Project Report

Sentiment Analysis on the Tweets of People of India about Farmers Protest Using Twitter API

Wednesday, 7th April 2021

Introduction

Every day, we generate huge amounts of text online, creating vast quantities of data about what is happening in the world and what people think. All of this text data is an invaluable resource that can be mined in order to generate meaningful business insights for analysts and organizations. But analyzing all of this content isn't easy, since converting text produced by people into structured information to analyze with a machine is a complex task. In recent years though, Natural Language Processing and Text Mining has become a lot more accessible for data scientists, analysts, and developers alike. In this project, I am going to analyse the sentiments on the Tweets of people of India who are giving feedback to Farmer Bill or Farmer's Protest.

Purpose

Today, social media covers a huge part of everyone's life. They are increasingly becoming the platform of communication for every means. Businesses can effectively utilize this by carefully listening and monitoring consumers. To properly understand customer needs, it is imperative to leverage Sentiment Analysis. Also, it can be used proactively to solve many business problems. It can also benefit Health Professionals, Policymakers, State and Central governments, and societal representatives.

Proposed Solution

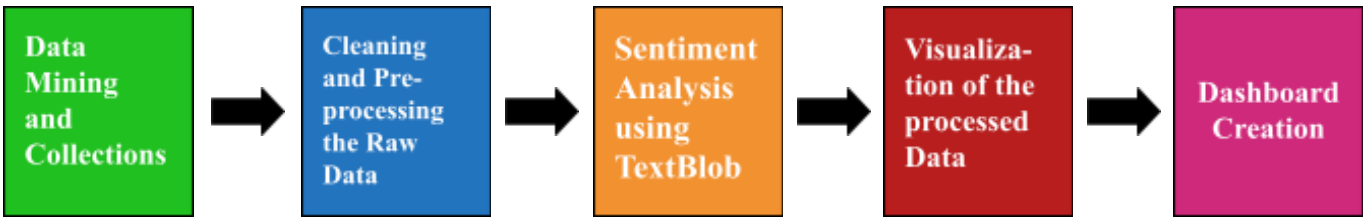
Using various Twitter and Python API's I fetched Tweets and performed Sentiment Analysis on the Tweets of the people of India to gain a wider public opinion on farmer's bill which is officially known as **Indian Agriculture Act 2020**.

This project will be completed into three phases:

1. **Building a Corpus:** I have used Tweepy Library to gather text data from Twitter's API with special keyword(#FarmerBill or #FarmerProtest).
2. **Analyzing text:** Sentiment analysis to determine the attitude of the mass is positive, negative or neutral towards the subject of interest. For that I have used TextBlob python library to calculate sentiments.
3. **Visualizing results:** Graphical representation of the sentiments is done using matplotlib and wordCloud. And this project is deployed using a platform named streamlit.

Theoretical Analysis

Block Diagram



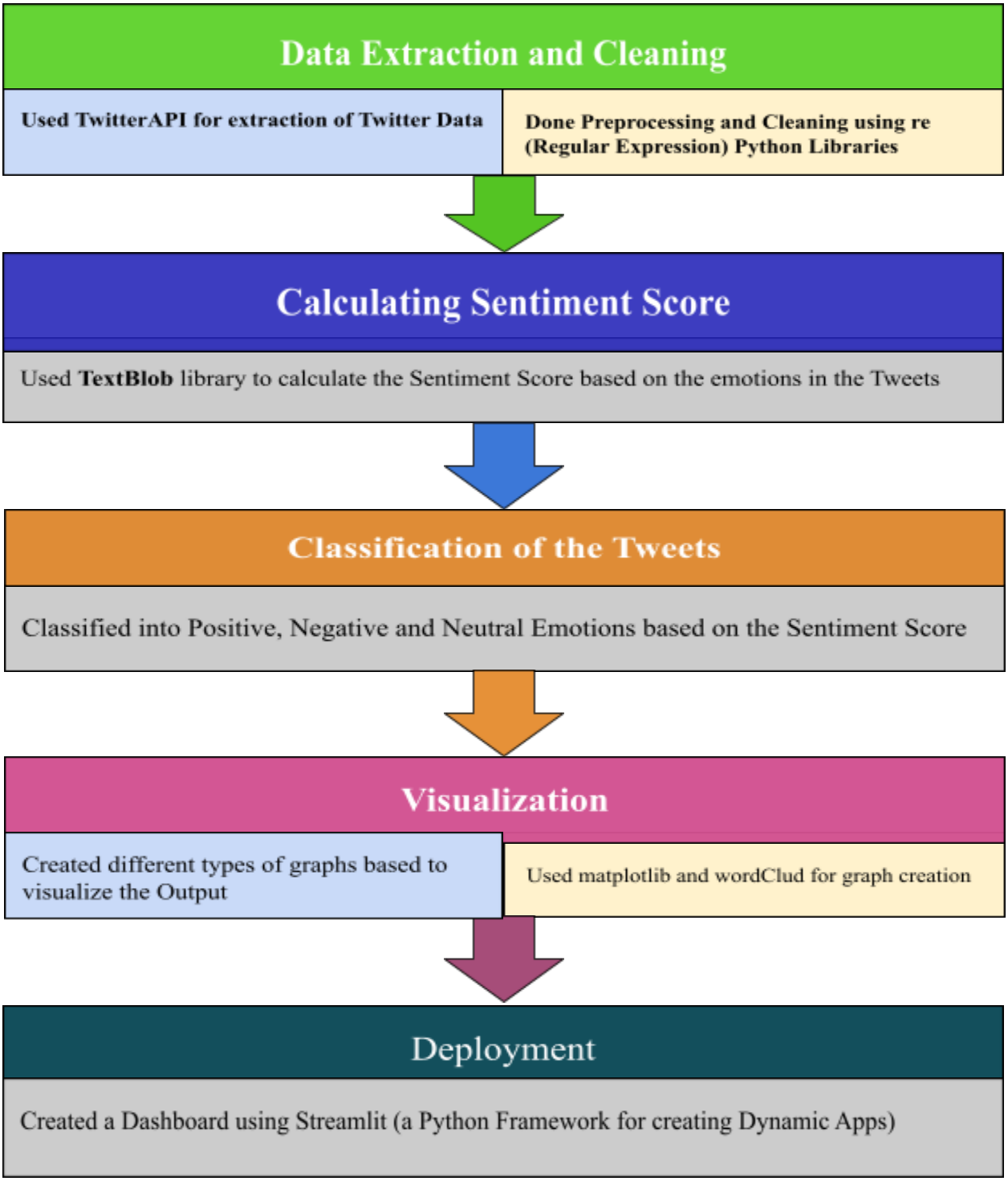
Required Libraries & Languages:

- Pycharm IDE
- Python 3.0 or above
- Tweepy Library
- Twitter API
- TextBlob Library
- Pandas
- WordCloud
- Streamlit library

Project Specifications

The project has been deployed on a Web server so any device that can support a Web browser with HTML5 support will be able to run the app. This project give results on Web Browser using Streamlit as a platform. A dashboard is created which can show visuals of this project using different graphs and wordCloud. A sidebar is created to choose which type of graph should show on the page at a time.

FLOWCHART


Data Extraction and Cleaning
Used TwitterAPI for extraction of Twitter Data
Done Preprocessing and Cleaning using re (Regular Expression) Python Libraries
Calculating Sentiment Score
Used TextBlob library to calculate the Sentiment Score based on the emotions in the Tweets
Classification of the Tweets
Classified into Positive, Negative and Neutral Emotions based on the Sentiment Score
Visualization
Created different types of graphs based to visualize the Output
Used matplotlib and wordClud for graph creation
Deployment
Created a Dashboard using Streamlit (a Python Framework for creating Dynamic Apps)

Source Code

In this project, code is written in python using Pycharm IDE as a platform.

Data Extraction:

```
#import required libraries
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import re
import streamlit as st

#Tweeter API Credentials
consumer_key="xxxxxxxxxxxxxxxxxx"
consumer_secret="XXXXXXXXXXXXXXXXXXXXXX"
access_token="xxxxxxxxxxxxxxxxxxxxxxxx"
access_token_secret="XXXXXXXXXXXXXXXXXXXXXXXXXX"
#API credential can't be given for security purposes

#Create the Authentication Object
auth=tweepy.OAuthHandler(consumer_key,consumer_secret)

#Set the access token and access token secret
auth.set_access_token(access_token,access_token_secret)

#Create the API Object while passing the Auth information
api=tweepy.API(auth)

#Extract the tweets from the tweeter user
results=api.search(q="#FarmersProtest", lang="en", result_type="recent",  count=10000)
```

Cleaning and Preprocessing:

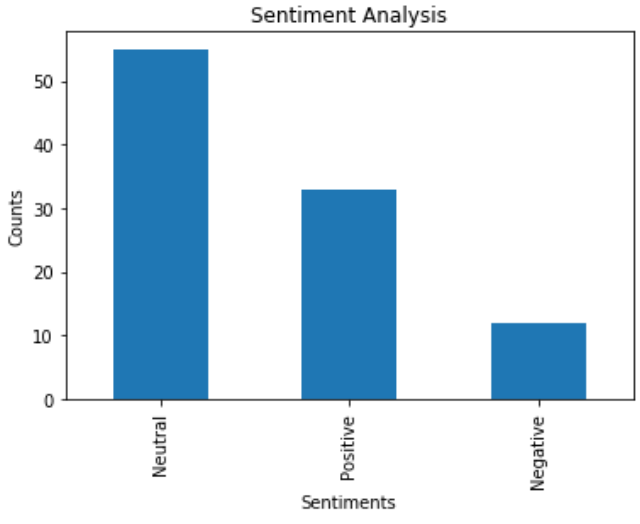
```
#Create dataframe with column named Tweets
df=pd.DataFrame([tweet.text for tweet in results], columns=['Tweets'])

#Creating function to clean tweets
def cleanTxt(text):
   text=re.sub(r'@[A-Za-z0-9]+','',text)
   text=re.sub(r'#','',text)
   text=re.sub(r'RT[\s]+','',text)
   text=re.sub(r'https?:\/\/\S+','',text)
   text=re.sub(r':','',text)
   return text
df['Tweets']=df['Tweets'].apply(cleanTxt)
```

Sentiment Analysis:

```
#Get subjectivity and polarity with the help of TextBlob
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#Adding two new columns in Dataframe named Subjectivity and Polarity
df['Subjectivity']=df['Tweets'].apply(getSubjectivity)
df['Polarity']=df['Tweets'].apply(getPolarity)
print(df)

#Creating function to get positive, negative or neutral values
def getAnalysis(score):
    if score < 0:
        return 'Negative'
    elif score==0:
        return 'Neutral'
    else:
        return 'Positive'

#Adding analysis column into dataframe
df['Analysis']=df['Polarity'].apply(getAnalysis)
print(df)
```

Visualization:

```
#Visualization of tweets
st.subheader('Bar Graph')
st.bar_chart(df['Analysis'].value_counts())
```



#Creating Image showing mostly used word in bold size

```
allWords=' '.join([twts for twts in df['Tweets']])
wordCloud=WordCloud(width=350,height=210,random_state=20,max_font_size=119).generate(
allWords)
st.image(wordCloud.to_array())
```

Analysis:

#Getting percentage of negative tweets

```
ntweets=df[df.Analysis=='Negative']
ntweets=ntweets['Tweets']
neg=round((ntweets.shape[0]/df.shape[0]*100),1)
st.subheader('Getting Percentage of calculated sentiments of tweets')
st.write('Negative Tweets(%): ')
st.write(neg)
```

#Getting percentage of positive tweets

```
ptweets=df[df.Analysis=='Positive']
ptweets=ptweets['Tweets']
pos=round((ptweets.shape[0]/df.shape[0]*100),1)
st.write('Positive Tweets(%): ')
st.write(pos)
```

#Getting percentage of neutral tweets

```
netweets=df[df.Analysis=='Neutral']
netweets=netweets['Tweets']
neu=round((netweets.shape[0]/df.shape[0]*100),1)
st.write('Neutral Tweets(%): ')
st.write(neu)
```

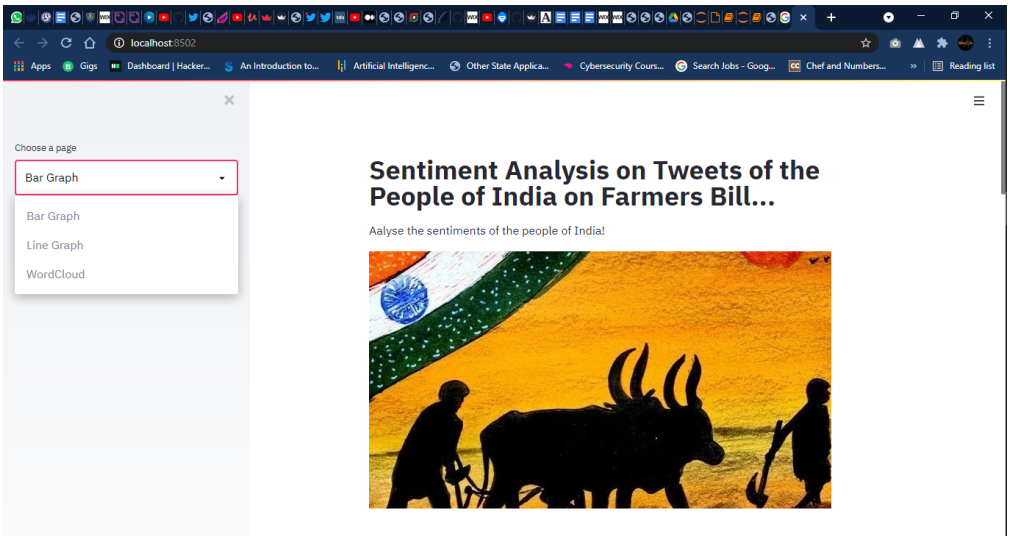
```
Negative:  8.0 %
Positive:  51.0 %
Neutral:  41.0 %
```

Some of the extracted Tweets:

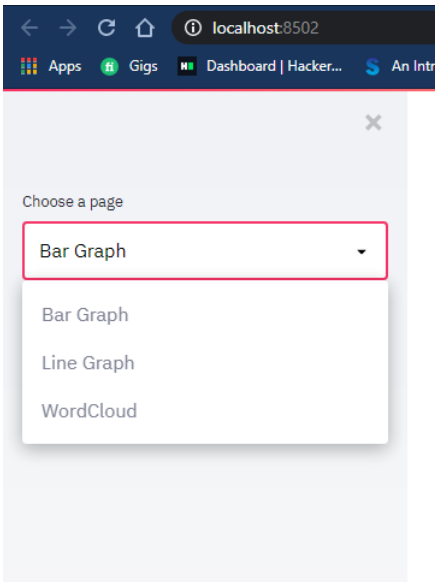
	Tweets	Subjectivity	Polarity
0	A quick reminder Farmers are still protesting...	0.500000	0.333333
1	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
2	_justice FarmersProtest goes on while Modi beg...	0.000000	0.000000
3	Lakha Sidhana has appealed everyone to reach ...	0.000000	0.000000
4	ager FarmersProtest corona spread kar rhya h...	0.000000	0.000000
5	Farmer Karamjit Singh s/o Major Singh of vill...	0.500000	0.062500
6	The peak of this protest can’t be when Rihann...	0.000000	0.000000
7	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
8	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
9	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
10	FarmersProtest enters day 133. \n\n1. Central...	0.250000	0.000000
11	India our FarmersProtest face a new edict by ...	0.454545	0.136364
12	“If we don't observe, report, and record the ...	0.125000	0.000000
13	_harry07 what are you? 😆\n\nAnti_Farmer_RSS \n...	0.000000	0.000000
14	Each day that passes by.......Making us more ...	0.616667	0.466667
15	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
16	| UP BJP leader Priyamvada Tomar resigns over...	0.000000	0.000000
17	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
18	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
19	Gautam Adani now one of world’s top 20 billio...	0.500000	0.500000
20	Lakha Sidhana has appealed everyone to reach ...	0.000000	0.000000
21	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
22	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
23	_Sekret While we worried about the media we sh...	0.700000	-0.400000
24	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
25	It’s still happening FarmersProtest	0.000000	0.000000
26	Hinduism the 3rd largest religion in the world...	0.000000	0.000000
27	When almost of all the people of their countr...	0.000000	0.000000
28	Good news coming in from FarmersProtest as Ad...	0.333333	0.350000
29	Irony \n\nA pro liar saying Farm laws ke agai...	0.000000	0.000000

Final Result shown on the Web Server

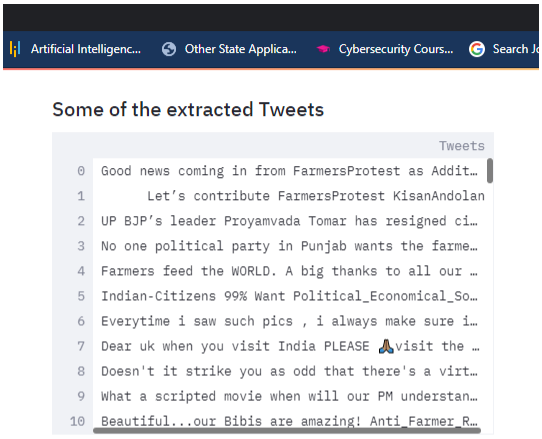
Dashboard

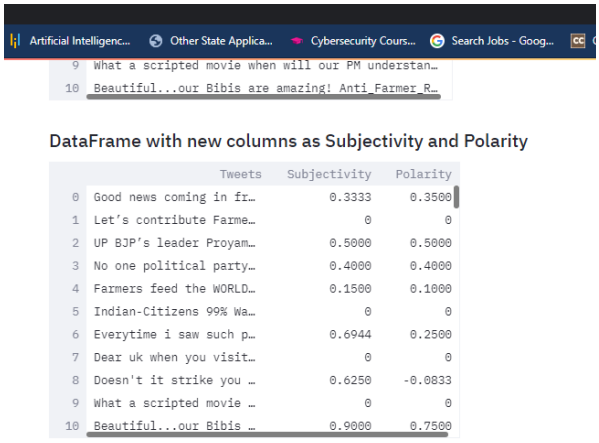


Sidebar

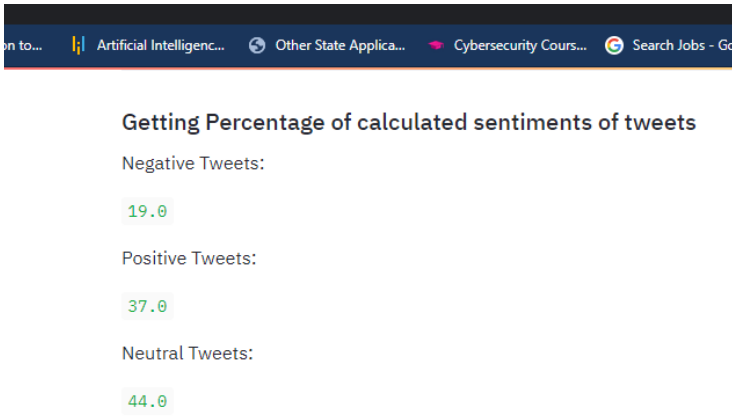


Displaying DateFrame Values

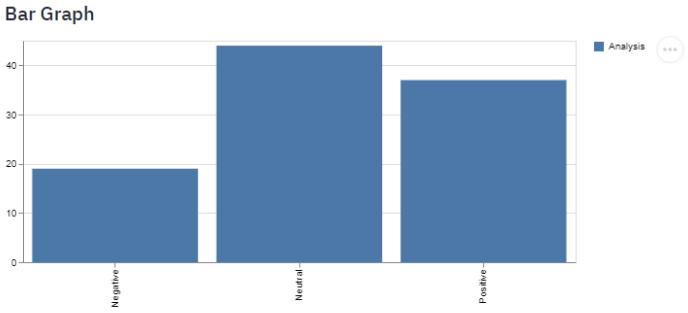


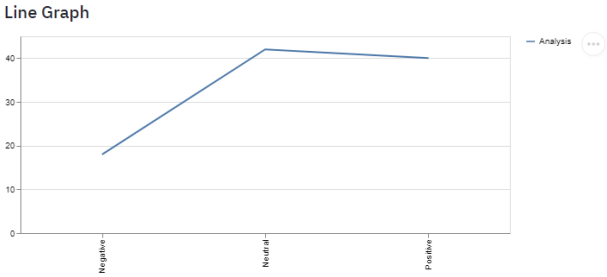


Percentage of Calculated sentiments of Tweets



Graphs





Conclusion

Sentiment analysis provides answers into what the most important issues are. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition that isn't always right. In this project, I have shown visuals of analysis which can be used to understand what people think about this Farm Bill and The Protest.

References

- https://towardsdatascience.com
- https://www.researchgate.net/publication
- https://www.sciencedirect.com/science/article

Reference Book

- Data Mining: Concept and Techniques by Jiawei Han, Michelin Kamber, Jian Pei
- The-Morgan-Kaufmann-Series-in-Data-Mining