

Abstract

In this globalization era technology, emergence of social networking that is increasingly becoming people's lifestyles, among other social networking. Twitter is a social media-type micro-blogging means having a proper service to find out what is happening in the virtual world, as well as the maximum can tweet 140 characters to search only the minimum of delay. To improve application performance, Twitter itself has features that support for any purpose. As well as eventually the widespread use of Twitter, the development also increased spam activity. The case is to be increased because the eraser can read data and tweets of others without having a Twitter account first. This final project will discuss how to detect spammers, in which case we require crawling the data to get the data it needs. The attributes used in this thesis such as, number of followers, following the number, the number of tweets, number of like, the number of URLs, mention the number, amount, hashtag, the number of international spamwords, and spamwords Indonesia. On detection of spammers to have a good performance, should be tested with the stages of preprocessing, such as Normalisasi Data, Reducing Data (Selection Attributes Manual (Manual Attribute Selection), Chi Square Attribute Eval, Gain Ratio Attribute Eval, and Info Gain Attribute Eval), Data and calculation accuracy discretization on the classification of the Random Forest, Support Vector Machine (SVM), Naïve Bayes, and J48. For the results obtained from the preprocessing stage, the results of this study can be seen from the quality of the source data using the data preprocessing techniques are appropriate and correct. Thus, in this final performance of the best value 88.34% obtained without preprocessing stage on Naïve Bayes classification

Keyword: *Twitter, Detection Spammer, Preprocessing Data*