Depression Detection on Twitter Social Media Platform using Bidirectional Long-Short Term Memory

Andre Agasi Simanungkalit [1], Warih Maharani[2], Prati Hutari Gani[3]

[1,2,3]Bachelor of Informatics Study Program, Informatics Faculty, Telkom University Bandung

Abstract

Depression is one of the mental disorders that are often experienced by a person in daily life. Social media platforms is a new thing as an alternative to tell stories and express current feelings by people today. Twitter is one of the social media that is often used to express feelings and opinions through tweets posts, including tweets that contain hate speech which indirectly shows symptoms of depressive disorder through statements uploaded. It also requires modeling that can recognize users with the potential to experience depression so that they can get initial treatment. This can be implemented using the BiLSTM (Bidirectional Long Short-Term Memory) method and the Word2Vec feature. It is also needs a modeling that can recognize the users who have the potential to experience depression so that they can get treatment at the beginning. This can be implemented using the BiLSTM (Bidirectional Long Short-Term Memory) method and the Word2Vec feature.

Keywords : Twitter, Mental Illness, BiLSTM, Depression

1. Introduction

1.1 Background

Depression is a disorder of a person's emotional characterized by negative thinking about everything (Sabarisman, 2016). Quoting from Jaka Arya Pradana (2016) who stated that depression is also referred to as an invisible disorder (Dirgayunita, 2016). Based on WHO data in 1980, almost 20% - 30% of hospital patients in developing countries experience emotional mental illnes, such as depression. This disorder can be experienced by all age groups (W. Yoeyoen, 2018). Data from *Riskesdas* 2018 results show that depressive disorders have started to occur since the age range of adolescents (15-24 years), with a prevalence of 6,2%. The prevalence pattern of depression is getting increase, the highest at 75+ years old at 8,9%, 65-74 years old at 8,0%, and 55-64 years old at 6,5%.

In the current era of the Industrial Revolution 4.0, technology continues to develop and supports life, one of which is in expressing oneself. There are many ways how people express themselves in this digital era, such as expressing the content of their feelings on social media platforms. The dominant social media platform chosen is Twitter. On this social media, depression person can pour his anxiety on a personal account or base account with a 280-character tweet that can certainly be uploaded every day (Zi'ni, 2020). In addition, another Twitter phenomenon that is currently well known is Autobase (Maulina, 2021). [5] It is an account that allows Twitter users to send messages or called *menfess* (mention and confess) in the form of information or questions anonymously through direct messages (dm) of the account's profile which will automatically be spread to their timeline (C. Sipahutar, A. Poerna, 2020). Reporting from the CSSMORA UINSA, various menfess websites that is commonly found on Autobase accounts is information about certain topics, general questions, and confide personal problems. One of the characteristics of a perpetrator who has a depressive mental disorder can be easily seen through his dominant post tweets using negative words, such as feelings of worthless, guilty, and hating himself.

---

[*] Corresponding author.
E-mail address: andreagasi@student.telkomuniversity.ac.id

Previously, research related to the detection of a person's depression has been widely carried out. In the study conducted between the Support Vector Machine (SVM) method which was compared with Bidirectional Encoder Representations from Transformers (BERT) and A Lite BERT (ALBERT), the performance of the BERT model has the highest value with an accuracy value of up to 75%. Another related study between the Naïve Bayes (NB) method and the Support Vector Regression (SVR) which showed that the performance of the SVR method had the highest accuracy value of 79,7% (Isnain et al., 2020). Then, Bidirectional Long Short Term Memory (BiLSTM) is a two-way of LSTM that is able to capture information without neglecting the context and meaning of a sentence. The Bidirectional Long Short Term Memory (BiLSTM) approach is one of the commonly used variants of the LSTM model. In this model variant, the incoming information is only concerned in the last word because it only reads sentences with a one-way version (Isnain et al., 2020). The BiLSTM model variant is a development of the LSTM model variant which technically uses two separate LSTM layers (forward and backward). These two layers read the incoming information from two directions at once and the output is generally combined into one. Through this layer, the model can learn past information (past) and future information for each input sequence. BiLSTM's ability to read incoming information from two directions makes the performance of this method more accurate and higher than all standard baseline models and LSTM because tweets can be understood sequentially every word with the highest accuracy reaching 94,12%. Another study related to the use of the BiLSTM method is in the detection of hate speech which resulted in the highest accuracy of 94.66%. Therefore, this study uses the BiLSTM method since the previous study has a high level of performance and accuracy. This research is expected to detect people who have depressive disorders through tweets of a Twitter user.

1.2 Topics and Limitations

The formulation of the problem is able to detect a person who is indicated to have symptoms of depression on the social media platform Twitter and factors that affect the BiLSTM model in detecting depression and recognize the level of accuracy given by BiLSTM method so that early treatment can be done on the user. Meanwhile, the limitations of the problem in this study involved Twitter tweet posts data used Indonesian language taken from user tweets on Twitter social media. The model used in detecting depression is Depression, Anxiety, and Stress Scales (DASS-42). It did not detect *emoji* and typo problems.

1.3 Objective

The objective of this study is to detect depression in a person on Twitter social media by using the BiLSTM method and Word2vec extraction. Table 1 presents the correlation between objectives, testing, and oconclusions.

Table 1. The Correlation between Objectives, Testing, and Conclusion

No	Objective	Testing	Conclusion
1	Depression Detection using Bidirectional Long Short Term Memory method	Testing performed by inputting a user's tweet dataset	Models can display results in detecting user tweets
2	What factors influence the model in detecting depression on Twitter social media	Through Confusion Matrix Evaluation	The results showed that the testing accuracy point was obtained by 70.59% with an F-1 Score of 80%
3	How is the level of accuracy in detecting depression using the BiLSTM method	Testing performed with and without using the BiLSTM method	Analysis shows that tuning BiLSTM parameters may affect results. In addition, split data experiments and Word2vec affect the results obtained.

1.4 Writing Organization

The Final Project Journal has the following sections:
- Chapter 1 is an Introduction that contains a description and problem formulation as the topic of Final Thesis.

- Chapter 2 is a literature review in explaining the theory and review of papers or journals relating final thesis raised and supporting its topic.
- Chapter 3 is an explanation of the built system and an overview of the Bidirectional Long Short Term Memory method.
- Chapter 4 is evaluation that will present the results and the analysis of the created system.
- Chapter 5 is the conclusion and suggestion of this study.