PROJECT REPORT

TABLE OF CONTENTS

ABSTRACT

Twitter is a stage generally utilized by individuals to express their feelings and show estimations on various events. Opinion investigation is a way to deal with break down information and recover opinion that it exemplifies. Twitter conclusion examination is utilization of feeling examination on information from Twitter (tweets), keeping in mind the end goal to separate conclusions passed on by the client. In the previous decades, the examination in this field has reliably developed. The purpose for this is the testing configuration of the tweets which makes the preparing troublesome. The tweet design is little which creates a radical new measurement of issues like utilization of slang, shortenings and so forth. In this paper, we plan to audit a few papers in regards to examine in opinion examination on Twitter, depicting the approaches received and models connected, alongside depicting a summed-up Python based approach. Twitter is a phase generally used by people to express their sentiments and show estimations on different occasions. Assessment examination is an approach to manage separate data and recoup sentiment that it represents. Twitter conclusion examination is a use of feeling examination on data from Twitter (tweets), remembering the ultimate objective to isolate conclusions passed on by the customer. In the earlier decades, the examination in this field has dependably created. The main contribution of this paper is the idea of portraying the use of two libraries ‘TextBlob’, ‘Tweepy’ for sentimental analysis on Twitter.

1. Introduction

Twitter is a prevalent microblogging administration where clients make status messages (called "tweets"). These tweets here and there express sentiments about various points. We propose a technique to naturally extricate opinion (positive, negative and neutral) from a tweet. This is extremely helpful in light of the fact that it permits criticism to be totaled without manual intercession. Shoppers can utilize slant examination to investigate items or then again benefits before making a buy. Advertisers can utilize this to investigate popular supposition of their organization and items, or on the other hand to investigate consumer loyalty. Associations can likewise utilize this to assemble basic criticism about issues in recently discharged products. There has been a lot of research in the zone of assessment characterization. Generally, its majority has concentrated on arranging bigger bits of content, similar to surveys. Tweets (and microblogs when all is said in done) are not quite the same as surveys principally in light of their motivation: while audits speak to outlined contemplations of creators, tweets are more easygoing and restricted to 140 characters of content. By and large, tweets are not as attentively formed as audits.

However, despite everything they offer organizations an extra road to accumulate input. There has been some work by scientists in the territory of expression level and sentence level assumption characterization as of late. Past inquire about on investigating blog entries incorporates. Past research in assumption investigation like [16]. They dissected the execution of various classifiers on motion picture surveys. Crafted [16] has filled in as a pattern. What's more, numerous creators have utilized the systems gave in their paper crosswise over various areas. [16] Likewise make utilize of a comparable thought as our own, utilizing star evaluations as extremity signals in their preparation information.

We demonstrate that we can deliver equivalent outcomes on tweets with far off supervision. Keeping in mind the end goal to prepare a classifier, directed adapting for the most part requires hand-marked preparing information. With the substantial scope of points talked about on Twitter, it would be exceptionally hard to physically gather enough information to prepare an opinion

classifier for tweets. Our answer is to utilize removed supervision, in which our preparation information comprises of tweets. This approach was presented by read a tweet demonstrates that the tweet contains positive estimation and shows that the tweet contains negative conclusion or a neutral tweet, With the assistance of the Twitter API, it is anything but difficult to extricate a lot of tweets in them. This is a noteworthy change over the numerous hours it might some way or another grasp to hand-mark preparing information. We run classifiers prepared on tweets information against a test set of tweets.

2. Literature Survey

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Starting from being a document level classification task [1], it has been handled at the sentence level and more recently at the phrase level. The emerging singularities of social media, such as Facebook, Twitter, LinkedIn, and Instagram, with each one has its own characteristics and its usages, are constantly affecting out societies. Facebook, for example, is considered as a social network where everyone in the network has a joint relationship with another one in the same network. Relationship in this case is undirected. Conversely, in Twitter everyone in the network does not necessarily have a reciprocated relationship with others. In this case, the relationship is either directed or undirected. In this paper, we focus on twitter for data analysis, where twitter is an online networking service that enables users to send and read character messages called "tweets".

Below we review several types of analyses that most researchers have used.:

Datasets: Datasets Analyzing structured data have been widely used. In such case, the traditional (RDBMS) can deal with the data. With the increasing amounts of unstructured data on various sources (e.g. Web, Social media, and Blog data) that are considered as Big Data, a single computer processor cannot process such huge amount of data. Hence, the RDBMS cannot deal with the unstructured data. To solve this problem a cross big data framework is usually employed, such as Apache Hadoop [4]. Hadoop also deals with structured and semi-structured data, XML/JSON files, for example. The strength of using Hadoop comes in storing and processing large volume of data. In [2], the mechanism of FireFox add-on and Clean Tweet filter was employed to remove users that have been on twitter for less than a day and they removed tweets that contain more than three hashtags.

Data Retrieval: Before retrieving the data, some questions should be addressed: What are the characteristics of the data? Is the data static, such as the profile user information "name, user Id, and bio"; or dynamic such as user's tweets, and user's network? Why is the data important? How is the data will be used? And how big the data is? It is important to note that it is easier to track a

certain keyword attached to a hashtag rather than a keyword not attached to it. Other studies, as in [5], have used GNU/GPL application like YourTwapperKeeper tool, which is a web-based application that stores social media data in MySQL tables. However, YourTwapperKeeper in storing and handling large size of data exhibits some limitations in using, as MySQL and spreadsheets databases can only store a limited size of data. Using a hybrid big data technology might address such limitations as we suggested above.

Ranking and Classifying Twitter Users: [3] There are different types of user networks. A network of users within a specific event (hashtag), a network of users in a specific user's account, and a network of users within a group in the network, that is, Twitter Lists. [2] Lists are used to group sets of users into up-to-date or other categories to better organize and filter incoming tweets.

Homophily: It is defined as the tendency that contacts among similar users occur at a higher rate than among dissimilar users, that is, similar users tend to follow each other. [2] It requires studying the static characteristics of twitter data, such as the profile name and the geographic feature of each user in twitter network. [3] It studies the homophily and geographical feature in twitter to investigate the similarity between users based on their location.

Reciprocity: [2] The characteristic nature of twitter as being both directed and undirected social network has made most studies analyze reciprocity. [3] Reciprocity is the property of following a user and being followed back that is mutual relationship.

3. Approach

3.1 Classification of Sentiment Analysis Approaches

[15]Starting from the analysis provided in the previous sections a classification of sentiment analysis approaches with respect to features/techniques and advantages /limitations is provided in following:

3.1.1 Machine learning:

- Bayesian Networks Naive Bayes Classification

Features/Techniques:

- Term presence and frequency
- Part of speech
- information
- Negations
- Opinion words and phrases

Advantages: The ability to adapt and create trained models for specific purposes and contexts Limitations: The low applicability to new data because it is necessary the availability of labeled data that could be costly or even prohibitive

3.1.2 Lexicon based:

- Dictionary based approach Novel Machine
- Learning Approach Corpus based approach Ensemble Approaches

Features/Techniques:

- Manual construction,
- Corpus-based
- Dictionary based

Advantages: wider term coverage

Limitations: finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words

3.1.3 Hybrid:

- Machine learning
- Lexicon based

Features/Techniques:

- Sentiment lexicon constructed using public resources for initial sentiment detection
- Sentiment words as features in machine learning method

Advantages: lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain

Limitations: noisy reviews

3.2 Our Approach using Naïve Bayes Machine Learning Method

Sentimental analysis will be performed after the data i.e., tweets are collected from the source (here twitter). The collected tweets need some pre-processing like cleaning, removing emoticons and unnecessary words. So that our tweets can be categorized into positive, negative and neutral.

3.2.1 Tweet Collection

Tweets are short length messages and have a greatest length of 140 characters. This constrains the measure of data that the client can impart to each message. Tweet collection includes gathering applicable tweets about the specific zone of interest. We are using Twitter's API for collecting large number of tweets. Before using twitter API, we must create a twitter account by providing basic details. To access data from twitter account and use it for our project, we have to login into dev.twitter.com website. On this website, we have to create an application by providing required details. After the creation of twitter API, we get the customer key, customer secret key, access token key and access secret key. These keys are utilized to confirm client when client need to get to twitter information. [7], [8]

For the analysis of tweets, we make a python script which will be utilized to bring tweets from twitter. Before making this content, we initially introduce a library in python called tweepy. In this

script we utilize all the keys, we initially make audience class which is utilized to stack the information from the twitter. presently to assemble information we first set up 'OAuth' convention. OAuth is a standard convention which is utilized for approval. It enables client to sign in any outsider sites by utilizing any informal community site account without uncovering passwords. OAuth gives security and authorization to client.

In this script we need to give all the keys which are given by twitter programming interface. To get the tweets for a specific point we import stream library from tweepy. When we run our script, we see tweets are transported in from twitter and we would then be able to utilize them for our purpose.

3.2.2 Pre-processing Tweets

The preprocessing of the tweets is a critical step as it chooses the effectiveness of alternate steps. It includes linguistic remedy of the tweets as wanted. The steps included should go for making the information more machine meaningful keeping in mind the end goal to decrease ambiguity in feature extraction.

Information gathered from twitter isn't fit for removing highlights. Generally, tweets comprise of message alongside usernames, void spaces, uncommon characters, stop words, emojis, shortenings, hash labels, time stamps, URL's and so forth. In pre-handling we first concentrate our principle message from the tweet, at that point we evacuate every unfilled space, stops words, hash labels, rehashing words, URL's and so on. Cleaning twitter information is vital, since tweets contain a few syntactic highlights that may not be valuable for examination. the pre-handling is done such that information spoke to just regarding words that can without much of a stretch group the class. [7]

- **URL’s**

  Individuals utilize twitter not just to express their feelings yet in addition for imparting data to others. Given the short greatest length of tweets, one method for sharing is utilizing joins. Tweets incorporate different connections or URLs and these don't contribute to the conclusion of the tweet.

- **Usernames**

Tweets regularly allude to different clients and such references start with the @ symbol. These again don't add to the notion and consequently are supplanted by the non-specific word USERNAME.

- **Repeated Characters**

Individuals utilize a great deal of easygoing dialect on twitter. For instance, 'Okay' is utilized as a part of the type of 'Okkaaayyyyy'. In spite of the fact that this suggests a similar word 'Okay', the classifiers consider these as two unique words. To enhance this and make words more like non-specific words, such arrangements of rehashed letters are supplanted by two events. In this manner “Okkaaayyyy” would be supplanted by “Okay”.

- **Twitter slang removal**

Words are exchanged by the genuine words that people speak to progress execution of the learning algorithms. The word ‘Hello’ is utilized by individuals utilizing some short shapes like heelloo, helo, hllo, etc. In the event that these are not mapped to the regular unique word, at that point preparing on them would not deliver great accuracy.

- **Stemming**

Stemming is the way toward decreasing a word to its root form. For instance, talking, talker, talker every one of these words are gotten from the root word talk. Henceforth, the stemmed type of all the above words is talk. Stemming decreases the element space the same number of determined words are decreased to a similar root frame. Numerous includes now point to a similar word and henceforth it increments the likelihood of the word. Stemming gives an increment in accuracy. By stemming, different determined words are mapped to their root words and this permits all the more coordinating between the tweets in the test and training set.

3.2.3 Feature extraction

3.2.3.1. Unigram

Unigrams are the least complex highlights that can be utilized for learning tweets. The pack of-words demonstrate is a ground-breaking procedure in assumption investigation. This procedure includes gathering all words in the record and utilizing them as highlights. The highlights can either be the recurrence of words, or essentially 1s to demonstrate if the word is available in the archive or on the other hand not.

3.2.3.2. Bigram

Bigrams are highlights comprising of sets of two neighboring words in a sentence. Unigram at times can't catch phrases and multi-word articulations, adequately dismissing any word arrange reliance. For instance, words like 'not glad', 'not great' plainly say that the conclusion is negative, in any case, a unigram may neglect to recognize this. In such cases, bigrams help in perceiving the right slant of the tweet. [7]

3.2.3.3. POS tagging

Part-of-speech tagging in semantics and data recovery is the way toward labeling each word in a sentence to a specific grammatical feature. There are numerous parts of speech for example, thing, descriptive word, pronoun, relational word, verb modifier, and so on. A word can take distinctive implications in various sentences, that is, a word can go about as a thing in one sentence, and as a descriptive word in another. Following is the list of POS and their full forms in the event that you need to know them in detail. [9]

POS Tag	Description	Example
CD	Cardinal Number	1, forth
DT	Determiner	The
FW	Foreign word	Le stylo

IN	Preposition	of,in
NN	Noun, singular	Door
PRP	Personal pronoun	I, he
MD	Modal	Could, should

3.2.4 Sentiment Classifiers

In Naive Bayes classifiers, every component impacts which stamp should be assigned out to a given data value. To pick a stamp for an information value, the naive Bayes classifier begins by finding out the prior probability of each name, which is directed by checking repeat of each name in the planning set. Presently, we have a training set, so we ought to just instantiate a classifier and describe test tweets. [6]

Mathematical Representation: Consider characteristics X1, X2....Xn to be restrictively free of each other given a class Y. This presumption gives us,

P(X1....Xn|Y) = Yn i=1 P(Xi|Y)

By Bayes hypothesis, we have,

P(Y |Xi) = (P(Xi|Y)P(Y))/ P(Xi)

Utilizing Bayes theorem in the previous equation, we discover the probability of predicting the class Y given the highlights Xi. The class that gives the most extreme likelihood that the given highlights foresee it, is the class that the tweet will have a place to. In this experiment, the Naïve Bayes Classifier from NLTK was utilized to prepare and test the information.

4. Twitter Sentimental Analysis with Python

4.1 Python

Python is a high-level programming language extensively used in the field of data science. It was created by Guido van Rossum during 1985- 1990. It is interpreted, interactive and object-oriented scripting language. It is known for its highly readable and easily understandable feature. It uses white space indentation to delimit blocks.

Python provides a vast library which can be used for various applications for example natural language processing, machine learning, data analysis etc. [13]

4.2 Natural Language Processing (NLTK)

Natural Language Processing Toolkit library is a part of machine learning in Python. [10] It is a practical introduction to programing for language processing. NTLK is a free, open source, community-driven project.

NLTK provides the base for text processing and classification. Tokenization, tagging, filtering, text manipulation operations can be performed using NLTK. It embodies various trainable classifiers into itself like Naïve Bayes Classifier.

4.3 Tweet Collection

In order to collect tweets from twitter, we need to create a Twitter App through our twitter account. For creating a twitter app, we need to follow the steps:

- Open link https://apps.twitter.com/ and click the button: ‘Create New App’
- Fill the details as required.
- The app is created and you will be redirected to the App page.
- Copy ‘Consumer Key’, ‘Consumer Secret Key’, ‘Access token’, ‘Access Token Secret’ from ‘Keys and Access Tokens’ tab of the App page.

This key will be used in the code for fetching real-time tweets through Twitter API [11].

4.4 Tweepy

Tweepy is a python client for the official Twitter API. We can install it through pip.

pip install tweepy

Tweepy supports oauth authentication. Authentication is handled by the tweepy.AuthHandler class. To begin the process, we need to register our client application with Twitter. We need to keep consumer token and secret handy from our application.

4.5 TextBlob

TextBlob is a python library built over top of NLTK and offers a simple API to perform NLP tasks. It is used for processing textual data. A big advantage of this is, it is easy to learn and offers a lot of features like sentiment analysis, pos-tagging, noun phrase extraction, etc. We can install it using following pip command:

pip install textblob

We need NLTK corpora using following command:

python -m textblob.download_corpora

Corpora is nothing but a large and structured set of texts.

[12] There are few advantages and disadvantages of using TextBlob. They are as follows:

Pros of TextBlob:

- Since, it is built on the shoulders of NLTK and Pattern, therefore making it simple for beginners by providing an intuitive interface to NLTK.
- It provides language translation and detection which is powered by Google Translate (not provided with Spacy).

Cons of TextBlob:

- It is little slower in the comparison to spacy but faster than NLTK. (Spacy > TextBlob > NLTK)

- It does not provide features like dependency parsing, word vectors etc. which is provided by spacy.

We follow these 3 major steps in our program:

- Authorize twitter API client.
- Make a GET request to Twitter API to fetch tweets for a particular query.
- Parse the tweets. Classify each tweet as positive, negative or neutral.

Now, let us try to understand the code:

- First of all, we create a **TwitterClient** class. This class contains all the methods to interact with Twitter API and parsing tweets. We use **__init__** function to handle the authentication of API client.
- In **get_tweets** function, we use:

```
fetched_tweets = self.api.search(q = query, count = count)
```

  to call the Twitter API to fetch tweets.

- In **get_tweet_sentiment** we use textblob module.

```
analysis = TextBlob(self.clean_tweet(tweet))
```

  TextBlob is actually a high-level library built over top of NLTK library. First, we call **clean_tweet** method to remove links, special characters, etc. from the tweet using some simple regex.
  Then, as we pass **tweet** to create a **TextBlob** object, following processing is done over text by textblob library:
  - Tokenize the tweet, i.e., split words from body of text.
  - Remove stop words from the tokens (stop words are the commonly used words which are irrelevant in text analysis like I, am, you, are, etc.)
  - Do POS (part of speech) tagging of the tokens and select only significant features/tokens like adjectives, adverbs, etc.
  - Pass the tokens to a **sentiment classifier** which classifies the tweet sentiment as positive, negative or neutral by assigning it a polarity between -1.0 to 1.0.

Here is how **sentiment classifier** is created:

- **TextBlob** uses a Movies Reviews dataset in which reviews have already been labelled as positive or negative.
- Positive and negative features are extracted from each positive and negative review respectively.
- Training data now consists of labelled positive and negative features. This data is trained on a Naive Bayes Classifier.

Then, we use **sentiment polarity** method of **TextBlob** class to get the polarity of tweet between -1 to 1.

Then, we classify polarity as:

```
if analysis.sentiment.polarity > 0:
    return 'positive'
elif analysis.sentiment.polarity == 0:
    return 'neutral'
else:
    return 'negative'
```

- Finally, parsed tweets are returned. Then, we can do various type of statistical analysis on the tweets. For example, in above program, we tried to find the percentage of positive, negative and neutral tweets about a query.

4.6 Matplotlib

Matplotlib is a plotting library for python programming language. Matplotlib is most used python package for 2-D graphics. It provides both a very quick way to visualize data from Python and publication-quality figures in many formats.

Matplotlib comes with a set of default settings that allow customizing all kinds of properties. You can control the defaults of almost every property in matplotlib: figure size and dpi, line width, color and style, axes, axis and grid properties, text and font properties and so on. While

matplotlib defaults are rather good in most cases, you may want to modify some properties for specific cases. [14]

5. Result and Discussions

Result 1: When the code is executed, it asks for keyword/tag to search in Tweets for categorizing the tweets into positive, negative and neutral.

```
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32)
[MSC v.1900 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

In [1]: runfile('E:/college/Term_02/Raed/2033/
Project_Sentiment_Analysis/Main_dev.py', wdir='E:/college/
Term_02/Raed/2033/Project_Sentiment_Analysis')
Connection is Successful!!!

Enter Keyword/Tag to search about:
```

Result 2: We need to enter the keyword. For example: Data Analytics

```
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32)
[MSC v.1900 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

In [1]: runfile('E:/college/Term_02/Raed/2033/
Project_Sentiment_Analysis/Main_dev.py', wdir='E:/college/
Term_02/Raed/2033/Project_Sentiment_Analysis')
Connection is Successful!!!

Enter Keyword/Tag to search about: Data Analytics
```

Result 3: After entering the keyword, we need to enter the number of tweets on which keyword will be searched.

```
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32)
[MSC v.1900 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

In [1]: runfile('E:/college/Term_02/Raed/2033/
Project_Sentiment_Analysis/Main_dev.py', wdir='E:/college/
Term_02/Raed/2033/Project_Sentiment_Analysis')
Connection is Successful!!!

Enter Keyword/Tag to search about: Data Analytics

Enter how many tweets to search: 1500
```

Result 4-1: Percentage of Positive, Negative and Neutral tweets are displayed.

```
Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32)
[MSC v.1900 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

In [1]: runfile('E:/college/Term_02/Raed/2033/
Project_Sentiment_Analysis/Main_dev.py', wdir='E:/college/
Term_02/Raed/2033/Project_Sentiment_Analysis')
Connection is Successful!!!

Enter Keyword/Tag to search about: Data Analytics

Enter how many tweets to search: 1500
Positive tweets percentage is: 75.00 %
Negetive tweets percentege is: 8.33 %
Neutral tweets percentage is: 16.67 %
```

Result 4-2: Five positive tweets are displayed on the output console as follows:

```
Positive tweets:

RT @Harry_Robots: #AI is “available anytime, polite and faster”.
#ArtificialIntelligence &gt;&gt; #Infographics &gt;&gt; via
@adrian_drones @ravikikan…
RT @INM7_ISN: Looking for two PhD students for exciting new
projects combining multi-modal imaging and machine-learning.

Join is to combi…
RT @Ronald_vanLoon: 12 #Python Resources for #DataScience
by @granvilleDSC @DataScienceCtrl |

Read more at https://t.co/MxaXSUXaBo

#BigDa…
RT @JCullenNow: There are some great new features in #ArcGIS
Business Analyst like the new population by generation #data
update. I did a q…
RT @Zac_Urback: For the 3rd straight year the @OHLSteelheads are
looking to expand our analytics department! If you are
interested in data…
RT @Sophia_May_00: #AI is “available anytime, polite and
faster”. #ArtificialIntelligence &gt;&gt; #Infographics &gt;&gt;
via @adrian_drones @ravikikan…
RT @DataAnalytics_1: #AI is “available anytime, polite and
faster”. #ArtificialIntelligence &gt;&gt; #Infographics &gt;&gt;
via @adrian_drones @ravikik…
RT @Ronald_vanLoon: #BI Success Beyond #DataVisualization
by @CIOwhitepapersreview @Sisense |

Read more at https://t.co/UozW0o8rBx

#DataV…
Get Smart with Pillow’s Data, Visibility and Control Tools
https://t.co/gRSvPTWsuQ #Analytics #data https://t.co/6zr4HuiLq3
```

Result 4-3: Neutral tweets are displayed. If tweets are less than 5 than all of them will be displayed on output console.

```
Neutral tweets:

RT @BerniceRogowitz: Needed: teaching "associate"  for a Vis
course in the Columbia University Applied Analytics Program this
fall. I wrote…
Business Analytics News: Data Analytics Trends - Blog - Stay
Informed | MicroStrategy https://t.co/vH8y4lQBNo, see… https://
t.co/yMUagLRfP9
```

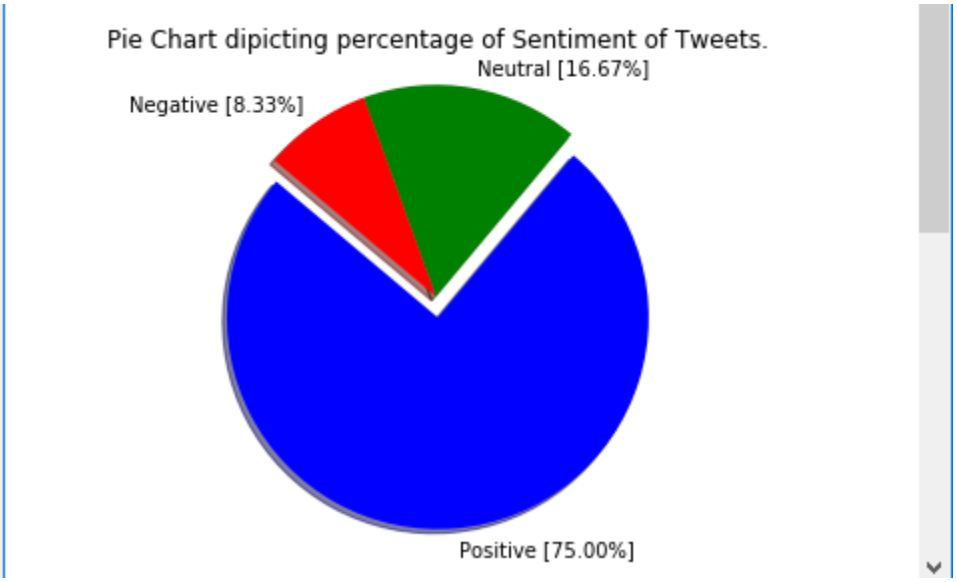
Result 4-4: Negative tweets are displayed on the output console.

```
Negative tweets:

RT @ashwinmegha: Big Data: Numbers mean nothing without context:
https://t.co/bQPrBytsWo

#SoftwareTesting #news #SoftwareDevelopment #outs…
```

Result 4-5: Pie chart is also displayed on the output console. Blue part depicts Positive tweets, Red depicts Negative tweets and Green part depicts Neutral tweets.



6. Conclusion

Twitter supposition investigation goes under the class of content furthermore, supposition mining. It centers around examining the suppositions of the tweets and encouraging the information to a machine learning model in request to prepare it and afterward check its precision, with the goal that we can utilize this model for later use as indicated by the outcomes. It contains of steps like information accumulation, content pre-preparing, conclusion identification, assumption characterization, preparing and testing the show. This exploration theme has developed amid the most recent decade with models achieving the proficiency of right around 85%-90%. Yet despite everything it does not have the measurement of decent variety in the information.

Alongside this it has a ton of utilization issues with the slang utilized and the short types of words. Numerous analyzers don't perform well at the point when the quantity of classes is expanded. Likewise, it's still not tried that how exact the model will be for points other than the one in thought. Subsequently supposition investigation has a bright extent of improvement in future.

In our project ***Twitter Sentimental Analysis,*** the stakeholder could be organizations looking for reviews of any product, hospitals for hospital feedback and review, store owners for store feedback and review, Hotel owners for hotel feedback and scope of improvement, Media industry for providing media reports about trending topics and its statistics and our team members including project leader and manager.

Reference

[1] http://www.ijcee.org/vol8/931-IT015.pdf

[2] Bastos, M. T., Travitzki, R., & Puschmann, C. (2012). What sticks with whom? Twitter follower- followee networks and news classification. Proceedings of 3th International AAAI Conference on Weblogs and Social Media—Workshop on the Potential of Social Media Tools and Data for Journalists in the News Media Industry

[3.] Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on twitter. Proceedings of the 20th International Conference on World Wide Web. ACM New York, NY, USA.

[4] Prajapati, V. (2013). Big Data Analytics with R and Hadoop. Packet Publishing.

[5] Bastos, M. T., Travitzki, R., & Raimundo, R. (2012). Tweeting political dissent: Retweets as pamphlets in #FreeIran, #FreeVenzuela, #Jan25, #SpanishRevolution and #OccupyWallSt. University of Oxford.

[6] http://ipullrank.com/step-step-twitter-sentiment-analysis-visualizing-united-airlines-pr-crisis/

[7]https://shachihkumar.files.wordpress.com/2014/02/twitter-sentiment-analysis-project-report.pdf

[8] http://dspace.thapar.edu:8080/jspui/bitstream/10266/4273/4/4273.pdf

[9] http://www.letscodepro.com/Twitter-Sentiment-Analysis/

[10] https://www.nltk.org/

[11] https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

[12] https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/

[13] https://pdfs.semanticscholar.org/c114/7f3d9b46ff0a0c7c43b668123cb15a26120d.pdf

[14] https://www.labri.fr/perso/nrougier/teaching/matplotlib/

[15] https://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf

[16] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.