

Spam Tweets and Account Detection on Twitter



Surendra Singh Gangwar
2016IPG-107

Dr. Santosh Singh Rathore
MTP Mentor

Problem Statement: To develop the system that can classify tweets and user accounts.

Keywords: Social network security, Twitter, Spam detection, Machine learning, Tweepy, User-Content features, Classification

Abstract

Twitter is one of the most famous social media platforms and because of its prevalence, spammers discover this stage to spam with clients. Twitter spamming is all the more compromising in light of the fact that its assortment of crowd, twitter clients length over all divisions of life for example it very well may be the educators or understudies, VIPs or politicians, marketers or clients or even the overall population. To stop spammers, Google SafeBrowsing and Twitter's BotMaker devices recognize and block spam tweets. These tools can block malicious links, however, they cannot protect the user in real-time as early as possible. Thus, industries and researchers have applied various approaches to remove the spam tweets from social network platforms. Some of them are just based on user-based features and others are based on tweet's

content-based features. In our work, we will try to make a model that will analyze both types of features to classify the tweets and users. The advantage of utilizing the tweet content feature is that we can distinguish the spam tweets regardless of whether the spammer makes another account which was impractical just with the client and tweet's content based features.

Hypothesis And Objectives

Our MTP project aims to complete the following objectives:

- We will try to develop a system where we can pass real time tweets to classify them as spam or non-spam using user and content based features.
- We will try to classify users for tweets as legit or spam users.
- We will analyze and compare the performances of different methods on our dataset preprocessed using Natural language processing.
- Extend this work to generate the more efficient model by bringing parameter tuning into consideration.

Tools Required

- **Programming Language:** Python
- **Libraries:** Tweepy, Scikit-Learn, Keras, Pandas, Matplotlib, Numpy, NLTK, Word Embeddings,
- **Dataset:** For our model we will collect all the tweets in the range of 3 months. We will crawl the tweets using Twitter Developer API using Tweepy library. All the tweets will be classified into spam non-spam tweets, later we will try to classify the user as legit user and spam user.