being processed *textblob* can only be processed with ENGLISH, that way we do *translated* using the *googletrans library* after the process of being passed the training data can be labeled. The training data will be processed using the LSTM architecture. The accuracy value in the analysis using the LSTM and LDA algorithms was 90.86%. The accuracy value in the method in this study was able to produce a pretty good score. But it has not been able to produce higher yields compared to other methods.

7. References

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A Review on Applications of ANN and SVM for Building Electrical Energy Consumption Forecasting. Renewable and Sustainable Energy Reviews, 33, 102–109. https://doi.org/https://doi.org/10.1016/j.rser.2014.01.069

Alamsyah, A., Prasetiyo, B., Hakim, M. F. A., & Pradana, F. D. (2021). Prediction of COVID-19 Using Recurrent Neural Network Model. *Scientific Journal of Informatics*, *8*(1), 98–103. https://doi.org/10.15294/sji.v8i1.30070

Alamsyah, W. a. (2017). *Recurrent Neural Network For Forecasting Time Series With* . Journal of Physics: Conference Series.

Al-Sultany, G. A., & Aleqabie, H. J. (2018). Events Tagging in Twitter Using Twitter Latent Dirichlet Allocation. International Journal of Engineering & Technology, 7(4.19), 884–888.

Azeharie, S., & Kusuma, O. (2014). Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis di Jakarta. 83–98.

Brownlee, J. (2017). Deep Learning for Natural Language Processing. 89 https://github.com/abhiramkadali/Machine_Learning-_Resources.

Gössling, S., Scott, D., & Hall, C. M. (2021). Pandemics, Tourism and Global Change: a Rapid Assessment of COVID-19. Journal of Sustainable Tourism, 29(1), 1–20. https://doi.org/10.1080/09669582.2020.1758708

Haridison, A., Sandi, J. R. A., Suprayitno, S., & Jaya, I. (2021). Regional Election and Covid-19: Evidence in Central Kalimantan. Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, 4(2), 1597–1606. https://doi.org/10.33258/birci.v4i2.1821

Haridison, A., Sandi, J. R. A., Suprayitno, S., & Jaya, I. (2021). Regional Election and Covid-19: Evidence in Central Kalimantan. Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, 4(2), 1597–1606. https://doi.org/10.33258/birci.v4i2.1821

Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., Liu, Z., & Zhang, J. (2021). LSTM Based Sentiment Analysis for Cryptocurrency Prediction BT - Database Systems for Advanced Applications (C. S. Jensen, E.-P. Lim, D.-N. Yang, W.-C. Lee, V. S. Tseng, V.

Kalogeraki, J.-W. Huang, & C.-Y. Shen (eds.); pp. 617–621). Springer International Publishing.

Idris, N. O., Widyawan, & Adji, T. B. (2019). Classification of Radicalism Content From Twitter Written in Indonesian Language Using Long Short Term Memory. ICICOS 2019 - 3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings, 0–4. https://doi.org/10.1109/ICICoS48119.2019.8982498

Karami, A., Bennett, L. S., & He, X. (2018). Mining Public Opinion About Economic Issues: Twitter and the U.S. Presidential Election. International Journal of Strategic Decision Sciences (IJSDS), 9(1), 18–28. https://doi.org/10.4018/IJSDS.2018010102

Li, D., & Qian, J. (2016). Text Sentiment Analysis Based on Long Short-Term Memory. 2016 1st IEEE International Conference on Computer Communication and the Internet, ICCCI 2016, 471–475. https://doi.org/10.1109/CCI.2016.7778967

Martellotta, F., Ayr, U., Stefanizzi, P., Sacchetti, A., & Riganti, G. (2017). On The Use of Artificial Neural Networks to Model Household Energy Consumptions. Energy Procedia, 126, 250–257. https://doi.org/10.1016/j.egypro.2017.08.149

Mihardja, H., & Harja, K. P. (2021). Acupuncture for Treating Hypertension and Type 2 Diabetes Mellitus as Comorbidities in Patients with COVID-19. Medical Acupuncture, 33(3), 197–202. https://doi.org/10.1089/acu.2020.1475

Muslim, M. A., Dasril, Y., Alamsyah, A., & Mustaqim, T. (2021). Bank predictions for prospective long-term deposit investors using machine learning LightGBM and SMOTE. *Journal of Physics: Conference Series*, *1918*(4). https://doi.org/10.1088/1742-6596/1918/4/042143

Naury, C., Hidayatullah, A. F., & Fudholi, D. (2021). Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. Jurnal Media Informatika Budidarma, 5, 24–33. https://doi.org/10.30865/mib.v5i1.2556

Naury, C., Hidayatullah, A. F., & Fudholi, D. (2021). Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. Jurnal Media Informatika Budidarma, 5, 24–33. https://doi.org/10.30865/mib.v5i1.2556

Prasetiyo, B., Alamsyah, Muslim, M. A., Subhan, & Baroroh, N. (2020). Artificial neural network model for banckrupty prediction. *Journal of Physics: Conference Series*, *1567*(3). https://doi.org/10.1088/1742-6596/1567/3/032022

Rai, A., & Borah, S. (2021). Study of Various Methods for Tokenization BT -Applications of Internet of Things (J. K. Mandal, S. Mukhopadhyay, & A. Roy (eds.); pp. 193–200). Springer Singapore.

Seo, S., Kim, C., Kim, H., Mo, K., & Kang, P. (2020). Comparative Study of Deep Learning-Based Sentiment Classification. IEEE Access, 8, 6861–6875. https://doi.org/10.1109/ACCESS.2019.2963426

Setijohatmo, U. T., Rachmat, S., Susilawati, T., Rahman, Y., & Kunci, K. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. Prosiding The 11th Industrial Research Workshop and National Seminar, 402–408

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Elsevier Science. https://books.google.co.id/books?id=1SylCgAAQBAJ

Zhou, Z., Zhang, X., & Sanderson, M. (2014). Sentiment Analysis on Twitter Through Topic-Based Lexicon Expansion (pp. 98–109). https://doi.org/10.1007/978-3-319-08608-8_9