Twitter Bot Classification Using the Random Forest Classifier Algorithm

M. Geofany Hermawan

Study Program of Informatics Engineering

Majoring in Information Technology

ABSTRACT

Twitter has become a popular social media platform widely used by millennials and adults. The percentage of Twitter users in Indonesia has also been rapidly increasing over the years. In 2019, the number of Twitter users in Indonesia reached 6.43 million or approximately 52% of the total social media users. However, Twitter has a high crime rate due to the presence of irresponsible individuals creating fake accounts to disseminate harmful information to other users. Therefore, the author conducted a research using Random Forest Classifier to determine whether a Twitter account is a bot or a human. In this study, the author used 1200 data as training and 10 testing data. Based on the labeled training data, classification was performed using the Random Forest Classifier method, and testing was conducted using the testing data. The results showed that the highest accuracy was achieved with an 8:2 data ratio, with an accuracy of 85%. The accuracy for the 9:1 data ratio was 81%, while for the 7:3 data ratio, it was 83%. The accuracy for the 6:4 data ratio was 82%, and for the 5:5 data ratio, it was 81%. In conclusion, the data used in this study was found to be significant as it influenced the accuracy results of the algorithm used. For future research, a larger dataset is needed to further improve the accuracy level.

Keywords : *Classification, Data Mining, Twitter, Bot, Random Forest Classifier.*