Abstract

Twitter is one of the most popular and easy-to-use social media to get information quickly. The *Retweet* feature is one of the reasons why information can spread quickly. A retweet occurs when followers retweet his followee tweet's. Therefore, most *tweets* spread from *followees* to their *followers*. In this research, modeling is done for retweet prediction based on *user-based* and *content-based* features using the *Ensemble Stacking* method through the *K-fold Cross Validation* process. In this case, This Stacking Ensemble is formed with 3 base-learners namely *Random Forest*, *Gradient Boosting*, and *Support Vector Machine* (SVM). While the *meta-learner* used is *Support Vector Machine* (SVM) This modeling shows the best results when it has been done *Imbalanced Class Handling* using the SMOTE Technique and *K-fold Cross Validation* with k = 10. F1-score results show 86.46%. With such results, it can be concluded that the modeling formed is able to improve the prediction results of the *base-learner*.

Keywords: twitter, retweet, ensemble stacking, k-fold cross validation, oversampling