Prediksi Retweet Berdasarkan Fitur Pengguna, Konten, dan Waktu Menggunakan Metode Evolutionary UnderSampling Boosting

Ghina Khoerunnisa[1], Jondri[2], Widi Astuti[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]ghinak@students.telkomuniversity.ac.id, [2]jondri@telkomuniversity.ac.id, [3]widiwdu@telkomuniversity.ac.id

Abstract

Twitter is one of the popular microblogs that allow users to write posts. Retweeting is one of the mechanisms for the diffusion of information on Twitter. One way to understand the spread of information is to learn about retweet predictions. This study focuses on predicting retweets using Evolutionary Undersampling Boosting (EUSBoost) based on user, content, and time-based features. We also consider the vector of text as a predictive feature. Models with EUSBoost are able to outperform models using the AdaBoost method. The evaluation results show that the best model can achieve an AUC performance score of 77.21% and a GM score of 77.18%. While the Adaboost-based models achieved AUC scores ranging from 68% to 69% and GM scores ranging from 62% to 63%. In addition, we found that there was no significant difference between using numeric features only and combining numeric and text features.

Keywords: information diffusion, retweet prediction, EUSBoost