Abstract- Determining a policy is often limited in a short time so that decisions are prone to inaccuracies and are ultimately judged to be less targeted. Therefore, there is a necessity to use data mining technology. Currently, especially due to the continuously increasing case of the spread of COVID-19 in Indonesia, in order to reduce the rate of spread of COVID-19, the government has established a COVID-19 vaccination and emergency Community Activity Restriction Implementation (PPKM) policies. For the success of the policies, the government is required to ascertain and understand the attitude of the society. Hence, the policies can be accepted and supported by the society. YouTube is one of the sources to discover people’s attitudes because in YouTube, people can express their opinions freely. In this study, a model based on Natural Language Processing (NLP) with the Deep Learning method was developed to analyze people’s attitudes from their writings or posts on social media. As for the algorithm stages, first the model analysis was created using the Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (Bi-LSTM) algorithms as a comparison. In the COVID-19 vaccination policy, the Bi-LSTM algorithm provided a better evaluation value, i.e., the accuracy value of 87.20%, recall of 89.15%, precision of 87.14%, and F1-score of 87.14%. In the emergency PPKM policy, the LSTM algorithm provided a better evaluation value of 95.13%, recall of 93.94%, precision of 94.08%, and F1-score of 94.01%. In addition, using the Maximum Marginal Relevance (MMR) method to obtain recommendations related to the COVID-19 vaccination policy, the result showed that the government needs to carry out re-socialization regarding vaccination and the impact of the vaccine so that the society can become more cooperative and the emergency PPKM policy needs to be reviewed because it has an impact on the society’s economy.

Keywords— bi-lstm; covid-19; lstm; maximum marginal relevance; nlp; public policy; social computing; social media;

I. INTRODUCTION

Coronavirus Disease or what is known as COVID-19 is a very dangerous disease that can even cause death. In Indonesia, the positive confirmed cases of COVID-19 have reached 3.08 million and the death cases of 80 thousand people were recorded by July 24, 2021 [1]. People who are confirmed positive, negative, and died due to COVID-19 in Indonesia continue to increase every day. To overcome the problem of the spread of COVID-19 in Indonesia, the government is currently making policies related to COVID-19 vaccination. The purpose of this vaccination is to prevent or overcome severe COVID-19 symptoms, protect others from getting COVID-19, and stop the spread of COVID-19. In addition, the impact of the spread of COVID-19 in Indonesia has caused the government to make an emergency Community Activity Restriction Implementation (PPKM) policy. The purpose of implementing this emergency PPKM policy is to control the increase in positive confirmed cases and deaths due to COVID-19 [2].

However, policies related to COVID-19 vaccination and emergency PPKM raise pros or cons or disagreement in the society. In accordance with a survey related to COVID-19 vaccination conducted in September 2020 involving more than 115,000 respondents from all provinces in Indonesia, almost 65 percent. (64.8 percent) of respondents expressed their desire to be vaccinated. Meanwhile, almost 8 percent (7.6 percent) of respondents stated that they did not want to be vaccinated [3]. The rest, or 27.6 percent, admitted to having doubts. Based on a survey covering the whole provinces in Indonesia from February 28, 2021 to March 8, 2021, nationally, there were around 44 percent who chose to undergo PPKM strictly despite their declining income and there were 40 percent who chose to stop PPKM despite increasing the risk of contracting Covid-19 [4].

For the success of the COVID-19 vaccination and PPKM policies in Indonesia, it is essential for the government and stakeholders to know and understand the public’s attitude towards these policies so that the government can determine the correct strategy to convince the public. In order to ascertain the public’s opinion regarding the COVID-19 vaccination program and the emergency PPKM, social
media can become the best source. One of the most popular social media today is YouTube. Social media users generally express their opinions through comments provided on YouTube. People’s comments, especially on news videos related to COVID-19 vaccination and emergency PPKM, are a form of channeling aspirations for the society.

This study focuses on developing a model to conduct sentiment analysis on comments on YouTube videos related to the COVID-19 vaccination program and emergency PPKM and conducting sentiment analysis using the Natural Language Processing (NLP) and Deep Learning methods as well as obtaining policy recommendations that can be used by the government using the Maximum Marginal Relevance method (MMR).

II. LITERATURE STUDY

A. Public Policy & Social Computing

Public policy is a form of a collection of several sets of actions, plans, laws and behaviors for decision making by the government and other political actors as an effort to influence, change, or overcome a problem or issue that is recognized as having entered the political realm by policy makers and larger community [5][6]. Meanwhile, social computing is a hybrid of the field of computer science that is connected with the analysis of social data (which includes politics, economics, socio-culture, and defense and security) and computing based on Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning technology [7]. In this case, the social data can come from various kinds of sources, such as social media (twitter, YouTube, Facebook, Instagram and others).

B. Deep Learning

Deep learning is part of machine learning which becomes the most popular research reference today. Deep learning uses Artificial Neural Network (ANN) which is an information processing engine modeled on the structure and actions of biological neural networks in the brain. Deep learning is dynamic because these algorithms have the ability to continuously improve performance and adapt easily to changes in patterns. Deep learning can be created with many layers. The layers used for the deep learning process consist of the Input Layer, Hidden Layer, and Output Layer [8]. Moreover, deep learning can work well for practical problems. Deep Neural Network (DNN) is an ANN that has many layers [9].

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is one type of Recurrent Neural Network (RNN) architectures which was developed with the aim of avoiding RNN constraints in long-term memory [10]. LSTM architecture has the ability to remember long-term information. In the RNN architectures, the network only uses 1 simple layer in its loop, i.e., a tanh layer. However, in LSTM architecture, there are 4 layers in the loop as shown in Fig. 1 [11-29].

D. Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) is a variant of Recurrent Neural Network (RNN). Bi-LSTM is utilized for sequence markers. Bi-LSTM is used to represent word features and Bi-LSTM output is employed as the input to a conditional random field. Forward and backward LSTM cells are combined respectively [30]. However, it needs a sentence level feature. Then, the final hidden states on the forward and backward movement of the LSTM cells are used and the output of the model is combined as displayed in Fig. 2. This feature is used as the input to the memory network, which is an approximation of the model itself [31-34].

E. Sentiment Analysis

Sentiment analysis is one of the fields of study to analyze opinions, sentiments, judgments and evaluations, attitude, emotional state of a person towards an entity such as products, services, individuals, organizations, events, topics, and their attributes. The purpose of sentiment analysis is to understand and classify the tendency of the content of the text in a sentence [16].

F. Maximum Marginal Relevance (MMR)

Maximum Marginal Relevance is one of the algorithms in text summarization that can summarize one or more documents to obtain a representative text or conclusion. This method is used with the aim of reducing redundancy in
sentence ranking setting in the case of multiple documents. The formula for obtaining the MMR value is presented in Equation (1). After carrying out the calculations, the sentence with the highest MMR value will be selected repeatedly into the summary until the desired number of words in the summary is reached [35].

\[ MMR = \arg\max[\alpha \cdot \text{Sim}(S_i, Q) - (1 - \alpha) \cdot \text{Sim}(S_i, S)] \]  

(1)

Note:
- \( \alpha \) = parameters affecting the degree of relevancy
- \( S_i \) = weight vector of the candidate word
- \( S \) = weight vector of words other than the candidate
- \( Q \) = weight vector of word from the query
- \( \text{Sim}(S_i, Q) \) = similarity value between sentence no. \( i \) and the query
- \( \text{Sim}(S_i, S) \) = similarity value between sentence no. \( i \) and the extracted sentence

III. PROPOSED RESEARCH CONTRIBUTION

This study developed a Deep Learning-based NLP model that can analyze people’s attitudes from their posts on social media in response to the COVID-19 vaccination program and emergency PPKM made by the government. In this case, the final result of this system can provide a recommendation in the form of generating automatic draft or blueprint document containing a government policy for the efficient success of the COVID-19 vaccination program and emergency PPKM in Indonesia.

![Fig. 3. Overall design of the developed system (Social Media Data, NLP, and Deep Learning)](image)

In this study, 7 videos related to the COVID-19 vaccination policy were used with a total of 1248 data and 4 videos related to the emergency PPKM policy were utilized with a total of 3382 data. This data was obtained from scraping YouTube video comments related to the COVID-19 vaccination program. This data collection was carried out for 5 days (23 July 2021 – 27 July 2021). The following are some of the stages carried out in the computing process.

A. Preprocessing Data until Sentiment Analysis

In data preprocessing, there were several stages including case folding, filtering, stop word removal, and tokenizing. Case folding is the process of converting the entire data into lowercase letters. The purpose of converting text data into lowercase letters is to facilitate further processing. Filtering stage is used to clean the data from punctuation which will then be converted into space characters.

The purpose of removing punctuation marks is to simplify the model training process. The third step, Stopword Removal, is a method to remove some words that are not related to sentiment value. Stopword is a collection of words that are not the features or unique words contained in a document. Finally, tokenizing is a method for parsing text content into words, terms, symbols, or elements that compose a text. The parse of the text content is referred to as a token. In this process, some characters such as spaces, periods (.), commas (,), and other characters used as separators of these words will be removed. The token list will be used for subsequent processing input for text representation.

B. Summary Process until Generating Public Policy Documents

At this stage, the number of positive and negative word dictionaries was calculated. After that, take the 5 largest documents which were calculated based on the number of positive sentences in the documents. Summary was carried out using the MMR algorithm, which is a summary extraction method that can be used to summarize single or multiple documents. The MMR summarizes a document by calculating the similarity between parts of the text. In document summarization, the document segmentation process is carried out into sentences and grouped according to the sentence class. MMR is also used by combining user-supplied queries. The results of the summary were used as the keywords to automatically create drafts or blueprints of public policy documents according to cases computed from social media.

IV. TEST RESULTS AND ANALYSIS

A. LSTM and Bi-LSTM Algorithm Testing

The dataset will be trained using a deep learning model with LSTM and Bi-LSTM algorithms for comparison. In this study, LSTM and Bi-LSTM algorithms were combined with Adam Optimizer. The test was carried out twice for each dataset. In the first test, a comparison of the performance of the LSTM and Bi-LSTM algorithms on datasets related to the COVID-19 vaccination policy was carried out. The results of the evaluation of the training data are presented in Figures 4 (a) and (b) while the results of the evaluation of the test data are presented in Figures 4 (c) and (d).
Fig. 4. LSTM vs Bi-LSTM (a) on Vaccination training data, (b) on COVID-19 Vaccination test data, (c) on emergency PPKM training data, (d) on COVID-19 emergency PPKM test data

B. MMR Algorithm Testing

The testing process was carried out in two stages. The first stage of testing the MMR algorithm on datasets was related to COVID-19 vaccination and the second stage of testing on datasets was related to emergency PPKM. In the first stage, the query of “covid vaccination policy” was given and in the second stage, the query of “impact of emergency PPKM” was provided. By performing calculations on the MMR algorithm, the results of the sentence extraction are presented in Table I.

<table>
<thead>
<tr>
<th>TABLE I MMR algorithm test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COVID-19 Vaccination</strong></td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Those who have been vaccinated can still catch Covid-19.</td>
</tr>
<tr>
<td>What will happen to those who haven’t been vaccinated? 😞</td>
</tr>
</tbody>
</table>

Based on Fig. 5, the results of the recommendations for public policy documents in the form of *.pdf files which are ready to be used as drafts to be given to stakeholders are shown. The following is an example of the results.
An example of a manually generated draft is presented in Fig. 6, which is easy to read and capable of being read and understood by the end user.

![Fig. 6. Public Policy Document Repair After by System](image)

According to the results of the evaluation of the LSTM and Bi-LSTM algorithms on the dataset related to COVID-19 vaccination presented in Fig. 4 (a & b), the LSTM and Bi-LSTM algorithms provided 100 percent accuracy, recall, precision, and F1-score values for training data as shown in Fig. 4(a). As for the test data of Bi-LSTM algorithm, the results provided a greater accuracy value than the LSTM method, i.e., the accuracy value of 87.20 percent, recall of 87.14 percent, precision of 87.14 percent, and F1-score of 87.14 percent as shown in Fig. 4 (b). The use of both LSTM and Bi-LSTM models showed that the Bi-LSTM algorithm tends to provide a slightly better evaluation value than the LSTM algorithm.

On the basis of the results of the evaluation of the LSTM and Bi-LSTM algorithms on the emergency PPKM-related dataset presented in Fig. 4 (c & d), the LSTM and Bi-LSTM algorithms provided accuracy, recall, precision, and F1-score values close to 100 percent for training data as shown in Fig. 4 (c). Meanwhile, the LSTM algorithm test data provided a greater accuracy value than the Bi-LSTM method, i.e., the accuracy value of 95.13 percent, recall of 93.94 percent, precision of 94.08 percent, and F1-score of 94.01 percent as displayed in Fig. 4 (d). The use of both LSTM and Bi-LSTM models showed that the LSTM algorithm tends to provide a slightly better evaluation value than the Bi-LSTM algorithm.

Then, in accordance with the test results, several words were obtained which were extracted from a collection of datasets related to the COVID-19 vaccination policy and emergency PPKM. Based on what has been obtained in Table 1, the COVID-19 vaccination policy received positive and negative sentiments related to the statement that COVID-19 vaccination makes us immune to the virus which causes people to have less confidence in the COVID-19 vaccination policy. The government should carry out re-socialization regarding vaccination and the impact of the vaccine so that people can become more cooperative. In the emergency PPKM policy, negative sentiments related to money were obtained because the existence of emergency PPKM resulted in some constraints for people’s work and there were obstacles to the society’s economy. It is recommended that the emergency PPKM policy be reviewed because it has an impact on the society’s economy.

V. CONCLUSION

Based on the results of this study, an analysis and development of models related to public sentiment towards the COVID-19 vaccination policy and emergency PPKM have been carried out which were expressed through comments on YouTube social media. In this study, the model was formed with a Deep Learning algorithm, specifically the Long Short-Term Memory and Bidirectional Long-Short-Term Memory algorithms which were used as comparisons.

In policies related to COVID-19 vaccination, the Bi-LSTM algorithm provided a greater accuracy value than the LSTM method, i.e., the accuracy value of 87.20%, recall of 87.14%, precision of 87.14%, and F1-score of 87.14%. In policies related to emergency PPKM, the LSTM algorithm provided a greater accuracy value than the Bi-LSTM method, i.e., the accuracy value of 95.13%, recall of 93.94%, precision of 94.08%, and F1-score of 94.01%. The results of recommendation documents related to government policies can have broad meanings, for example the necessity to carry out re-socialization regarding vaccination and the impact of the vaccine so that the society can become more cooperative and the emergency PPKM policy that needs to be studied more deeply because it has an impact on the society’s economy.

REFERENCES


