SYRACUSE UNIVERSITY

Twitter Sentiment Analysis with Geographic Visualization

Author:
Chetanraj KADAM - 250256631
Pushkar TATIYA - 212517742
Sakshi SALOKHE - 201026946
Varun MULAY - 721214633

Professor:
Martin HARRISON

May 2, 2018

Contents

Abstract

The application is designed in order to provide sentimental analysis of tweets related to a particular hashtag and to display parts of the world or a particular area did most of the tweets come from. The homepage will display the top 10 trending hashtags. From the displayed 10 hashtags, the user can click on any hashtag of his choice. Based on the hashtag chosen, all the tweets related to that particular hashtag will be extracted. Sentimental Analysis is performed on the extracted tweets to find if the tweets are positive, negative or neutral. After performing Sentiment Analysis, the location tag of each tweet which were extracted, will be used to plot locations of the tweets over the map with different colours for positive, negative and neutral sentiments.

1 Introduction

Sentiment Analysis is the process of determining the emotional tone behind a series of words, used to gain an understanding of the the attitudes, opinions and emotions expressed within an online mention. It is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organisations across the world. Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. Few real life examples of Sentiment Analysis being applied are: The Obama administration used sentiment analysis to gauge public opinion to policy announcements and campaign messages ahead of 2012 presidential election.

The ability to quickly understand consumer attitudes and react accordingly is something that Expedia Canada took advantage of when they noticed that there was a steady increase in negative feedback to the music used in one of their television adverts.

In the similar way, the application that we have developed will be used to do sentiment analysis over tweets. The basic motive of this application is to find out the sentiment analysis of tweets for a particular hashtag and to find out the region where these tweets came from based on their categories (Positive, Negative and Neutral). The advantage of this application is to give the user the flexibility to choose hashtag from the top trending ones and select the number of tweets that he wants to analyse.

2 Specifications

2.1 Source of data

We used data extracted from twitter which included the tweets, user location and coordinates of the user who tweeted.

2.2 API used

We used twitter app API for extraction of Twitter data and Google maps API to plot the coordinates on the map.

2.3 Libraries used

1. Tweepy

2. Pandas

3. Textblob

4. Numpy

5. Geocoder

6. Bokeh

3 Algorithm

1. We use TextBlob for sentiment analysis which is built on the basis of Naive Bayes algorithm.

2. TextBlob breaks the tweet into words and Naive Bayes algorithm processes every word and assigns polarity to the tweets.

3. If the polarity of the word is greater than 0 (polarity ¿ 0), then the word is considered to be Positive.

   If the polarity of the word is equal to 0 (polarity = 0), then the word is considered to be Neutral.

   If the polarity of the word is less than 0 (polarity ¡ 0), then the word is considered to be Negative.

4. Naive Bayes works on the basis of the formula :

$$P(X \mid c) = \frac{P(c \mid X) \; P(c)}{P(X)}$$

5. Use of Pandas to read and write the data into csv file.

6. Bokeh and Google maps API was used for plotting the coordinates on the map along with colors assigned according to the sentiment for the tweet.

4 Procedure

1. When the program is executed, the user will get a list of top 10 trending hashtags for that particular day.

2. User can select the hashtag and number of tweets for that hashtag for which he wants to perform the analysis.

3. Sentiment Analysis will be performed on the data extracted. A pie chart is shown denoting the percentage positive, negative and neutral tweets. The information about the sentiment for every tweet is maintained in a CSV file.

4. Location data is extracted for every tweet and stored in the form of latitude and longitude.

5. The CSV file containing the latitude, longitude and sentiment label is then plotted on the map to get the final output.
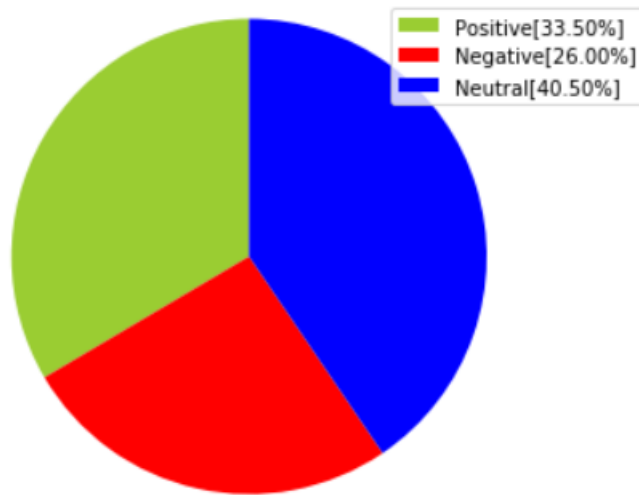
5 Output

```
1 #DíaDelTrabajo
2 #RCBvMI
3 Casemiro
4 #f82018
5 Román
6 #LegendBamBamDay
7 #عاده_اجتماعيه_دمرتنا
8 Oculus Go
9 Rıza Kayaalp
10 Universo 2000


serial number of the hashtag to be analysed:
1
you have selected  #DíaDelTrabajo
Enter the number of tweets to analyze200
```
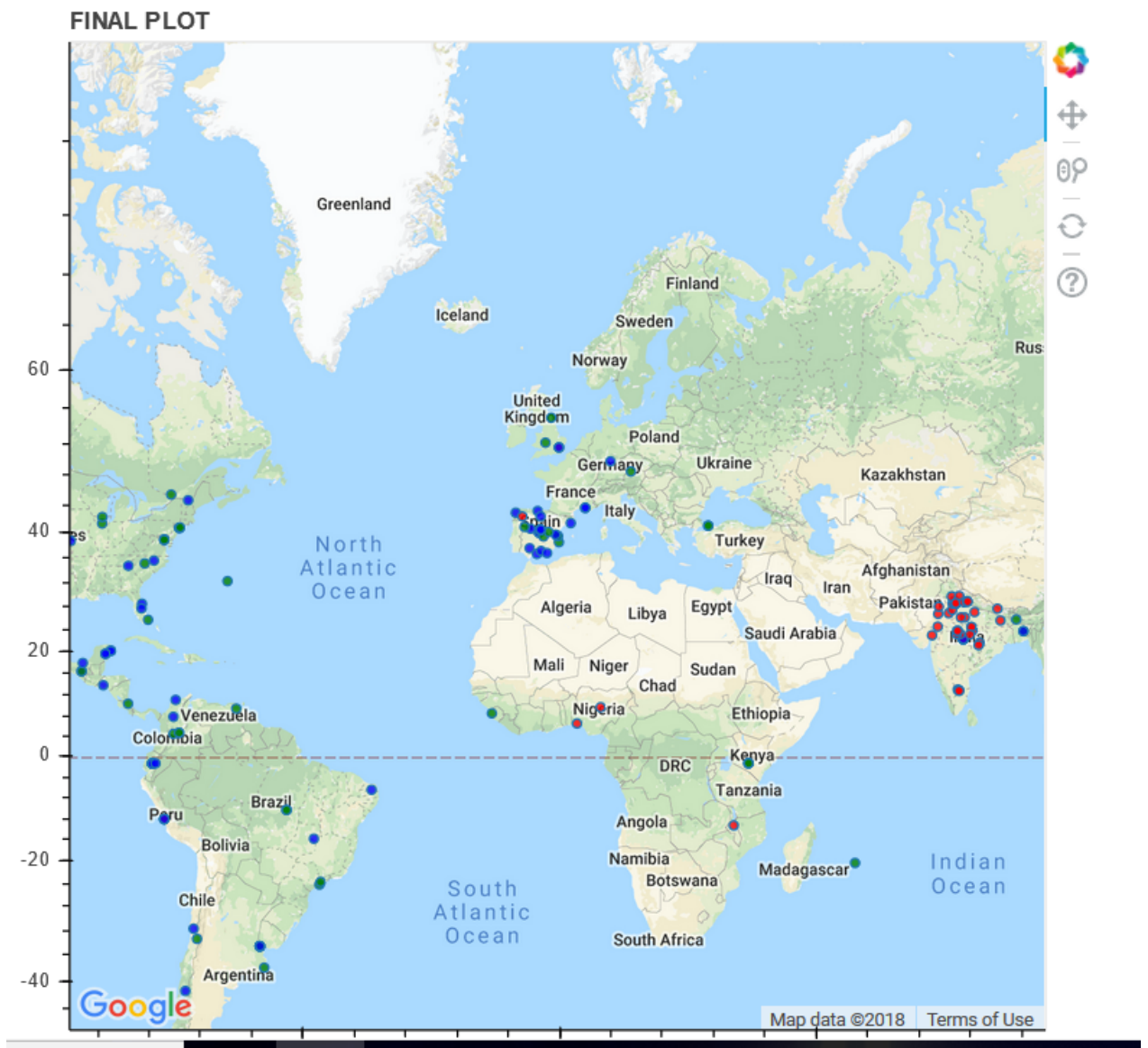
7

SENTIMENT ANALYSIS ON THE SELECTED HASHTAG

BokehJS 0.12.15 successfully loaded.

FINAL PLOT

6   Conclusions

Geographic visualization allows visual access to huge amounts of data which is otherwise hard to analyze. Data visualization gives us easily digestible visuals by providing the main gist of the huge amount of data on a particular topic. In particular, geographic visualization helps you understand trends, patterns and to make correlations. This type of visualization provides regional quantitative information. This project aims to scrape tweets data of trending hashtags and map them these tweets by their geographic location on the world map. Also, additionally, we do sentiment analysis on these hashtags and present it on a pie chart.

7 Major Contribution:

During the process of building the project, the task of extraction of top trending tweets and extracting the top 10 out of them and displaying it and taking user input was done by Pushkar Tatiya and Sakshi Salokhe. The sentiment Analysis part of the project was done by Varun Mulay. The task of plotting the location data on google maps by preprocessing the stored data was done by Chetanraj Kadam, Pushkar Tatiya and Sakshi Salokhe. The task of integrating the project and the documentation part was done by all the team members collectively.

References

[1] Visualizing Geocode Data
    http://support.gnip.com/articles/visualizing-twitter-geo-data.html

[2] Extract data from twitter
    http://www.mikaelbrunila.fi/2017/03/27/scraping-extracting-mapping-geodata-twit

[3] Plotting the data on map.
    https://bokeh.pydata.org/en/latest/docs/user_guide/geo.html