Capstone Project Proposal

Title:- Twitter Sentiment Analysis

Rationale

This project addresses the problem of sentiment analysis in twitter; that is classifying tweets according to the sentiment expressed in them: positive, negative or neutral. Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users out of which 100 million are active users and half of them log on twitter on a daily basis - generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analysing the sentiments expressed in the tweets. Analysing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

Introduction

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analysing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favourable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis. One such study was conducted by Tumasjan et al. in Germany for predicting the outcome of federal elections in which concluded that twitter is a good reflection of offline sentiment.

Literature Survey

In today's world, micro-blogging sites has become a platform for individuals or organizations across the world to express their opinions, sentiment and experience in the form of tweets, status updates, blog posts, etc. This platform has no political and economic restrictions. A stream of tweets on electronic products from the twitter micro-blogging site are then subjected to preprocessing and classified based on their emotional content as positive, negative and neutral. The performance of the unsupervised algorithm is then analyzed.

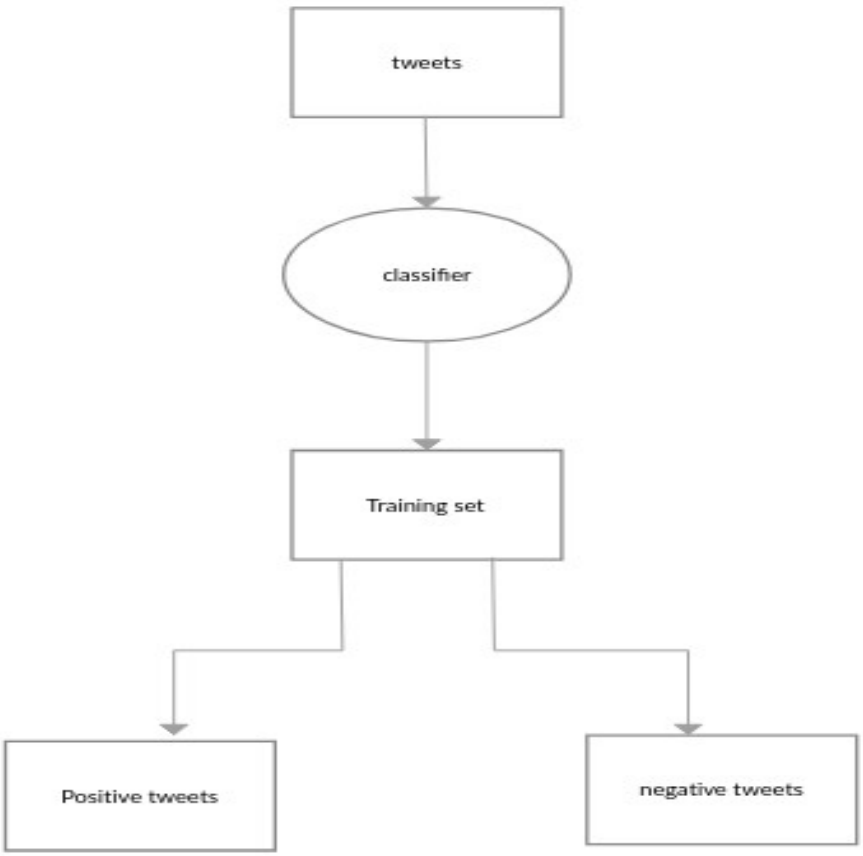
Problem Definition

When we want to know any sentiments about any person, product, organization, company, trends, we require manual asking type procedure or via spending lots of efforts in searching on google. It is not possible to know the sentiments by asking from the people about person, product, organization, company, trends, nor searching it on the google or any other search engines, thats why we have proposed a application that will tell the sentiments about all the things.

Proposed methodology of solving identified problem

This project of analyzing sentiments of tweets comes under the domain of “Pattern Classification” and “Data Mining”. Both of these terms are very closely related and intertwined, and they can be formally defined as the process of discovering “useful” patterns in large set of data, either automatically (unsupervised) or semiautomatically (supervised). The project would heavily rely on techniques of “Natural Language Processing” in extracting significant patterns and features from the large data set of tweets and on “Machine Learning” techniques for accurately classifying individual unlabelled data samples (tweets) according to whichever pattern model best describes them.

DFD Diagram



Resources and consumables required

Sr. No.	Name of Resource	Specification	Quantity	Remarks
1.	Computer System	8GB Ram and i5 processor	3	
2.	Operating System	Ubuntu	-	
3.	Software	Chrome Browser	1	
4.	Language	Python 3.6	1	
5.	Framework	Django	-	
6.	Database	MySQL	-	

Sr. No.	Main Activities	Activity Performed (Workwise)	Planned start date	Planned end Date	Name of Responsible Team Members
1.	Information Gathering	Brief study about Capstone Project Planning	26-06-2019	02-07-2019	
		Research started for searching 3 different topic's for selection	03-07-2019	10-07-2019	
		3 topics submitted to our faculty	11-07-2019	18-07-2019	
		1 topic got selected by our faculty	19-07-2019	26-07-2019	
		Information gathered on the selected topic	27-07-2019	03-08-2019	
		The requirements to build the project are specified in SRS	04-08-2019	11-08-2019	
2.	Planning and Designing	Guide allocated on project	12-08-2019	19-08-2019	Annu Poddar
		In brief Discussion on project ideas with guide	20-08-2019	27-08-2019	Sunil Thakur
		Presentation Given to guide on app prototype & features	28-08-2019	04-09-2019	Usama Tahseen Ul Haque
		Planning started on making workflow of project	05-09-2019	12-09-2019	
		Designed the system workflow (Data Flow Diagram)	13-09-2019	20-09-2019	
		Submitted System workflow to guide	21-09-2019	28-09-2019	
		System workflow got approved by guide	29-09-2019	05-10-2019	
3.	Implementation	Started extracting basic functionality of app	06-10-2019	13-10-2019	

required from SRS document		
Made an Algorithm	14-10-2019	21-10-2019
Started making a prototype in app development	22-10-2019	29-10-2019