PAPER • OPEN ACCESS

Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset

View the article online for updates and enhancements.

Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset

Ahmad Fathan Hidayatullah[1], Siwi Cahyaningtyas[2], Anisa Miladya Hakim[3]

Department of Informatics, Universitas Islam Indonesia
Jalan Kaliurang KM 14,5 Sleman Yogyakarta Indonesia

E-mail: [1]fathan@uii.ac.id, [2]siwictyas@gmail.com, [3]anisanisa.nh@gmail.com

Abstract. Due to the utilization of Twitter by Indonesian politicians ahead of the Indonesian presidential election 2019, Indonesian people have given diverse responses and sentiments to the politicians. This study aims to classify sentiment on Indonesian presidential election 2019 tweet data by using neural network algorithms and obtain the best algorithm. In our study, we train our dataset using some variants of deep neural network algorithms, including Convolutional Neural Network (CNN), Long short-term memory (LSTM), CNN-LSTM, Gated Recurrent Unit (GRU) -LSTM and Bidirectional LSTM. Moreover, as a comparison with our deep learning model, we also train our dataset using other traditional machine learning algorithms, namely Support Vector Machine (SVM), Logistic Regression (LR) and Multinomial Naïve Bayes (MNB). Our experiments showed that Bidirectional LSTM achieved the best performance with the accuracy of 84.60%.

1. Background

Currently, with the popularity of social networking media which has become a powerful tool to influence people and share opinion to the public. For example, Twitter is one of social media that often be used by politician for political campaign in Indonesia [1,2]. Among several social networking sites, Twitter is one of the most well-known tools used by politicians [2]. Twitter has been used by many Indonesian politicians to attract public sympathy and increase popularity ahead of the election [3]. Through Twitter, politicians post their opinion, thoughts and activities to attract and influence people to vote them in the general election. In addition, Twitter has been utilized as a data source to analyze public sentiment in the election in some countries such as Indonesia, India, Germany, United States, UK and Bulgaria [1,4–8].

The Indonesian presidential which held on April 17, 2019, is getting more interesting considering the two presidential candidates in this year are the same name in the previous presidential election in 2014. In the 2019 Indonesian presidential election, the battle of public figures and supporters of presidential candidates on Twitter is very exciting to discuss. It can be seen from several Twitter hashtags from two different sides such as #2019GantiPresiden (means: 2019 replace the president) and #Jokowi2Periode (means: Jokowi 2 periods). The tweets posted by those politicians will certainly get diverse responses and sentiments from netizens. Therefore, knowing the sentiment towards the tweet posted is necessary. The goal of sentiment analysis is to identify the sentiment polarity of a given sentence or statement into three polarity categories, namely positive, negative, or neutral [9,10].

Knowing the sentiment polarity of tweets posted by politician can illustrate the public response toward a specific politician.

However, sentiment analysis task in for short texts like Twitter is quite challenging because of the limitation of contextual information in the texts [11]. Therefore, several algorithms have been continuously developed to obtain the best sentiment analysis model result. Several previous studies have proven that traditional machine learning method such as Support Vector Machine (SVM) and Naïve Bayes Classifier (NBC) have shown great performance for the sentiment analysis task. Almatrafi et al [4] proposed a location-based sentiment analysis system using Twitter dataset to discover several topic trends during Indian general elections in 2014 using Naïve Bayes Classifier. Hidayatullah and Sn [3] applied both Support Vector Machine and Naïve Bayes Classifier to conduct sentiment analysis and category classification on Indonesian politicians using Twitter dataset. The study revealed that Support Vector Machine using term frequency as a feature obtained better accuracy than Naïve Bayes. Hasan et al [12] also employed both SVM and Naïve Bayes algorithms for Twitter sentiment analysis task. They combined the two algorithms with sentiment lexicon. In addition, the sentiment analysis task was conducted by comparing several sentiment lexicons with three libraries, including SentiWordNet, W-WSD and TextBlob.

Currently, with the increasing numbers and size of data, researchers have been developing a deep neural networks approach to be applied in the sentiment analysis task. Santos and Gatti [13] conducted perform sentiment analysis by applying deep convolutional neural networks architecture joined with three different level representations, such as character-level, word-level and sentence-level representations. The proposed method has been applied on two different corpora, namely Stanford Sentiment Treebank and Stanford Twitter Sentiment corpus. Ali, et al [14] carried out sentiment analysis using movie reviews dataset by utilizing four deep learning algorithms, namely Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long short-term memory (LSTM) and CNN_LSTM. The experiments have shown that CNN_LSTM outperformed other deep learning methods. Jianqiang, et al [10] proposed a method by combining Glove (Global Vectors) as word representation and Deep Convolutional Neural Networks (DCNN) to conduct sentiment analysis utilizing Twitter dataset. Their study revealed that the Glove-DCNN method performed better than bag-of-words model with Support Vector Machine (SVM) classifier. Wang et al [11] performed sentiment analysis by applying two Deep Learning algorithms, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The study revealed that the model from the combination between CNNs and RNNs achieved higher accuracy than other existing models.

This study applies several different deep neural network algorithms to classify the sentiment on Indonesian Presidential Election 2019 Twitter dataset. In addition, this work aims to classify the sentiment of the dataset into positive and negative. To obtain the best model, we train our dataset using some different deep neural network algorithms, including Convolutional Neural Network (CNN), Long short-term memory (LSTM), CNN-LSTM, Gated Recurrent Unit (GRU)-LSTM and Bidirectional LSTM. As a comparison, we also train our dataset using other well-known traditional machine learning algorithms, namely Support Vector Machine (SVM), Logistic Regression (LR) and Multinomial Naïve Bayes (MNB).

The remainder of this paper is organized as follows: Section 2 describes deep learning algorithms; in Section 3, we explain our research methodology; Section 4 describes result and discussion of our study; In Section 5, the conclusion of this study is presented.

2. Deep Learning Algorithm

2.1. Convolutional Neural Network (CNN)
Convolutional Neural Network (CNN) is one of deep learning method that is developed from multilayer perceptron (MLP). CNN consists of three main layers such as convolution, pooling and fully connected layer. The convolution layer is the primary CNN building block which consists of a set independent filters. Pooling layer aims to reduce the complexity of convolved layer representation. The vector results

from several convolutions and pooling operations on the multilayer perceptron are known as fully-connected layers that are used to do a particular task such as classification.

2.2. Long-short Term Memory (LSTM)

LSTM is an improved model of RNN that is developed to overcome long time-dependencies problem. An LSTM cell consists of three gates: input gate, forget gate and output gate. These gates are produced by a sigmoid function over the input sequences and the previous hidden state [15]. These gates receive information across LSTM through cell state. The fundamental component of LSTM is cell state as a transport information from memory from previous block to memory from current block.

2.3. Gated Recurrent Unit (GRU)

GRU (Gated Recurrent Unit) is similar to LSTM that is able to deal with long term dependencies, but they have fewer parameters [16]. GRU has gating units like LSTM, except memory cells. GRU aims to overcome the vanishing gradient problem by using update and reset gate. The update gate assists the model to identify the past information from previous steps should be passed along to the future. The reset gate is utilized from the model to choose the amount of the past information to forget.

2.4. Bidirectional LSTM

Bidirectional LSTM (BLSTM) is an improved model of LSTM that connects two hidden layers to the same output. The objectives of BLSTM are to enhance performance of the model on sequential problem. Different from LSTM that using input are only from the past, BLSTM using inputs into two ways, past to future and future to past.

3. Methodology

3.1. Data Collection

Dataset collected by two different periods of time, before and after the presidential election. As for the period before the presidential election, we crawled the tweet five times, in January, February, March and April 2019. The data were crawled the day after the debates of presidential candidates and vice presidential candidates. We used tweepy as a library provided by Python obtained 115,931 raw tweets.

3.2. Data Labelling

Manually labelling data is expensive and time-consuming. To reduce process on labelling data, we can use pseudo-labelling to labelling our data. Pseudo-labelling is a semi-supervised technique to labelling data with still requires labelled data. Pseudo-labelling implements machine learning algorithms for building models. The model then used to generate a pseudo-label as a class of unlabelled data. In this work, we used Multinomial Naïve Bayes as a model. First, we used 1369 raw Twitter sentiment dataset that have been labelled manually as a training set [3]. Training set that was labelled used to train our model using Multinomial Naïve Bayes. Model that build by training set then used to predict labels (pseudo-label) on unlabelled data. Finally, we concatenate the training set with the pseudo-labelled set and use both datasets to retrain our model. The process is schematically shown in Figure 1.
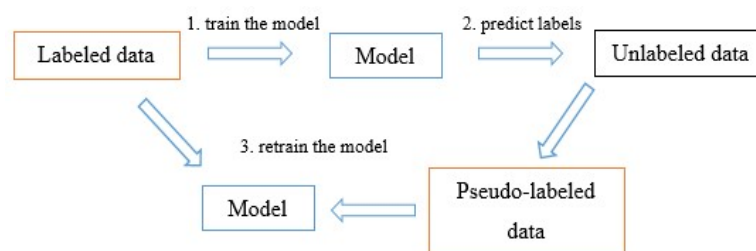
Figure 1. Pseudo-labelling.

3.3. Pre-processing

We applied some pre-processing techniques to our dataset. We omitted several characters from the Twitter data, including ASCII characters; username, hashtags, URLs, retweet; punctuations; duplicate characters in a word; excessive space; stop words; and tweet duplication. In addition, we also performed other pre-processing tasks such as case folding; word normalization; and stemming.

3.4. Sentiment Analysis Model Variations

In this work, we created our sentiment analysis model by performing several experimental model variations. We built the first model variations using three traditional machine learning algorithms, such as SVM, Logistic Regression and Multinomial Naïve Bayes. We also compared the accuracy result by applying Term Frequency-Inversed Document Frequency (TF-IDF) as a feature and not applying TF-IDF on those three algorithms. The second sentiment analysis model variations were built using several deep neural network algorithms, including Convolutional Neural Network (CNN), Long short-term memory (LSTM), CNN+LSTM, Gated Recurrent Unit (GRU)+LSTM and Bidirectional LSTM.

4. Result and Discussion

In this study, we used Holdout Method to split the data into 2 parts, 80% for training and 20% for testing. In addition, we also added validation data which obtained from 10% of the training data. Table 1 shows the accuracy results from traditional machine learning algorithms. Based on the accuracy results, SVM outperformed Multinomial Naïve Bayes and Logistic Regression with the accuracy of 84.04%. In addition, application of TF-IDF can improve the accuracy of all the methods.

Table 1. Results of traditional machine learning algorithms

| Method	TF-IDF Feature	Accuracy (%)
SVM	Yes	**84.04**
	No	83.23
Logistic Regression	Yes	83.15
	No	82.74
Multinomial Naïve Bayes	Yes	82.13
	No	82.08

Table 2 shows the accuracy results from deep learning algorithms. We performed training using five different deep neural network algorithms for our dataset with 5 epochs and batch size of 128. In the experiments, we split our dataset into training and testing data with the ratio of 80:20. In addition, we took 10% of our training set for validation set. Based on the accuracy results, Bidirectional LSTM obtained the best result of 84.60%. As we can see, the accuracy of Bidirectional LSTM is higher 0.56% compared to SVM. This achievement shows that Bidirectional LSTM has better performance than the SVM algorithm.

Table 2. Results of deep learning algorithms

Method	Accuracy(%)
LSTM	84.20
CNN	84.05
CNN+LSTM	84.30
GRU+LSTM	84.50
Bidirectional LSTM	**84.60**

Figure 2 shows the layer architecture of Bidirectional LSTM as the best model. Firstly, we put the embedding layer as our input layer which receive three arguments, input dimension, output dimension and input length. Input dimension is the size of vocabulary in the dataset. Output dimension defines the size of the output vector for respective words from this layer. Input length assigns the length of input

sequences. The next layer is a dropout layer which aims to overcome overfitting of the model. Dropout layer regularizes a deep neural network by randomly removing inputs to a layer using probabilistic calculation. After that, we added two stack Bidirectional LSTM layers with 256 hidden nodes. In addition, we added a recurrent dropout parameter argument into both Bidirectional LSTM layers to drop the linear transformation of the current state. Moreover, we put dropout layer after the two Bidirectional LSTM layer. Finally, we put the dense layer as the output layer.
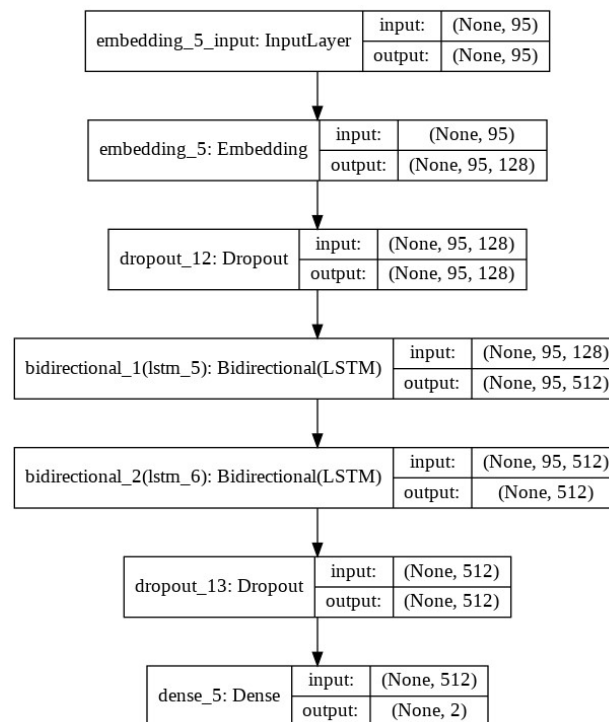
Figure 2. Bidirectional LSTM Architecture

5. Conclusion
In this work, we have presented sentiment analysis on Indonesian presidential election 2019 Twitter dataset using deep neural network algorithms. Overall, our experiments showed that deep neural networks algorithms had better accuracy compared with other three traditional machine learning algorithms (SVM, Logistic Regression and Multinomial Naïve Bayes). In addition, as for deep learning approach, we proposed five different methods, including LSTM, CNN, CNN+LSTM, GRU+LSTM and Bidirectional LSTM. According to our experiments, Bidirectional LSTM outperformed other deep learning methods with the accuracy of 84.60%.

Acknowledgement

References
[1]   Alashri S, Parriott E, Kandala S S, Awazu Y, Bajaj V and Desouza K C 2018 The 2016 US Presidential Election on Facebook: An Exploratory Analysis of Sentiments *Proc. of the 51st Hawaii Int. Conf. on System Sciences* pp 1771–80
[2]   Bansal B and Srivastava S 2018 On predicting elections with hybrid topic based sentiment

analysis of tweets *Procedia Computer Sci.* **135** pp 346–53

[3]   Hidayatullah A F and Sn A 2014 Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada Twitter *Seminar Nasional Informatika 2014 (semnasIF 2014)* pp 115–22

[4]   Almatrafi O, Parack S and Chavan B 2015 Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014 *Proc. of the 9th Int. Conf. on Ubiquitous Information Management and Communication - IMCOM '15* (Bali, Indonesia: ACM Press) pp 1–5

[5]   Budiharto W and Meiliana M 2018 Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis *J. of Big Data* **5**

[6]   Heredia B, Prusa J D and Khoshgoftaar T M 2018 Location-Based Twitter sentiment Analysis for Predicting the U.S. 2016 Presidential Election *The Thirty-First Int. Florida Artificial Intelligence Research Society Conf. (FLAIRS-31)* pp 265–70

[7]   Burnap P, Gibson R, Sloan L, Southern R and Williams M 2016 140 characters to victory?: Using Twitter to predict the UK 2015 General Election *Electoral Studies* **41** pp 230–3

[8]   Smailović J, Kranjc J, Grčar M, Žnidaršič M and Mozetič I 2015 Monitoring the Twitter sentiment during the Bulgarian elections *2015 IEEE Int. Conf. on Data Sci. and Advanced Analytics (DSAA)* pp 1–10

[9]   Tang D, Wei F, Qin B, Liu T and Zhou M 2014 Coooolll: A Deep Learning System for Twitter sentiment classification *Proc. of the 8th Int. Workshop on Semantic Evaluation (SemEval 2014)* (Dublin, Ireland: Association for Computational Linguistics) pp 208–12

[10]   Jianqiang Z, Xiaolin G and Xuejun Z 2018 Deep Convolution Neural Networks for Twitter sentiment analysis *IEEE Access* **6** pp 23253–60

[11]   Wang X, Jiang W and Luo Z 2016 Combination of Convolutional and Recurrent Neural Network for sentiment analysis of short texts *Proc. of COLING 2016 The 26th Int. Conf. on computational linguistics: Technical papers* pp 2428–37

[12]   Hasan A, Moin S, Karim A and Shamshirband S 2018 Machine learning-based sentiment analysis for Twitter accounts *Mathematical and Computational Applications* **23** p 11

[13]   Dos Santos C N and Gatti M 2014 Deep Convolutional Neural Networks form *Proceedings of COLING 2014, the 25th Int. Conf. on Computational Linguistics: Technical Papers* (Dublin, Ireland: Dublin City University and Association for Computational Linguistics) pp 69–78

[14]   Ali N M, El Hamid M M A and Youssif A 2019 Sentiment analysis for movies reviews dataset using deep learning models *Int. J. of Data Mining & Knowledge Management Process* **09** pp 19–27

[15]   Yin W, Kann K, Yu M and Schütze H 2017 Comparative Study of CNN and RNN for Natural Language Processing *Preprint arXiv:1702.01923 [cs]*

[16]   Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation *Preprint arXiv:1406.1078 [cs, stat]*