

Abstract

Along with the use of the internet as tools to share information, exchange and share opinions can also be performed. The opinions that exist on the internet, not all opinions are useful so that sometimes complicate the parties who need the related information. The opinions that are not useful can be describe as Spam Opinion. Spam opinion in internet can occur in any case, one of them in the forum. Spam opinion in the forum can be the junk opinions that useless in a topic being discussed.

Text Classification is one of the solution that can solve these problems. One of Text Classification method is the Support Vector Machine (SVM). The classification focused on determining a post that junk post or not. In practice, carried out weighting the data set by using the TF and TF-IDF. The TF-IDF weighting is produced better results when compared to TF alone. Both weighting was used to calculate the similarity of the existing posts which are associated with the topic being discussed on the forum using the cosine similarity. The evaluation result shows that Support Vector Machine method is able to soleve the junk post problems. The process of data preprocessing such as stopword removal may affect the classification results, while the value of the parameter C can not affect the classification results significantly.

Keywords: text classification, opinion spam, Support Vector Machine, TF,TF-IDF, cosine similarity