Personality Detection On Twitter User With RoBERTa

Rianda Khusuma[1*], Warih Maharani [2], Prati Hutari Gani [3]

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: [1*]riandakhusuma@student.telkomuniversity.ac.id, [2]wmaharani@telkomuniversity.ac.id,
[3]pratihutarigani@telkomuniversity.ac.id
Correspondence Author Email: riandakhusuma@student.telkomuniversity.ac.id

1. INTRODUCTION

In the modern era, one of the things that have become the primary research in companies is the relationship between individual personality and the work environment [1]. Personality can be described as a distinctive and moderately stable style of thinking, manners, and vigorous response that distinguishes a person's adaption to circumstances [2]. A personality test is needed to determine an individual's personality. Generally, company personality tests are still in manual forms, such as interviews. Assessing a person's personality can be seen from the social media they use. Currently, social media is a platform that Indonesians widely use. Social media is generally used for many things, including to provide the latest news about oneself. In Indonesia, 191.4 million people are social media users. Twitter is one of Indonesia's most widely used social media, with a total of 18.45 million users in early 2022 [3]. Twitter is an online communication and social networking service on which users publish and interact with messages comprehended as "tweets." On Twitter, users can use each other and spread information further through "retweeting" [4]. Detecting a person's personality on social media can be done using the Big Five Personality method. In this model, personality can be classified into 5 types, namely Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N) [5]. To support the formation of a personality detection tool, Machine Learning is also needed [6]. One of the known used machine learning models is RoBERTa.

Personality identification using machine learning has been widely researched before. Like previous research [7], it can be proven that by using machine learning, a system to detect personality can be built. In this study, the MBTI personality classification was carried out on the Twitter dataset, and several models were applied to carry out the classification. It was found in this study that the SVM model was able to outperform other classifiers. In this study [8], which formed the RoBERTa model from the BERT model architecture, it was proven to be able to obtain state-of-the-art results on GLUE, RACE, and SQuAD, without multi-task finetuning for GLUE or additional data for SQuAD. In the GLUE Task, RoBERTa accomplishes state-of-the-art results on all 9 of the GLUE task development sets. In the SQuAD task, RoBERTa is able to achieve an accuracy of up to 94.6%, which is better than BERT and XLNet, and in the RACE task, the RoBERTa model is able to outperform the BERT and XLNet model and achieved an accuracy of 83.2%. Research [9] uses the RoBERTa model to detect sarcastic tweets in English based on the RoBERTa pre-train model. The results of this study can achieve a new state-of-the-art from the iSarcasm dataset, which produces an f1-score model of 0.526. Based on previous research [10], using the dialogue-based personality dataset, FriendsPersona. The results show that RoBERTa has the highest accuracy value, with 4 of the 5 big five personalities obtaining the highest accuracy value compared to other models, namely the ABCNN, ABLSTM, HAN, and BERT models. The RoBERTa model obtained the highest accuracy score on the AGR personality of 59.72%, EXT of 60.62%, OPN of 65.86%, and NEU of 61.07%, while for CON personality with an accuracy score of 60.13%, it belongs to the ABCNN model. Based on research [11] uses datasets from Facebook and Twitter. It can be concluded that the averaging model from the BERT, RoBERTa, and XLNet models and the NLP feature can improve prediction accuracy and obtain an accuracy of 77.34% and an f1-score of 0.749.

Based on some of the research results above, using machine learning, a model can be formed that can classify the personalities of Twitter social media users. With reference to the research results above, it can be seen that the A Robustly Optimized BERT Pretraining Approach (RoBERTa) model has an excellent ability to perform NLP tasks. Based on some of the results of previous studies, there has yet to research using the RoBERTa model as a classifier to classify the personalities of Indonesian Twitter social media users, so in this study, the personality detection of Indonesian Twitter users was carried out with RoBERTa. It is hoped that this research will lead to the use of personality detection tools by companies when recruiting new employees through prospective employee social media.