

Abstract

Information technology develops very rapidly so that it helps human needs to obtain information and communication facilities. In social media such as Twitter, it is very easy to obtain the latest information such as Political, Health and other issues, with one of the distribution features, called retweet, information will be very easy to move from one user to another. This research seeks to build a retweet prediction system from user content based on Content Similarity and Content Based features using the Support Vector Machine classification method. To split the dataset used k fold cross validation with value $k=10$. The final results in this research got an accuracy average value of 71.84%.

Keywords: Twitter, Retweet, Content Similarity, Support Vector Machine