Classification of Depression Based on Social Media Posting Patterns Using RoBERTa Method

1st Aflah Iram Athena
School of Computing,
Telkom University,
Jl. Telekomunikasi No.1, Bandung, Indonesia
aflahathn@student.telkomuniversity.ac.id

2nd Warih Maharani *
School of Computing,
Telkom University,
Jl. Telekomunikasi No.1, Bandung, Indonesia
wmaharani@student.telkomuniversity.ac.id

***Abstract*— This research focuses on the effectiveness of RoBERTa deep learning model in classifying depression based on social media posting patterns. Platform X provides a potential tool that can be used for detecting early signs of mental health disorders such as depression, through analysis of social media behavior. This research uses data from X users in Indonesia who have completed the DASS-42 questionnaire. The dataset contains 88,842 rows representing 138 active users, which include two labeled categories: positive and negative. This research explores two data preprocessing to compare model performance. The first approach, half preprocessing, involves case folding, cleansing, and tokenization, but excludes normalization, stopword removal, and stemming. The full preprocessing approach includes all steps of preprocessing techniques, including case folding, cleansing, tokenization, normalization, stopword removal, and stemming. The half preprocessing approach achieved highest accuracy at 58.07% with f1 score 57.14%. On the other hand, the full preprocessing approach resulted in a lower accuracy at 53.81% and f1 score 52.38%. The outcomes of this research underscore the importance of preprocessing approaches in optimizing the RoBERTa model's performance. These results provide insight into refining data preparation strategies, which can lead to the development of more advanced models for mental health classification. By understanding the impact of different preprocessing techniques, future research may concentrate on developing models with enhanced capability to manage noise in the data.**

Keywords—RoBERTa, Depression, Classification, X, Social Media

I. INTRODUCTION

Platform X is a social network that used by millions of people around the world to share information directly. In 2023, This platform has around 368 million monthly active users, making X is one of the most influential social media [4]. X is a favorite platform for millennials where ages between 25-34 dominates these platforms, with the majority of users having higher education and upper-middle income. In a psychological context, these platforms reflect how users feel, with posts about something that describing their mood, including early signs of mental health disorders such as depression [3].

Depression or Depression Depressive disorder is a common mental disorder. The typical signs of depression involve sleep disturbance, anxiety, and loss of interest in favorite activities for long periods of time [7]. The World Health Organization (WHO) reported that about 280 million people around the world indicated depression, which can afford daily activities [1]. Regarding X, users express a depressions pattern, such as the use of certain words, frequency of posts, and engagement in negative discussions, highlighting the platforms as a key in reinforcing and classifying behavioral patterns related to depression [2].

Previous research has applied transformer-based models, including RoBERTa, to classify signs of depression from social media text. For instance, Poświata and Perełkiewicz [18] utilized RoBERTa specifically trained to detect depression levels from social media posts. However, this research generally focuses on English-language data and have not extensively explored RoBERTa for detecting depression on social media platforms in other languages, such as Indonesian. Moreover, while some studies have integrated additional features to enhance depression classification accuracy, the specific factors influencing model performance in different linguistics have not been thoroughly investigated [19].

Therefore, this research focuses on addressing these gaps by exploring the RoBERTa model in classifying depression based on Indonesian-language social media. Furthermore, this research analyzes the factors affecting model performance in depression classification, using the DASS-42 questionnaire as a validation reference. RoBERTa was chosen based on several advantages that have been demonstrated in the literature [4][8][13]. First, RoBERTa uses more advanced pre-training techniques and larger datasets, which allow models to learn better representations. Secondly, RoBERTa not only depends on simple tokenization but also considers inclusive context in the sentence, thus improving accuracy in the classification of emotions and mental states [8].

RoBERTa itself is a method in NLP (Natural Language Processing) that provides advanced performance in various classification tasks [4]. It provides an innovation called dynamic masking approach to handle data noise, making it a great choice for classification using large text data, including tweets in Indonesian [5]. In several trials, RoBERTa gained high F1 scores, such as 94.6% on SQUAD (Stanford Question Answering), 89.4% on MNLI-m, and 90.2% on SST-2, proves expertise in classification text [8].