

Abstract

Tweet is a message containing information shared by Twitter users. Tweets can be shared with other users by retweets, so this feature plays a very important role in disseminating information. This study discusses retweet prediction using user-based feature and content-based feature, as well as ANN as a classifier. The research problem studied here is how to overcome data duplication and class balance by using undersampling and oversampling. The evaluation results show that the classification process reaches an F1 score of 86% on the model by undersampling and removing duplication in the data.

Keywords: prediction, retweet, undersampling, oversampling, ANN.