Depression Detection on Social Media X by Text Analysis with Attention-based CNN-BiLSTM using FastText

1st Chikandra Permata Zahira
School of Computing
Telkom University
Bandung, Indonesia
chikandrapz@student.telkomuniversity.ac.id

2nd Erwin Budi Setiawan
School of Computing
Telkom University
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

***Abstract*—The exponential growth of the digital domain has rendered contemporary society reliant on social media. Consequently, the manner in which many individuals engage with social media can manifest indications of distress, such as depression. Social media X is a popular platform that can contain all the outpourings of its users called tweets. With the increasing cases of depression, it is important to be able to detect depression early. This research contributes to combining a hybrid deep learning method to detect depression on social media X with TF-IDF as a feature extraction that plays a role in measuring the importance of words in each user's tweet, FastText as feature expansion to improve word representation and finding semantic similarities, and attention mechanisms as optimization in adding weights. With a total of 50,523 tweet data, a similarity corpus of 100,594 was constructed. Based on the result, using the attention mechanism the BiLSTM model achieved 84.25% accuracy, a 2.03% increase from the baseline, and the CNN-BiLSTM hybrid model achieved 83.43% with an accuracy increase of 0.99% from the baseline.**

Keywords: attention mechanism, depression, FastText, hybrid deep learning, TF-IDF.

I. Introduction

In contemporary society, social media has become a ubiquitous aspect of daily life, with a notable impact on stress and depression levels [1]. People use a variety of social media platforms, such as Facebook, X, TikTok, and Instagram, to disseminate their perspectives, concepts, experiences, and emotions. Currently, X is among the most prevalent social media platforms, providing users with a platform to publish original written content [2]. However, a considerable number of users have also reported feelings of unease and distress associated with the content, including symptoms of depression. Depression is a global mental health concern that in its most severe form results in self-mutilation or self-harm [3]. A 2022 survey, the Indonesia National Adolescent Mental Health Survey (I-NAMHS), revealed that 17.95 million adolescents in Indonesia had a diagnosed mental disorder. Among them, 1.0% had major depressive disorder, One of the most common mental health conditions experienced by Indonesian adolescents [4].

The publicly created user data on social media is of significant importance to the field of health technology, as the patterns that can be discerned from such data will prove to be highly useful. One such application is the automatic detection of an individual's mental health status, including conditions such as depression [2]. The presence of feelings of worthlessness, hopelessness, disinterest in pleasant things, and persistent sadness are indicative of depression, which can potentially result in suicidal ideation [5]. The number of cases of depression-induced suicide has increased in recent years [6]. However, depression is also a challenging condition to diagnose, with current approaches relying on scales that do not permit the direct and rapid measurement of its severity [7]. Depression is regarded as a dangerous illness, as it can have adverse effects on both mental and physical health [8]. In light of recent developments in artificial intelligence applications, the affective computing community has witnessed a surge in interest surrounding the creation of automated systems capable of detecting depression [9]. Automated depression detection systems offer a valuable tool for diagnosing depression and facilitating early intervention [10].

The attention mechanism in the Convolutional Neural Network-Bidirect Long Short-Term Memory (CNN-BiLSTM) hybrid deep learning model is investigated in this study. The attention mechanism allows the deep learning model to concentrate on the most important parts of the input sentence, making it easier to identify depression levels from social media text data [11]. Furthermore, FastText embedding as word embedding facilitates the rapid and effective learning of words, thereby enhancing the capacity for text learning, which is a crucial aspect [19]. Additionally, the author employs word weighting through Term Frequency-Inverse Document Frequency (TF-IDF) as feature extraction, a technique not utilized in some of the referenced studies.

This research was developed by referencing a number of literature sources from previous studies that exhibited similarities in terms of the methods and research objects employed. In the research conducted by Joel Philip et al. [11], linguistic characteristics like the use of pronouns that refer to oneself

were employed to detect signs of depression. Additionally, various deep learning architectures were utilized, demonstrating that the CNN-BiLSTM Attention (CBA) model achieved the highest performance with an accuracy score of 96.71%, an AUC-ROC of 0.85, and an MCC of 0.77 on the CLEF 2017 dataset. Nevertheless, this research has not investigated the potential benefits of utilizing word embedding techniques.

The research of Tapotish Ghosh, et al. [12] proposed an attention-based model, namely BiLSTM-CNN, for the detection of Bangla text related to depression on social media. This model attained an accuracy of 94.3%, with a sensitivity of 92.63% and a specificity of 95.12%. However, this research does not use TF-IDF as feature extraction. Research by Vankayala et al. [13] the FCL model or FastText Convolutional Neural Network with Long Short- Term Memory used in detecting depression can produce high accuracy. The FastText classification algorithm demonstrated a 90% accuracy rate in detecting depression from Twitter data [14]. Unfortunately, both studies did not use the attention mechanism.

The primary contribution of this research is the integration of a CNN-BiLSTM hybrid deep learning model with an attention mechanism, along with TF-IDF for feature extraction and FastText for feature expansion, to detect depression in an Indonesian tweets dataset. To the best of our knowledge, no other research has done this. In addition, there is still a lack of word embedding using datasets from Twitter or currently X, on the detection of depression with Indonesian datasets based on a hybrid deep learning approach. So this research aims to detect depression in X by employing a combination of deep learning technique of CNN and BiLSTM based on attention mechanism, feature extraction with TF-IDF, and FastText feature expansion.

A variety of scenarios will be tested, including the selection of the data split ratio and the identification of the optimal n-gram for feature extraction. In some scenarios, the maximum feature setting for the model combination is compared. Additional scenarios utilize the model to expand features, employ FastText, and select the top-ranked corpus with optimal performance. The final scenario explores the use of an attention mechanism as a means of optimization. The structure of this paper is as follows: Section 2 details the methodology for detecting depression with Hybrid Deep Learning and FastText expansion. Section 3 will provide an analysis of the results and discussion. Section 4 will conclude the research conducted.

II. Research Methodology

The following flowchart illustrates the design of a depression detection model system using TF-IDF feature extraction, FastText feature expansion, and CNN-BiLSTM hybrid deep learning with attention, as shown in Fig 1.

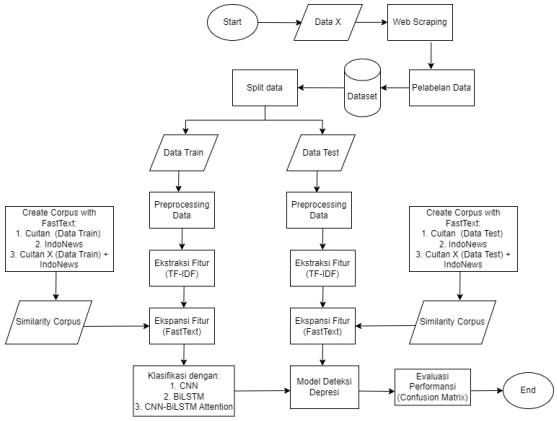

Fig. 1. Proposed flowchart

A. Crawling

Data crawling is an automated method of systematically gathering information from multiple sources, including websites, social media, and other online platforms. The data collected is usually a large amount, and it can be used for various purposes such as analysis, research, or application development. In this research, the data used is data sourced from platform X through data crawling. Of course, this crawling process uses an API that is already available, so that the crawling process becomes easier. Crawling data is focused on words or sentences that tend to lead to symptoms of depression. Such as expressions of feelings or moods that are not okay, sad, and not passionate. Therefore, several groups of words were used in the crawling process such as ”capek,” ”sedih,” ”sengsara,” ”putus asa,” and ”gelisah,” with a total of 26,523 data.

Furthermore, 24,000 data were added from GitHub on link https://github.com/marshall1632/Mental-health-and-suicide-ideation-assement-with-social-media-/tree/master in English. The dataset is derived from individual posts on the Twitter social media platform related to depression or non-depression. Then translated into Indonesian. So, the total dataset used in this research is 50,523 data.

B. Data Labeling

Data labeling refers to the process of categorizing or tagging raw data with specific labels or classifications. This is done after the data has been collected. In this research, which aims to detect depression, data labeling is done by assigning the values of 1 to indicate depression and 0 to indicate non-depression.

This labeling process is based on the results of three people's discussions and perception of each user's tweet based on keywords. Each user's tweet is declared depressed or not depressed based on the majority vote. Table I presents the results of data labeling in the context of depression detection by analysis of tweets on social media platform X.

C. Preprocessing Data

Following the data splitting process, the data must first be subjected to preprocessing. In some cases, the inclusion of unused symbols and other extraneous components in the acquired data can impede the development of a robust model. Therefore, it is essential to remove any unnecessary data to