## Abstract

Depression is a mental illness that is commonly found in the world and is associated with being the main cause of disability. Inadequate treatment services for mental illness in the world provide motivation for prevention efforts. Research in the field of mental health, which lacks quantitative data due to the complexity of mental illness, has made social media a potential source of data in creating efforts to prevent mental illness, especially depression. In this study, sentiment analysis was used to detect depressive behavior on Reddit social media. Analyzes of pronoun pronouns and the word patterns that follow were performed to look at language patterns in the depression data. Next, several experiments were carried out to get the best model performance, such as comparison of preprocessing types, comparison of the Information Gain and Categorical Proportional Difference feature selection, and the search for the best smoothing parameters in Multinomial Naïve Bayes. The results showed that there was an increase in model performance due to stopword removal and word reduction by stemming. In the feature selection comparison, Information Gain produced the best feature subset as much as 40% of the total features and managed to provide the best accuracy increase by 5.59% to 87.82% and an increase in f1-score by 3.91% to 89.74%. Finally, tuning the smoothing parameter with an alpha value of 0.1 on Multinomial Naïve Bayes resulted in an increase in the best accuracy of 2.54% to 84.77% and an increase in the f1-score by 1.56% to 87.39%.

Keywords: Reddit, Sentiment Analysis, Information Gain, Categorical Proportional Difference, Multinomial Naïve Bayes