Abstract

Rapid growth of social networking have had an immense effect on today’s general public and Web stage. Social networking sites are developing in both size and prevalence with a high rate in recent years. Twitter is one of the quickest developing Social Networking Sites. With the measure of information developing in Twitter lately, detection of spam in real time has become a challenging task for researchers as well as for Twitter itself. Enormous work is being done towards spam detection. The work done previously was not giving the appropriate results in the context of content based spam discovery on Twitter. In this paper accuracy is analyzed by using Classical approaches like Naïve Bayes and Random forest algorithm. It is observed that
these algorithms are not giving accurate results. With a specific end goal to increase the accuracy of spam detection Random forest with Feature Subset Selection have been used. Here the aim is to propose a Feature Subset Based Classification Approach where a set of features will be tested using Random Forest Classifier for twitter spam detection. In this paper the capabilities of Random Forest Classifier has been extended for detecting spam by including Feature Subset with it.

References

- J. a. Gama, I. Zliobait?e, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on
Feature Subset Selection for Twitter Spam Detection


Index Terms

Computer Science

Security

Keywords

Labeled Dataset Feature Subset Selection random Forest