

Abstract

This research focuses on evaluating the efficacy of Random Forest and Decision Tree, in detecting depression on tweets and interaction patterns on X social media. Depression as a global health problem often happens because of individuals' online behavior. This study uses data from X social media users in Indonesia who have filled out the DASS-42 questionnaire with an analysis approach that includes crawling data that includes tweets and interactions on X. The purpose of this research is to more accurately and comprehensively identify signs of depression by analyzing the interaction patterns of users on social media platforms through the integration of several many methods for feature extraction and preprocessing situations. The methods used include data preprocessing, feature combination using TF-IDF, Bag of Words, and Word2Vec and model evaluation utilizing metrics such as Precision, Recall, Accuracy, and F1-score. The findings of this research show that Random Forest performs better than Decision Tree, with a combination of TF-IDF, BoW, Word2Vec and TF-IDF, Word2Vec features obtained an accuracy of 0.60. Although Random Forest is superior, both models are difficult to identify the positive class of depression which can be seen from the relatively low F1-score and recall values. Other factors affecting model performance include lack of data relevance, low interaction rate, and limited feature extraction.