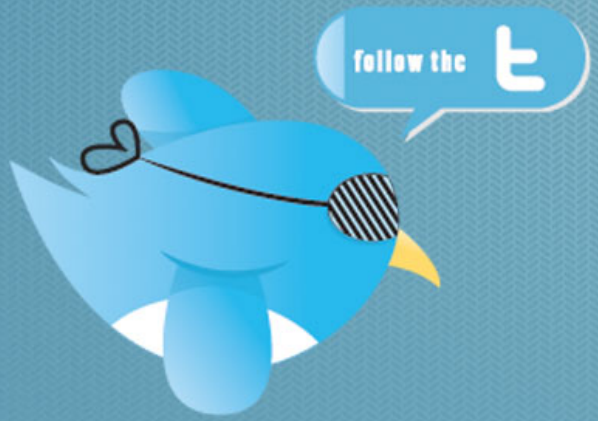


Sentimentalytics

Twitter : Social Media and Text Analytics

Eldridge Adriel Gomes - 171046003
Surya S - 171046022

Guide : Deepak Rao

Introduction

The objective of this project is to analyze the sentiment of twitter users through their tweets to understand whether the reaction of the people is positive, negative or neutral towards a particular event.

Specifications

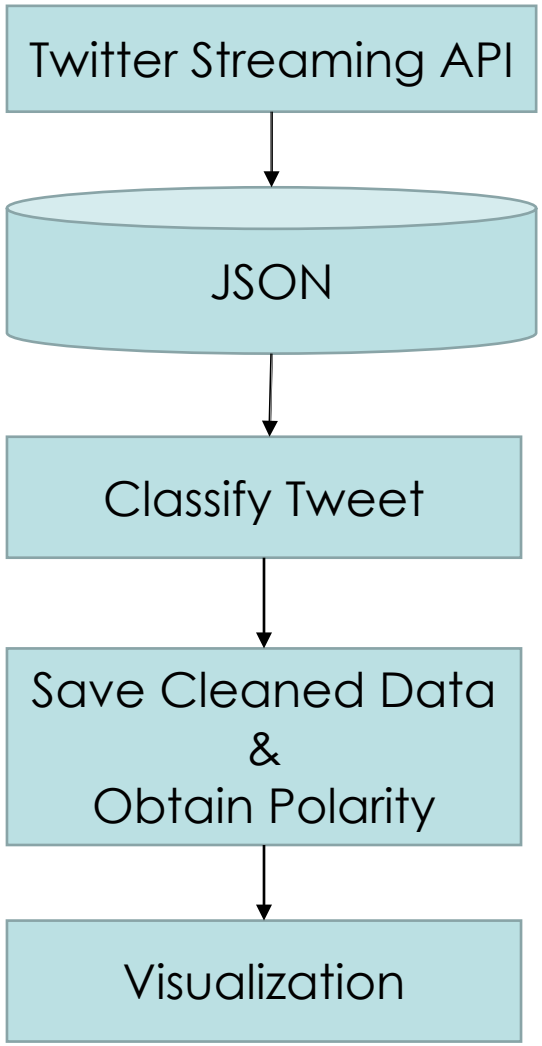
Software

Operating System	OS X, Windows, Ubuntu
Programming Language	Python 2.7 or above
Python Packages	Numpy, Pandas, Matplotlib, Tweepy, NLTK, TextBlob
Storage	Spreadsheet

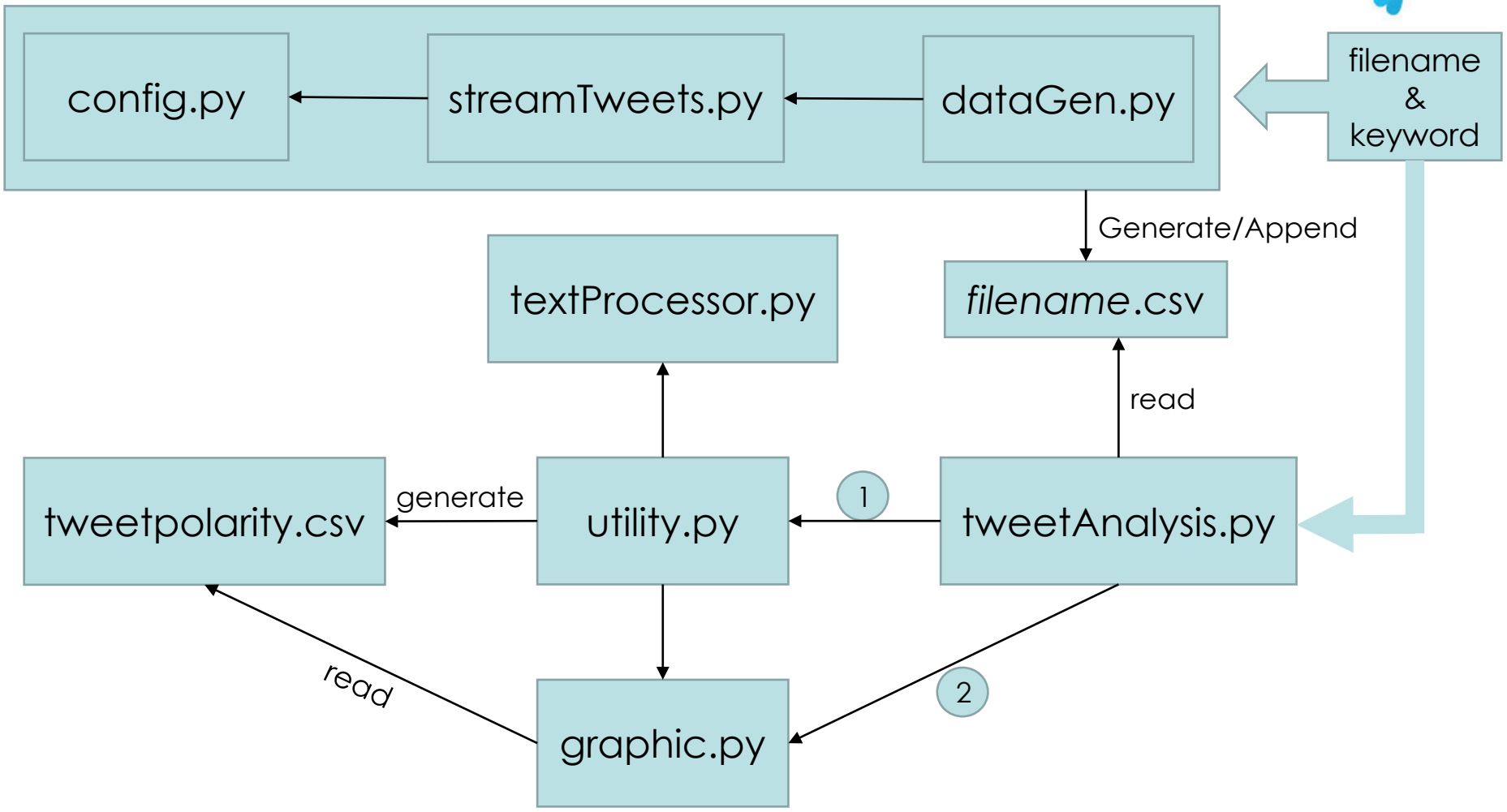
Hardware

Processor	64 Bit Intel or AMD CPU
Primary Memory	4 Gigabytes or above
Secondary Memory	50 Gigabytes or above
Network Connection	2 Megabits per sec or above

Flow Diagram



Architecture



Modules

- **twitterStream.py**: Function and class to gather tweets using twitter api and save in JSON format.
- **textProcessor.py**: Functions to clean tweet and classify the tweet polarity using TextBlob.
- **utility.py**: Function to save cleaned data and polarity in a csv file, other functions.
- **graphic.py**: Functions for different kinds of charts.

Streaming API

Twitter's Streaming API is a push of data as tweets happen in near real-time. Users register a set of criteria (keywords, usernames, locations) and as tweets match the criteria, they are pushed directly to the user.

Streaming API

Consumer Key is essentially the API key associated with Twitter. This key (or 'client ID', as Twitter calls it) is what identifies the client.

Consumer Secret is the client password that is used to authenticate the client with the authentication server.

Access Token is issued to the client once the client successfully authenticates itself (using the consumer key & secret) and defines the privileges of the client.

Every time the client wants to access the end-user's data, the **Access Token Secret** is sent with the access token as a password.

TextBlob

TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

The **textblob.sentiments** module contains sentiment analysis implementation, NaiveBayesAnalyzer (an NLTK classifier trained on a movie reviews corpus).

Problems Faced

Only the tweets in English will be classified.

The classification of tweets will be done based on probabilities of positive, negative and neutral words which is defined by the training dataset.

The model could fail on various constructs of the English language such as Sarcasm, Idioms and Phrases, Negated Sentences etc.

Results

A .csv file is generated which has the tweets, timestamps and the polarity of the tweets.

Various graphs are plotted using the data from this file which is used to infer the reaction of the people towards a particular event.

Contributions

Eldridge:
Design, Streaming Data, Utility Functions, Text Processing and Classification.

Surya:
Graphs, Utility Functions ,Text Processing and Classification.

Thank You! :)