Software Requirements Specification

for

Twitter Sentiment Analysis

Prepared by Mr. Anurag Pramod Jadhav

Date = 19-06-2022

Table of Contents

Table of Contents ii

1. Introduction 1

1.1 Purpose 1

1.2 Document Conventions 1

1.3 Intended Audience and Reading Suggestions 1

1.4 Product Scope 1

1.5 References 1

2. Overall Description 2

2.1 Product Perspective 2

2.2 Product Functions 2

2.3 User Classes and Characteristics 2

2.4 Operating Environment 2

2.5 Design and Implementation Constraints 2

3. External Interface Requirements 3

3.1 User Interfaces 3

3.2 Hardware requirements 3

3.3 Software Interfaces 3

4. System Features 4

4.1 System Feature 1 4

5. Other Nonfunctional Requirements 4

5.1 Performance Requirements 4

5.2 Safety Requirements 5

5.3 Security Requirements 5

6. Other Requirements 5

# Introduction

Data analysis is the process of applying organized and systematic statistical techniques to describe, recap, check and condense data. It is a multistep process that involves collecting, cleaning, organizing and analyzing. Data mining is like applying techniques to mold data to suit our requirement. Data mining is needed because different sources like social media, transactions, public data, enterprises data etc. generates data of increasing volume, and it is important to handle and analyze such a big data. It won't be wrong to say that social media is something we live by. In the 21st century social media has been the game changer, be it advertising, politics or globalization, it has been estimated that data is increasing faster than before and by the year 2020; about 1.7 megabytes of additional data will be generated each instant for each person on the earth. More data has been generated in the past two years than ever before in the history of the mankind. It is clear from the fact that the number of internet users are now grown from millions to billions.

Database which is opted for the proposed study is from Twitter. It is now day's very popular service which provides facility of microblogging. In this people write short messages generally less than 140 characters, about 11 words on average. It is appropriate for analysis as the number of messages is large. It is much easier task as compared to searching blogs from the net. The objective of the proposed analysis, 'Sentiment Analysis', is the analysis of the enormous amount of data easily available from social media.

Algorithm generates an overall sentiment score from the inputted topic in terms of positive, negative or neutral, further it also works on finding the frequency of the words being used. Word cloud that is a pictorial representation of words based on frequency occurrence of words in the text is also generated. Calculation is actualized utilizing R attributable to its component rich, thorough and expressive abilities for measurable information.

## Purpose

This project of analysing 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 semi automatically(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 un labelled data samples(tweets)according to

whichever pattern model best describes them.

## Document Conventions

Database:

A system intended to organize, store, and retrieve large amounts of [data](http://en.wikipedia.org/wiki/Data) easily; a repository of information.

Events:

Collection of data and analyzing sentiments through it.

In this document we had describe about features, modules and technology is used in application. Arial font because it is easy to read .

## Intended Audience and Reading Suggestions

This SRS is intended for audiences, including the customer, as well as the project manager.

* + - The customer will use this SRS to verify that the developer team has created a product that is acceptable to the customer.
    - The project manager of the developer team will use this SRS to plan milestones and a delivery date, and ensure that the developing team is on track during development of the system.

The Software requirement specification document is for both type of audience technical and non- technical to understand what technology is used in this application and feature of this application.

## Product Scope

This project will be helpful to the companies, political parties as well as to the common people. It will be helpful to political party for reviewing about the program that they are going to do or the program that they have performed. Similarly companies also can get review about their new product on newly released hardware or softwares. Also the movie maker can take review on the currently running movie. By analyzing the tweets analyzer can get result on how positive or negative or neutral are peoples about it

## References

<https://www.tutorialspoint.com/twitter-sentiment-analysis-using-python-programming>

# Overall Description

## Product Perspective

This project of analysing 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 semi automatically(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.

## Product Functions

* Collection of tweets.
* Sorting them based on Naïve bayes theorem and Natural language processing concept.

## User Classes and Characteristics

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing(NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction. The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.

Opinion: A conclusion open to dispute(because different experts have different opinions)

View: subjective opinion

Belief: deliberate acceptance and intellectual assent

Sentiment: opinion representing one's feelings

Sentiment Analysis is a term that include many tasks such as sentiment extraction, sentiment classification, subjectivity classification, summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as products, individuals, topics, organizations, and services.

## Operating Environment

The essential program in a computer that maintains disk files, runs applications, and handles devices such as the mouse and printer.

## Design and Implementation Constraints

Here the main focus is on design and software life cycle. First thing is to identify entity, attribute and relationship. We should also specify the phases from which user must go and what is flow of data and does the data is showing results properly. Here we need to also remember about the method and attribute to access data and represent it in pie chart format.

There are some diagrams which help to design software life cycle and they are:

• Data Flow Diagram

• UML Diagram

• Database design

• E R Diagram

# External Interface Requirements

## User Interfaces

The interface of the software will provide options for a relatively easy data input processes text-boxes that will be properly labeled. It will also have a user-friendly view of the whole system with simple and easy undertaking of action-driven processes as command buttons are functionally labeled. With all these, target users of this website will relatively find it not difficult to use it.

## Hardware Requirements

To be able to run the system, the minimum requirements of the hardware for this system are:

 CPU 2.0 GHz or CPU (laptops) Core 2

 CPU (desktops) RAM 2 GB RAM

 HDD 60 GB min

 7200 RPM6 GB or at least 10% free space (whichever is greater)

## Software Interfaces

* Language- Python
* Libraries – TextBlob, Matplotlib
* Server Google

# System Features

## System Feature 1

The system will operate in a graphical environment where the users interact with the program by viewing data in form of pie chart. It will be fully operational on any personal computer with a Windows operating system.

# Other Non functional Requirements

Non-functional requirements is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

-User friendly

-System should provide better accuracy

-To perform with efficient throughput and response time

## Performance Requirements

* Getting accurate data analysis.
* Representation of data done in pie chart to be accurate.

## Safety Requirements and Security Requirements

➢ Using https internet communication protocol

➢ Application is not connected to any other application

# Other Requirements

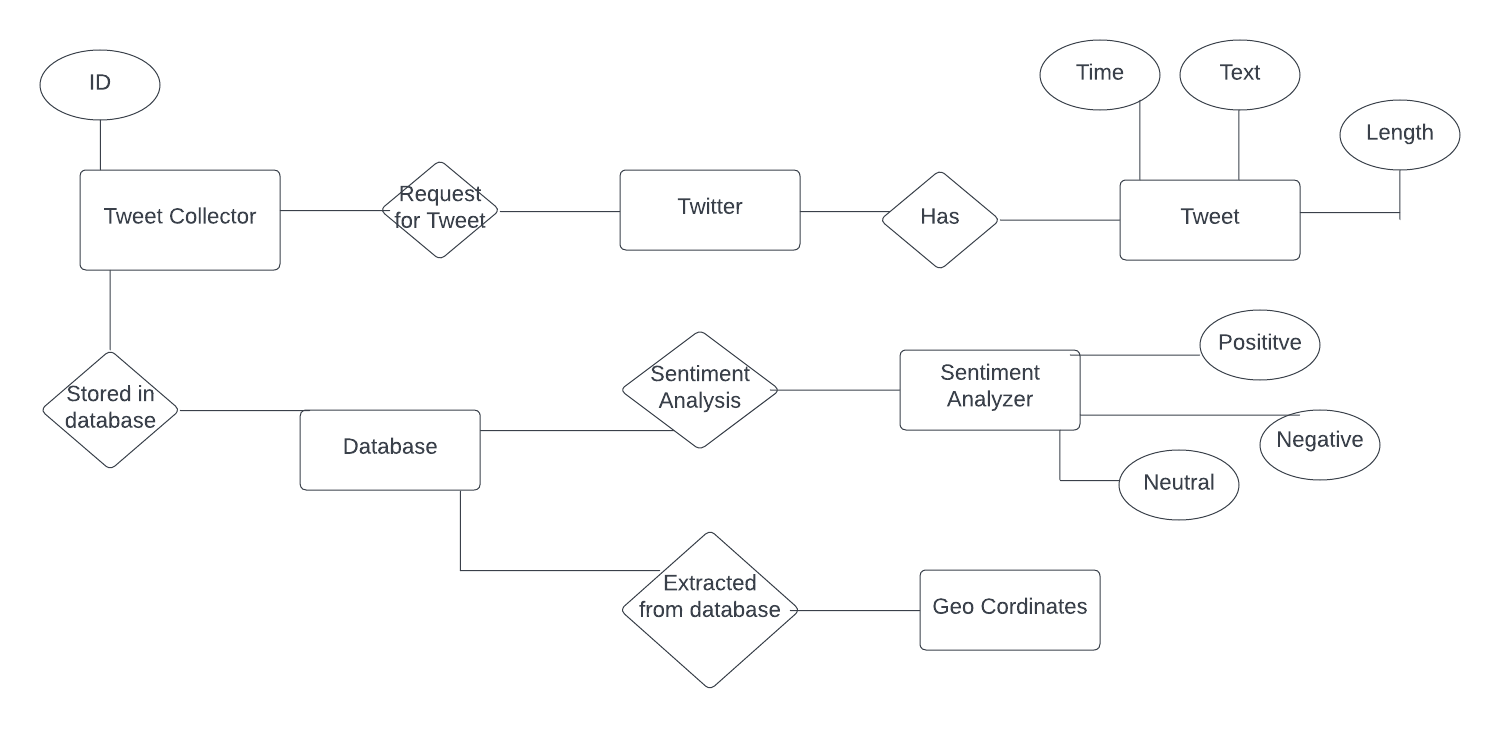
None

**ALGORITHM**

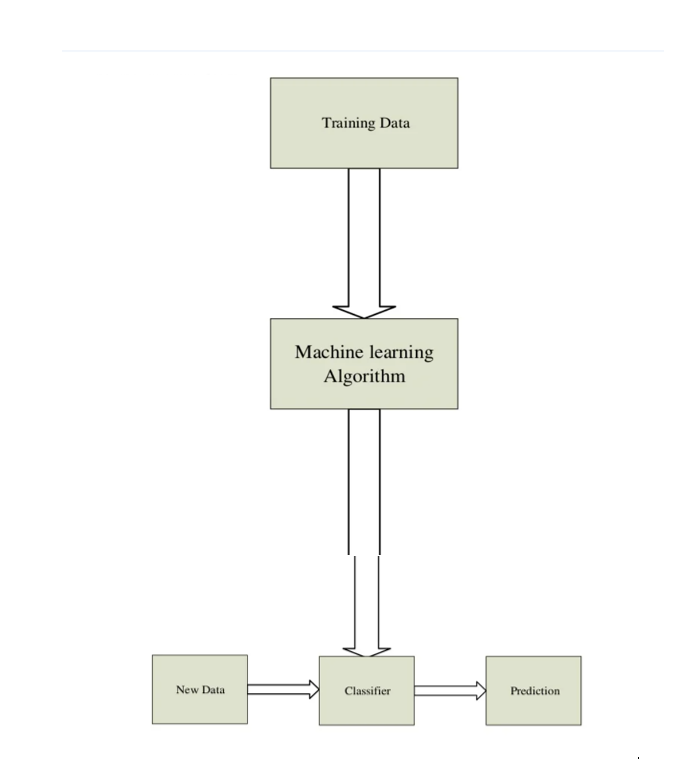
1. Get authentication for Twitter Developer API
2. Install and Import tweepy, textblob, matplotlib and csv
3. Getting Tweets with Keyword or Hashtag and number of tweets to be fetched from user.
4. Tweepy module will fetch the tweets from twitter.
5. Cleaning Tweets to Analyze Sentiment
6. Sentiment Analysis.
7. Plotting the pichart.

Diagrams –

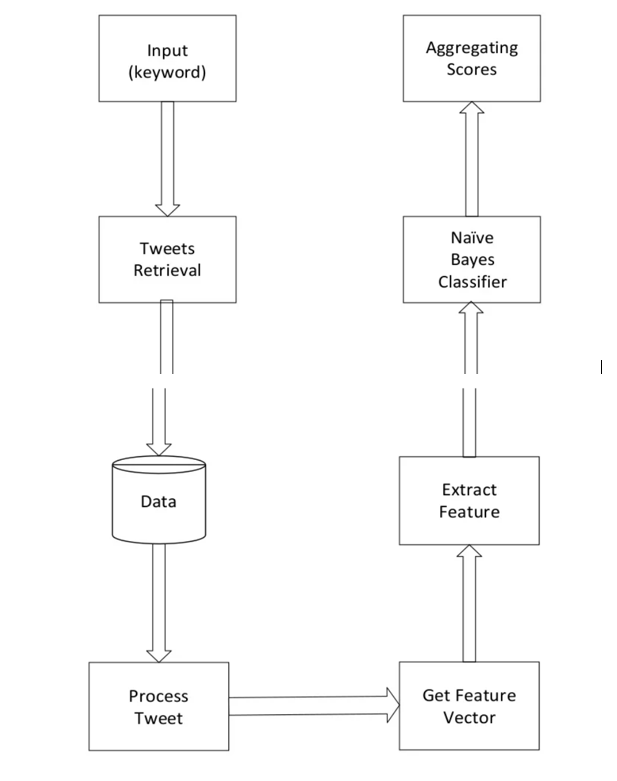
1. ERD (Entity Relationship Diagram)



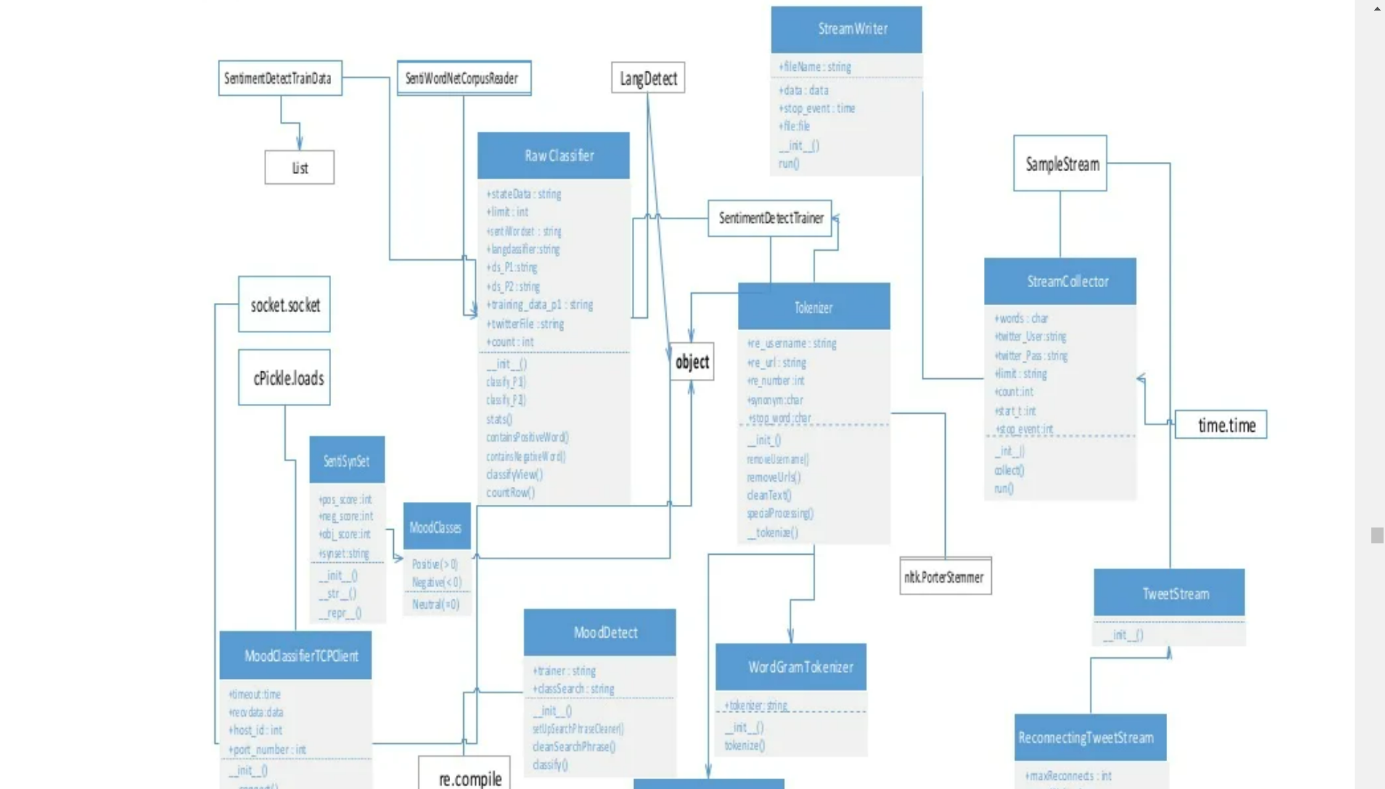
1. DFD ( Context Level)



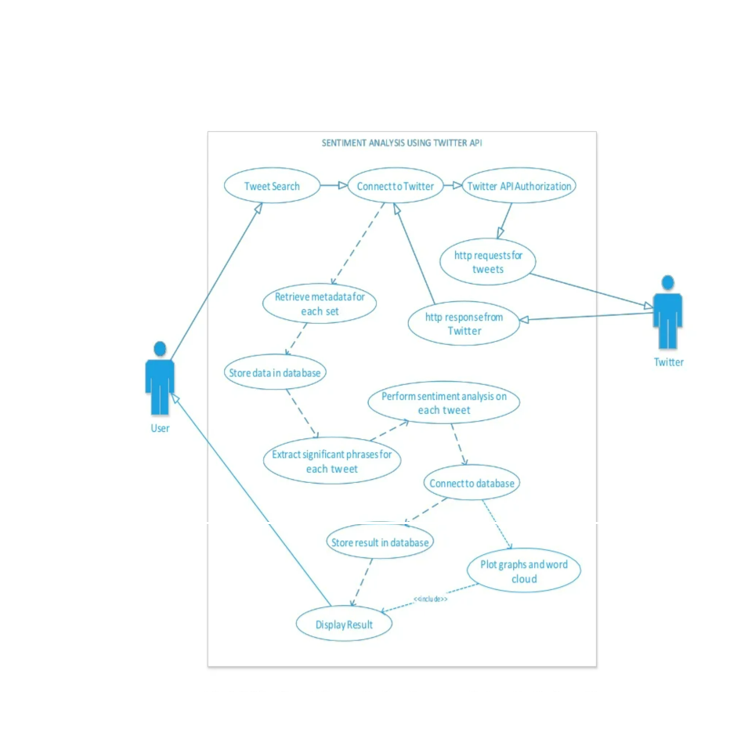
DFD (Level 1)



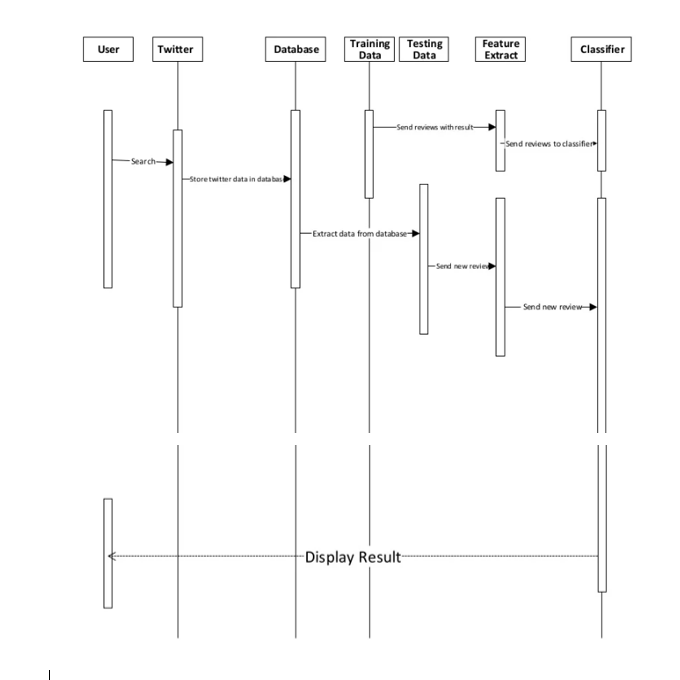
1. Class Diagram



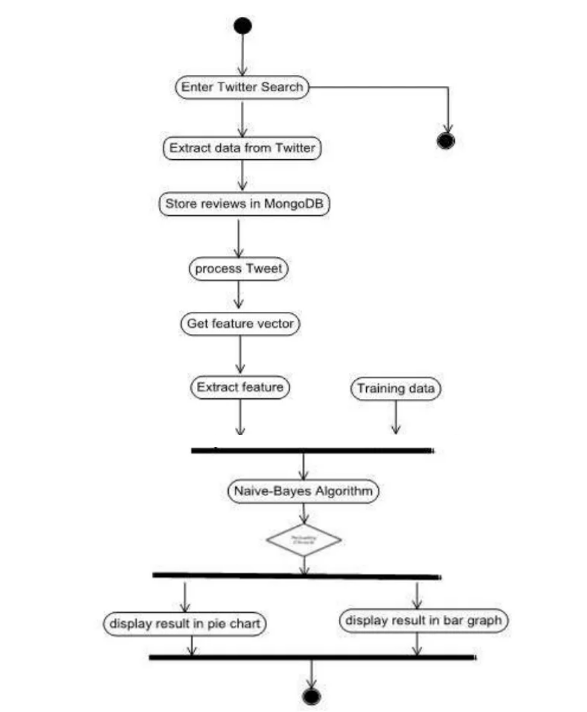
4.Use Case Diagram



5.Sequence Diagram



# Activity Diagram



GUI Screen Shots –

