

Implementation of the Random Forests Method on the Retweet Classification Model based on Content

Akmal Ariq Santoso¹, Jondri², Kemas Muslim Lhaksana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹akmalariq@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id,

³kemasmuslim@telkomuniversity.ac.id

Abstract

Twitter as one of the biggest social media on the internet has been used as the center of information exchange on mainstream media. As this paper was written Covid-19 information sporadically propagated through twitter. To help spread validated information to the masses we need to understand which factors are relevant and support the information diffusion. In this paper author tried to find similarities between tweets by using TF-IDF, author also applied content features from tweet's meta-data to random forests classifier to predict which tweets users might retweet. The result of the shows that by using content features, machine learning models can predict retweets from users. The proposed method of combining content features from twitter metadata and TF-IDF leads to a better model than the stand-alone features with 69.97% of accuracy.

Keywords: Retweet Prediction, TF-IDF, Random Forests
