



Spotting Spammer using Group Spammer Behaviour Analysis

Final Year Project (2016-2020)

Department of Computer Science , COMSATS University Islamabad, Lahore Campus

Introduction

Online reviews are considered the best source of customer's opinion about a product. Spammers have started working in groups, which has become more damaging. Solutions to find such groups are very limited. We solve this by identifying which reviews belong to which candidate group of spammers.

Motivations

- Get the original opinion and the truth about a particular product.
- Develop a system for betterment of businesses.

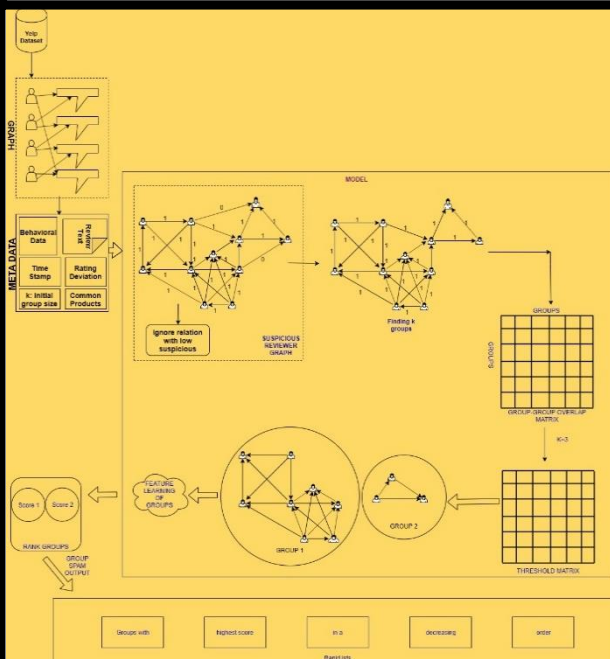
Objectives

- Identifying group spammers.

History

Group spammers give a lot of fake reviews to misguide users and very little work is done on this.

System Architecture



Future Directions

The project is limited by its scope right now but it can be helpful to researchers for building a well-organized prototype. Location, time, IP-Addresses can be used in longer run for more precise results. The deployment of the project on a web portal was also made, which is an example of how the project can be used for future references.

Conclusion

A model is developed which spots spammer analysis using largescale real-life dataset with high accuracy fake review labels and find a group of them working together. Random Forest Classifier gave us the best model with TF-IDF Vectorizer using k=7 and Unigram + Bigram + Trigram with Precision, Recall and AUROC being 1.0, 1.0 and 0.74 respectively.

Results

Spam Detection Models	Precision	Recall	F-Measure	Accuracy
Random Forest	0.96	0.96	0.95	0.96
K-Nearest Neighbour	0.94	0.95	0.94	0.95
Multinomial Naïve Bayes	0.91	0.90	0.90	0.90
Xtreme Gradient Boost	0.96	0.96	0.95	0.96
Long Short-Term Memory	0.86	0.93	0.89	0.93
Artificial Neural Network	0.93	0.93	0.89	0.93

Group Members

Kafeel Butt

kafeelbutt45@gmail.com

Omer Tariq

om3r.tariq97@gmail.com

Sheikh Muhammad Qasim

q.shaikh14@gmail.com

Supervisor: Dr. Hamid Turab Mirza

