Sentiment Analysis of twitter feed

Project based learning

Project Goals

- The goal of this project is to create an application that can analyze the sentiment of tweets in real time and provide the user with interesting and informative data. The application takes keywords from the user and searches the twitter feed for tweets that contain those particular hashtags.

Description about the Data Sets
User Profile

- User profile data is in JSON format.
- Followers count: The number of followers this account currently has.

"followers_count": 21

- Friends count: The number of users this account is following (AKA their “followings”). "friends_count": 32
- Id: The integer representation of the unique identifier for this User. "id":6253282
- Location
- Language.

Description about the Data Sets
Tweet data

- Tweet data is in JSON format
- Id : A unique IRI for the tweet.

When storing Tweets, this should be used as the unique identifier or primary key. "id": "tag:search.twitter.com,2005:3477692434099773 44“

- Actor: An object representing the twitter user who tweeted. The Actor Object refers to a Twitter User, and contains all metadata relevant to that user.
- Body: The tweet text.
- Posted time: The time the action occurred, e.g. the time the Tweet was posted. postedTime": "2013-06-25T17:12:52.000Z“
- Location: contains country code, name, coordinates.

Progress of the Project

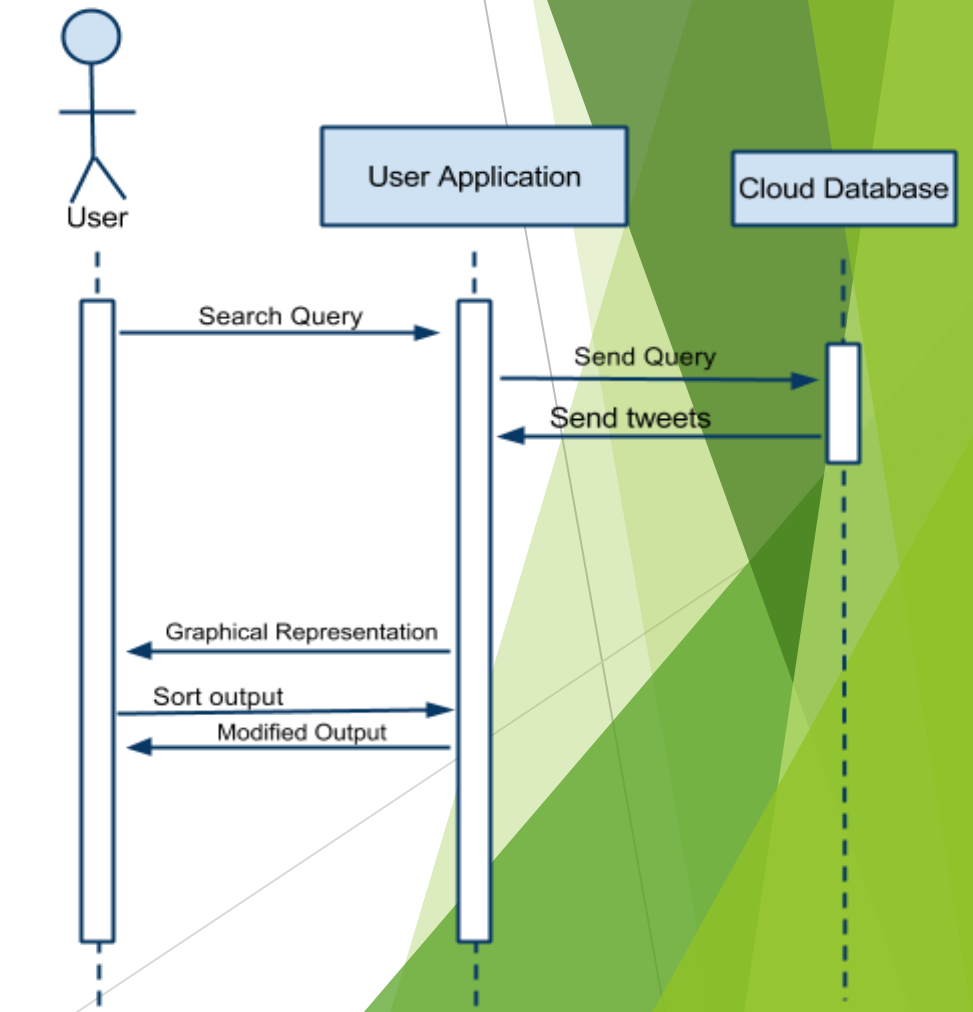
- Software Requirements document completed
- Twitter Streaming API is ready
- Class diagram and sequence diagram created
- Project architecture created

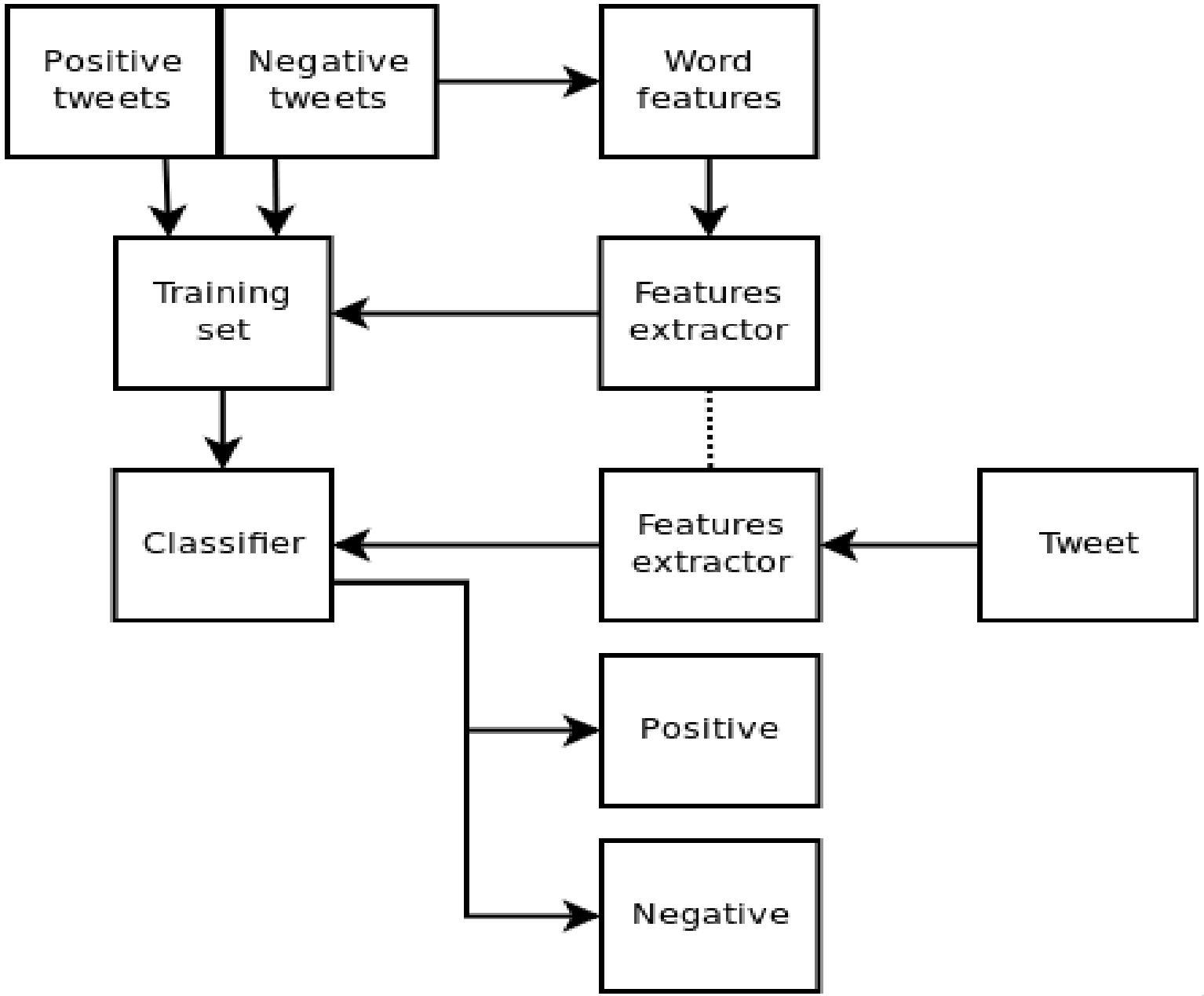
Description about the SRS

- The Software Requirements have been dictated by the SRS. The project must have all the functionalities detailed in the SRS.
- The User Interface must accept the search query from the user and perform the processes required to generate an interactive graphical representation of the results.
- The results must show the sentiment analysis of the tweets that correspond to the search query.
- The project must run optimally and according to the requirements specified in the SRS.

Description about the Design with Class Diagram and Sequence Diagram

- The class diagram and the sequence diagram are given below and in the design document.
- The show the flow of control in the program and the return of results.




Positive tweets
Negative tweets
Word features
Training set
Features extractor
Classifier
Features extractor
Tweet
Positive
Negative

Implementation

- All 3 modules are programmed in python language.
- Implemented 2 sentimental analysis algorithms (Basic Weighted analysis and Naive- Bayes analysis).
- Naive- Bayes classifier is implemented using NLTK library for python.
- Twitter data is retrieved using Twitter Streaming API.
- Training data is stored as a .csv file (Comma Separated Values).
- Twitter data is stored in JSON format.
- The 2 algorithms take this twitter data and return the overall sentiment in terms of positive or negative sentiment.

Results

- We can analyse the overall sentiment of multiple tweets which contain the users interests.
- Using this application we can analyse the popularity of various events, people and products.
- This application can be used to monitor the success and failure of products and market research.

```
https://t.co/qCUYDnHh8g - Tickets are now on sale for BERNIE SANDERS DEBATE-WATCHING PARTY!
Sentiment of the tweet: positive
@anastasia_omo let me guess, Bernie Sanders 2016? Haha
Sentiment of the tweet: positive

The number of Positive tweets are :  80

The number of Negative tweets are :  16
Please enter the file name of the data file(.json)
```

```
" RT @jk_rowling: How horrible.  Voldemort was nowhere near as bad. https://t.co/hFO0XmOpPH " : Had score  -6.0
" RT @Phxlippe: Trump and Katie Hopkins will be the worst right wing combination since Aiden McGeady and Tony Hibbert smh " : Had score  -3.0
" @LindseyGrahamSC   TRUMP: Dear establishment conservatives, WE HATE YOU. You're all habitual liars &amp;amp; only want self-enrichment " : Had score
 -3.0
" RT @EyeOnFreedom: Proof that @Reince and the #WashingtonCartel do not give a damn what @GOP voters want!  https://t.co/ia39EQawPa " : Had score
 -3.0


The overall popularity of this mention is  NEGATIVE :  -0.840855106888

The number of Positive tweets are :  175

The number of Negetive tweets are :  246
krunal@krunal-Inspiron-3542:~/twitter-api-sentiment-analysis$
```

Sentiment analysis for candidates of 2016 US elections

Conclusion and future work

Sentiment analysis is a very wide branch for research. We have covered some of the important aspects. We plan ahead to improve our algorithm used for determining the sentiment value.

Also the project as of now can also be expanded to other social media platform usages like movie Reviews(IMDB reviews), personal blogs.

Sarcastic comments are the ones which are very difficult to identify. Tweets containing sarcastic comments give exactly opposite results owing to the mindset of the author.

Bibliography

- An Introduction to Text Mining using Twitter Streaming API and Python -a blogpost
- Text Classification for Sentiment Analysis - Naive Bayes Classifier
- Natural Language Toolkit (NLTK) –toolkit for python
- "Sentimental Analysis", Inc. [Online]. Available:

http://www.cs.uic.edu/~liub/FBS/sentiment-analysis