TWITOSPHERE

Team Members:-

Srishti Verma - E18CSE181

Tanishi Tyagi - E18CSE186

Tushar Mittal - E18CSE191

Vishvesh Gupta - E18CSE213

Mentor:

Dr. Tanveer Ahmed

Introduction

Our project is an application of sentiment analysis of Twitter. Sentiment analysis is a great tool with various industrial use cases. It can identify and analyse many pieces of text automatically and quickly. Through this project we are trying to do the sentiment analysis of live tweets fetched from Twitter using python libraries. It will determine whether a tweet is positive, negative or neutral. We will combine NLP (natural language processing) and machine learning techniques to analyse and classify live tweets as positive, negative or neutral. By listening to and analysing comments on Twitter, local government departments can gauge public sentiment towards their department and the services they provide and use the results to improve services such as parking and leisure facilities, local policing, and the condition of roads. Businesses can compare their results with those of their competitors to better understand people's attitude to their business. They can identify where they may be excelling or identify where there's room for improvement compared to the competition.

Data Set:

The data set used by us is a pretrained data set in python libraries.

The library used is based on NLTK and its name is TextBlob.

This library has pretrained data set based on movie reviews.

Problem statement:

Sentiment analysis is considered a relatively new topic in the domain of micro-blogging and there is a lot of scope in this field. There are a lot of other websites that offer reviews on different topics, but we primarily chose twitter because of the limit of 140 characters per tweet through which the user is forced to express his/her opinion in a very compressed form.

For sentiment analysis we are using a python library, TextBlob, which is derived from NLTK and uses supervised learning techniques such as Naïve Bayes. It classifies the parsed tweet, fetched using twitter API, into 3 categories i.e. Positive, negative and neutral.

Why the project was chosen

We decided to do this project because sentiment analysis is a tool that can help any organisation or a group which values public sentiment or attitude for their success and growth. We see millions of people on social media sites writing blogs etc to share their experiences and views on topics related to every possible thing be it reviewing businesses or a place, writing about current affair or a recent movie that was released. By analysing these blogs and posts affected industries can analyse and work upon the issue being discussed, for example if we government departments can gauge public sentiments towards their department and services they provide and use the results generated by our application to improve services like parking

and leisure facilities, local policing etc. Businesses like restaurants, shopping complexes etc can segregate negative reviews and work upon it. Similarly, universities could also analyse student posts and tweets regarding their needs and feedbacks and develop in the right direction.

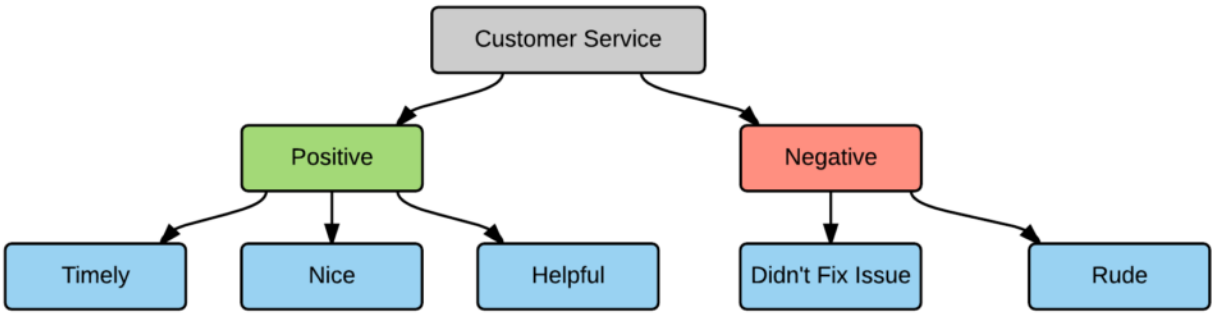
Background study

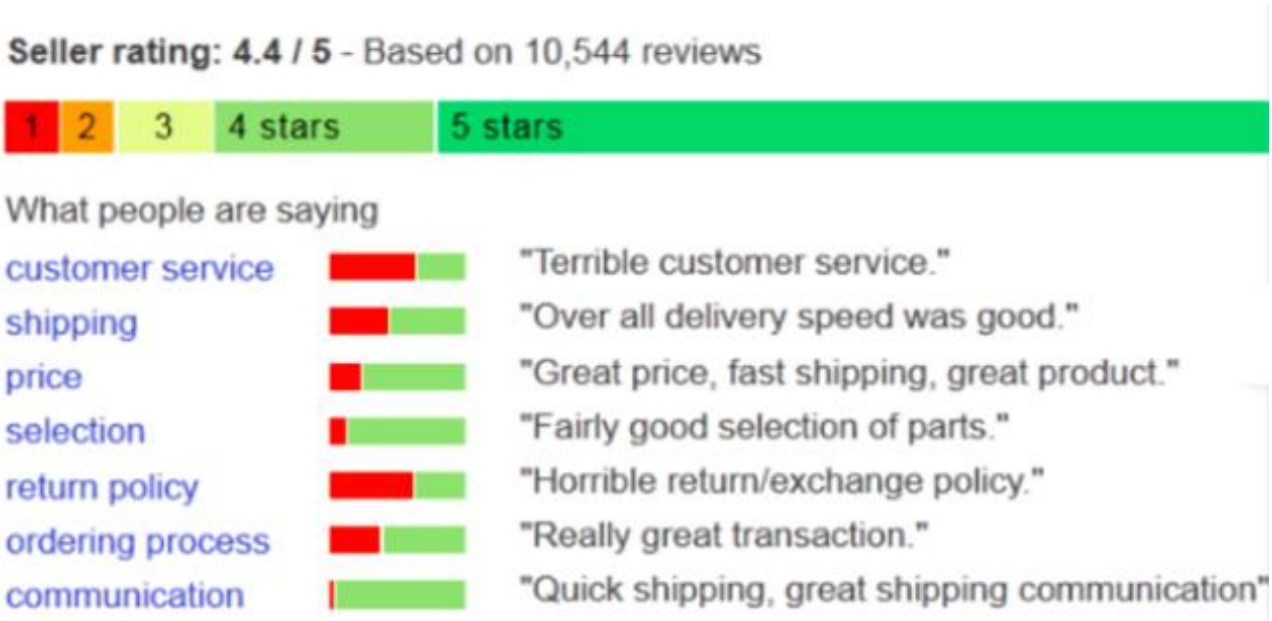
Sentiment Analysis is used my many companies now-a-days to review their customer feedback and get an idea about how their product is doing in market.

Some famous examples of this is Uber, Google, and most of the MNCs.

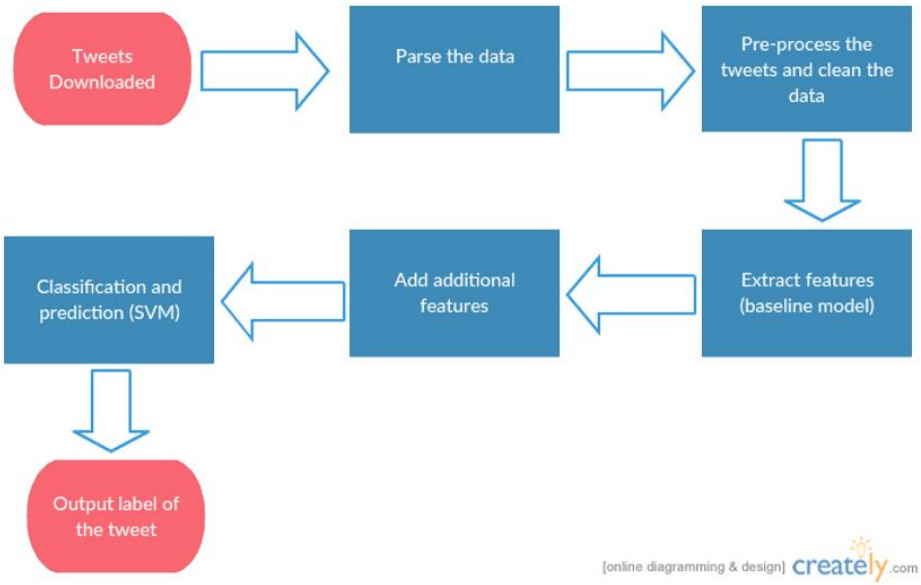
This technique is also used in Politics to analyse how a certain policy is doing amongst public.

Shown below is a real-life example sited from towardsdatascience.

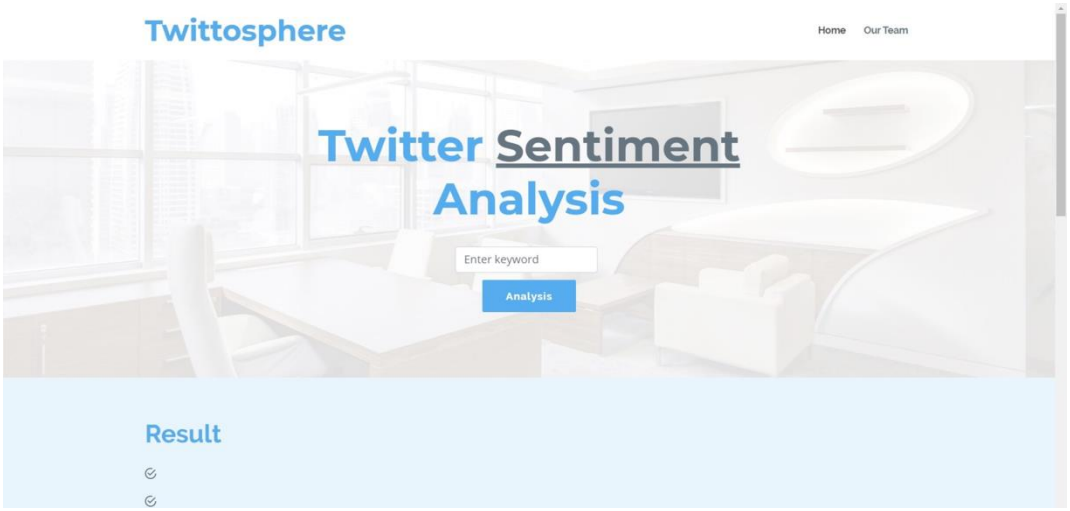


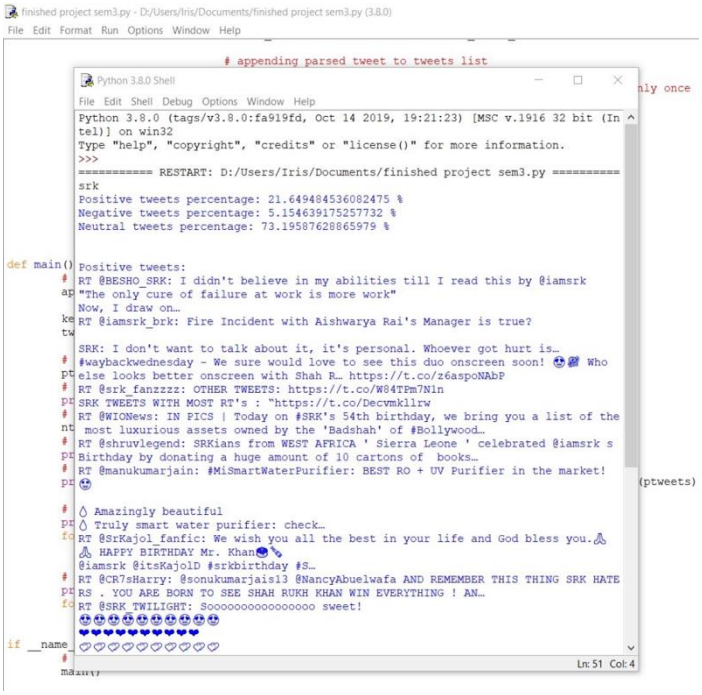


Project design diagram



Implementation Details





Functionality

1. Any keyword is entered in the frontend of our project.
2. The keyword is then passed to the python backend.
3. Tweets are fetched from twitter using their API (Tweepy).
4. The fetched tweets are then passed to then passed to the library for analysis.
5. After the analysis, the percentage of positive, negative and neutral tweets are returned
6. Everything is then pushed back to the frontend.

Limitations

Sentiment analysis tools can identify and analyse many pieces of text automatically and quickly. But computer programs have problems recognising things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognise these can skew the results. With short sentences and pieces of text, for example like those you find on Twitter especially, and sometimes on Facebook, there might not be enough context for a reliable sentiment analysis. However, in general, Twitter has a reputation for being a good source of information for sentiment analysis, and with the new increased word count for tweets it's likely it will become even more useful. Also the program can be less accurate than we expect it to be, the main reason for this would be testing of our program with less complex data due to time constraints. Regardless, we will try our best to achieve and showcase what we have planned

Project Development Time Details

1st week of September 2019, dated : (1 / 09 / 19 - 6 / 09 / 19) :
Finalising the project after discussion with mentor, seniors and other knowns

1st week of September 2019, dated : (6 / 09 / 19 - 7 / 09 / 19) :
Working on software require specification (srs)and documentation of our project A shallow study about the topic and required resources. Learning about the libraries involved.

2nd week of September 2019, dated : (8 / 09 / 19 - 11 / 09 / 19) :
Learning about extracting twitter API. Gaining knowledge regarding API.

2nd week of September 2019, dated:(12 / 09 / 19 - 14 / 09 / 19) :
revising python done in first semester, learning about tweepy (twitter API client for python)

3rd week of September 2019, dated :(15 / 09 / 19 - 18 / 09 / 19) :
Learning about Textblop (python library used)

4th week of September 2019, dated : (21 / 09 / 19 - 28 / 09 / 19) :
Frequently contacting mentor regarding any changes in project

1st week of October 2019 (1 / 10 / 19 - 5/10 / 19) :

Twitter changes API policies, making necessary amendments in project. Giving a market use to our product

1st week of October 2019 (6 / 10 / 19 - 7 / 10 / 19):

Starting collecting data using twitter API

2nd week of October 2019 (8 / 10 / 19 - 14 / 10 / 19):

basic prototype ready and will be discussed with mentor

Final week:

project final presentation Incase we somehow fail in delivering project on time due to some unforeseen and some unavoidable events, a second approach will be adopted wherein some modules of the project will be cut down.

Learning and Reflection from the Project

We've covered multiple courses in the past (both technical and non-technical). Our project combines both theoretical and practical knowledge gained. A brief of which is as follows. All the coding part of our project is done in python which was covered in semester 1. We are using built-in python modules (TextBlop, etc) in our project. TextBlop internally uses mathematical techniques such as Naive Bayes classifier to predict the polarity and subjectivity of the fetched tweet. Appropriate data handling techniques are also used to make our project more optimized. We've also been taught entrepreneurship and marketing is a very essential part of it. Through our project we can test whether people are liking the product released based on the type of feedback provided by them on twitter. To predict whether the fetched tweet is positive, negative or neutral multiple logical operations will be used. Development is a very essential part of CSE engineering and fetching details using an API plays a major role it, via our project we will also get to learn some basics about it.

Conclusion

- Sentiment analysis is a tool that evolved as AI and machine learning became more sophisticated and continues to do so. Through our application Twittospere we tried to achieve a model of twitter sentiment analysis with more precise analysing and classifying skills than the other existing ones.
- However, phrases suggesting sarcasm, irony, negation etc are hard to segregate into positive or negative with the available training data. Even a person can sometimes have little trouble identifying such sentences.

References

- Cuauhtemoc Luna-Nevarez. “Chapter 32 AnExploratory Analysis of Consumer Opinions, Ethics, and sentiment of neuro marketing: an abstract” Springer Nature, 2018
- Geeksforgeeks.com
- Lexalytics.com