Twitter Sentiment Analysis for Election Prediction
(Twitter sentiment Analyzer)

Authored by: SEGP_GROUP_5
Haider Iqbal, Manzoor Hussain, Izzat Amin, M. Waseem, Ali Rizwan

Contents

1. Introduction

1.1 Purpose

The Purpose of writing this document is to describe the requirements and give clear descriptions of a system "Twitter sentiment Analyzer". It will provide complete details of scope, Graphical User Interface, features, dependencies, assumptions taken, and constraints under which system will operate. System will illustrate all functional and non-functional requirements of the project and also enlighten the technique of expressing them. System's flow, hardware & software interface requirement.

1.2 Scope

As now-a-days, future predictions are really important, every company predicts about its future. Apart from that, they even predict the upcoming famous products. It has vast range in stoke market too. The "Twitter sentiment analyzer" has also a basic idea of prediction but it's "Election".

This product will completely show the daily positivity track of the all the 3 parties, PTI, PML(N), PPPP. And the basic idea of sentiment analysis can be shifted to any other project easily.

1.3 Definitions/ acronyms & abbreviations

Acronyms	Full Form
User	Any user which use the program
Tweepy	Twitter Library
crawler	To fetch tweets from twitter
NLTK	Natural Language tool kit
Stream	To fetch data from twitter
POS	Parts of speech
URL	Link in tweets of some attached picture or video etc.
CSV file	Comma separated values file
JSON	JavaScript Object Notation
Hashtag	phrase preceded by a hash sign (#) to identify messages on a specific topic
NFR	Non-Functional Requirements
PTI	Pakistan Tahreek-e-Insaf

PML (N)	Pakistan Muslim League (Nawaz)
PPPP	Pakistan People's Party Parliamentarian
All 3 parties	PTI, PML(N), PPPP

1.4 References

[1] J. Ramteke, S. Shah, D. Godhia and A. Shaikh, "Election result prediction using Twitter sentiment analysis - IEEE Conference Publication", *Ieeexplore.ieee.org*, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7823280/. [Accessed: 15- Jan- 2018]

[2] P. Sharma and Teng-Sheng Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter - IEEE Conference Publication", *Ieeexplore.ieee.org*, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7840818/. [Accessed: 15- Jan- 2018]

1.5 Overview

The remaining part of SRS describes the technique to express the requirement including its different types (i.e. functional and non-functional requirements) along with its software and hardware interfaces.

2. Overall Description

2.1 Product Perspective

The system will be a desktop application using JSON format to store the tweets coming from all the popular pages of all 3 parties. The text separated from tweet will stored in CSV.

2.2 User characteristics

Any user can use the product as it'll be giving statistics regarding all the 3 parties.

2.3 Constrains

Software Requirement	Hardware Requirement
Operating System: Any	Processor: any
User Interface: Desktop Application	Hard Disk: Not Specific
Language: Python	RAM: 512 MB (recommended)
Database: JSON file	Screen: any
Library: NLTK, Test Blob	Internet Connection required

2.4 Apportioning of Requirements

```
        System will be made available for user, where he can get
the daily track of positivity of all 3 parties, he just needed to
have internet connection. The tweet crawler files will also need
to add as windows services which developer team can do for user
at time of deploying.
```

3. Basic Flow

```
        The basic flow of project consists of following
```

- Data Mining
- Natural Language Process
  - o Pre-Processing
  - o Data Labeling

3.1 Data Mining

```
        This phase is completely relevant to tweets fetching in
which the tweets will be converted into JSON format. Then the flow
will move to next phase, Natural Language processing.
```

Phase 1 : Data Mining

Fetching Data

Hash tags

Text

Phase 2 : Natural Language Processing
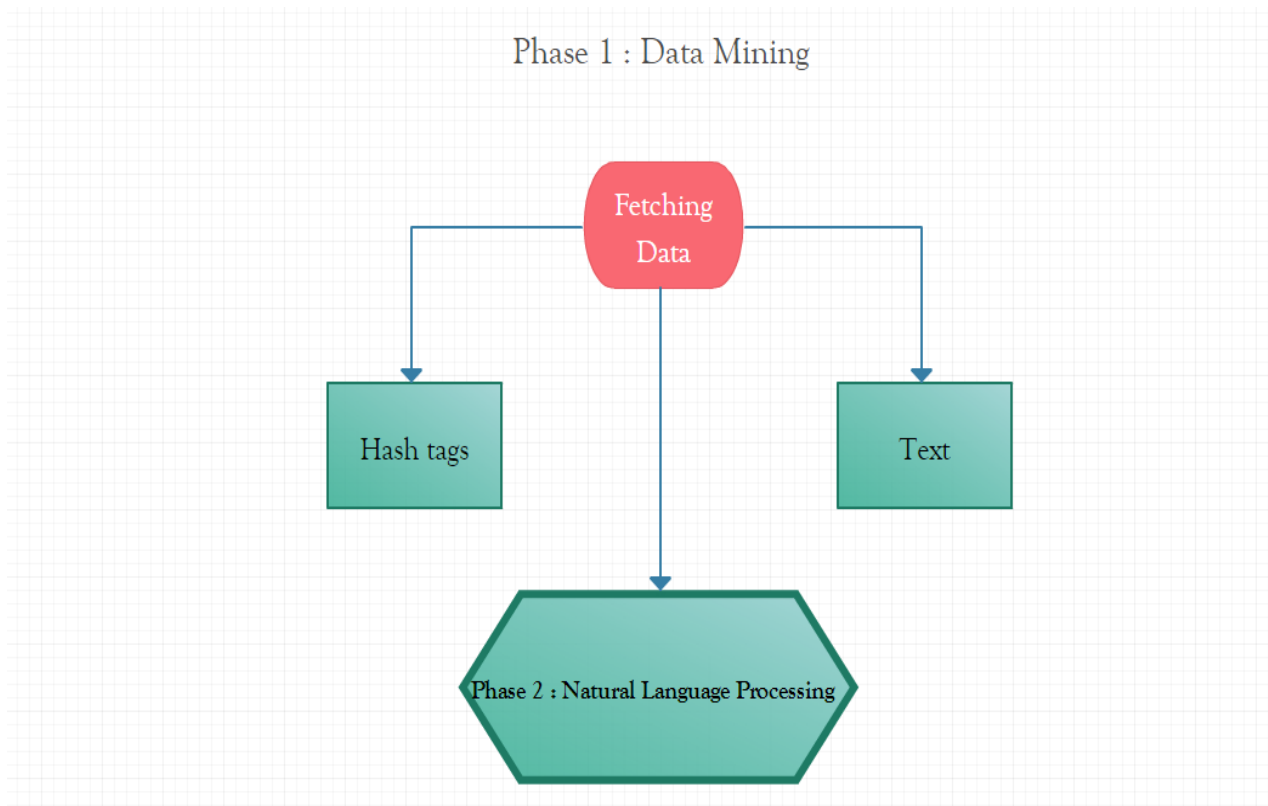
3.2 Natural Language Processing

3.2.1 Pre-Processing

This is sub phase of natural language processing related to cleaning the noise from the tweets, e.g. removal of URL from tweets, Removal of annotations, separations of text and Hashtags. The 2nd part of this phase is POS tagging.

3.2.2 Sentiment Analysis

This is the most important phase of the project. It has 2 parts, first involves data labeling and the second is the actual sentiment analysis. Data labeling involves the identification of sentence according to parts of speech (e.g. Verbs, nouns etc.). While, Sentiment analysis use algorithms to check the polarity of the sentence. There are multiple libraries e.g. VADAR, Test blob etc., but test Blob will be used.

Phase 2 : Natural Language Processing
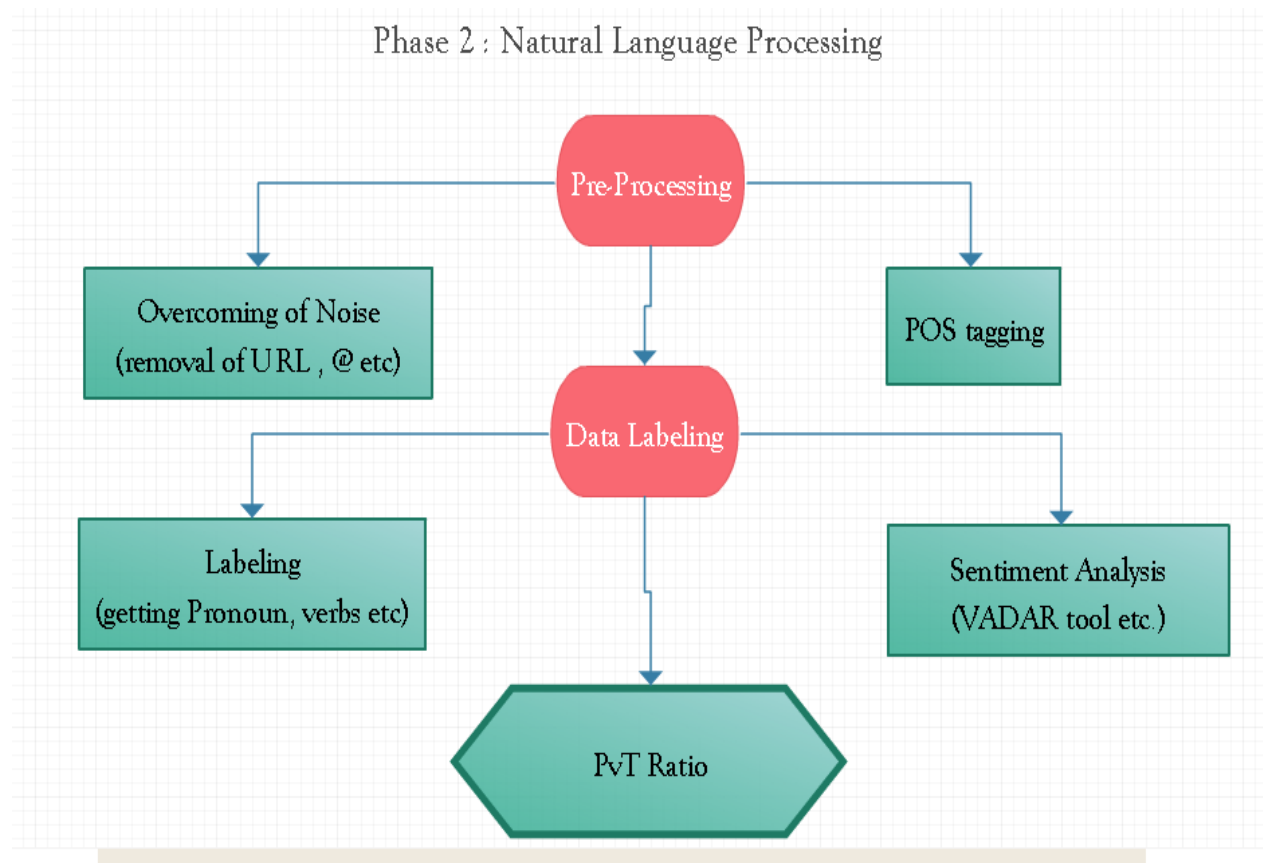
3.3 Machine Learning

As there is always room for improvement and enhancement so machine learning factor can also be added. Different algorithms of machine learning can be used e.g. SVM, Naïve Bayes etc.

4. Use Cases

4.1 Tweet Fetching

Name	Tweet Fetching
Participating Actors	N/A
Goal	Fetch a tweet from particular page
Trigger	Crawler will be run
Basic Flow	i)Crawler should be run. ii)Tweet from specific page should be get
Qualities	Tweet should be saved in JSON format

4.2 Removing Noise

Name	Removing Noise
Participating Actors	N/A
Goal	Remove the URLs etc. from tweet
Basic Flow	i)Tweet's text will be got. ii)Noise will be removed.
Qualities	Tweets after cleaning will be saved in CSV

4.3 POS Tagging

Name	POS Tagging
Participating Actors	N/A
Goal	Finding the part of speeches of sentences
Basic Flow	i)Sentence will be loaded from CSV. ii)Parts of speech will be identified

4.4 Data Labeling

Name	Data Labeling
Participating Actors	N/A
Goal	sentence should be labeled
Trigger	POS tagging should be done
Basic Flow	i)POS tagging performed. ii)Sentence should be labeled.
Qualities	Data should be labeled according to POS tagging

4.5 Sentiment Analysis

Name	Sentiment Analysis
Participating Actors	N/A
Goal	Polarity of sentence should be checked
Basic Flow	i)Sentence should be loaded. ii)Polarity of sentence should be checked.
Qualities	Polarity will help checking of popularity

5. Specific Requirements

5.1 External Interface Requirements

5.1.1 Hardware interfaces

The minimum hardware required for "Twitter Sentiment Analyzer" is cited below. However, better the hardware, and internet connection better will be the performance of System. Because internet connection is very important for fetching the tweets which is the most important part of product.

a) Processor: 1.8 GHz or faster. Dual-core or better recommend.
b) Memory: Minimum of 512 MB RAM is recommended.
c) Secondary Storage space: minimum 10 MB.

5.1.2 Software interfaces

"Twitter sentiment Analyzer" is tested on only Microsoft Windows. It'll need internet connection, tweets will be fetched, cleaned, labeled and then sentiment analysis will be done.

5.1.3 Communications interfaces

"Twitter Sentiment Analyzer" have to use internet connection for fetching tweets.

6. User interfaces

6.1 GUI

As the project is not a development and much research based. Analysis are involved so, it doesn't have heavy Graphical User interface like development product. Well, at least have simple GUI which shows the Graph of popularity of all 3 parties (PTI, PML(N), PPPP) on daily bases based on sentiment analysis.
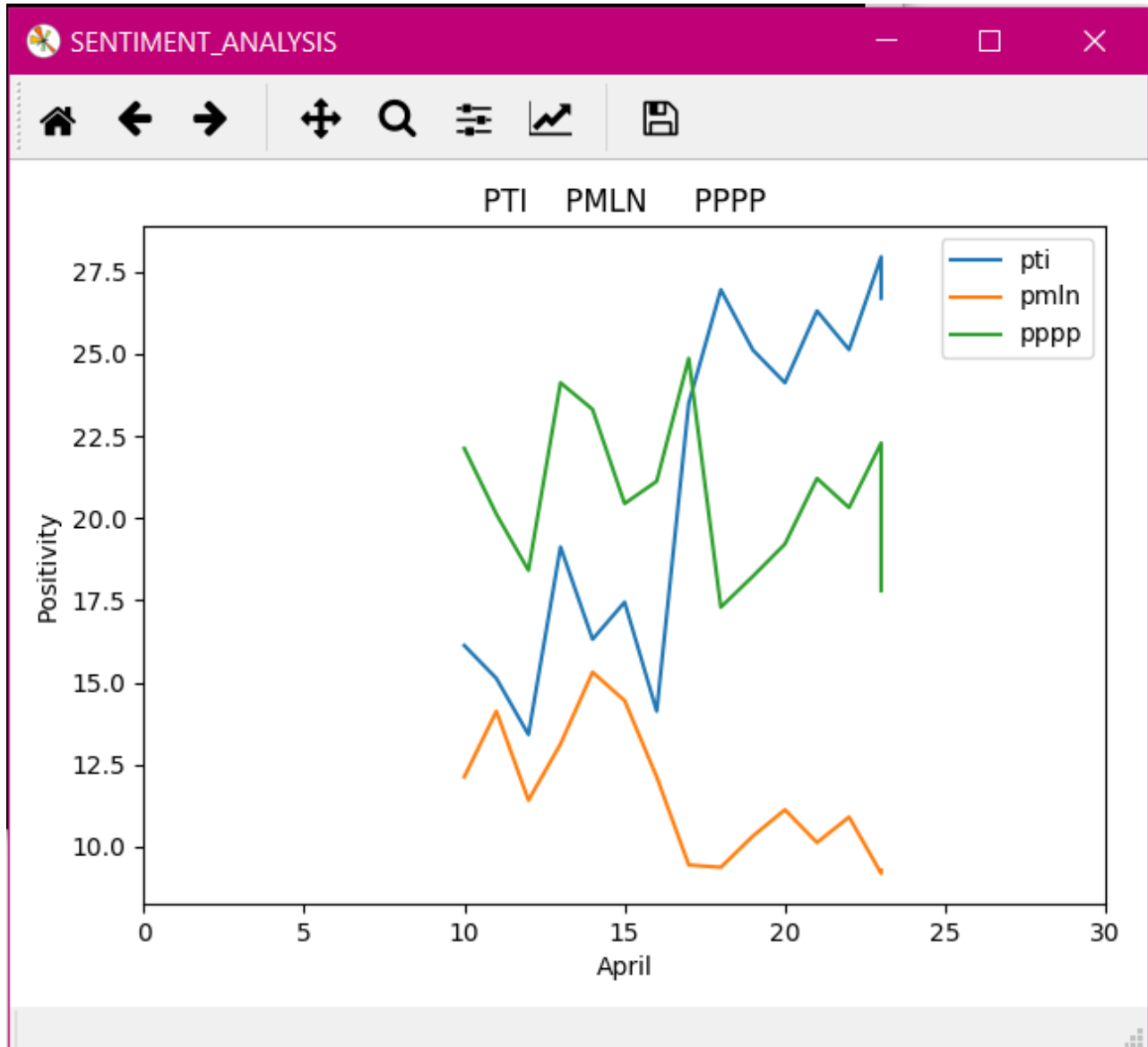
7. Non-functional requirements (NFR)

7.1 Reliability

Our software is reliable. We have gathered data from reliable sources. During all process reliability was our major concern. We also have focus on reliable and accurate results. The software will not fail in usual cases i.e. when work load is low. But with the greater workload it may slows down because of processing a long data. Internet connection will also matter.

7.2 Availability

The software will be available for user anytime and anyone can check accuracy, Internet will be needed. Because we doing live streaming that's why lot of data is coming from twitter accounts. Anyone can use it for his/her own tweets predictions by providing account credentials.

7.3 Security

Our Software will make sure that only authorized person can have used it. For that purpose, He/she get permission from software administrators. If administrators provide access, he/she may use this software.

7.4 Maintainability

In case of any bug or error, User will not be able to solve the problem. For the solution of problem, he has to contact with the developing team to fix that problem and give better suggestion about updated version.

7.5 Scalability

It will provide services regarding prediction of election on major level. So, for halve, prediction is not as efficient as we are expecting. In near future changes can be made easily according to the change in requirements. This software will be able to provide services all the time.

7.6 Flexibility

This software is very easy to use and understand. You have to just give credentials of your tweeter accounts then it will start working. For all inexperienced and experienced user. It will let the user to perform task easily without being stuck.

8. Operating environment

The system will be runs in any operating system environment with relevant IDE of Python. But the software is only tested in Windows operating system.

9.Dependencies and Assumptions

9.1 Dependencies

We assume that administrators should not be biased and maintain and manage the data correctly and sincerely.

9.2 Assumptions

The coding and scripting must be error free.

The system should be very user friendly so that user can use it easily and understand it.

The information of all the Parties should store in JSON file.