Social networking websites are used by millions of people around the world. People express their views, opinions and share current topics. Millions of data generated every day. It's a good platform to connect with the people. Now a day's spammers used this platform to advertise spam content on the social networking websites.

The proposed system used to classify tweets into different groups as spam and non spam tweets. The system will use 120 character tweets for analysis purpose. Various active and verified twitter accounts would be chosen to extract the tweets. Each tweet is to be classified into 2 category-spam and non-spam. These classified tweets then are used to train the various machine learning techniques. Words of each tweet considered as features and a feature vector was created using bag-of-words approach in order to create the instances. The data will be trained using SVM (Support Vector Machine).
References

2. “Machine Learning Techniques in Spam Filtering” Konstantin Tretyakov, kt@ut.ee Institute of Computer Science, University of Tartu Data Mining Problem oriented Seminar, MTAT.03.177,
6. “A Performance Evaluation of Machine Learning-Based Streaming Spam Tweets Detection” Chao Chen, Jun Zhang, Member, IEEE, Yi Xie, Yang Senior Member, IEEE, @2015
8. “Spam Filtering Techniques and MapReduce with SVM: A Study” by Amol G. Kakade, Prashant K. Kharat, Anil Kumar Gupta, Tarun Batra ©2014 IEEE

Index Terms

Computer Science   Algorithms

Keywords

Spam tweets, svm, kernel functions, SVMLIB