Sentiment Analysis Using Machine Learning

Yallanki Vikas[1], K. Nagendra chary[2]

[1,2]UG Scholar, CSE Department, SNIST, Hyderabad, Telangana, India

Assistant Professor, Department of Computer Science and Engineering, SNIST, Hyderabad, Telangana, India

Abstract - **Social medias are one of the main platforms for promoting, commenting and sharing information, ideas, interests, and different kinds of expression through digital groups and social networks. The present generation is full of Micro-blogging and millions of customers share and propose their perspectives and critiques on diverse everyday existence regarding them through social media systems like Twitter, snapchat and Facebook and Instagram. Data from those web sites may be used for advertising or social studies. People describe their feelings, hate, love and their views at the social media dialogue forums or social campaigns. The enterprise agencies hire researchers to analyze the unrevealed statistics approximately their merchandise and services. The programs of sentiment evaluation embody the regions like social occasion planning, election campaigning, healthcare monitoring, purchaser merchandise and attention services.**

I.INTRODUCTION

Sentiment Analysis or opinion mining is a broadly researched subject matter as many patron orientated groups are inquisitive about receiving remarks and comparing their overall performance. One manner to correctly expect patron remarks is to generate a version and educate it the use of the records of patron's behaviours or preferences. This is a precious factor for monetary and marketplace research. there are a lot of useful resource at the internet for sentiment evaluation due to fast boom withinside the online pool of facts this is now complete of critiques like social networking websites, online blogs, forums, etc. Our cause changed into to create a machine or a device that may perceive and classify a sentiment represented in a digital textual content withinside the maximum correct manner. In this project, we use the Bernoulli, NUSVC, Pattern classifiers and sentiment lexicons to examine their overall performance in figuring out the sentiment embedded withinside the textual data.

Scope and Proposed Model

Sentiment Analysis was initially started with the idea of machines understanding the feelings of humans from their texts and comments. when that idea was successfully implemented the organizations wanted to take advantage of a popular new analysing method to reach out to customers. Thus, they can acquire the feedback from people without actually collecting it from them. This project focuses on developing a machine learning model to establish the connection between the sentiments and machines. To do this, we will be using wide range of libraries and Jupyter notebook to implement the code. Similar to some projects like house price prediction and stock market analysis, this project analyses the sentiment of the given tweet or article by training and testing the model.

II.LITERATURE SURVEY

Machine Learning

Machine learning is known as the branch of artificial intelligence which deals with the machines learning and predicting the outcomes on their own without any involvement of human beings. The machine learning models learn on their own using the labelled or un labelled data and predict or classify the new or raw data that requires to classify and use for an organization. There are wide variety of algorithms which the models uses to learn and predict, the algorithms having highest accuracy will be choose as the required one.

Python

Python is a programing language which is simple and easy to learn than many other languages. it is more flexible, portable and logic-oriented language. Modules which are required are imported at the starting code of the model and later used at the preprocessing stages.

NLTK

Natural language processing toolkit is used for preprocessing the twitter datasets collected so that only the words having polarity will be stored together. All the other neutral words will be eleminated through a series of steps in preprocessing of sentences.

Scikit learn

The scikit learn library contains all the machine learning models and algorithms in it. Once we import the scikit learn module, we can use any algorithm required for the model. Training and testing take place using train test split method and we can even calculate accuracy from it.

Pandas

It is an open-source python package which is used for handling machine learning tasks dealing with multi-dimensional datasets. Pandas is used when simple version of Numpy module is required. Pandas can be used for slicing and selecting the data from datasets.

Word Cloud

Word cloud is used for a pictorial representation of the words for better understanding of the highest repeated words. Words can be separated and grouped together by using their polarity. All the words will be represented in a single picture according to their polarity with high frequency.
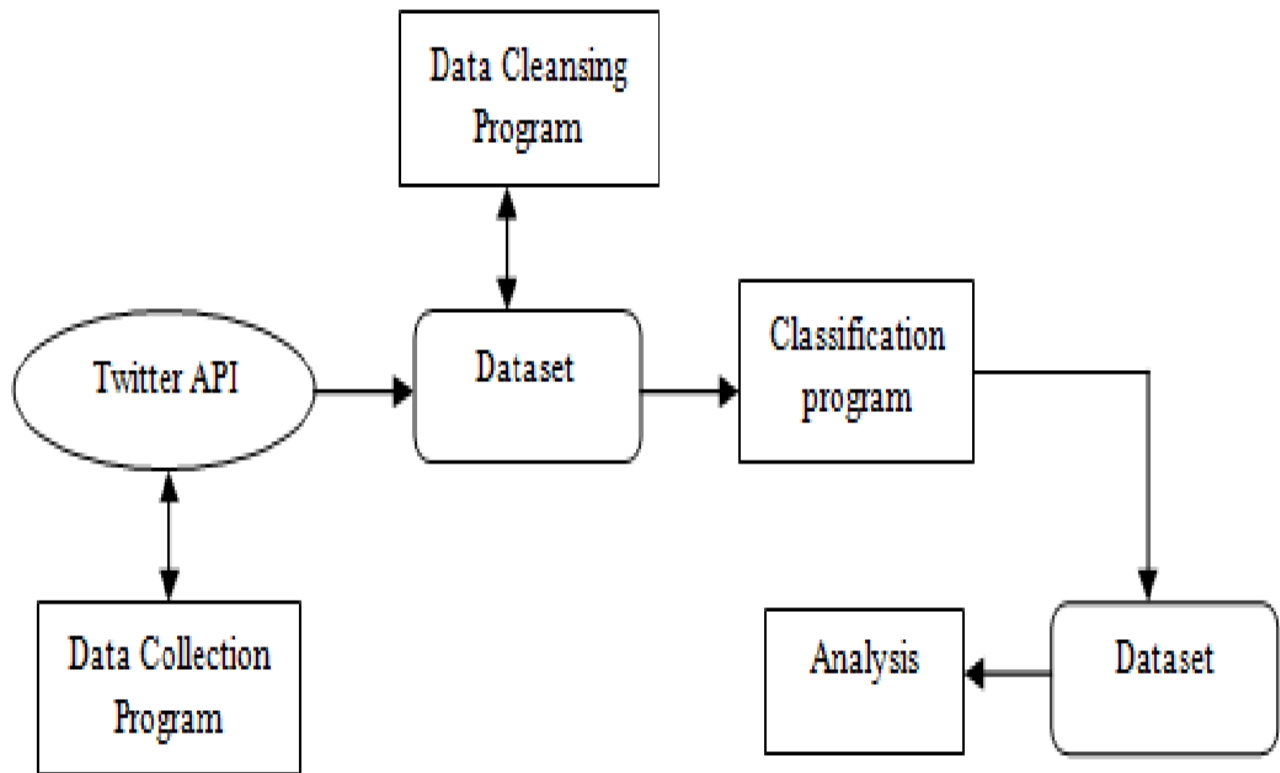
III.SYSTEM ARCHITECTURE

Architecture

Fig. 1 System Architecture diagram

The Architecture diagram shows the process of sentiment analysis through various stages. The datasets can be included directly from the available data sets or the can be collected from the Twitter when required. After all the preprocessing is completed a new data set will be generated which contains only the main words which have a polarity, all the other words will be deleted at various stages. The analysis will be performed on the final dataset formed with words having polarity.

Data Flow Diagram
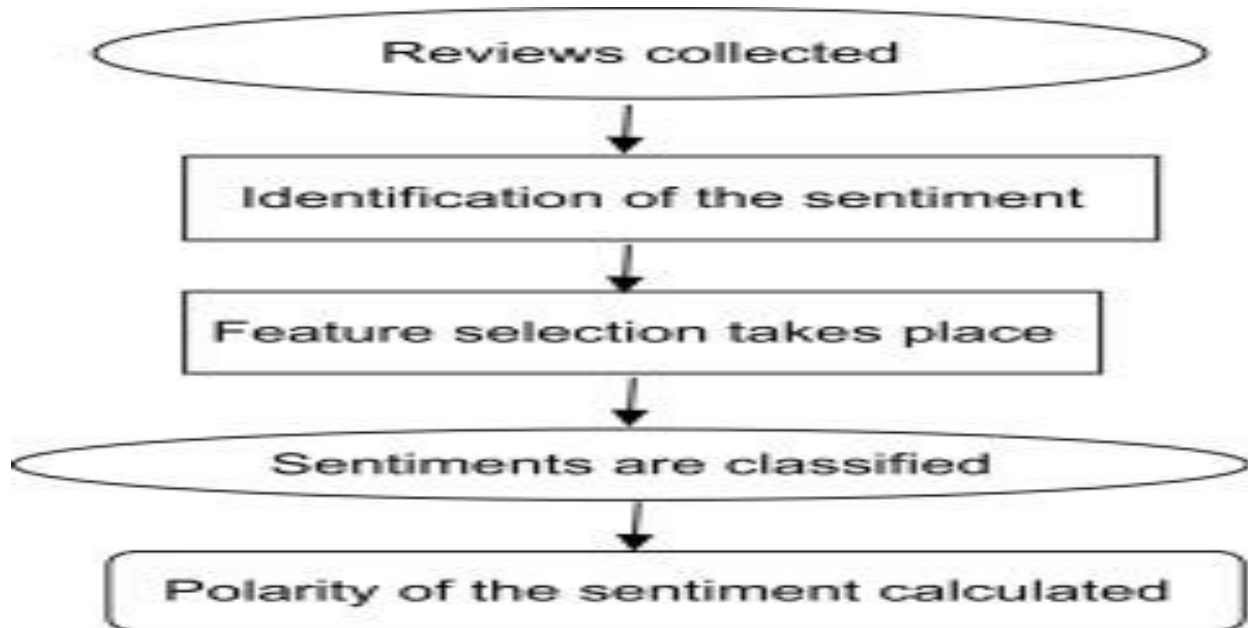
The DATA_FLOW DIAGRAM depicts the flow in which steps and actions take place in the total sentiment analysis process. It clearly shows the data flow in the step wise process for clear understanding of the system process design. When the polarity of the statement is obtained, it is classified as positive or negative or neutral according to its opinion.
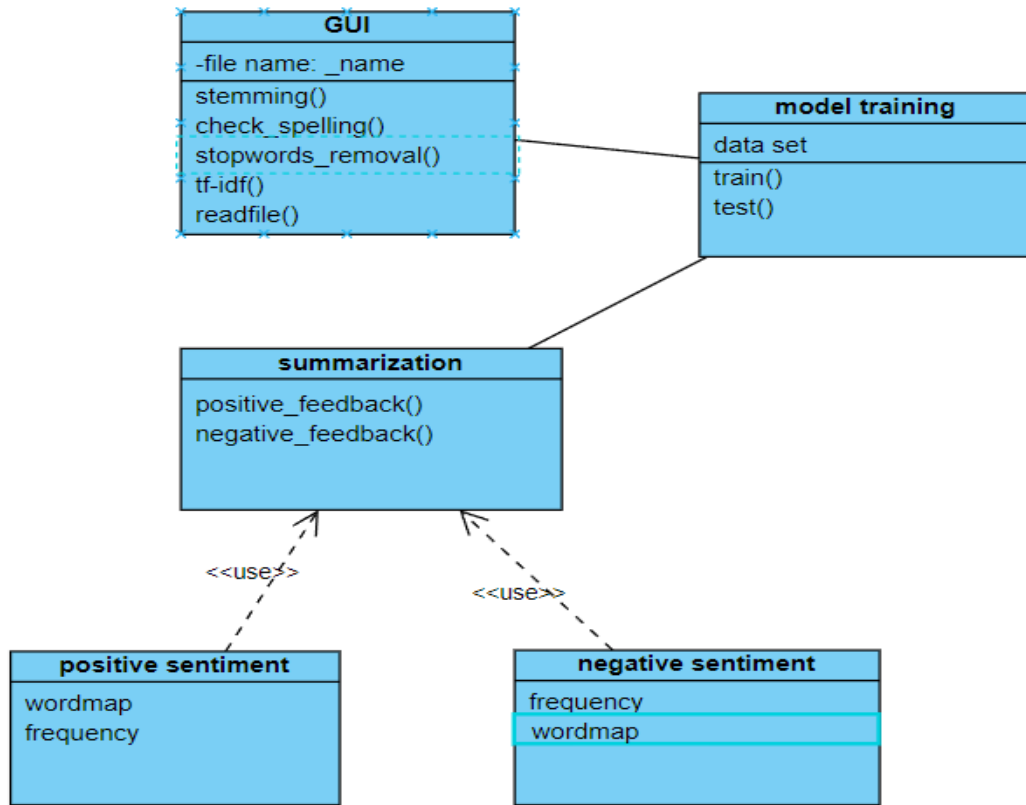
## IV.SYSTEM DESIGN

Use-case Diagram
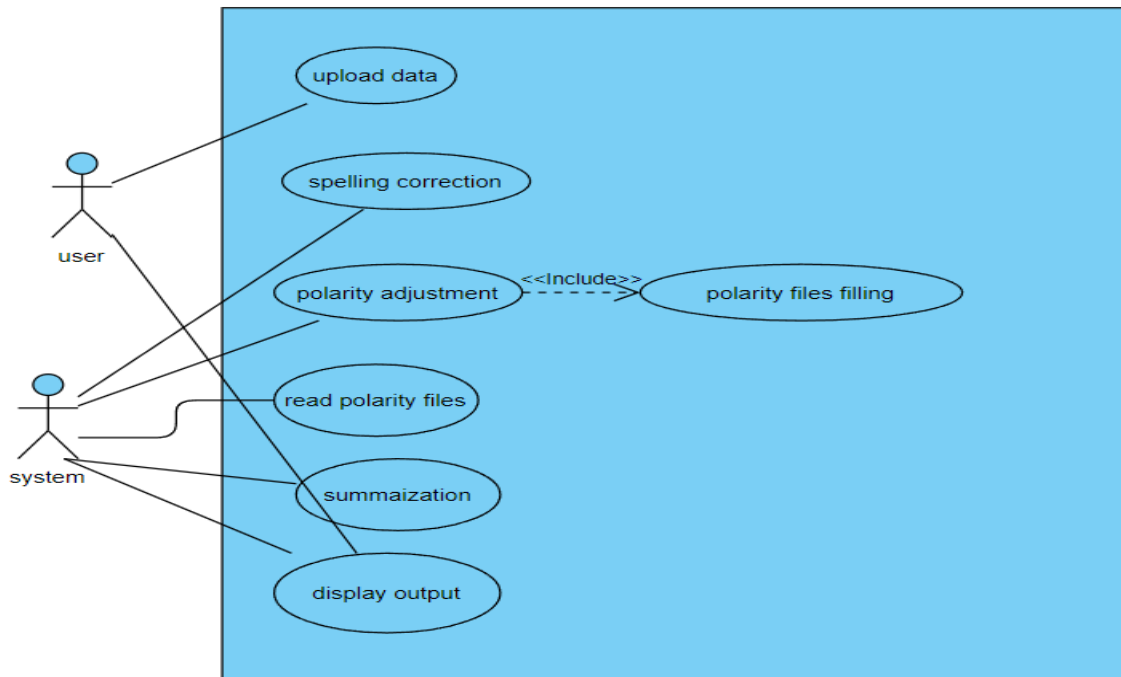
In the use case diagram the actors are users and system, in which user can upload or provide URL to web system through the command and then he can personally know the sentiment of the tweet or the article, and can also know the type of the tweet as a comment or a compliment or a supporting statement.

Class Diagram

A class diagram depicts the attributes and behaviors of a class, as well as the system's restrictions. Class diagrams are commonly used in the design of object-oriented systems because they are the only case diagram that can be immediately converted to object-oriented languages.

Activity Diagram
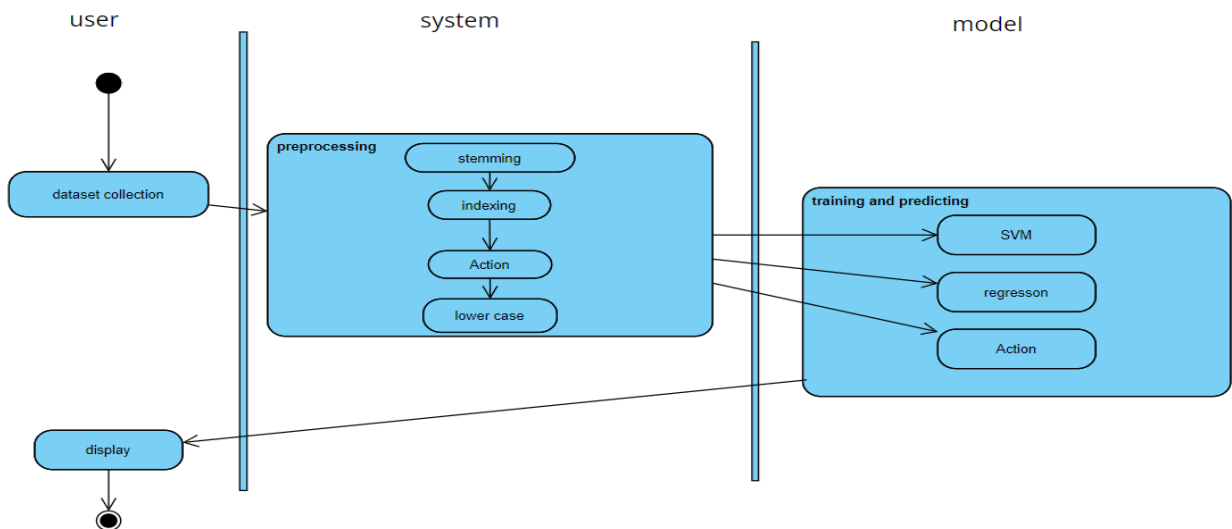
The above is the activity diagram of the user, and how he interacts with the system and its actions and activities according to the conditions. when the model is finally ready with the

Sequence Diagram
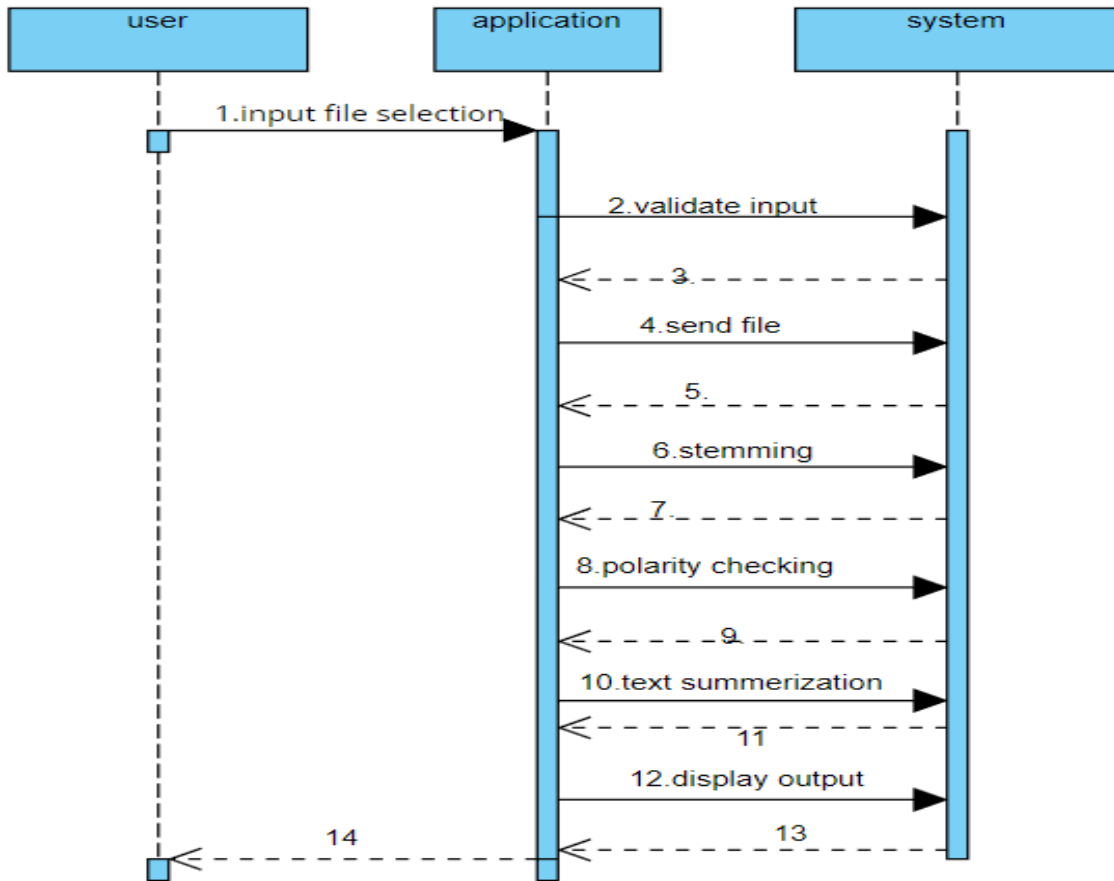
The sequence diagram shows the sequence of actions of the user, application, system. based upon this sequence of actions, a sequence diagram had drawn.

V. SYSTEM IMPLEMENTATION

	id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	
0	1340539111971516416	Rachel Roh	La Crescenta-Montrose, CA	Aggregator of Asian American news; scanning di...	2009-04-08 17:52:46	405	1692	3247	False	2020-12-20 06:06:44	Sa sai pas trea
1	1338158543359250433	Albert Fong	San Francisco, CA	Marketing dude, tech geek, heavy metal & '80s ...	2009-09-21 15:27:30	834	666	178	False	2020-12-13 16:27:13	V w bee wron
2	1337858199140118533	eliᴌᴛᴇᴜ 👌	Your Bed	heil, hydra ✋ ☺	2020-06-25 23:30:28	10	88	155	False	2020-12-12 20:33:45	#cor #S #Astra #Pfi
3	1337855739918835717	Charles Adler	Vancouver, BC - Canada	Hosting "CharlesAdlerTonight" Global News Radi...	2008-09-10 11:28:53	49165	3933	21853	True	2020-12-12 20:23:59	F im Senat when
4	1337854064604966912	Citizen News Channel	NaN	Citizen News Channel bringing you an alternati...	2020-04-23 17:58:42	152	580	1473	False	2020-12-12 20:17:19	Expla again

Figure 1 : twitter api vaccines review Database

```
In [23]: text_df = df.drop(['id', 'user_name', 'user_location', 'user_description', 'user_created',
                'user_followers', 'user_friends', 'user_favourites', 'user_verified',
                'date', 'hashtags', 'source', 'retweets', 'favorites',
                'is_retweet'], axis=1)
         text_df.head()
```

Out[23]:

	text
0	Same folks said daikon paste could treat a cyt...
1	While the world has been on the wrong side of ...
2	#coronavirus #SputnikV #AstraZeneca #PfizerBio...
3	Facts are immutable, Senator, even when you're...
4	Explain to me again why we need a vaccine @Bor...

```
In [24]: print(text_df['text'].iloc[0],"\n")
         print(text_df['text'].iloc[1],"\n")
         print(text_df['text'].iloc[2],"\n")
         print(text_df['text'].iloc[3],"\n")
         print(text_df['text'].iloc[4],"\n")
```

```
Same folks said daikon paste could treat a cytokine storm #PfizerBioNTech https://t.co/xeHhIMg1kF

While the world has been on the wrong side of history this year, hopefully, the biggest vaccination effort we've ev… https://t.
co/dlCHrZjkhm

#coronavirus #SputnikV #AstraZeneca #PfizerBioNTech #Moderna #Covid_19 Russian vaccine is created to last 2-4 years… https://t.
co/ieYlCKBr8P

Facts are immutable, Senator, even when you're not ethically sturdy enough to acknowledge them. (1) You were born i… https://t.
co/jqgV18kch4
```
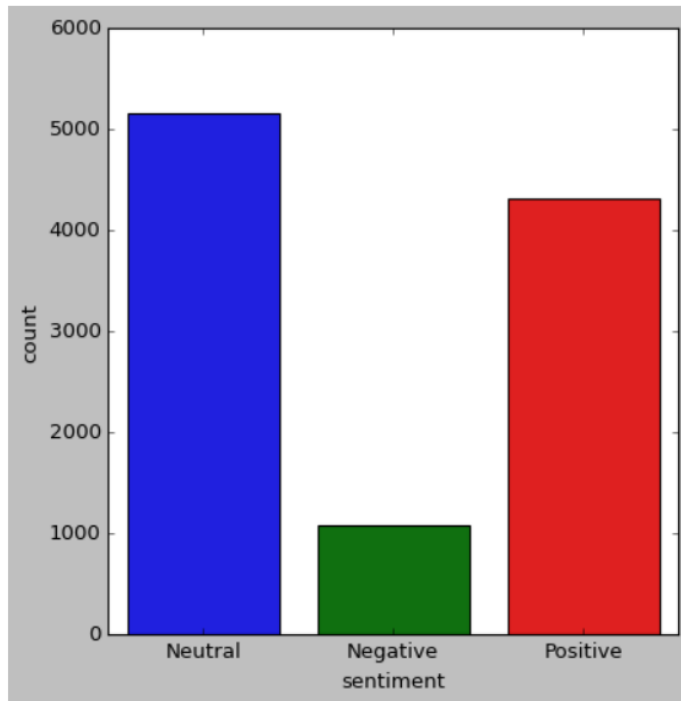
Figure 2 jupyter notebook Code Structure

VI.RESULT AND OUTPUTS

```
In [72]: fig = plt.figure(figsize=(6,6))
         sns.countplot(x='sentiment', data = text_df)
```

Out[72]: <AxesSubplot:xlabel='sentiment', ylabel='count'>

Graph of sentiments

```
In [55]: print("Best parameter:", grid.best_params_)

         Best parameter: {'C': 10}
```

```
In [60]: y_pred = grid.predict(x_test)
         logreg_acc = accuracy_score(y_pred, y_test)
         print("Test accuracy: {:.2f}%".format(logreg_acc*100))
         print(y_pred)

         Test accuracy: 87.58%
         ['Neutral' 'Positive' 'Negative' ... 'Negative' 'Positive' 'Positive']
```

```
In [61]: print(confusion_matrix(y_test, y_pred))
         print("\n")
         print(classification_report(y_test, y_pred))

         [[ 105   87   34]
          [   7 1005    9]
          [  14  111  737]]
```

	precision	recall	f1-score	support
Negative	0.83	0.46	0.60	226
Neutral	0.84	0.98	0.90	1021
Positive	0.94	0.85	0.90	862

Prediction and accuracy details

```
In [65]: t=ed.summary
         print(t)
```

Kalākaua, King of HawaiiThe 1881 world tour of King Kalākaua (pictured) made him the first monarch to circumnavigate the globe.
He visited American legislators, had an audience with Pope Leo XIII in Rome, and met with European and Asian heads of state.
In between negotiations, Kalākaua and his companions visited tourist sites and attended local Masonic lodge meetings.
As a result of his visit with Thomas Edison, Iolani Palace became the first building in Hawaii with electric lighting.
The King's amiable personality generated worldwide goodwill, and he succeeded in increasing Hawaii's labor force with Japanese
workers.

```
In [69]: xy=TextBlob(t)
         s=xy.sentiment.polarity
         if s==0:
             print("neutral")
         elif s>0:
             print("positive")
         else:
             print("negative")
```

positive

Article sentiment prediction
Just by providing the URL of an article we can automatically get its sentiment as the model gathers the data from that URL on its own and then analyses the article and provides the opinion for it as positive neutral or negative

VII.CONCLUSION

Asking people for feedback about products or services and then analysing them to improve the service or the product will take large amount of time and manpower for the organizations, But by using sentiment analysis or opinion mining of the tweets or discussions of that

product, we can reduce the data collection and improve the performance of that respective product or service within no time

REFERENCE

[1] https://scikit-learn.org/stable/
[2] https://www.nltk.org/howto/wordnet.html
[3] https://monkeylearn.com/blog/sentiment-analysis-machine-learning/
[4] https://www.kaggle.com/datasets/gpreda/pfizer-vaccine-tweets
[5] https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
[6] https://matplotlib.org/stable/plot_types/stats/errorbar_plot.html
[7] https://www.ibm.com/cloud/learn/natural-language-processing