AIP Conference Proceedings

RESEARCH ARTICLE | JULY 12 2024

Preface: Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022 FREE

Check for updates

AIP Conf. Proc. 3077, 010001 (2024)
https://doi.org/10.1063/12.0024850

View Online
Export Citation

21 July 2024 04:35:40

APL Energy
Latest Articles Online!
Read Now
AIP Publishing

AIP Publishing

PREFACE: Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022

"Development of Digital and Green Technology on Post Pandemic Era"

It is with great pleasure to welcome you to Transdisciplinary Symposium on Engineering and Technology (TSET) 2022 hosted by Universitas Dian Nusantara on September 21, 2022. The event aims to a venue for engineers, researchers, scholars, and policy makers to explore the challenges and opportunities from the post pandemic era on civil engineering, mechanical engineering, electrical engineering and computer science. For civil engineers, they will play a significant part in the recovery since design and construction services will be needed in the future, and they need to develop new construction methods, materials, and technologies in order to build a sustainable and resilient infrastructure. For engineers, they need to start thinking about the long-term change of their operations and adapt to the "new normal" that has emerged because of the epidemic. We welcome all parties to share their research and thoughts in the symposium.

Participants of the symposium were invited to submit their papers and disseminate them through oral presentation covering such scope as civil engineering, mechanical engineering, electrical engineering and computer science. To enrich the discussion under the theme of "Development of Digital and Green Technology on Post Pandemic Era", we invited speakers with reputable expertise, namely Prof. Josaphat Tetuko Sri Sumantyo, Ph.D. from Chiba University, Japan; Prof. Dr. rer. nat. Evvy Kartini, M.Sc. from National Nuclear Energy Agency of Indonesia; Prof. Dr. Ir. Bambang Sugiarto,, M.Eng. from Universitas Indonesia, Indonesia; and Sulfikar Amir, Ph.D. from Nanyang Technological University, Singapore. In addition to presenting their research results, the participants of the symposium were also encouraged to submit their papers to be proposed for publication to American Institute of Physics (AIP), one of the world's top publishers as conference proceedings. There were 125 manuscripts submitted to the committee comprising 99 papers of Biology, Chemistry, Computer Science and Technology, and Engineering.

Finally, on behalf of the editors of TSET 2022, I would like to extend my most sincere gratitude to the organizing committee, co-hosting institutions, and most importantly, participants, speakers, presenters, and authors of the symposium. I do hope the proceedings bring significant contribution, particularly to the field of advances of sustainable engineering. I look forward to seeing you all at the upcoming symposium.

The Editors,
Ade Gafar Abdullah
Desi Ramayanti
Henri Septanto
Yohanes Galih Adhiyoga

21 July 2024 04:35:40

Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022
AIP Conf. Proc. 3077, 010001-1–010001-1; https://doi.org/10.1063/12.0024850
Published under an exclusive license by AIP Publishing. 978-0-7354-4964-0/$30.00

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 12 2024

Committees: Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022 FREE

Check for updates

AIP Conf. Proc. 3077, 010002 (2024)
https://doi.org/10.1063/12.0026137

View Online
Export Citation

AIP Publishing

21 July 2024 04:35:51

TSET 2022 COMMITTEE

Advisory Board
Prof. Josaphat Tetuko Sri Sumantyo
Prof. Sulfikar Amir
Prof. E.S. Margianti
Prof. Bambang Sugiarto
Prof. Raihan
Prof. Moermahadi Soerja Djanegara
Prof. Evvy Kartini
Prof. Memen Kustiawan
Prof. Jony Oktavianto
Prof. Intiyas Utami
Dr. Awaludin Martin

Scientific Committee
Prof. Ade Gafar Abdullah
Prof. Suharyadi
Dr. M. Hasannudin Toyieb
Dr. Dewi Anggraini
Magito

Organizing Committee
Conference Chair:
Ir. Margono Sugeng, M.Sc.

Co-Conference Chair:
Ir. Komarudin, ST., MT

Members:
Assoc. Prof. Isma Widiaty
Dr. Yohanes Galih Adhiyoga
Dr. Ari Purwanti
Dr. Didin Hikmah Perkasa
Kornelia Johana, S.I.Kom., M.I.Kom

21 July 2024 04:35:51

Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022
AIP Conf. Proc. 3077, 010002-1–010002-1; https://doi.org/10.1063/12.0026137
Published under an exclusive license by AIP Publishing. 978-0-7354-4964-0/$30.00

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 12 2024

Convolutional neural networks for text classification: A study on public activity restriction FREE

H. Anggit Taba; Hari Suparwito

Check for updates

AIP Conf. Proc. 3077, 040016 (2024)
https://doi.org/10.1063/5.0201145

View Online
Export Citation

21 July 2024 04:36:37

AIP Publishing

AIP Advances
Why Publish With Us?
25 DAYS average time to 1st decision
740+ DOWNLOADS average per article
INCLUSIVE scope
Learn More
AIP Publishing

Convolutional Neural Networks for Text Classification: A Study on Public Activity Restriction

H Anggit Taba[1, a)] and Hari Suparwito[1, b)]

[1] *Informatics Department, Sanata Dharma University, Yogyakarta, Indonesia*

[a)] anggittaba7@gmail.com
[b)] Corresponding author: shirsj@jesuits.net

Abstract. The Public Activity Restriction (PPKM) has become a trending topic on Twitter. There are so many opinions that it is difficult to classify which opinions are pros and cons with PPKM. This study aims to determine the pros and cons of PPKM through a Machine Learning approach using the Convolutional Neural Network (CNN) algorithm. Data have been collected from November 2021 to February 2022. We crawled tweets and obtained 68,953 data. We deliberately choose data labelling without human intervention. We label the tweet data using Vader, a python library, to determine whether the tweets have a negative, positive, or neutral connotation. We proposed using the Wikipedia training Fasttext model to deliver word embedding. The crucial step is to train datasets to get the predictive model. The implementation of the CNN algorithm focuses on layer architecture and parameter tuning variations. Accuracy results of 88% were obtained by using two convolution layers, ReLU and Softmax. Two Pooling techniques, MaxPooling and AveragePolling, were used to reduce the matrix size. It shows how to use machine learning approaches for predicting qualitative data in text processing.

21 July 2024 04:36:37

INTRODUCTION

The implementation of Community Activity Restrictions (PPKM) raises pros and cons. According to research data conducted by the Central Statistics Agency (BPS), 11.53 million people (5.53 per cent) of the working-age population are affected by COVID-19 [1]. 5 out of 10 companies admit that they have problems marketing their products. At the same time, 2 out of 10 companies face difficulties in paying debts, paying workers' wages, paying bills, and producing raw materials [2]. There are, however, community groups who agree to apply PPKM because this is considered effective in suppressing the spread of COVID-19. Based on the results of the SRMC survey reported on the saifulmunjani.com website, nationally, 44% chose to run PPKM strictly even though they had to sacrifice declining income [3].

PPKM impacts several aspects of life, becoming a hot topic discussed on social media, such as Twitter. People use Twitter to express their responses to government regulations, whether they agree or not with PPKM [4]. The government and industry have previously used opinions in the form of public tweets and researchers to gain knowledge in solving problems, including describing actual human behaviour [5]. Then, judging by the number of reviews and comments, Twitter significantly impacts organizations in automatically identifying positive and negative remarks by conducting sentiment analysis [6]. Reviews and comments on social media about PPKM become material for sentiment analysis. Sentiment analysis is a fairly popular topic in the field of Natural Language Processing (NLP) [7] which is quite influential in determining public opinion [8]. Sentiment analysis studies opinions, attitudes, and emotions that exist in an entity. These entities can represent individuals, events, or a topic that can be in the form of a review [9]. This information can help the organization make a good decision [10].

Proceedings of the Transdisciplinary Symposium on Engineering and Technology (TSET) 2022
AIP Conf. Proc. 3077, 040016-1–040016-8; https://doi.org/10.1063/5.0201145
Published under an exclusive license by AIP Publishing. 978-0-7354-4964-0/$30.00

Research on sentiment analysis in Indonesian texts has previously been carried out using several methods such as Naïve Bayes [11], Support Vector Machine, and Decision Tree [12]. In this study, we proposed Convolutional Neural Network (CNN) as a method to perform text processing. As we know, CNN is one of the most effective image classification methods and has a convolutional layer to extract information from text chunks [13-15]. To our knowledge, there is little research done on sentiment analysis using CNN on Indonesian text [6]. Therefore, this study conducted a sentiment analysis of the pros and cons of PPKM on Twitter using the CNN algorithm.
We also proposed using Vader's methods for labelling data in this study. As Vader has only an English dictionary, we first tried translating the Indonesian text to English using google translate. Here, we do not use a human intervention (linguist) to determine whether the texts have negative, positive or neutral connotations.

The rest of the paper is structured as follows. Section 2 discusses the materials and methods. Section 3 describes the sentiment analysis results on PPKM using CNNs classifiers, and we discuss the results. Finally, section 4 concludes the paper and presents future implications.

MATERIALS AND METHODS

Data Collection

In this study, we used a dataset from the twitter.com website using the Twitter API with the crawling method [16] and a tool called tweepy with restrictions on the keywords "ppkm" and "ppkm rules". The data was taken randomly from November 4, 2021, to February 23, 2022. The data is in the form of tweets from the Twitter social media user community who argued about government policies regarding PPKM regulations. The number of data obtained is 68,953 data tweets. The data contains the results of tweets from these social users on PPKM policies stored in CSV format.

21 July 2024 04:36:37

Methods

Data labelling

Data labelling was carried out automatically without human intervention. To provide data on this label, utilize Natural Language Processing (NLP) using a Python library called Vader. The steps performed on the Vader library are: Tweets are collected in a single variable list. The tweet is then translated into English then Vader will analyze the tweet. Each tweet that has been explored is assigned a polarity, whether that data is Positive or Negative.

Data balancing

A data balancing process is carried out to overcome imbalanced data labelling results. Imbalanced data make the classification be built to ignore minority classes. To balance the data, we viewed the results of the number of sentiment labelling, and then sampling was carried out randomly according to the number of minority labelling class data. Next, reduce the majority sentiment class according to the number of types in the minority at random.

Text preprocessing

In the preprocessing steps, data become more structured before the classification process. Figure 1 explains the preprocessing data. Several techniques include text cleansing, case folding, normalization, stemming, tokenizing, and stopwords.

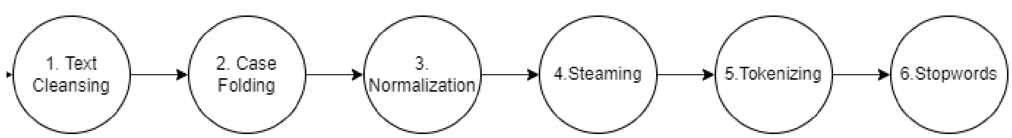


FIGURE 1. The preprocessing steps in NLP.

Text cleansing

The initial text still has a lot of noise, such as URLs, emojis, symbols, etc. All these words would be removed. In text cleansing, we perform reading sentiment text data thoroughly. Sentences containing noise would be removed. After the process is complete, the saved dataset no longer makes noise.

Case folding

Text data is managed into a standard form in the case folding process to obtain the same format. The original or standard text form is to change the existing words into lowercase entirely. The step in case folding is: to read sentiment text data thoroughly. Words containing capital letters would be changed to lowercase. After the process, the data stored is now lowercase.

Normalization

In the normalization stage, sentences that have shortened or extended word spellings become common words, according to the Kamus Besar Bahasa Indonesia (KBBI). The steps in the normalization are: read the results of the previous case folding. Match it with the standard word dictionary. If the word has similarities to the standard dictionary, the word would be changed to the standard form of the word. If there is no similarity, then the word would not be converted into the standard form.

Stemming

In the stemming process, the text would be changed to find the essential words by removing the suffix and prefix. The stemming process involved: rereading the words in the sentence and then comparing them with the words in the normalized dictionary. If the word is the same as that found in the dictionary, then the word is included in the normalization. If the word still has an affix, the affix would be removed, and the text would be normalized according to the dictionary.

Tokenizing

In tokenizing, entire sentences will be separated by detecting spaces. That is in the sentence so that later the sentence can be in token form. In this step, we did: read text data as a whole. If space or word separator is detected, it will be split. Sentences are now stored in the token form of word fractions.

Stopwords

The stopwords goal is to reduce the number of words in a document which can later affect the speed of NLP performance. The process is: reading the existing word dataset. If the word has similarities as in the NLP library, the text is deleted. If there is no similarity, then the text is deleted.

Data splitting

After performing the data preprocessing, data are divided into two, namely training and testing datasets with 75% and 25%, respectively. We used the python library to perform data splitting by using the function_train_test_split in the sklearn library.

Features extraction

Feature extraction, in this study, used Fasttext, a Python library for the dataset with word embedding. At first, pre-trained data is needed. However, a model like the one provided by Google has limitations only in English. So, we can take advantage of the Wikipedia site, which has article sources in more than 250 languages from around the world to be adapted to this research.

21 July 2024 04:36:37

Steps in building a model with Fasttext are: importing pre-trained models by utilizing the article dataset from Wikipedia. Declaration of parameters that will be used in training the Fasttext model. Training Fasttext model to produce word embedding using training data documents, and finally, save the training model with Fasttext for use in classification with CNN.

Experiments

Some experiments were performed to obtain the optimal model prediction by changing the activation layer. The experimental parameters can be seen in Table 1 below:

TABLE 1. CNN parameters for each experiment.

Parameters	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Conv 1D, Layer 1	Sigmoid	Tahn	Relu	Relu
Polling Layer 1	MaxPolling	MaxPolling	MaxPolling	MaxPolling
Conv 1D, Layer 2	Sigmoid	Relu	Softmax	Relu
Polling Layer 2	AveragePolling	AveragePolling	AveragePolling	AveragePolling
Dense Layer 1	Sigmoid	Relu	Relu	Relu
Dense Layer 2	Sigmoid	Tahn	Relu	Relu
Dense Layer 3	Sigmoid	Relu	Relu	Relu
Learning rate	0.00001	0.00001	0.00001	0.00001

The first step in this research is to label the dataset using the Vader module contained in the NLTK library in the Python programming language. In the labelled dataset, the previously unstructured tweets would be processed into more structured data. The tweets are classified into negative and positive classes. The preprocessing results will divide the data for the training set and testing set using a ratio of 75:25, where 75% is for the training dataset and 25% is for the testing dataset. The training data is then used for feature extraction using the FastText word embedding method. Furthermore, the matrix from the feature extraction results will be used for modelling with a CNN algorithm.

Four experiments were carried out in modelling the CNN by changing the parameters to obtain the optimal model. The classifier model will then be evaluated with a confusion matrix to get the accuracy value. Two values were produced: the training and validation value and the accuracy value. From the accuracy value obtained, it can be seen how the performance of the CNN algorithm in conducting sentiment analysis on the topic of Enforcement of Public Activity Restrictions.

21 July 2024 04:36:37

RESULTS AND DISCUSSION

Tweets obtained from Twitter social media would be given a sentiment label. First, we translated tweets into English by using Google Translate Apps. The translation results will then be labelled with a sentiment label using Vader's SentimentIntensityAnalyzer. The dataset labelling process used the Vader library. This library is performed to determine the sentiment of a text based on the sentiment dictionary. The result of this process is the polarity value of each existing tweet. When the polarity value is more than 0.5, the sentiment is one or positive, while if the polarity value is less than 0.5, the sentiment is negative.

The labelling results contain 29,409 negative sentences, 11,045 positive sentences, and 25,441 neutral sentences. In this study, we only used tweets with positive and negative labels. Tweets with neutral labels were removed. Neutral labelling is not used because the process of assessing a sentence to be a neutral sentence is a challenging thing. The labelling results obtained negative sentiments of 29,409 and positive 11,045 data tweets. This data is unbalanced (see figure 2 & 3).

FIGURE 2. Negative and positive tweets.

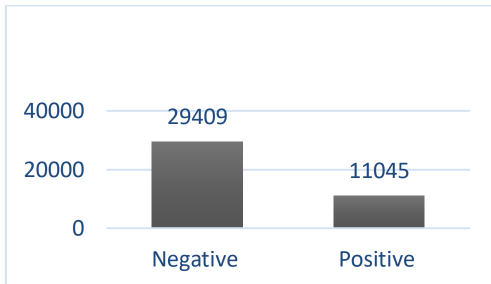


FIGURE 3. Imbalanced data, where the negative sentiment has 72% of the data while the positive is only 28% of the data.

To overcome the imbalance dataset, we used a sampling technique by sampling 10,000 sentences with positive and 10,000 sentences with negative labels [17]. Then we randomized the data so that the data became more varied. The data then needs to be preprocessed to become more structured before the classification process is carried out. The techniques that will be passed in the preprocessing include text cleansing, case folding, normalization, steaming, tokenizing and stopwords. Furthermore, we performed the feature extraction process, where the process will change the tweet dataset into a matrix size.

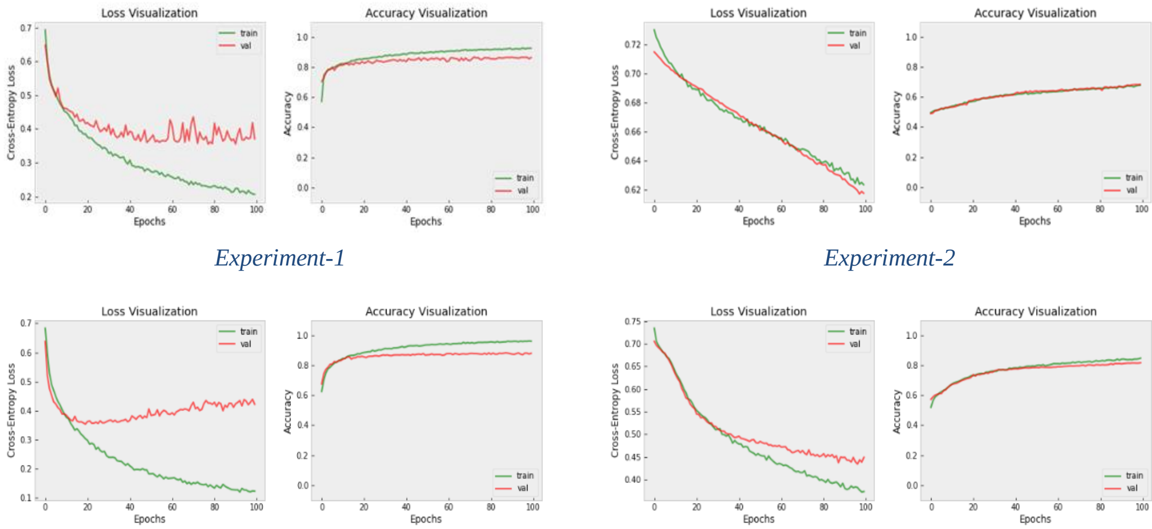
The next step is to complete the classification process using the CNN method. The data used for the classification process is the training dataset. In the classification process, three main layers will be used. The first layer is the convolutional layer, the second layer is the polling layer, and the last is the fully connected layer. Table 2 below shows the results of the experiment. All performance measurements from the experiment were measured using the Confusion Matrix by changing the activation layer. Performance measurement from the third experiment obtained the optimal model classifier prediction results.

TABLE 2. Training dataset accuracy results.

	Label	Precision	Recall	F1 score	Accuracy
Experiment 1	Positive	0.88	0.849	0.864	0.866
	Negative	0.852	0.882	0.867	
Experiment 2	Positive	0.693	0.772	0.707	0.699
	Negative	0.706	0.678	0.691	
Experiment 3	Positive	0.875	0.88	0.877	0.916
	Negative	0.877	0.872	0.875	
Experiment 4	Positive	0.767	0.934	0.843	0.824
	negative	0.914	0.713	0.801	

The classifier model that has been built is then compiled with Adam's optimization with the parameters 'binary_crossentropy' and metrics 'accuracy' with 'learning rate = 0.0001'. When we applied the testing dataset to model classifier prediction, the accuracy dropped slightly to 86%.

21 July 2024 04:36:37



Experiment-1

Experiment-2

Experiment-3

Experiment- 4

FIGURE 4. Lost and accurate visualization from the experiments.

To show whether Vader's method of labelling data is accurate, the following result compares the actual labelling of data and the prediction of data labelling carried out by Vader. It can be seen in the actual and predicted labelling that using the third test experiment with ReLu and Softmax activation functions can predict the labelling results well, as shown in Figure 5.

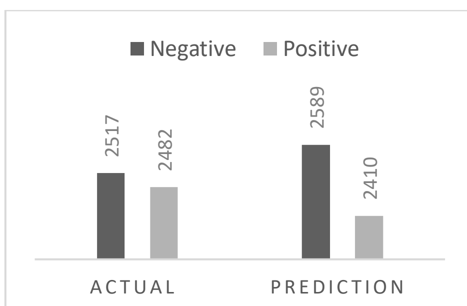


FIGURE 5. The comparison between the actual and the prediction labelled data using Vader methods.

After testing four experiments using the Convolutional Neural Network method with the PPKM topic, the accuracy results are pretty optimal. The results of the comparison of each experiment can be seen in Table 2. The evaluation results obtained in this study using 2 convolution layers, 2 polling layers, and the addition of 2 normalization batches, 2 dropouts, and fully-connected layers can be said to be quite good.

In Figure 4, we had the experiment with 2 different learning rate tunings. The second and fourth experiments with lr = 0.00001 and the first and third experiments with lr = 0.0001. In the third experiment, the average values for Train Acc, Loss Train, Validation Acc, and Loss Acc are still the most optimal compared to other experiments. However, if we look at the comparison graph of the training model results, the first and third experiments with lr = 0.0001 can be said to be overfitting. This situation can occur because it uses a large learning rate; the model quickly learns all the details, including the noise contained in the data, and tries to include all the data into the model training process.

Overfitting can be overcome by using the correct learning rate and parameter combinations. As in experiment 2 and experiment 4, it can be seen that the graphs for Train Acc, Loss Train, Validation Acc, and Loss Acc show that

21 July 2024 04:36:37

the lines are pretty close and stable. Therefore, it can be said that the addition of batch normalization and dropout and the use of learning rates with the proper parameter tuning on the layer used can affect maintaining the stability of the model so that it does not experience overfitting [18]. In determining the activation function in the combination of 2 convolution layers, it is also crucial to note that this needs to be done to overcome the unstable CNN model performance.

The results obtained in the labelling process with Vader, the actual data from the data testing data, will be compared with the predicted data, which is the data obtained by modelling using CNN. As in the example of the actual labelling results with Vader and the predicted labelling results with CNN, as shown in Figure 5. The actual and predicted labelling are from the third test experiment using the ReLu and Softmax activation functions. These parameters can predict the labelling results of tweets about PPKM rules using CNN well. The actual and prediction results using Vader and the label prediction results using CNN did not have a significant difference, less than 0.015. This proves that the CNN algorithm can predict sentiment analysis of tweets about PPKM well. In this study, single data testing was not carried out because this study focuses on testing whether using CNN can perform sentiment analysis well or not by comparing the actual data from labelling with Vader with predictive data from modelling results with CNN [19].

CONCLUSION

The study on Public Activity Restriction (PPKM) with the machine learning approach using the CNN algorithm has been done. The performance shown by 91% and 86% accuracy for training and testing datasets, respectively, using the CNN algorithm in the topic of PPKM rules on Twitter has a good sentiment analysis. The optimal result is 81% from the third test experiment using two convolution layers, i.e., ReLU and Softmax, 2 Polling applied: MaxPolling and AveragePolling, BatchNormalization Layer 1 and 2, Dropout layer 1 is 0.3, Dropout layer 2 is 0.2, and learning rate = 0.0001. The addition of batch normalization and dropout and the use of learning rates by tuning the proper parameters on the layer used can affect maintaining the stability of the model so that it does not experience overfitting.

Further studies include labelling data after text preprocessing so that the text becomes more structured, and it is hoped the data labelling process using Vader can be more optimal. More experiments are needed by performing hyperparameter tuning in the CNN model architecture so that the CNN model can automatically choose the most optimal architecture and minimize overfitting. The use of a large dataset requires many slang words in making a normalization dictionary so that the resulting text does not have much noise so that the model's The study on Public Activity Restriction (PPKM) with the machine learning approach using the CNN algorithm has been done.

21 July 2024 04:36:37

ACKNOWLEDGMENTS

The authors thank the Center for Research and Community Service of Sanata Dharma University for funding this research.

REFERENCES

1. Badan Pusat Statistik, “Keadaan Ketenagakerjaan Indonesia Februari 2022,” no. 34. (2022).
2. Badan Pusat Statistik, “Perilaku Masyarakat Pada Masa PPKM Darurat.” (2021).
3. T. Krisdiyanto, “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers,” J. Has. Penelit. Ilmu Komput. dan Teknol. Inf., vol. **7**, no. 1, pp. 32–37 (2021).
4. S. F. Pane, “Pemodelan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter,” J. T. Informatika, P. P. Indonesia, vol. **05**, no. 01, pp. 12–20. (2022).
5. M. S. Saputri, R. Mahendra, and M. Adriani, “Emotion Classification on Indonesian Twitter Dataset,” in *2018 Int. Conf. Asian Lang. Process.*, pp. 90–95 (2018).
6. H Juwiantho, EI Setiawan, J Santoso, and M. H. Purnomo “Sentimen Analysis Twitter Bahasa Indonesia Berbasis WORD2VEC Menggunakan Deep Convolutional Neural Network,” vol. **7**, no. 1, pp. 181–188 (2020).

7. S. Rani and P. Kumar, “Deep Learning Based Sentiment Analysis Using Convolution Neural Network,” Arab. J. Sci. Eng., vol. **44**, no. 4, pp. 3305–3314 (2019).
8. S. A. Anwar Hridoy, M. T. Ekram, M. S. Islam, F. Ahmed, and R. M. Rahman, “Localized twitter opinion mining using sentiment analysis,” Decis. Anal., vol. **2**, no. 1 (2015).
9. W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” Ain Shams Eng. J., vol. **5**, no. 4, pp. 1093–1113 (2014).
10. A. S. M. Alharbi and E. de Doncker, “Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information,” Cogn. Syst. Res., vol. **54**, pp. 50–61 (2019).
11. L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, “Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier,” Int. J. Inf. Eng. Electron. Bus., vol. **8**, no. 4, pp. 54–62 (2016).
12. K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” J. Inf. Syst. Manag., vol. **3**, no. 1, pp. 1–7, (2021).
13. S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, “CNN for situations understanding based on sentiment analysis of twitter data,” Procedia Comput. Sci., vol. **111**, pp. 376–381, (2017).
14. S. Rani and P. Kumar, “Deep Learning Based Sentiment Analysis Using Convolution Neural Network,” Arab. J. Sci. Eng., vol. **44**, no. 4, pp. 3305–3314 (2019).
15. U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, “Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM),” Wirel. Pers. Commun., pp. 1–9 (2021).
16. B. R. Aditya, “Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining,” J. INFOTEL **-** Inform. Telekomun. Elektron., vol. **7**, no. 2, p. 93 (2015).
17. A. Sonak and R. A. Patankar, “A Survey on Methods to Handle Imbalance Dataset,” Int. J. Comput. Sci. Mob. Comput., vol. **4**, no. 11, pp. 338–343, (2015).
18. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov,. “Dropout: a simple way to prevent neural networks from overfitting,” J. of Mach. Learn. Res., **15**(1), pp.1929-1958 (2015).
19. S. Elbagir, and J. Yang. Twitter sentiment analysis using natural language toolkit and VADER sentiment. Paper presented at the Proceedings of the international multiconference of engineers and computer scientists. (2019).

21 July 2024 04:36:37