Depression Detection Using Hybrid Model BiLSTM – CNN with Glove as Feature Expansion in X

1st Bintan Dinar Artamevia
School Of Computing
Telkom University
Bandung, Indonesia
bintandinar@student.telkomuniversity.ac.id

2nd Erwin B udi Setiawan
School Of Computing
Telkom University
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

Abstract*— Depression is a significant global health issue, with increasing prevalence among various age groups, especially severed by the COVID-19 pandemic. Early detection of depression is crucial for effective intervention because of its traditional methods may be time-consuming. This study proposes a hybrid deep learning model, BiLSTM-CNN, combined with GloVe word embeddings and TF-IDF feature extraction, to detect depression from textual data in the Indonesian language. The dataset used in this study is collected from X consist of 50.523 tweets then manually labeled using a majority vote system. Various scenarios were evaluated, including testing the best split ratios, n-gram, maximum feature, then apply GloVe as feature expansion with three different built corpuses consist of 50.523 data from tweet, 100.594 data from indonews, and the combination of both. The result of this study is that BiLSTM model earned the highest accuracy of 84.2% which has increased by 0.67% from base model because of the model architecture, while the proposed BiLS`TM-CNN hybrid model attained an accuracy of 83.79% that incline 0.244% compared to the base model.

Keywords*— Depression Detection, BiLSTM-CNN, GloVe, TF-IDF

I. INTRODUCTION

Mental illness can result in a loss of energy and interest that leads to a negative impact on an individual's workplace behavior, even increasing the risk of suicide [1]. Depression is a mental disorder characterized by feelings of hopelessness, lack of motivation, negative mood fluctuations, and a diminished interest in daily physical, mental, and social activities. This condition can lead to significant emotional distress and may result in physical changes within the body of the affected individual [2]. This particular condition may suffered by all age groups. In adult, depression can be more dangerous than other ages. Furthermore, women are more likely suffer depression than men. One out of three women will show the depression symptoms in her life [3]. During the COVID-19 pandemic, the number of depression patients increase significantly to 25% [4]. Approximately 280 million people worldwide suffer from depression [2]. More than 800,000 death cases are caused by depression anually. Between 15 – 29 years old, depression be a second most common cause of death [5]. If it is not treated appropriately, depression can affected suicide which a fourth most common cause of death based on WHO [6].

Moreover, the individuals experiencing depression are less likely to seek the professional help due to the stigma associated with mental illness which resulting a significant proportion of patients with depression do not receive the appropriate treatment or sufficient time for recovery [1]. Moreover, 70% of them not receiving appropriate treatment due to limitations in resources and medical care [7]. Depression can be diagnosed by medical history, pyschological evaluations, lab tests, or physical exam [8]. Interview and assessment system such as Hamilton Rating Scale for Depression is the most used tools for depression detection. This traditional methods are mostly time consuming, ineffective, and expensive [2]. Thus, Early identification in a early curable staged and proper treatment can significantly decrease an individual's mortility [9].

In the current technology-driven era, individuals exhibiting psychological symptoms, such as depression, are increasingly active on social media platforms [9]. It is supported by statistics on social media usage reported by DemandSage, over 63% of the global population uses social media. One of the most widely used social platforms today is X, formerly known as Twitter. A growing number of users are utilizing X as a medium for expressing their opinions, communicating with others, and sharing their feelings [5]. Consequently, data from social media has the potential to be a significant source of information for detecting depression, as users tend to be more honest and open about their emotions and thoughts [8]. The data retrieved from social media can be classified by the sentiment expressed using natural language processing.

Based on the informations above, there are several studies proposing the depression detection system using machine learning. Thus, study [2] analyze the comparation between textual model CNN, audio model CNN, and hybrid data model which is a combination of audio and text features. On hybrid data model, LSTM and BiLSTM layer are applied. Through this study, they obtained the highest accuracy 98% for the audio model CNN while the textual model got 92% for accuracy. For the hybrid data model, they reach 80% accuracy for the LSTM and 88% for BiLSTM. Though the preprocessing in the BiLSTM model takes more time, this model actually has a higher learning rate among all models. In other research, they also propose the hybrid deep learning using feature-rich CNN as a features extraction and BiLSTM to processing the textual data. In addition, they compare the proposed model with the CNN and RNN baseline. This study attained 94% for the hybrid model accuracy, received the highest accuracy compared to the CNN and RNN baseline model [10]. Instead of using the hybrid model as used in study [10],