TWITTER SENTIMENTAL ANALYSIS USING WEBAPP

Project Report

Submitted in fulfillment of the requirement for
the degree of Bachelor of Technology
In
Computer Science and Engineering and Information Technology

By

Kanishk Kala (161466)

Under the supervision of
Dr. Himanshu Jindal

To

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, HimachalPradesh

Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled "Twitter Sentimental Analysis Usimg WebApp" in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/Information Technology submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from July 2019 to May 2020 under the supervision of Dr. Himanshu Jindal (Assistant Professor, Grade-II, Computer Science &Engineering Department).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature)

Kanishk Kala (161466)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.
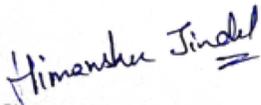
(Supervisor Signature)

Supervisor Name: Dr. Himanshu Jindal

Designation: Assistant Professor (Grade-II)

Department name: Computer Science and Engineering and Information Technology

Dated:

II

ACKNOWLEDGEMENT

TABLE OF CONTENT

Chapter	Page No.

TABLE OF CONTENT

LIST OF FIGURES

ABSTRACT

The primary objective that this project embarks upon is to provide a user that does not have a sound technical knowledge of implementing sentiment analysis algorithms and analyzing data so as to facilitate his knowledge on the subject of his choice. We aim to provide the user with a simple webapp wherein he could just type the keyword he wants to get an opinion about and the app runs complex sentiment analysis algorithms in the backend and eventually present the user with a detailed analysis in the form of charts and figures.

Every second, on average, around 6,000 tweets are tweeted on Twitter (visualize them here), which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year. Since twitter has become the hub of people expressing their ideas, reviews and opinions about every product, services or events, it is the primary source of our dataset.

CHAPTER 1: INTRODUCTION

1.1) Introduction :

Sentimental Analysis is the technique of analyzing the polarity of the words in a sentence by using complex natural language processing algorithms hence classifying into different categories according to their sentiment score. These categories are positive, negative and neutral.

The dataset that serves as input consists of sentences collected by different platforms issued by users al around the globe. Twitter is one such platform that could serve as a provider of potential database to carry out such a task.

1.1.1) Social Media

Ever since the dawn of internet, people throughout the world have been more connected that ever. Now they already had a way to communicate with each other faster than any other way possible. All they eve needed was a platform that could provide such features. Hence came the age of social media.

Social media is group of application that provides a platform to people all over the world to express their ideas, opinions, information, reviews, photos and basically anything they want to over the internet. It first started with emails and then gradually moved on the messengers, chat rooms and now it has evolved to a level where one could go live and chat with the world or make a private video call, post their everyday chores or upload images of every task that they undergo. It is a part of their virtual life now. The trend shows that the use of social media over internet has increased by 37 percent to121 billion minutes in 2012, compared to 88 billion minutes in 2011[20].

This is mainly due to the fact that now since everyone has a smart device in their pocket inter mobility is no longer an issue. If a person has access to a network he has access to his virtual life.

But the use of social media is now no longer limited to just chatting with friends or just having a leisure time. It is the biggest potential business hub where certain sectors of market depend heavily upon for it's growth and livelihood.

Since the whole world is on social media now, it serves as the best way to advertise your

product. The customer reach and potential to increase your sales ratio is infinite. Big companies find their target clients and communicate with them on first person basis hence increasing their client base and through a medium of referrals and attractive offers lure their customers in.

It is often seen that these social media sites also serves as a platform to raise propaganda messages by various organizations around the world. Sometime it is in favour of the user sometimes it isn't. But regardless of the fact no one can deny that the reach it gives to that propaganda voice is like no other. But it's not always this mad, many times it serves as the only way to spread essential information by targeting a huge audience that traditional ways of doing the same could never do.

Online notoriety is one of the most valuable resources for brands. An awful survey via web-based networking media can be expensive to an organization if it's not taken care of appropriately and quickly. All things considered, 80% of purchasers get guidance on items from web based life.

1.1.2) Twitter

Twitter is the most famous micro blogging platforms that allows user to express their opinions, feelings, disapprovals, information and data in the form of short messages known as tweets. These tweets however has a restriction of upto prevalent ongoing micro blogging characters only. Although twitter too is a part of this social media cluster we are talking about but it has some unique features that makes it way more significant that any of the sites or app in its competition.

Twitter is the hub of intellectuals and subject matter experts who are always tweeting their views about what's going around in the world. These subject may have any alignment in nature. They subjects may range from being as important as politics or a major environmental issue to a minor inconvenience in a tech newly launched. But the main thing is that since it is being given by verified experts in their respective fields, it is more likely to be accurate. So the followers of these accounts when find the tweet appealing to them, they retweet it along with their own opinion attached to it hence creating a chain of information sharing. This leaves twitter with a user base of over 500 million with their exponential message hence creating a database large enough that it could be used for any modern day analysis that requires realtime human reflections.

However, the miniaturized scale blogging stage has been built up a couple years" time for advancing remote exchange site by utilizing an outside smaller scale blogging stage as Twitter promoting. The moment of sharing, intuitive, network situated highlights are opening internet business, propelled another splendid spot which it tends to be demonstrated that small scale blogging stage has empowered organizations do mark picture, item significant deals channel, improve item deals, converse with the shopper for a decent communication and different business exercises included. stated, truth be told, the company"s assembling such items have begun to survey postulations small scale websites to get a feeling of general assessment for an item. Ordinarily these organizations study client responses and answer to clients on miniaturized scale online journals.

Moreover, it is expressed that AI strategies can create a fixed number of the most consistently happening well known words which allocated a whole number an incentive for the benefit of the recurrence of the word in Twitter[20].

Supposition Mining suggests the extensive locale of normal language taking care of, content mining, computational semantics, which incorporates the computational examination of evaluations, sentiments and emotions conveyed in content despite the way that, view or outlook subject to feeling as opposed to reason is as often as possible casually insinuated as an inclination .

Subsequently, crediting to an equivalent for feeling mining or estimation examination communicated that conclusion investigation territories including law, legislative issues, advancement and advancing. In earlier days various web based systems administration have given web customers street for opening up to communicate and confer their thoughts and experiences.

1.1.3)   Why is Twitter Sentiment Analysis important?

Almost 80% world's advanced information is unstructured, & information got off the internet based life sources are no exemption. Luckily, on account of the improvements in Machine Learning & NLP, it is presently conceivable toward making models that gain off the models & are utilized to process & sort out content information.

Twitter sentiment analysis frameworks permit sorting huge arrangements of tweets & recognize extremity off every announcement naturally. Furthermore, the best part, it's quick and basic, sparing groups significant hrs & permitting them to concentrate on errands where they can have a greater effect.

Listed below are some of the major merits of twitter sentiment analysis:

- **Scalability**: suppose you have to dissect many tweets referencing a brand. While you could do that physically, it would take a long stretch of time of manual preparing and would wind up being conflicting and difficult to scale. By using this algorithm, you can robotize this errand and get savvy brings about an extremely brief timeframe.

- **Real-Time Analysis:** This algorithm is basic to see abrupt moves in client states of mind, identify if pundits and grumblings are expanding and make a move before the issue raises. You can be checking your image continuously and get important bits of knowledge that permit you to make changes or upgrades when required.

- **Consistent Criteria**: breaking down sentiment in a book is an abstract undertaking. At the point when done physically, a similar tweet might be seen diversely by two individuals from a similar group, and the outcomes will most likely be one-sided.

1.1.4) Common examples of tweets:

- This tweet, for instance, shows that quick transportation is one of the most esteemed aspect for Amazon clients:
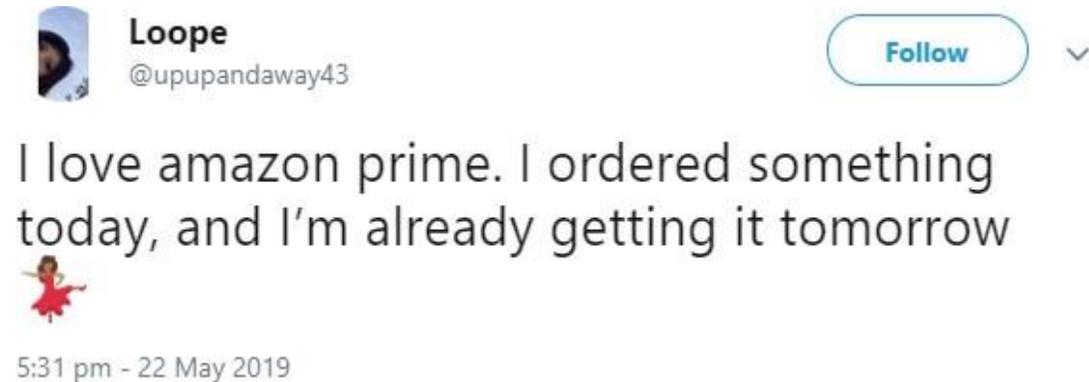

Fig 1.1 : Example1

- The tweets bellows shows a dissatisfied customer reacting to uber:

Fig 1.2 : Example 2

- Below is example of a tweets that advertises the importance of machine learning:
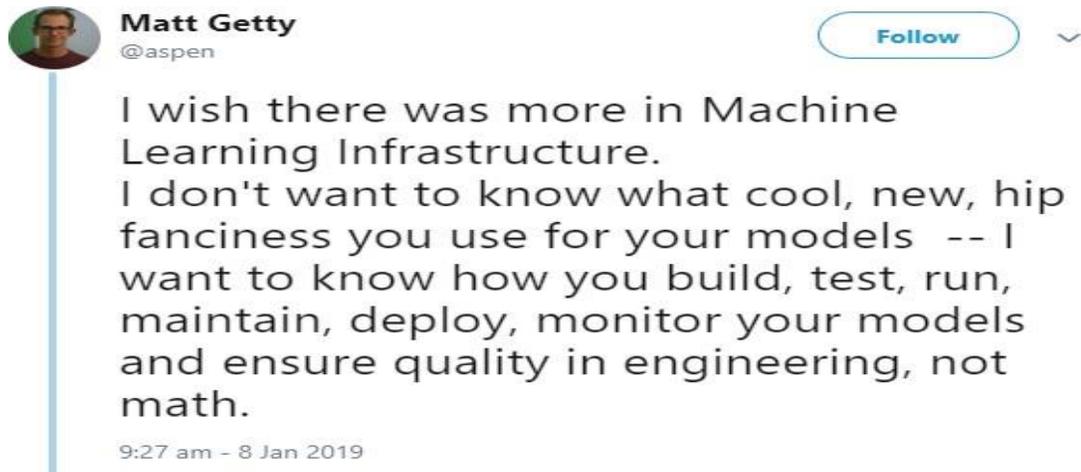
Fig 1.3 : Example 3

- This figure below displays time variant tweet count and it's analyzed sentiment for Donald Trump:
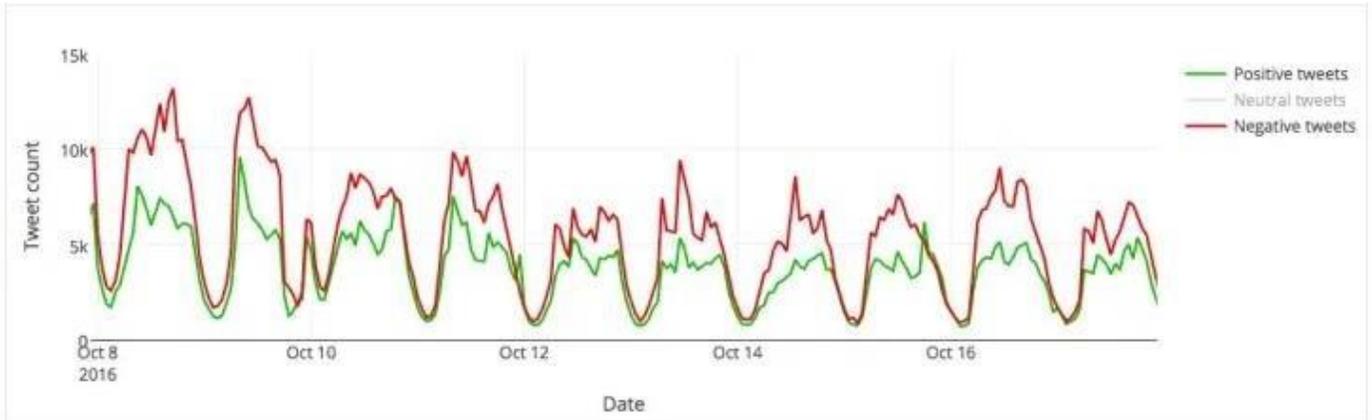
Trump tweet count by sentiment.

Fig 1.4 : Example 4

- In contrast, figure below displays time variant tweet count and it's analyzed sentiment for Hilary Clinton:
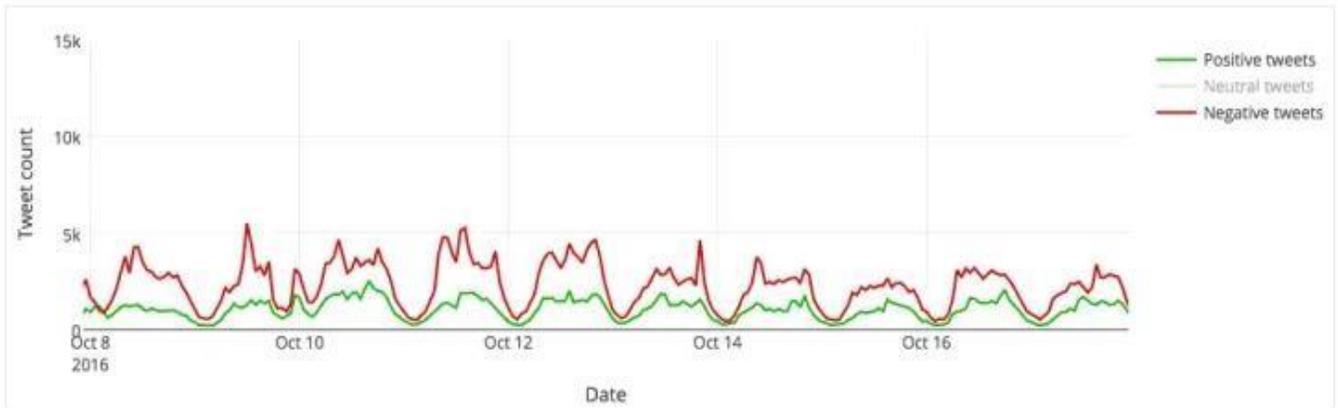
Clinton tweet count by sentiment.

Fig 1.5 : Example 5

1.1.5) Limitations

Sentiment analysis instruments can distinguish and break down numerous bits of content consequently and rapidly.

In any case, PC programs have issues perceiving things like mockery and incongruity, refutations, jokes, and embellishments - the sorts of things an individual would experience little difficulty recognizing. Furthermore, neglecting to perceive these can slant the outcomes.

'Frustrated' might be named a negative word for the reasons for sentiment analysis, yet inside the expression "I wasn't disillusioned", it ought to be delegated positive.

We would think that its simple to perceive as mockery the announcement "I'm truly cherishing the tremendous pool at my inn!", if this announcement is joined by a photograph of a little pool; while a computerized sentiment analysis device presumably would not, and would no doubt arrange it for instance of positive sentiment.

With short sentences and bits of content, for instance like those you find on Twitter particularly, and now and then on Facebook, there probably won't be sufficient setting for a dependable sentiment analysis. Notwithstanding, when all is said in done, Twitter has gained notoriety for being a decent wellspring of data for sentiment analysis, and with the new expanded word mean tweets it's possible it will turn out to be much increasingly helpful.

Along these lines, computerized sentiment analysis instruments do an extremely incredible activity of examining content for conclusion and mentality, yet they're not great.

At the point when you're utilizing a device like Typely to investigate your content to check whether it passes on the sentiment you need for your perusers/crowd, join the outcomes it gives you with your human judgment to distinguish anything the instrument will be unable to effectively decide.

Typely features states in your content by positive and negative sentiment, making it excessively simple for you to see where your report is either communicating precisely the sentiments you need it to, or where you may need to roll out certain improvements.

1.2) Problem Statement :

Large chunks of data are available over these social media platforms. This data could be put to good use by analyzing it and predicting results and deriving further opinions. It could be pretty handy in following trades:

1.2.1) Science and Technology: How target audience is reacting to the technology could be predicted even before the tech is out.

1.2.2) Stock Market: The prices of shares are always fluctuating and its almost difficult to predict the success of an investment on a particular stock.

1.2.3) Politics: There is always a lot of hassle on twitter during election season. By analyzing those tweets one can get a clear idea of popularity of a party/ candidate and predict the results and hence decide a more efficient campaign strategy.

1.3) Objectives:

The primary objective that this project embarks upon is to provide a user that does not have a sound technical knowledge of implementing sentiment analysis algorithms and analyzing data so as to facilitate his knowledge on the subject of his choice.

1.4) Methodology :

(1) Twitter data is mined through twitter API for PHP . Data is extracted in JSON format.

(2) The extracted data is analysed through another php script "Twitter Sentiment Analysis" which in turn uses Datumbox api which is a machine learning framework and offers classification services accessible via RESTFUL api. A polarity score is generated for each tweet in order to classify it.

(3) The polarity score is used to display the analysis in the form of charts. This is achieved using Canvas.js library framework for JavaScript.

1.5) **Organization :**

The chapters of the work are divided into five chapters:

Chapter 1:Introduction -

The introduction highlights the current scenario of social media in contemporary world setting. It lays down the importance of social media platform in our life and how they serve as an integral part for providing dataset for our project.

Chapter 2 : Literature Review -

Literature review walks through the various literary work done by subject matter experts that utilized and inspired some or other part of this project. These papers are taken from a collection of scholarly work found over ACM, IEEE and internet libararies.

Chapter 3: System Development -

This chapter lays down how step by step how the system is developed for successful working of the project. It also gives a list of hardware and software requirements.

Chapter 4 : Results and Performance Analysis -

This chapter contains the test plan and realtime working of our apps in different platforms. It also shows the difference in result obtained by both the platforms

Chapter 5 : Conclusion -

This chapter contains the summary of the project as well as the conclusion drawn by it's working.

CHAPTER 2 : LITERATURE SURVEY

Following papers were taken into consideration:

i) Opinion mining and sentiment analysis[1]:

This paper gives a view of techniques that help us to understand opinion related machine learning. It aims on overcoming the limitations presented by conventional approaches of gathering information and predicting results by providing us with knowledge of computation of opinion, sentiment and polarity in texts.

ii) Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis [2]

This paper is aimed on imparting knowledge of new techniques of analyzing phrase level sentiments. First of all it is determined whether or not a phrase is polar or neutral. If found polar, polarity of the sentiment is the analyzed by evaluating sentiment score. The system is given the freedom to automatically determine the contextual polarity.

iii) Enhancing deep learning sentiment analysis with ensemble techniques in social applications [3]:

This paper proposes a the algorithms that utilizes modern day deep learning approaches to carry out sentiment analysis. It prescribes a six step process of adding deep learning algorithm to conventional sentiment analysis algorithms to make them more effective and thereby reducing human effort. It automates the process and also makes sure that the system is not neutral biased.

iv) Facebook Impact and Sentiment Analysis on Political Campaigns:[4]

Given below is the formula used to determine the sentiment index of the political data obtained via various social networking platforms.

$$Sentiment\ Index = ln\left[\frac{1 + TOTAL^{POSITIVE}}{1 + TOTAL^{NEGATIVE}}\right]$$

The investigation showed that the feelings of voters can be found yet not their goal to cast a ballot. It was discovered that the ideological group that won had an awful observation via web-based networking media while the one having great couldn't win the Mexican political race.

v) SentiML: functional annotation for multilingual sentiment analysis:[5]

This method utilized an approach in which there are different levels of abstraction involved.

i)   Target

ii) Modifier

iii) Appraisal group.

Target refers to articulation that the sentiment alludes to. Modifier refers to articulation passing on the sentiment & examination bunch incorporates the arrangement of targets & modifiers. The importance of a sentence gets turned around when some words were available which are called as modifiers. For grouping the distinguished characteristic is arranges into it"s classifications.

vi) Evaluation of Features on Sentimental Analysis:[7]

This includes 4 stage technique. At the initial point of process the dataset is preprocessed, the first thing that takes place is the words are changed to lowercase for ease. At that point the words finishing with punctuation marks are changed over back to unique structure like don"t - > don't, any non ASCII character is expelled.

This is trailed by expulsion of stop words (eg. a,an the) as the don't pass on any component so evacuation of stop words is liked. Second arrangement includes parceling information into preparing and test information. For preparing information in the wake of performing stemming include determination is to be acted in which different measurable strategies are applied to check whether notion of audit can be extricated from the include of words in each example.

vii) Review of Sentimental Analysis Methods using Lexicon Based Approach :[8]

This methodology remembers computation for tendency of suppositions for words by analyzing their arrangementin the archive . Dictionary methodology word references are made naturally and physically , with the assistance of basic words we can grow the rundown of words. The examination is centered mostly around utilizing modifiers as indicators of the semantic arrangement of content. It begins with accumulation of modifiers and their (SO) Sentiment Orientation into a word reference and from that point designs are coordinated for wistful estimation.

viii) Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach: [9]

The process begins with the scrapping of tweets which are made to pre-process and then sending them to a classifier calculation. It includes five stage system. Initially an organized information is framed, Then normal sentence structure words like action words ,relational word are evacuated. This is trailed with progression in which letters finishing with "ing","ize","ed" are diminished to their original word. This is steaming. The term recurrence – converse term recurrence score is determined for each term and a 2D lattice is made in which lines speak to reports and sections speak to the word removed from archive after the above preprocessing.

ix) Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm:[10]

Naïve Bayes technique for arranging depends on probability of grouping in which Bayes Theorem is utilized & hypothesis all out likelihood is utilized. The research urges it"s appropriateness for enormous datasets. It begins with tokenization then expulsion of stop words & after that content change. At that point highlights are chosen for pieces of record that are for + & - letters. These parts are participated in a manner so the likelihood is limit of subsequent sentence existing in both positive, negative terms.

x) Sentiment Analysis using Neuro-Fuzzy and Hidden Markov Models of Text:[11]

The information are achieved via various records informational index marked in a single source document known as "corpus". When joined content is changed over into cluster of letters and further advance was to arrange this variety assortment. Presently a last piece of estimation of enrollment level of every letter is determined using a diagnostic recipe before being prepared by neural system.

xi) Rating Prediction Based on Social Sentiment From Textual Reviews:[12]

This research indicates two kinds of lists:

i)   Word List: Consists of words belonging to all polarities..

ii) Topic list: Consists of topics that will be used as the primary sentiment.

Model includes accompanying advances: Initially pre-handling errands, for example, stop word expulsion, commotion word evacuation are performed and afterward remaining item includes are removed utilizing Latent Dirichlet Allocation.

xii) Sentiment analysis using support vector machines with diverse information sources [13]

This research provides us with an approach to use sentiment analysis by using support vector machines. It helps the system to work with advance lemmatized version of text. Experiments done on various datasets show high accuracy rate as compared to traditional sentiment analysis algorithms.
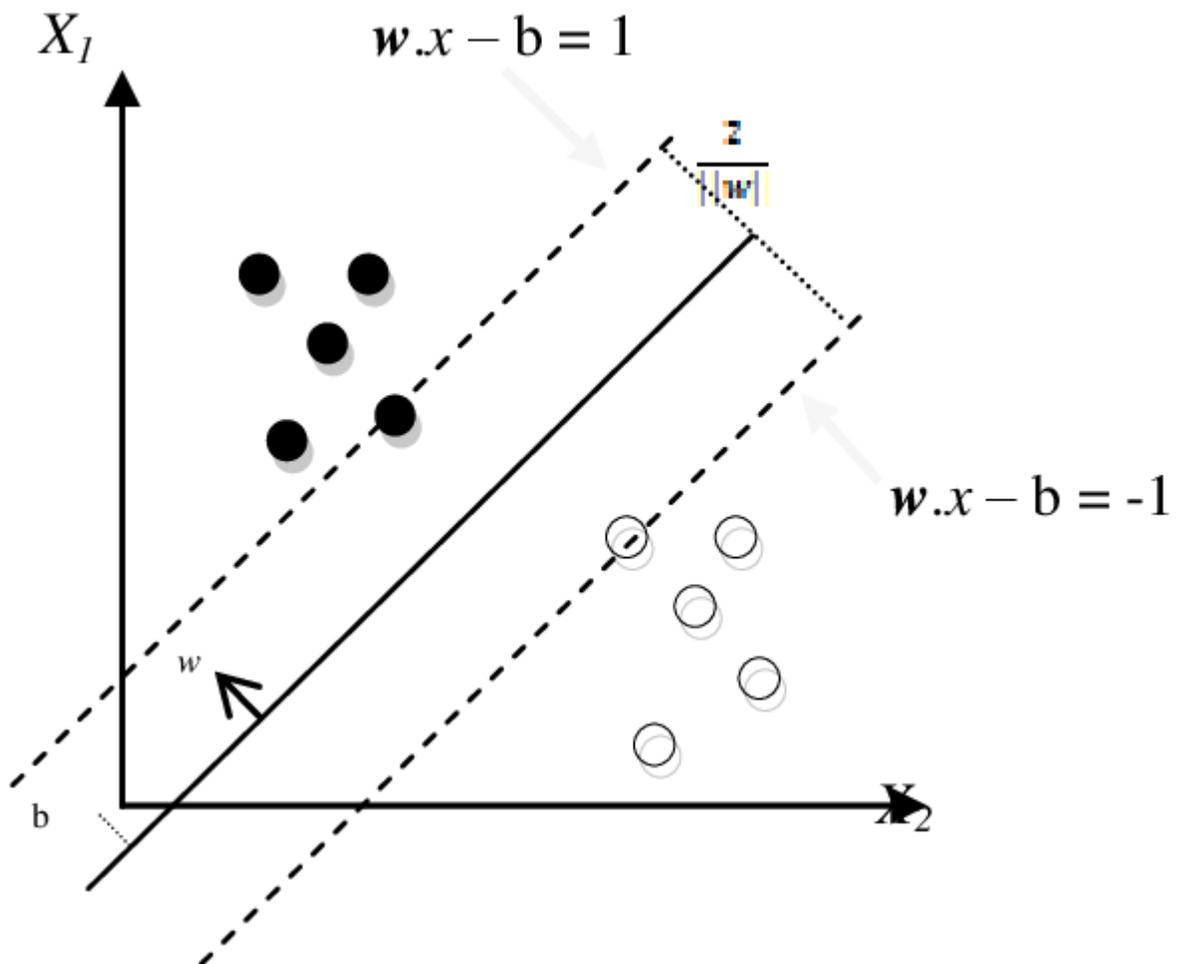
Fig 2.1: Sentiment analysis using SVM

xiii) Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches:

[18]

System fuses blend of information from a few electronic business regions which are from the outset preprocessed in order delete letters such as „or", „that" & postfixes like „ing" „ ness" lessened. It is sent into classifier which channels a peculiarity such as locale, thing, highlight & modifier. The highlights are referenced before gave dataset whose uttermost point are accounted for and right outcome the outcome are displayed or the client gets again the data.

xiv) Sentimental feature selection for sentiment analysis of Chinese online reviews:[19]

System joins isolating substance in divides, from there progression of wistful types are selected. The decision relies on the similarity with planning test set. A heap for component are resolved & every game plan of that type is connected in accordance to the weight to create social occasion consisting content close in accordance of weight. Henceforth study is sorted out subsequently & limit is analyzed by remembering the level of the size.

CHAPTER 3: SYSTEM DEVELOPMENT

3.1) System Requirement:

Hardware:

System type: 64 bit

ProcessorSpecs: Intel Core i3 (or above)

Software:

OperatingSystem: Windows 10

Python

Google Chrome

XAAMP

MySQL

PHP 7.4.5

3.2) Model Development :
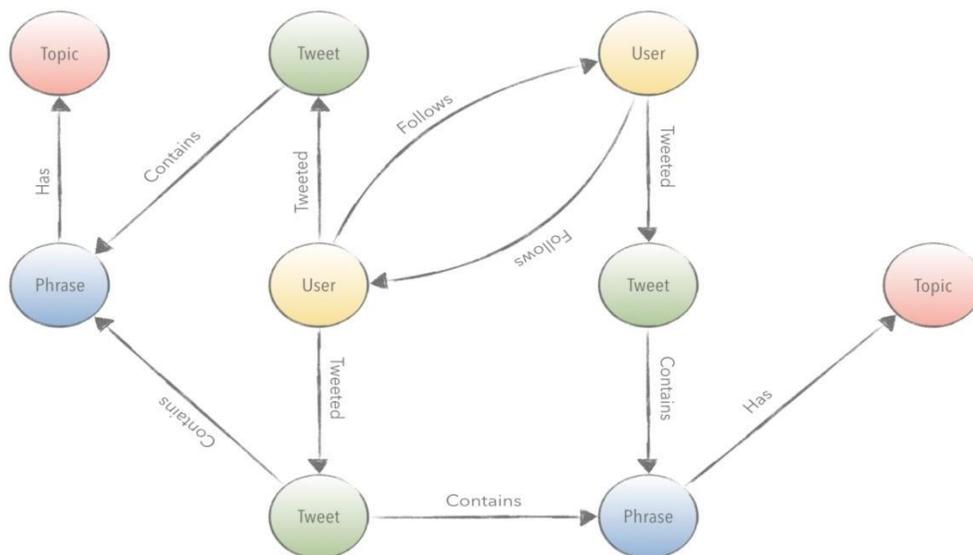
Fig(3.1): System development
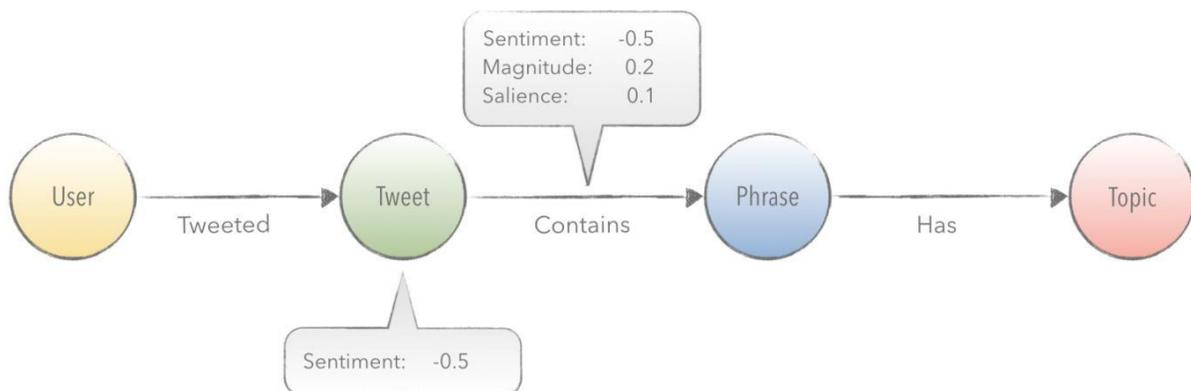
Fig(3.2): System development

3.2.1) Dataset: The tweets extracted using twitter api for php serves as input for the app.

3.2.2) Processing: API is backed up by the Datumbox Machine Learning Framework. It gives a variety of different functionalities. Functions used in the project utilizes powerful machine learning algorithms accessible via REST Api. TSA function analyzes tweets in accordance to their sentiment score and assigns them a polarity.

3.2.3) Fetch & Store: The tweets are mined using twitter API for php. We need a twitter developer account on authentication credentials to carry out this process.

3.2.4) Data Visualization: The processed data is analysed and the result is displayed in the form of chart and figures. We use Canvas.js javascript based library for visualization of data on our webapp.

3.2.5) Working of DatumBox Twitter Sentiment Analysis Classifier:

So as to recognize the Sentiment of the tweets we utilized our Machine Learning system to assemble a classifier fit for identifying Positive, Negative and Neutral tweets. Our preparation set comprised of 1.2 million tweets uniformly conveyed over the 3 classifications. We tokenized the tweets by extricating their bigrams and by considering the URLs, the hash labels, the usernames and the emojis.

So as to choose the best highlights we utilized a few distinct algorithms and toward the end we picked the Mutual Information. At long last in the wake of playing out a few tests with different models and arrangements we chose the Binarized Naïve Bayes as the best performing classifier for the specific issue (for some odd reason Naïve Bayes beat SVM, Max Entropy and different classifiers which are known to perform typically superior to NB). To assess the outcomes we utilized the 10-crease cross-approval strategy and our best performing classifier accomplishes a precision of 83.26%.

3.3) Libraries Used:

1) Twitter for PHP

2) DatumBox Machine Learning library

3) Canvas.js

4) Matplotlib

3.3.1)Twitter for PHP: Twitter for PHP provides PHP script for carrying out all the technicalities and formalities we need to get data from twitter.

● **Classes :**

i) API   :Api class utilized methods that define parameters used to call twitter app.
ii) OAuthHandler : This class authenticates the twitter developer accounts by verifying credentials such as api keys and access tokens.
iii) cursor : This class is used to retrieve tweets corresponding to the data keywords in the program.

● **Methods utilized:**

i) set_access_token: This method contains the credential keys to the twitter developer acoount.
ii) Cursor : This method contains the cursor class used to retrieve data from twitter.
iii) items() : This method organizes data in accordance to result obtained per tweets.

3.3.2.) DatumBox: The Datumbox Machine Learning Framework is an open-source framework written in Java which allows the rapid development of Machine Learning and Statistical applications. It provides extensive machine learning support to many apps. We are utilizing it's twitter sentiment analysis framework in our project.

3.3.3.) Canvas.js: CanvasJS is an easy to use HTML5 and Javascript Charting library. It supports all sorts of smart devices and is supported by all the browsers in market. The most attractive feature of Canvas.js is the fact that it could easily be attached with any code snippet just by typing in a few lines of JavaScript code.

3.3.4.) Matplotlib :

Matplotlib is a python library used to diplay visualizations such as charts and figures in the program.

- **Classes utilized:**
i)   pyplot: This class is used to plot the data in charts.

- **Methods Utilized:**

i)   pie   : This method produces pie chart in accordance to data provided with legends.

ii) legend() : This method is used to control the demographics of chart in accordance to data .

iii) title()   :Provides title to the chart.

iv) axis()   : This method is used to define axis for the chart..

CHAPTER 4

RESULTS & PERFORMANCE

4.1) Testing of the program:

We shall conduct the analysis of a particular keyword using both python based app and our modified webapp.

The aim is to draw comparison between the performance as well as output structure of the analysis obtain.

We will try to justify later in conclusion why our modified webapp is a better app that the traditionally used python based app.

4.1.1) Test Case 1:

For this test, we have analyzed the polarity and sentiment of **"iphone"**.

❖ **Python based sentiment analysis:**

● **Dataset:**

● Tweets from various users are retrieved and stored in csv format.

	text	has_hasht	retweet_c	followers	following	user	screen_name
0	RT @rcastradiouk: #win the #ALLNEW #iPhoneSE2 #Apple #iPhone SIMFREE. To ε	1	576	0	46	Hannah ðᵞ	Hannah_Austin_9
1	For Nintendo lovers. 2 wallpapers for you ðŸ»	1	0	5006	199	Frenchitoɪ	frenchitouch
2	#Apple better release their new #iPhone soon or Iâ€™m switching to #Samsung	1	0	149	221	Ya Doctoo	WTSethi
3	"The quality of our home videos has dropped so much with lots of crappy actors	0	0	18118	3140	MARK 'Dsɪ	MarkOtabor
4	RT @AppleTerminal: News:	1	1	12	362	Danny Raf	munozdanny1
5	"Nigeria does not have Coronavirus. It's just a way of stealing money" P... https:	0	0	2621	2586	Sleek Gist	Sleek_Gist
6	Now on eBay Apple iPhone 7 - 32GB 128GB - Unlocked Smartphone - GSM Only A	1	0	657	970	NYC Globɑ	NYCGlobalBrands
7	https://t.co/LsXVQgjxnK	1	0	61	1085	Been Dop	1030_in_CF
8	News:	1	1	7800	5	Apple Ter	AppleTerminal
9	Flash sale ðŸš¨ ðŸš¨	0	0	1575	1510	YÅ«suf ÙŠ	_Tiga_b
10	#iPhone #CryptoNews #SaudiArabia	1	0	206	660	GrazingClɪ	GrazingClouds44

Fig 4.1 : Dataset scrapped from twitter

- Structure of data after preprocessing, cleaning, removing stop words and tokenization.

0	['#win', '#ALLNEW', '#iPhoneSE2', '#Apple', '#iPhone', 'SIMFREE.', 'enter', '#retweet', '&', '#followus.', '#Competition', '#giveaway.', 'Entrants', 'mustâ€¦']
1	['Nintendo', 'lovers.', 'wallpapers', '#nintendo', '#ios', '#wallpaper', '#iphone', 'https://t.co/d3BlXJI9PG']
2	['#Apple', 'better', 'release', 'their', '#iPhone', 'soon', 'switching', '#Samsung', 'S20+.']
3	['"The', 'quality', 'home', 'videos', 'dropped', 'much', 'with', 'lots', 'crappy', 'actors', 'directors"', 'actor', 'Edochieâ€¦', 'https://t.co/CZMI8iT9tu']
4	['News:', 'sensor', 'that', 'connects', 'your', 'iPhone', 'that', 'will', 'cost', 'claims', 'detect', 'someone', '#COVID19', 'themâ€¦']
5	['"Nigeria', 'does', 'have', 'Coronavirus.', '"It's', 'just', 'stealing', 'money"', 'P...', 'https://t.co/kWecljZPEW', 'https://t.co/E0zTbLLjEa']
6	['eBay', 'Apple', 'iPhone', '32GB', '128GB', 'Unlocked', 'Smartphone', 'Only', 'A1778', 'https://t.co/ayo4dor6js', '#iphone', '#android', '#ebay']
7	['https://t.co/LsXVQgjxnK', 'Protect', 'Vibes', 'Protect', 'Phone!', 'Only', '#TikTok', '#Apple', '#iPhone', '#phonecaseâ€¦', 'https://t.co/4rCU8EDXlQ']
8	['News:', 'sensor', 'that', 'connects', 'your', 'iPhone', 'that', 'will', 'cost', 'claims', 'detect', 'someone', '#COVID19', 'https://t.co/gz32FRPbmg']
9	['Flash', 'sale', 'Registered', '2016', 'Mercedes', 'Benz', 'C300', 'Price:', 'million', '(Negotiable)', 'Retweet', 'customers', 'https://t.co/63HM9LSkVc']
10	['#iPhone', '#CryptoNews', '#SaudiArabia', '#technology', '#lawenforcement', 'F.B.I.', 'Said', 'Find', 'Links', 'Between', 'Pensacola', 'Gunmanâ€¦', 'https://t.co/Y1iKFUxkei']

Fig 4.2 : Preprocessed dataset

- **User Interface:**
  - Since there is no ui as such, following is a screenshot of rather complex code structure that the user has to run in python ide (Sypder in this case) and then perform analysis

Fig 4.3 : User Interface

● **Result :**

After the extraction and preprocessing of data, we use TextBlob library to perform sentiment analysis of the tweets and classify them according to their respective polarity. A detailed analysis figures are displayed in console along with a pie chart of same created using MatplotLib library.

Following are the screenshot of result obtained in ide console.

```
Enter Keyword/Tag to search about: iphone
```

Fig 4.4 : Result

```
General Report:
Weakly Positive

Detailed Report:
5.00% people thought it was positive
22.00% people thought it was weakly positive
0.00% people thought it was strongly positive
6.00% people thought it was negative
11.00% people thought it was weakly negative
0.00% people thought it was strongly negative
56.00% people thought it was neutral
```

Fig 4.5 : Result

How people are reacting on iphone by analyzing 100 Tweets.

Positive [5.00%]
Weakly Positive [22.00%]
Strongly Positive [0.00%]
Neutral [56.00%]
Negative [6.00%]
Weakly Negative [11.00%]
Strongly Negative [0.00%]
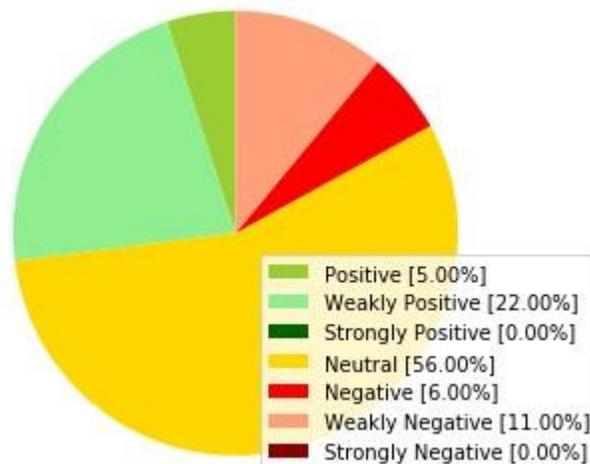
Fig 4.6 : Chart analysis

❖ **WebApp based sentiment analysis:**

● **User Interface:**

WebApp consists of a very user-friendly user interface developed using html, css, js and bootstrap. All you have to do is enter the keyword and hit enter/fetch tweets button below.
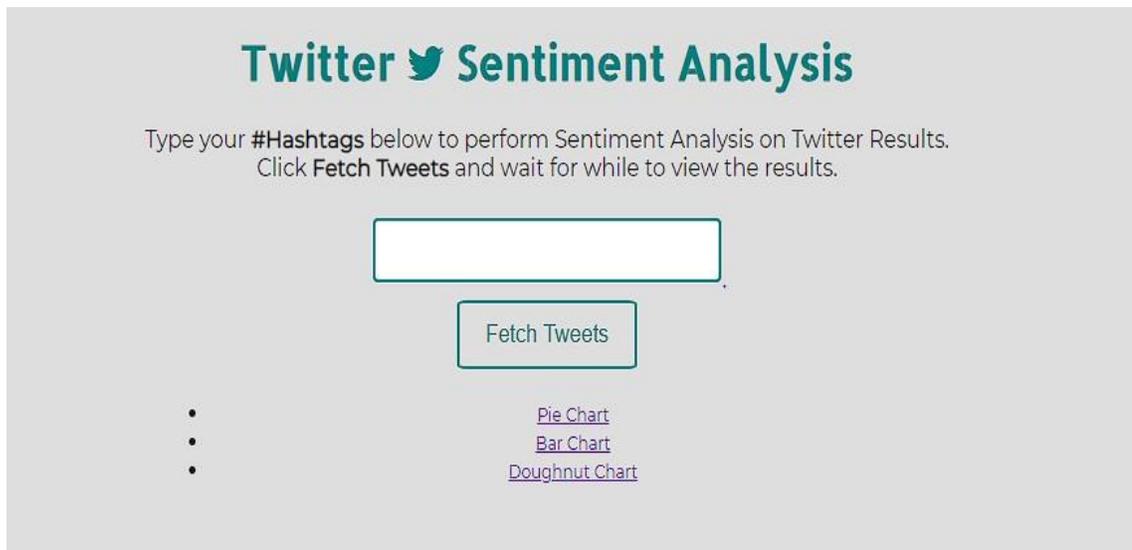
Fig 4.7: WebApp UI
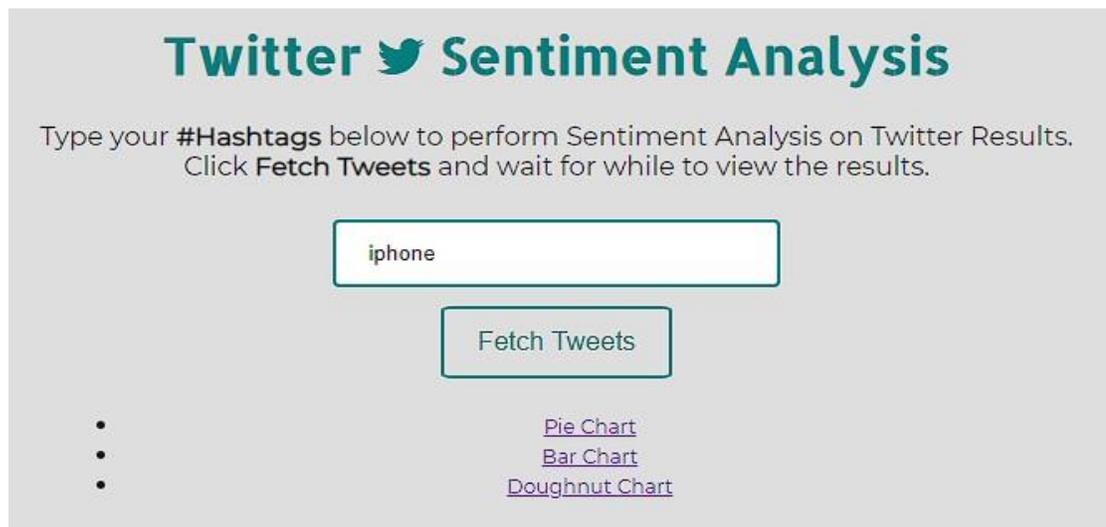
Fig 4.8: WebApp UI

● **Result:**

Following screenshots depicts how interactively and vividly the results are displayed in our webapp.

• First we see the result obtained in the form of various relevant chart and figures.
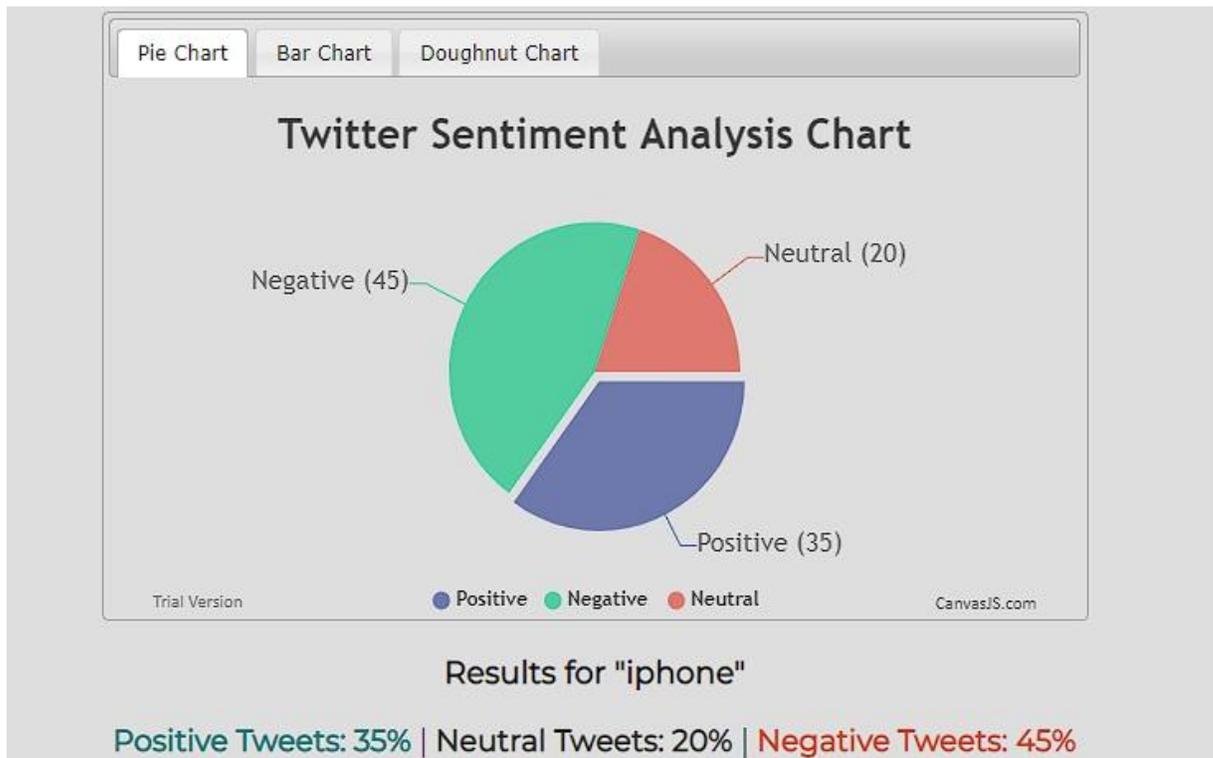
Fig 4.9: Pie chart
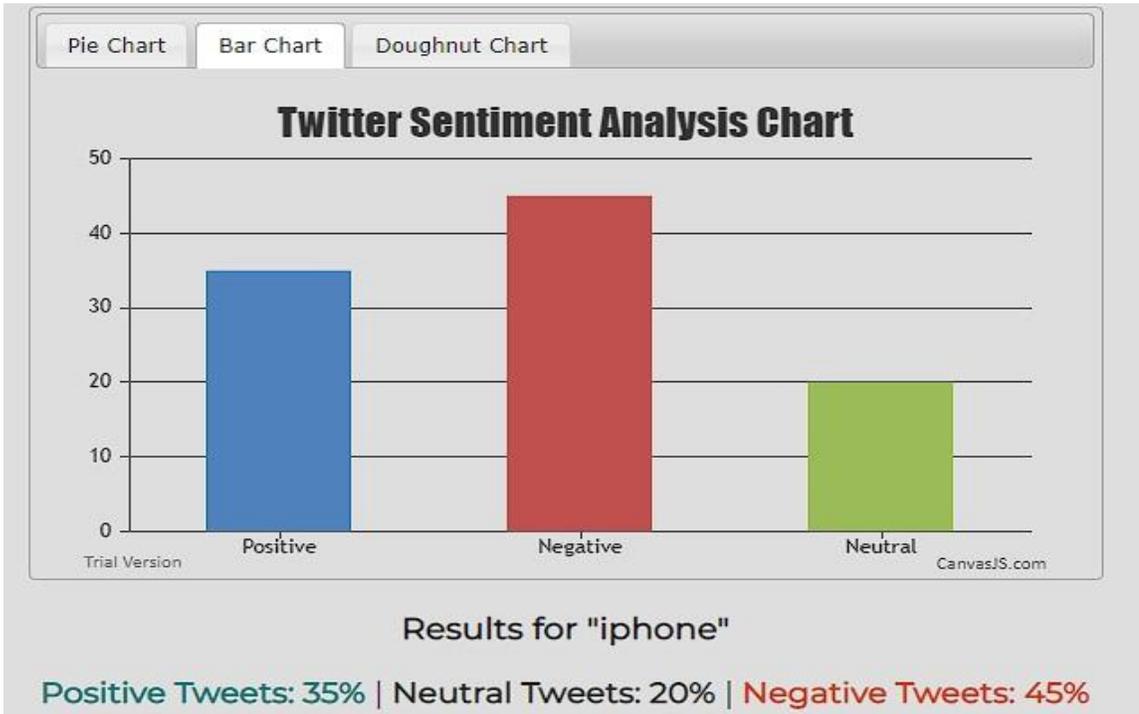
Fig 4.10: Bar Chart
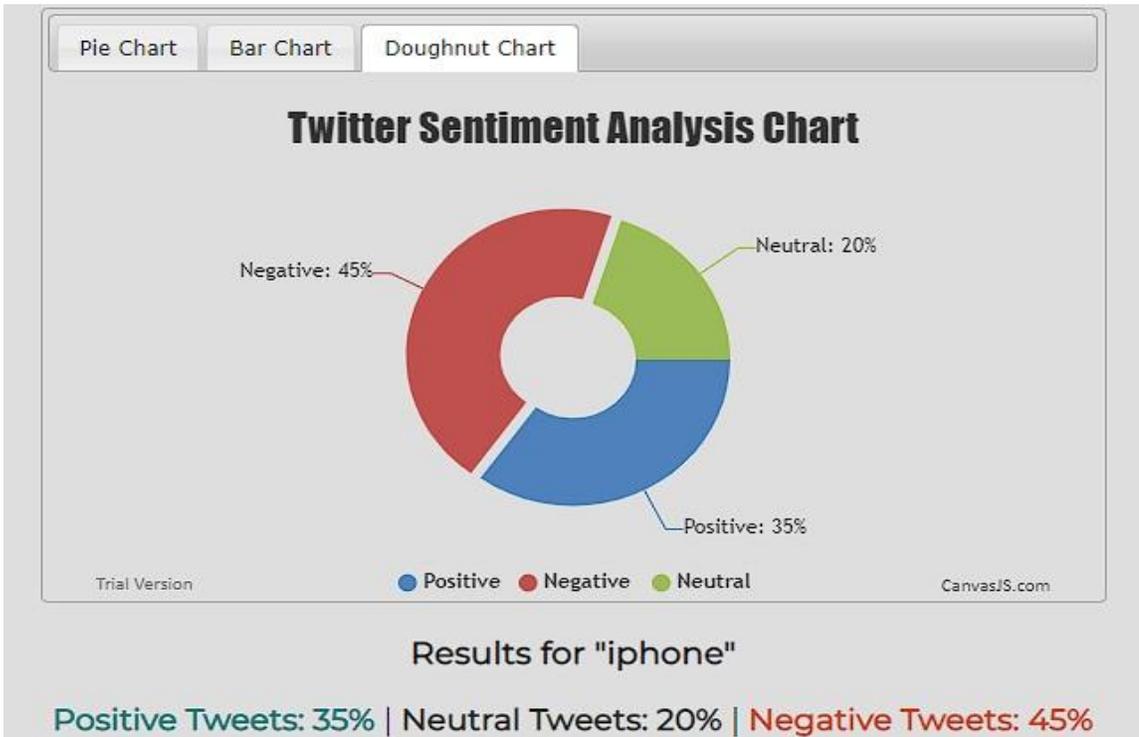
Fig 4.11: Doughnut Chart

- The following screenshots displays some of the tweets that were used to analyse the result so obtained.These tweets are also coloured in accordance with their respective polarity.
- Green is used for positive tweets.
- Red is used for negative tweets.
- White is used for neutral tweets.

Fig 4.12: Tweets

Fig 4.13: Tweets

Mak	@LeoMessi10i_ Im getting an iphone 11 and theres a good deal for a 64GB one 125 Gig of data unlimited mins and tex... https://t.co/ZjMJB4HTrG	🐦	positive
𐀃𐀃𐀃𐀃	And yes it's iPhone 6	🐦	positive
Asake🦋	@alabathefourth @AlliamZaying Tbvh The best phone for selfie is iPhone 7	🐦	positive
김비누멋휑	@jeonjell_o Oh... ok sorry But I think end is coming to my iPhone	🐦	negative
Anniss A	@ShopeeID A. Iphone Min @ShopeeID #NgabuburitBarengShopee	🐦	negative
Big Ole Table	I don't sleepwalk, I just wake up to about 9 different Basketball-Reference tabs on my iPhone that I don't remember looking up at 4am	🐦	negative

Fig 4.14: Tweets

4.1.2) Test Case 2:

For this test, we have analyzed the polarity and sentiment of **"amazon"**.

❖ **Python based sentiment analysis:**

● **Dataset:**

● Tweets from various users are retrieved and stored in csv format.

	text	has_hasht	retweet_c	followers	following	user	screen_name
0	RT @paul_cude: One young dragon's selfless determination w	1	53	566	337	Samuel Si	EmpireFiction
1	RT @BRasher_RACER: #Amazon is in advanced talks to buy @zc	1	1	1721	5001	Lindsay Sh	lsheltongross
2	RT @samcap924: Hey #WritingCommunity --My #sequel came	1	2	1662	110	Reveal Ho	Reveal_House
3	RT @Shut_downAmazon: #Amazon brainwashing tactics are or	1	63	78	1202	L Anthony	LAnthonyDyer1
4	Anker PowerCore Fusion 5000 PD, 18W USB-C Portable Charger	0	0	3625	3568	HeSheShc	hesheshopping
5	ðŸ¥ PlayStation Plus Abbonamento 12 Mesi | Codice	0	0	208	400	The Seller	maurygol91
6	Isn't this absolutely adorable? What a unique way to store toil	0	0	2934	2933	Denise Sa	DeniseSanger
7	RT @_Manish_pal: #Uttarakhand is burning but all of us are sile	1	32	28094	14424	ðŸ‡®ðŸ‡³[	DeepsUnique3434
8	#amazon 80% off Deal price: 43.99 Code: C6GTGEOW EXP: 2020,	1	0	5955	1169	NaturalHa	LatinaR2
9	RT @SheilaT2121: Want an easy, short read that packs a punch?	1	1	16991	188	Ben Davis	bendavisnbcs
10	Blasting Blueface, Jeremih - Close Up (Club) right now on	1	0	103	475	Pretty Dor	XperienceDope

Fig 4.15 : Dataset scrapped from twitter

● Structure of data after preprocessing, cleaning, removing stop words and tokenization.

0	['young', "dragon's", 'selfless', 'determination', 'will', 'tested', 'like', 'never', 'before', 'https://t.co/5ZNqPiazaB', '#fantasybooks', '#YAFantasyâ€¦']
1	['#Amazon', 'advanced', 'talks', 'move', 'that', 'would', 'expand', 'reach', 'autonomous-vehicle', 'technology.', '#AVs', '#moâ€¦']
2	['#WritingCommunity', '--My', '#sequel', 'came', 'today', '#Amazon', 'print!', 'could', 'your', 'help', 'with', '#WriterLift!', '#Foâ€¦']
3	['#Amazon', 'brainwashing', 'tactics', 'only', 'going', 'there', 'isnâ€™t', 'enough', 'money', 'replace', 'liveâ€¦']
4	['Anker', 'PowerCore', 'Fusion', '5000', 'USB-C', 'Portable', 'Charger', '2-in-1', 'with', 'Power', 'Delivery', 'Wall', 'Charger', 'iPhone', 'https://t.co/jJqil2nla3']
5	['PlayStation', 'Plus', 'Abbonamento', 'Mesi', 'Codice', 'download', 'Account', 'italiano', 'âšiï,41,99â,¬âšiï,', 'https://t.co/fR3vQtZgpV']
6	["Isn't", 'this', 'absolutely', 'adorable?', 'What', 'unique', 'store', 'toilet', 'tissue', 'your', 'beachy', 'bathroom.', 'this', 'amazâ€¦', 'https://t.co/lwPYV274ad']
7	['#Uttarakhand', 'burning', 'silent.', 'spoke', '#Australia', '#Amazon,', 'keeping', 'Jungle.â€¦']
8	['#amazon', 'Deal', 'price:', '43.99', 'Code:', 'C6GTGEOW', 'EXP:', '2020/7/19', 'https://t.co/hw86butNAN']
9	['Want', 'easy,', 'short', 'read', 'that', 'packs', 'punch?', 'Kindle', 'paperback', 'available', 'through', 'Amazon!', '#grief', '#shortreads', '#amazonâ€¦']

Fig 4.16 : Preprocessed dataset

- **User Interface:**
  - Since there is no ui as such, following is a screenshot of rather complex code structure that the user has to run in python ide (Sypder in this case) and then perform analysis

Fig 4.17 : User Interface

- **Result :**

After the extraction and preprocessing of data, we use TextBlob library to perform sentiment analysis of the tweets and classify them according to their respective polarity. A detailed analysis figures are displayed in console along with a pie chart of same created using MatplotLib library.

Following are the screenshot of result obtained in ide console.

```
Enter Keyword/Tag to search about: amazon
```

Fig 4.18 : Result

```
General Report:
Weakly Positive

Detailed Report:
16.00% people thought it was positive
20.00% people thought it was weakly positive
1.00% people thought it was strongly positive
1.00% people thought it was negative
9.00% people thought it was weakly negative
1.00% people thought it was strongly negative
52.00% people thought it was neutral
```

Fig 4.19 : Result

How people are reacting on amazon by analyzing 100 Tweets.

Positive [16.00%]
Weakly Positive [20.00%]
Strongly Positive [1.00%]
Neutral [52.00%]
Negative [1.00%]
Weakly Negative [9.00%]
Strongly Negative [1.00%]
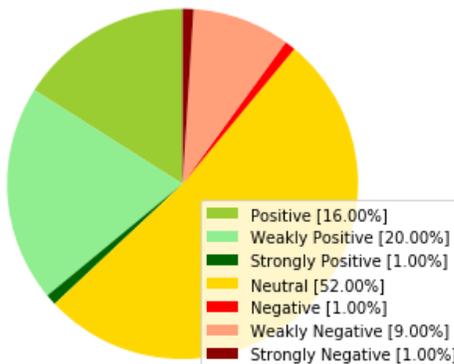
Fig 4.20 : Chart analysis

❖ **WebApp based sentiment analysis:**

● **User Interface:**

WebApp consists of a very user-friendly user interface developed using html, css, js and bootstrap. All you have to do is enter the keyword and hit enter/fetch tweets button below.
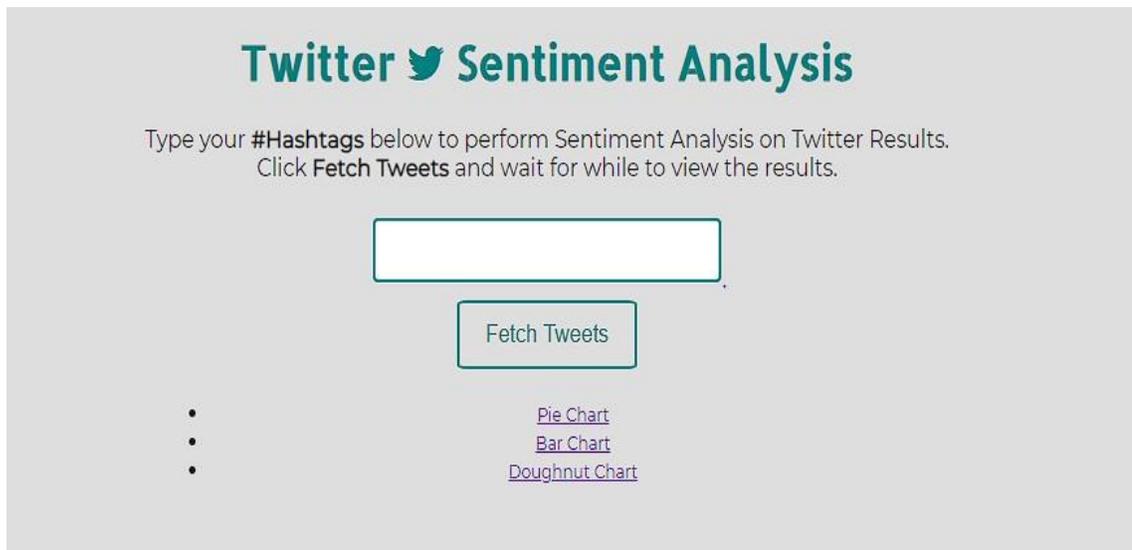

Fig 4.21: WebApp UI
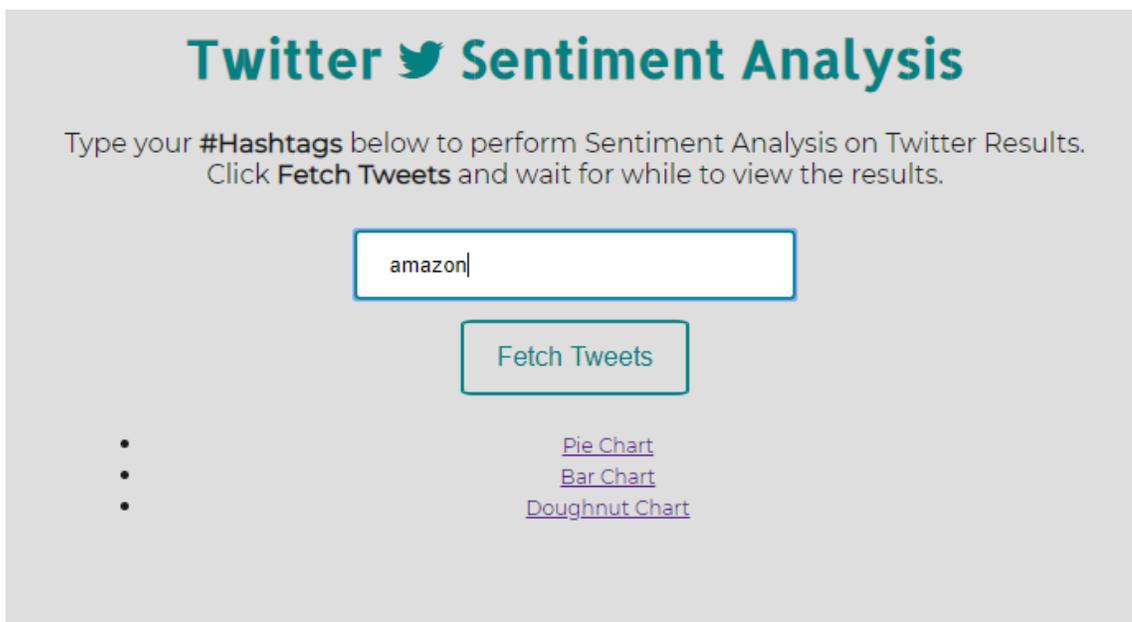
Fig 4.22: WebApp UI

● **Result:**

Following screenshots depicts how interactively and vividly the results are displayed in our webapp.

● First we see the result obtained in the form of various relevant chart and figures.
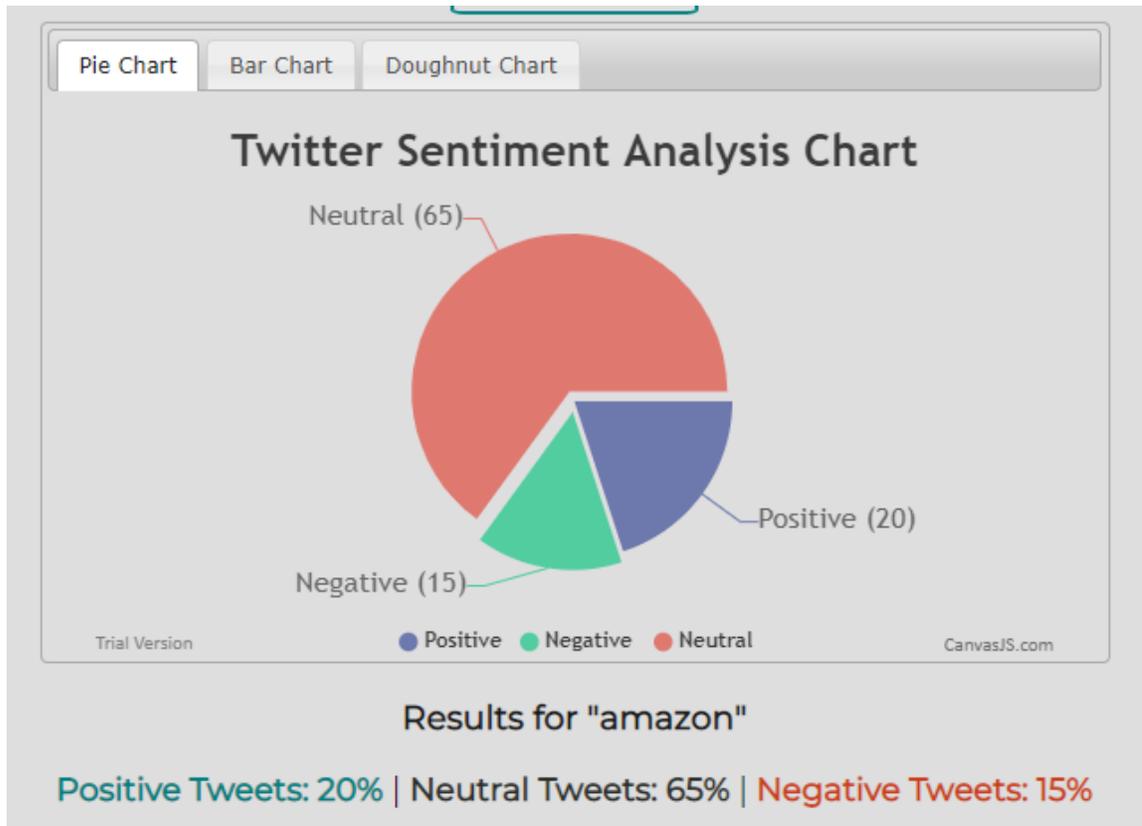
Fig 4.23: Pie chart
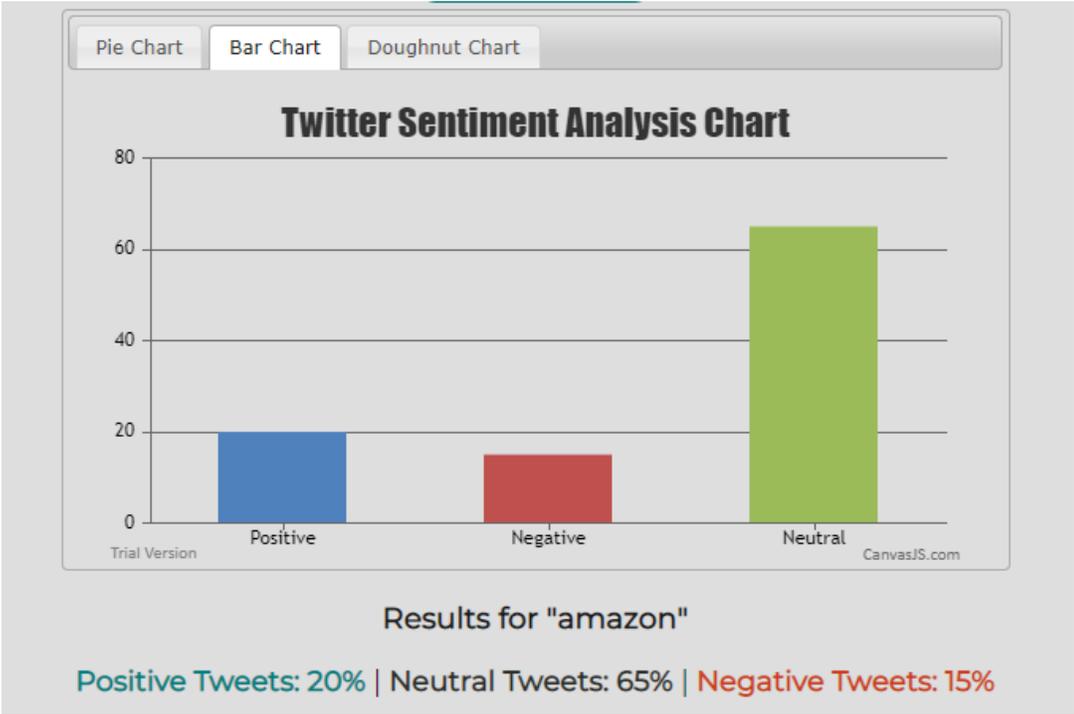
Fig 4.24: Bar Chart
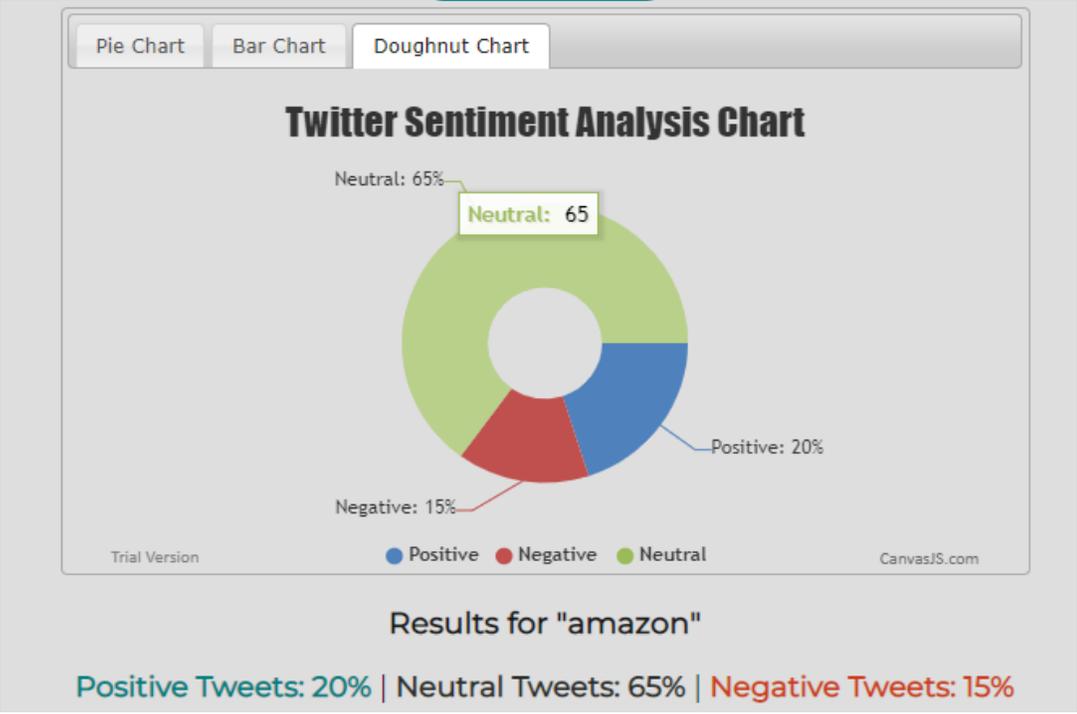
Fig 4.25: Doughnut Chart

- The following screenshots displays some of the tweets that were used to analyse the result so obtained.These tweets are also coloured in accordance with their respective polarity.
- Green is used for positive tweets.
- Red is used for negative tweets.
- White is used for neutral tweets.

Fig 4.26: Tweets

Fig 4.27: Tweets

Aaron	RT @realDonaldTrump: .@Twitter is now interfering in the 2020 Presidential Election. They are saying my statement on Mail-In Ballots, which...	🐦	neutral
Monica Michel	VersionTECH. Wired Gaming Mouse, Ergonomic USB Optical Mouse Mice with Chroma RGB Backlit, 1200 to 3600 DPI for Lap... https://t.co/qhRkKOVVS2	🐦	neutral
💜 Kelly Miller 💜 Regency Romance Author	RT @AnxietyandLiz: @luanabracklow Hiya! My debut novella! UK: https://t.co/QGji5KeQOf... US: https://t.co/7IObyVYhVf It's FREE on #Kindl...	🐦	neutral
Shivz [4am In Miami Out Now]	RT @50cent: my man @SnoopDogg Hustle Harder Hustle Smarter - https://t.co/0NVuEOqQ4T #bransoncognac #lecheminduroi https://t.co/b3z6mZeg...	🐦	neutral

Fig 4.28: Tweets

CHAPTER 5: CONCLUSION

5.1) Conclusions:

In conclusion, we have discussed in detail how our webapp is not only a far more convenient way for carrying out twitter sentiment analysis but also is a platform which could be utilized by individual with no knowledge of how to program an algorithm and run it in an ide.

In traditional methodology the sentiments were mined through surveying every person from various socioeconomics. It required a lot of human time and labour hence decreasing the productivity that could've been.

As opposed to that in the event that a similar work is done from the biggest interpersonal organization, at that point it could be proficiently done likewise, we can break down the pattern in equivalent to assortment of tweets from past posts is as yet possible. Also since this algorithm is being used in backend of a powerful webapp, this facilitates common man to easily access the information at hand.

As we have seen their happens a tremendous variety of extremity and opinions gathered after some time that helps in administering the mind-set of worldwide individuals across various topographical areas. Also a few words like investigation demonstrated zero negative proportions.

Above mentioned result and analysis clearly indicates that our webapp is far more convenient to carry out twitter sentiment analysis in comparison to the traditional python based app that requires sophisticated knowledge of coding as well as statistics in order to discover meaningful data from the result obtained.

Every person has access to a to network and is more than curious to find things out for himself. This webapp will provide a platform to everyone where they can easily analyse how to product/services they wish to avail is actually turning out to be by checking out the performance review already existing users have given on twitter. They do not need to consult a professional to get simple information about the product nor they do need to fall under misinformation or online scams. Now they have to power to verify the information by simply entering the keyword and analyzing the polarity themselves.

Big business ventures could also use the app for keeping a track of the polarity of their product in the market. Keeping a track of twitter allows the organization to be aware of their brand popularity in the opinion of their target audience, they could also analyze how well their competition is doing as compared to them.

5.2) Future Work:

As you might have noticed that we are still getting a large number of tweets classified as neutral in polarity. This is due to the reason that the tweets are based on language that is common to our tongue. This language does not follow the traditional rules of grammar nor do they pertain to a single style of writing. It is full of slangs that might mean different but express an entirely different emotion and hence might mislead the classifier into classifying it to a polarity it does not belong to or most the times end up in expressing conflicting polarity hence classifying it into neutral.

This represents a small part of a bigger problem at hand. This issue affects the accuracy of webapp.

Hence in order to overcome this issue, we need to have an ever more broad extensive support of machine learning frameworks that support web based scripting language such as jQuery, Javascript or PHP. In the current scenario, we have sentiment nodejs module, datumbox machine learning framework and a couple of twitter sentiment analysis libraries that we can utilize to carry out the task at hand.

But there needs to be more research and development in developing more efficient unsupervised machine learning models based on neural networks and provides support to web based apps too so that more such apps could be created and made accessible to general public for their own use.

5.3) Applications:

Fundamentally, the sentiment analysis app proves to be a valuable asset in predicting the image of the brand and it's products. It permits the brand to:

- Analyze the image of their brand in the opinion of the public;

- Predicting and informing the brand beforehand about it's future downfall;

- Discover examples and patterns;

- watch out for the introduction by the influencers.

This permits us to conform to the situation in like manner and give the item a legitimate introduction.

In general, sentiment analysis can be utilized to:

- Mechanize media observing procedure and the going with ready framework

- Screen notices or audits of the brand on various stages (web journals, online life, survey destinations, discussions, and so forth.)

- Order direness of notices as per the significance scoring (i.e., which stage, sort of client is crucial to the brand)

The entire activity comprises of two phases:

- At the underlying stage, the organization responds to the approaching outcomes and adjusts.

- After some time, supposition investigation can change the strategy from responding to dealing with the observation.

Following sentiment in client created substance can permit you to explore through the beginning times of item advancement.

At the point when you've propelled another item or highlight for your organization, it's conceivable you'll burn through a large number of dollars in client research to attempt to make sense of on the off chance that it that item will be effective and a hit with purchasers.

Things being what they are, imagine a scenario where data-mining could assist you with tuning in to what clients were stating about your item from the second it propelled.

Finnish outdoor equipment giant Suunto did only that when they propelled their Spartan Ultra watch. The pack had a couple of early innovation issues, and clients weren't modest about sharing their musings:
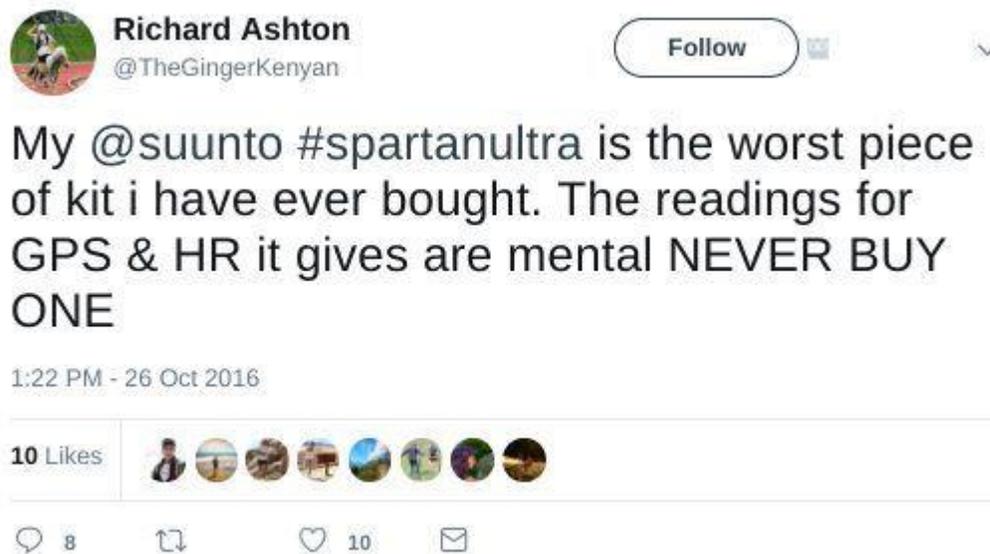
Fig 5.1: Applications
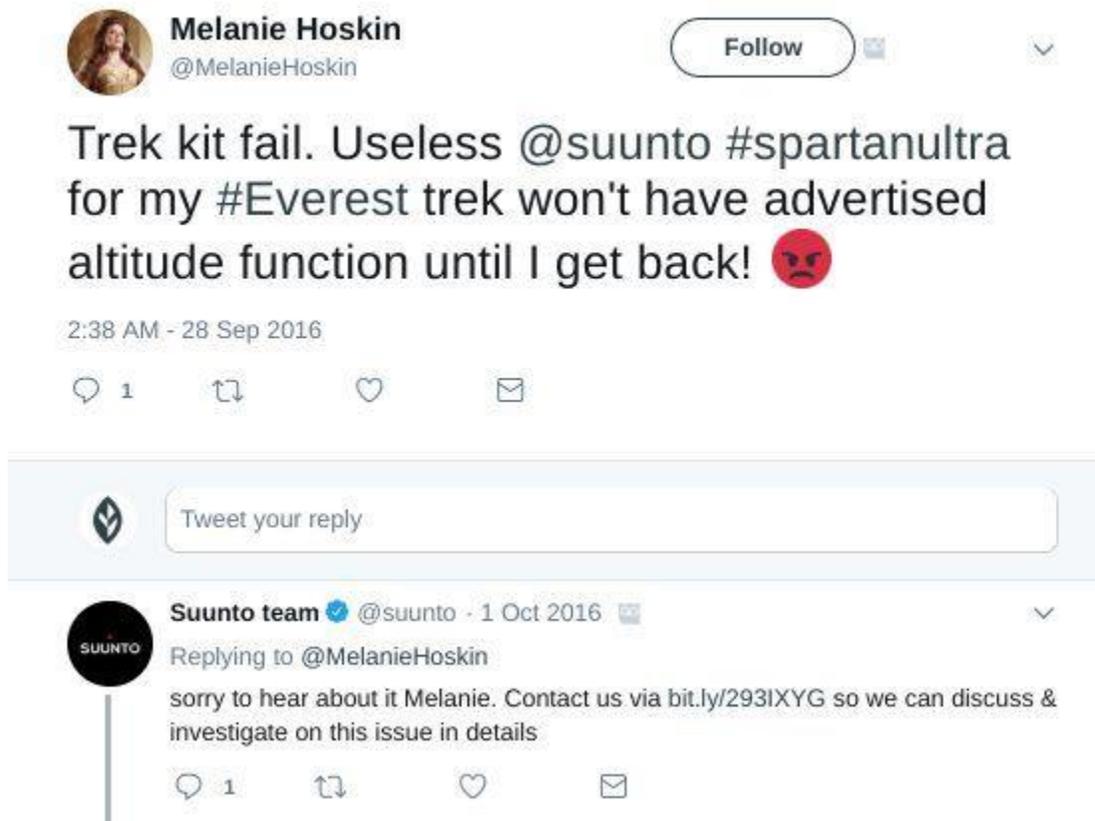
Fig 5.2: Applications

Since the organization had the option to follow up on this negative input in a flash, just as use it to help change the new item, harm to the brand was kept to a base.

Fig 5.3: Applications

Fig 5.4: Applications

The way to client created content is recollecting that a certain something. In the event that a client sets aside the effort to compose something about your image, you can wager they need to be heard.

Tune in, switch things up, and let individuals realize they're being heard.

REFERENCES:

[1]   :B. Pang and L. Lee(2008), "Opinion mining and sentiment analysis" . Foundations and Trends in Information Retrieval 2(1-2), pages 1-135.

[2]   Recognizing contextual polarity in phrase level sentient analysis: Theresa Wilson, Jaynce Wiebe, Paul Hoffman, 2005

[3]   Enhancing deep learning sentiment analysis with ensemble techniques in social applications: Oscar Araque, IgnacioCorcuera-Platas, J. FernandoSánchez-Rada, Carlos A.Iglesias, Expert systems with application, Volume77, 1 July,2017

[4]   : Rodrigo Sandoval-Almazan, David Valle-Cruz , "Facebook impact and sentiment analysis on political campaigns on Political rallies",19th Annual International Conference -May 2018.

[5]   : Marilena Di Bar,S Sharoff, Martin Thomas , "SentiML: functional annotation for multilingual sentiment analysis", ACM, Florence, Italy, September 2013.

[6]   : Justin Martineau, and Tim Finin, "Delta TFIDF An Enhanced Element Space for Notion Examination",In Third International ICWSM Conference, Pg 258-261, November 2009.

[7]   : Shahana P.H,Bini Ommanb , "Evaluation of Features on Sentimental Analysis", International Conference on Information and Communication Technologies, Pg 1585-1592, 2014.

[8]   : Rahul Rajput, Arun Kumar Solanki, "Review of Sentimental Analysis Methods using Lexicon Based Approach", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 5, Issue. 2, February 2016, pg.159 – 166.

[9]   : Shubham Goyal, "Sentimental Analysis of Twitter Data using Text Mining and Hybrid Classification Approach", International Journal of Advanced Research, Ideas and

Innovations in Technology, Volume2, Issue 5, 2016.

[10]   : Gurneet Kaur, Abhinash Singla, "Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 1, January 2016.

[11]   : Zhen Hai, "Analyzing Sentiments in One Go: A Supervised Joint Topic Modeling Approach", IEEE Transactions on Knowledge and Data Engineering, Vol 29, pg: 1172-1185.

[12]   : Yazhi Gao, Wenge Rong, Yikang Shen, Zhang Xiong, "Convolutional Neural Network based sentiment analysis using Adaboost combination", International Joint Conference on Neural Networks (IJCNN), 2016, Pg 1333-1338.

[13]   Sentiment analysis using support vector machines with diverse information sources Tony Mullen and Nigel Collier National Institute of Informatics (NII) Hitotsubashi 2-1-2, Chiyoda-ku Tokyo 101-8430 Japan

[14]   : Xiaojiang Lei, Xueming Qian and Guoshuai Zhao, "Rating Forecast Dependent on Social Slant From Literary Surveys",Volume 18 Issue 9, September 2016 .

[15]   : Bouazizi, M., & Ohtsuki, T.A Pattern-Based Approach for Multi-Class Sentiment   Analysis   in   Twitter.   *IEEE   Access.* https://doi.org/10.1109/ACCESS.2017.2740982

[16]   : K. R. Aida-zade, S. S. Rustamov, E. E. Mustafayev, and N. T. Aliyeva, "Human-Computer Dialogue Understanding Hybrid System, " IEEE Xplore, International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey, Jule 2012, pp. 1-5.

[17]   :Najkar, Negin & Razzazi, Farbod & Sameti, Hossein. (2010). "A novel approach to HMM based speech recognition system using particle swarm optimization". Mathematical and Computer modeling. 52. 1910-1920. 10.1016/j.mcm.2010.03.041.

[18]   : Manvee Chauhan ,Divakar Yadav , "Sentimental Analysis of Product Based Reviews Using Machine Learning Approaches", Journal of Network Communications and Emerging Technologies (JNCET) Volume 5, Special Issue 2, December (2015).

[19]   : Lijuan Zheng, Hongwei Wang, Tune Gao, Sentimental feature selection for sentiment analysis of Chinese online reviews, International Journal of Machine Learning and Cybernetics, Springer

[20]   :https://www.nielsen.com/in/en/insights/news/2012/social-media-report-2012-social-media- comes-of-age.html,[Accessed: May 10"19]

[21]   : https://monkeylearn.com/sentiment-analysis, [Accessed: May 10ʹ19]

KKR2

9 | Submitted to Blackpool and The Fylde College, Lancashire
Student Paper
<1%

10 | Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede. "Lexicon-Based Methods for Sentiment Analysis", Computational Linguistics, 2011
Publication
<1%

11 | Submitted to Laureate Higher Education Group
Student Paper
<1%

| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 17/07/2020

Type of Document (Tick): | PhD Thesis | M.Tech Dissertation/ Report | B.Tech Project Report ✓ | Paper |

Name: Kanishk Kala   **Department:** Information Technology   **Enrolment No** 16166

Contact No. 8076594663   **E-mail.** kanishkk97@gmail.com

Name of the Supervisor: Dr Himanshu Jindal

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): TWITTER SENTIMENT ANALYSIS USING WEBAPP

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

☐ Total No. of Pages = 59

☐ Total No. of Preliminary pages = 8

☐ Total No. of pages accommodate bibliography/references = 5

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at ………………..(%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor)   **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
	☐ All Preliminary Pages		Word Counts	
Report Generated on	☐ Bibliography/Images/Quotes		Character Counts	
		Submission ID	Total Pages Scanned	
	☐ 14 Words String		File Size	

Checked by
Name & Signature   **Librarian**

……

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com