|  |
| --- |
| Table of contents |

1. What is sentiment analysis?
2. Major steps associated with sentiment analysis
3. What is the need for processes like sentiment analysis?
4. How sentiment analysis actually works?
5. What are the benefits of sentiment analysis?
6. How accurate and practical is sentiment analysis?
7. Challenges faced in sentiment analysis
8. Some prerequisites for sentiment analysis
9. Different steps required for sentiment analysis
10. Python codes along with sample testing

* Testing for code 1
* Testing for code 2

1. Decoding the codes
2. Applications for sentiment analysis
3. Conclusion
4. References
5. **What is sentiment analysis?**

Starting with a technical definition, we can see the term ‘Sentiment analysis’ as the procedure of 'computationally' deciding if a bit of writing is sure, negative or nonpartisan. It is often referred to as sentiment mining or opinion mining, inferring the feeling or frame of mind of a speaker. If we have to put everything into simple words, this can be described as the interpretation as well as classification of emotions (i.e. either positive, negative or neutral) within any textual data using different techniques of analysis.

The main goal of this procedure is to allow businesses to analyse the customer’s sentiment or emotion towards any product, service and even brand which can be done using online conversations as well as feedbacks. Understanding emotions of people or customers is quite essential in terms of businesses as each and every one of the customers can express his/her opinion, view, feeling and thought more openly. In order to meet customers’ needs certain methods like survey responses and conversations on forums and social media are used to analyse the consumers’ feedback and that too automatically. Individuals share learning, encounters and contemplations with the world by utilizing social media like online journals, discussions, wiki, audit locales, informal organizations, tweets, etc.

This whole procedure of analysing the textual data and implying its meaning can be understood by taking a simple example. If we talk about ourselves i.e. human beings, we can classify any piece of text or writing or conversation as positive, negative or neutral. Like if we say ‘That movie was pretty good’ or ‘He is not a good-natured person’ then we are inferring something positive, negative or neutral and same is inferred by the individual who’s listening. Now how this is done? This is done by extracting and processing the words which were used during the conversation, mainly important words which imply that this thing which was said is positive or the same thing which was said is negative. After finding such words and decoding the average interpretation of that sentence gives us an overall sentiment of that particular conversation. The same thing can be applied to an application which uses this technique to decode what actually any writing or text is trying to say. For example, any consumer checking out a product on Amazon used sentiment analysis in order to automatically imply and analyse more than 2500 reviews and ratings which led him to the conclusion that many consumers were pretty happy about the service of that particular product but many also complained about the pricing and quality of the same product.

Sentiment analysis has changed the way wherein individuals impart and impact social, political and financial conduct of other individuals on the web. With the blast of client created feelings there is the need by organizations, government officials, specialist co-ops, social analysts, scientists and different on-screen characters to examine them so as to execute better choice decisions. We’ll be looking at different forms and models of sentiment analysis that focus on identifying intentions as well as detecting polarity (positive, negative and neutral) and feelings as we proceed further into this report. Some of the most popular kinds of sentiment analysis which are worth mentioning are fine-grained sentiment analysis, emotion detection, aspect-based sentiment analysis and multilingual sentiment analysis. Our main objective of the report is to fetch and parse tweets from twitter and analyse its sentiment along with the polarity, subjectivity and graph with the use of python programming language.

1. **Major steps associated with sentiment analysis**

The process of sentiment analysis involves 5 different steps to analyse data. These steps are listed below:

* Collection of data
* Preparation of text
* Detection of sentiment
* Classification of sentiment
* Output presentation.

Now let us dive deep into these mentioned steps.

1. **Collection of data**: The very basic step of sentiment analysis is to gather some information from the content generated by user which is contained in blogs, forums and even social networks. This type of data is disorganized and communicated in various ways by utilizing various vocabularies, slangs, setting of composing and so on. Manual investigation is inconceivable. In this manner, content examination and common language preparing are utilized to separate and order.
2. **Preparation of text:** consists in cleaning the removed information before examination. Non-literary substance and substance that are unessential for the investigation are recognized and killed.
3. **Detection of sentiment:** the extricated text of the audits and conclusions are examined. Sentences with abstract articulations (conclusions, convictions and perspectives) are held and sentences with target correspondence (realities, truthful data) are disposed of.
4. **Classification of sentiment:** in this progression, emotional sentences are arranged in positive, negative, great, awful; like, detest, however order can be made by utilizing different focuses.
5. **Output presentation:** the fundamental target of sentiment analysis is to change over unstructured content into some significant or meaningful data. At the point when the investigation is done, the content outcomes are shown on diagrams like pie outline, bar diagram and line charts. Likewise, time can be examined and can be graphically shown developing an estimation course of events with the picked worth (recurrence, rates and midpoints) after some time.
6. **What is the need for processes like sentiment analysis?**

* **In the world of business:** In promoting field organizations use it to build up their methodologies, to comprehend clients' sentiments towards items or brand, how individuals react to their crusades or item dispatches and why customers don't get a few items.
* **Various political needs:** In political field, it is utilized to monitor political view, to recognize consistency and irregularity among articulations and activities at the administration level. It tends to be utilized to foresee race results too!
* **General public actions:** Sentiment analysis is likewise used to screen and break down social marvels, for the spotting of possibly perilous circumstances and deciding the general state of mind of the blogosphere.

1. **How sentiment analysis actually works?**

There are certain number of methods and algorithms which are directly associated with Natural Language Processing (NLP) and NLP being the most crucial part of sentiment analysis has mainly three types of algorithms which we’ll be using i.e. automatic, hybrid and rule-based approach.

* Automatic approach includes systems which learn about the data from the technique of machine learning and completely rely on them.
* Rule-based approach includes systems which use a set of manually crafted rules and protocols in order to perform the sentiment analysis process.
* Hybrid approach basically combines both the above-mentioned approaches.

Sentiment analysis is another field of research conceived in Natural Language Processing (NLP), targeting recognizing subjectivity in content and additionally extricating and grouping assessments and opinions. Supposition investigation concentrates individuals' estimations, sentiments, dispositions, assessments, evaluations and feelings towards administrations, items, people, associations, issues, points, occasions and their qualities.

In sentiment analysis content is arranged by the accompanying various criteria:

* the extremity of the notion communicated (into positive, negative and unbiased)
* the extremity of the result (for example improvement versus passing in restorative writings)
* agree or can't help contradicting a point (for example political discussions)
* good or awful news
* support or resistance
* pros and cons

For the sentiment analysis usage distinctive conclusion order approaches and devices are utilized.

1. **What are the benefits of sentiment analysis?**

In a survey it was found that, if we talk about the world’s data, 80 percent is unorganized or unstructured, if we put into other words. Since there is a massive creation of data, mostly text, in the form of conversations, emails, chats, articles, documents etc, it turns into a challenging task to analyse and understand each and every data, each and every day, which also is a time-consuming as well as expensive process. Whenever we talk about this unstructured data, sentiment analysis plays a very crucial role in making that data organized and that too by using automatic tags.

Some of the main benefits of sentiment analysis include:

* No one can imagine processing that unorganized data manually and sorting that data one by one in order for that data to make sense. Sentiment analysis just comes to the rescue whenever we talk about a very huge amount of data that is to be converted into something sensible and that too in an efficient and cost-effective way.
* Sentiment analysis allows us to deal with real-time scenarios and situations using various analysis models with which we can easily predict (say) any social media trend which is important or an angry customer who is about to stop using some services or products of any business.
* Although consumers don’t agree with each and everything all the time because of different personal experiences, beliefs, feelings and thoughts, companies can still use a sentiment analysis system which is centralized and can apply that scenario or criteria to all the data which’ll help in improving accuracy as well as gaining better insights.

1. **How accurate and practical is sentiment analysis?**

If we ask ourselves the same question that whether the process of sentiment analysis is feasible or practical or not, then the short answer is **yes**. This process is very practical and can be done by anyone with some prerequisites, which are:

* **Basic knowledge of programming:** This simply means that you need not to have any programming language-based knowledge like one must know **python** or one must know **java**. This knowledge which I’ve mentioned deals with the basic fundamentals of programming which one must always be aware of. Without these basics one would not be able to interpret the code which’ll generate our main output.
* **Some tools which must be installed onto the system:** There are certain tools which are required to proceed with this project. These tools are tweepy, textblob, nltk, python, twitter, and a twitter developer account. How to install these tools is also mentioned in this report.

1. **Challenges faced in sentiment analysis**

Let us take a look at some of the challenges which are faced by data scientists as well as computer scientists in the field of sentiment analysis:

1. **Emojis:** There are basically two types of emojis i.e. Eastern and Western emojis. Western emojis are the ones which contain only one or two encoded characters whereas Eastern emojis are the ones with a combination much longer than the western ones. We know that today, we often use emojis (basically both western and eastern) to express an emotion towards some textual data especially in tweets. In order to decode the expression of these emojis, a lot of pre-processing is needed so as to fully understand the nature of emoji. If we fail to analyse the emoji’s true nature, the polarity of the text might be different than the actual one.
2. **Comparisons:** Treating comparisons in texts is also a challenge in sentiment analysis. For e.g. This mouse is better than older one. In this sentence it’s clear that we can’t just classify it as positive, negative or even neutral.
3. **Irony:** We have often seen people tweeting with the use of positive words but at the same time inferring the negative sentiments which they are feeling and trying to express. Now these types of situations are nothing but challenges for the machines which find difficulty in detecting the expressed feeling without understanding the context of the given situation. For e.g. Did you enjoy your meal? (Replying in a sarcastic or ironical way) Yeah, sure. So delicious!
4. **Neutral:** Sometimes it is hard to classify as something neutral while performing a sentiment analysis operation. In order to get accurate results, the definition of neutral or neutrality is a challenge in itself. For e.g. I wish this smartphone had a better camera. This sentence in itself is difficult to categorize as it is including a comparison. However, ‘I wish the smartphone had more space’ generally comes under the neutral category. Clearly, we can see that defining something as neutral is a confusing task for the machine.
5. **Some prerequisites for sentiment analysis**

Before proceeding with the main sentiment analysis task, we are required to have some tools installed onto our system.

* **Tweepy:** It is a client in python which is used to access the twitter Application programming interface.

To ensure that tweepy is installed onto your system you can use:

**pip install tweepy**

* **TextBlob:** It is a library in python which is used to process texts.

To ensure that textblob is installed onto your system you can use:

**pip install textblob**

* **Twitter developer account:** We can fetch tweets from twitter only if we have a twitter developer account. We must first register and then create an app which will be used further in our main code which is to be written on python.
* **Python:** Python is the most important part of this process as our main code will be written in and run on python. Go to <https://www.python.org/downloads/> and download and install the latest version i.e. 3.7.4 from the website.

We also need to ensure that NLTK corpora is installed onto the system, which can be installed by using:

**python –m textblob.download\_corpora**

NLTK corpora is defined as a large set of texts which is also structured.

1. **Different steps required for sentiment analysis**

We follow these steps in our program:

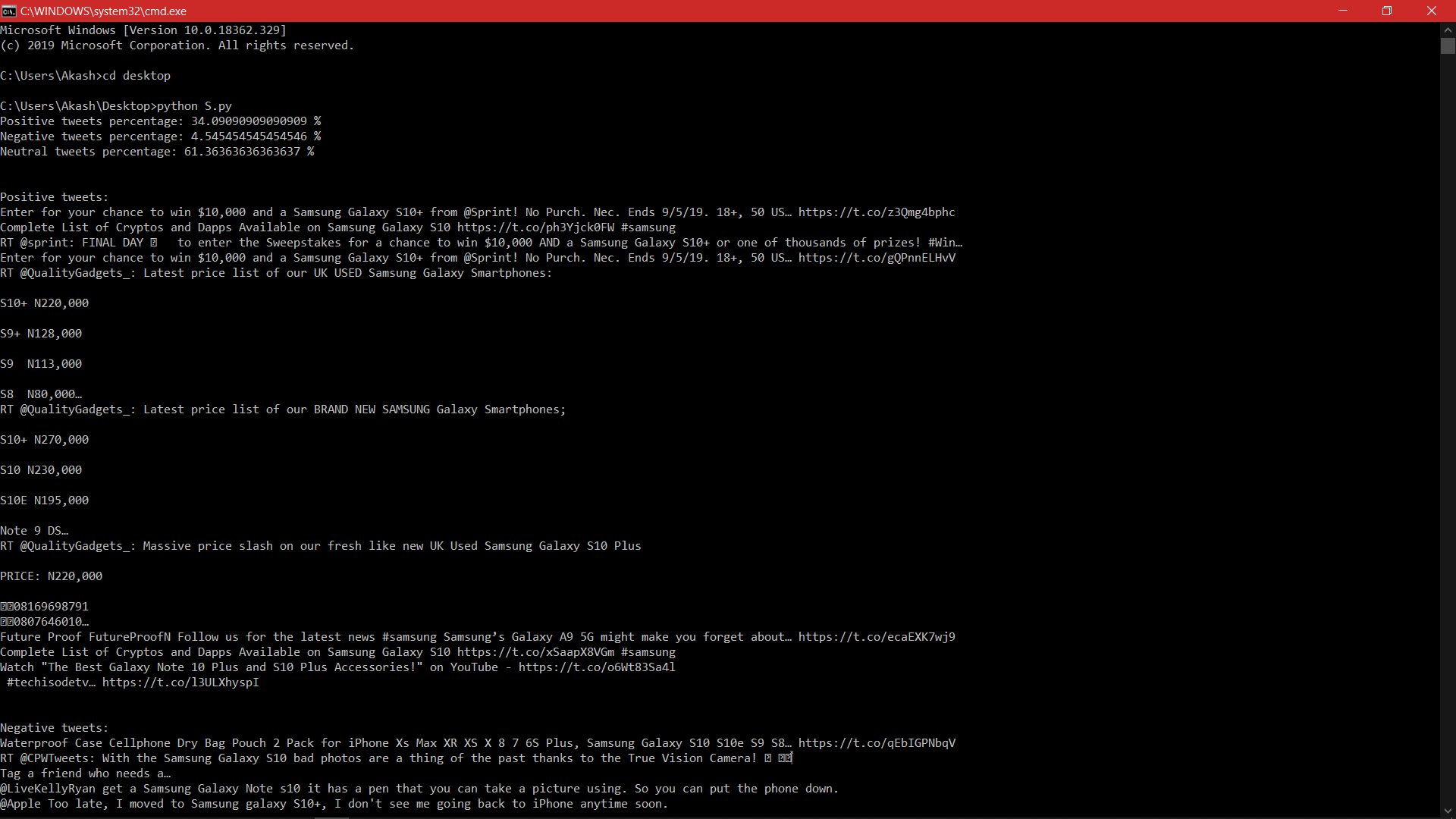
1. Open <https://developer.twitter.com/en/apps> and got to the ‘create a new app section’.
2. Fill in the application with all details.
3. After the creation of the app, just go to app page.
4. Go to the tab labelled as ‘keys and access tokens’.
5. Now we need certain values like both the consumer keys i.e. consumer and consumer secret key and both the access tokens i.e. access token and access token secret. These keys will be used and embedded into our python code.
6. Generate first the consumer key and then also the consumer secret key within the twitter app.
7. Copy both the keys to your main python code in order to make the code recognize the tweets which are to be parsed.
8. Run the code and it’ll examine the tweets and put it into the category of either positive, negative or neutral.
9. **Python codes along with sample testing**

**Testing for code 1**

We can provide anything as a search query like any movie or any smartphone and get an output with some positive, negative and neutral sentiments about that search query or basically a sort of ‘review’.

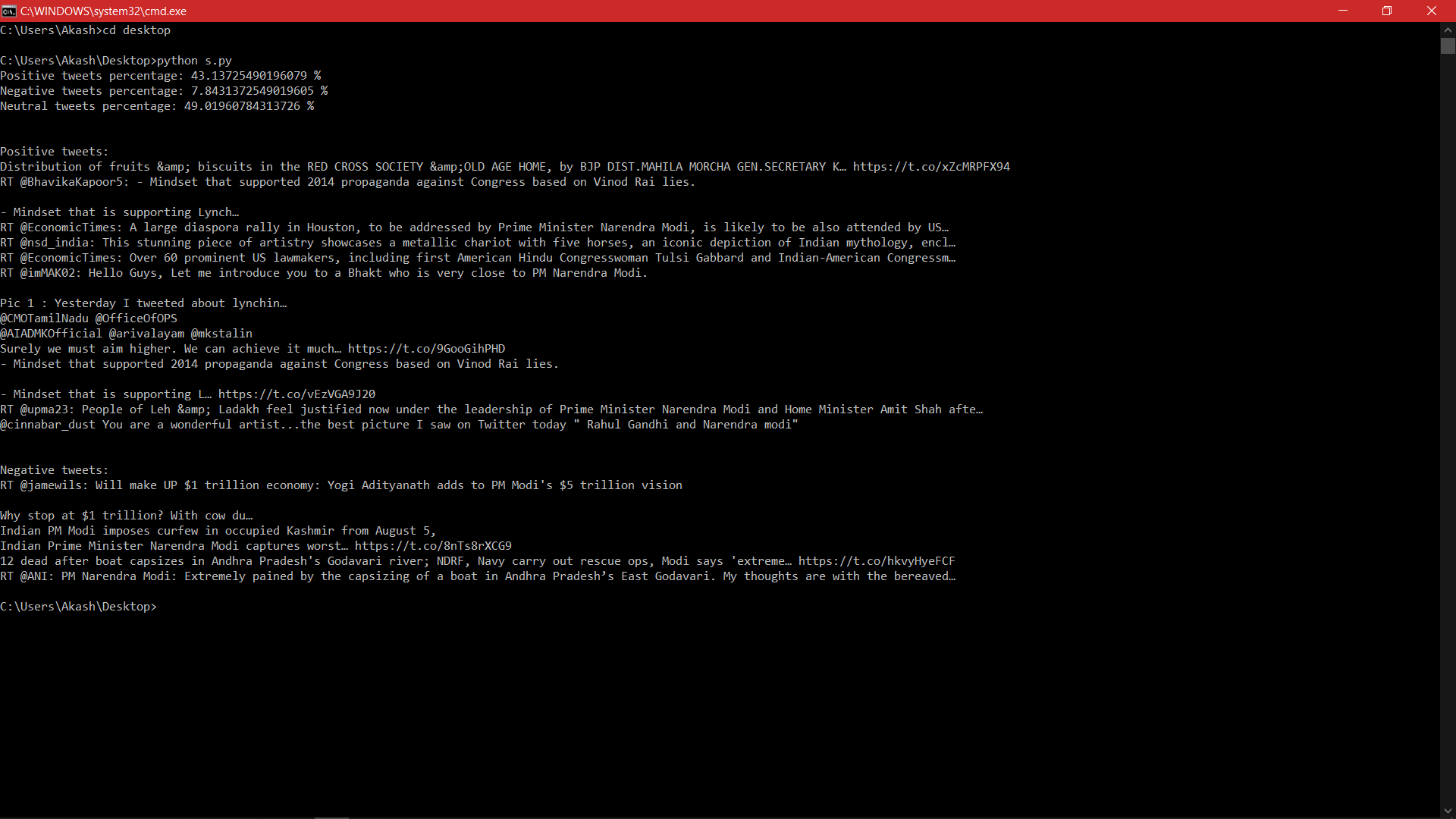
1. **Sample input I**

Below is the generated output which took ‘Samsung galaxy S10’ as the search query and generated some tweets with positive reviews and negative reviews for the same.



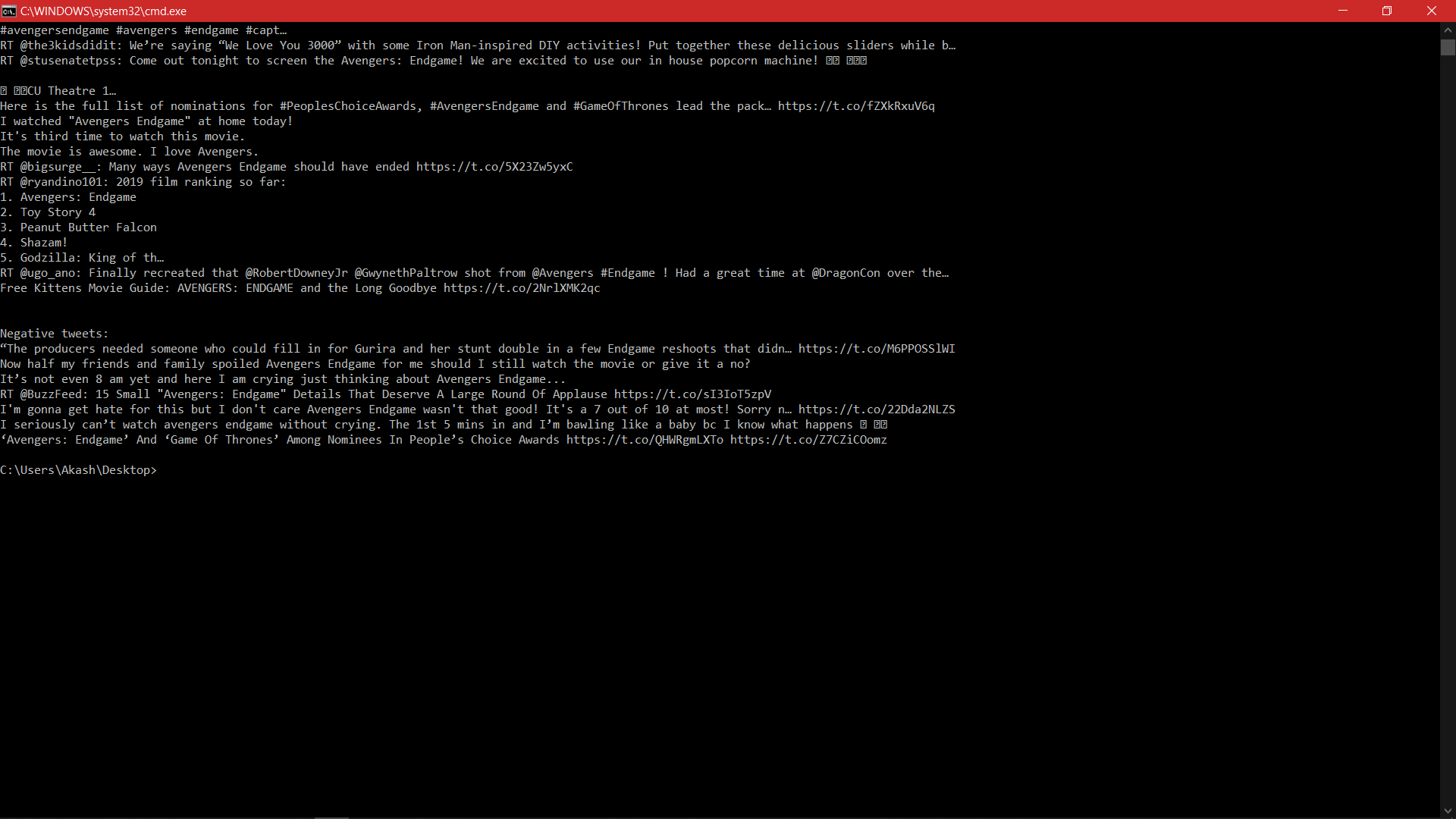
1. **Sample input II**

Below is the generated output which took ‘Narendra Modi’ as the search query and generated some tweets with positive reviews and negative reviews for the same.



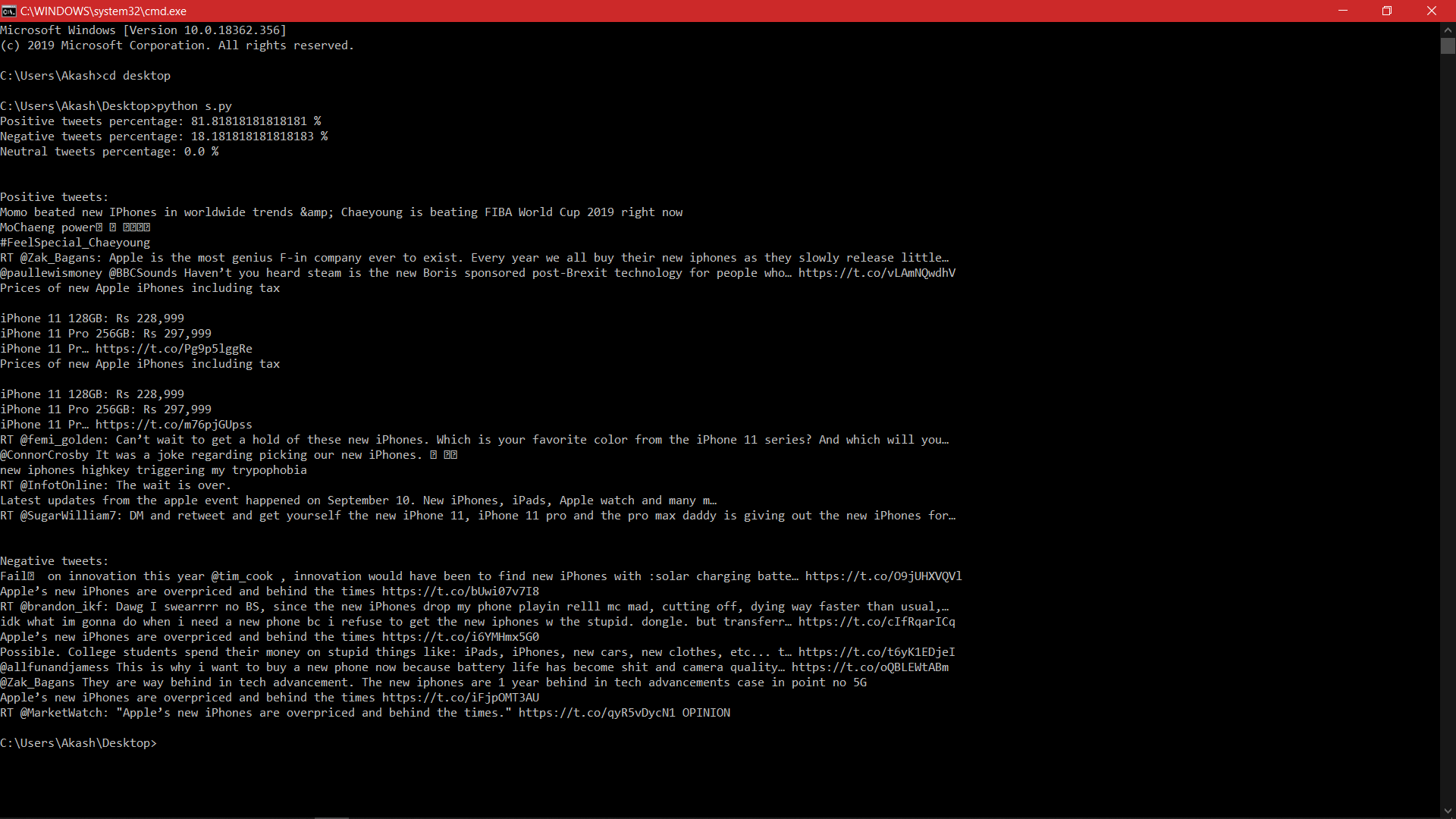
1. **Sample input III**

Below is the generated output which took ‘Avengers Endgame’ as the search query and generated some tweets with positive reviews and negative reviews for the same.



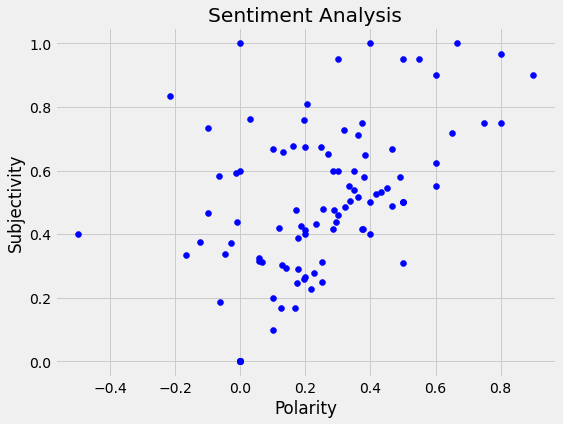
1. **Sample input IV**

Below is the generated output which took ‘New iPhones’ as the search query and generated some tweets with positive reviews and negative reviews for the same.

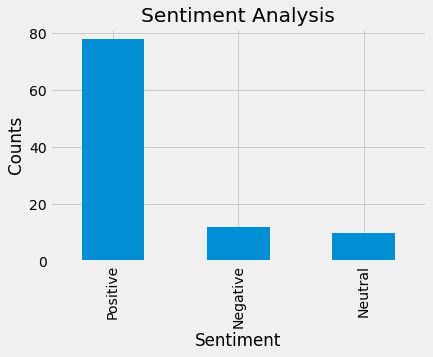


**Testing for code 2**

Below is the **Subjectivity vs Polarity** plotting.



Below is the **Bar graph** of the counts of sentiments, which shows total percentage of positive, negative and neutral tweets.



1. **Decoding the codes**

The code written in python can be interpreted like this:

* First, we make a class called ‘twitterclient’. This class will contain every one of the strategies to associate with API and reading and examining the tweets. We us \_init\_ capacity to deal with the confirmation of API customer.
* After getting the tweet we use the textblob module.
* TextBlob is a high-level state library worked over top of NLTK library. First, we call clean\_tweet technique to expel joins, exceptional characters, and so forth. From the tweet utilizing some straightforward regex.
* At that point, as we pass the tweet to make an object of textblob, following preparing is done over content by the textblob library.
* Break the tweet into tokens, for example split the words from assortment of content.
* Separate the stopwords and the tokens. (stopwords are the words which are unessential in content examination as am I, you, are and so forth.)
* Do grammatical feature labeling of the tokens and select just huge highlights/tokens like modifiers, intensifiers and so forth.
* The tokens are then passed to a notion classifier which groups the tweet as positive, neutral or negative by relegating it an extremity between - 1.0 to1.0.

The means by which sentiment classifier is made:

* Textblob utilizes a motion picture audits dataset in which surveys have just been named as positive or negative.
* Positive highlights as well as negative highlights are separated from every positive audit and negative audit individually.
* Training information currently comprises of named positive highlights and negative highlights.

At that point, we arrange extremity as:

in the event that polarity is greater than 0:

return as positive

or if the polarity is equal to 0:

return as unbiased

otherwise

return as negative

* Tweets after getting parsed are finally returned. At that point, we can do different sort of measurable examination on the tweets i.e. in above code, we attempted to discover the level of positive tweets, negative tweets and unbiased tweets about an inquiry.

1. **Applications of sentiment analysis**

Sentiment analysis has a variety of applications in different fields, some of the most important of them are listed below:

* **Social media monitoring:** This is basically a way in which business are using sentiment analysis technique for keeping a track of what the customers are actually saying and talking about.
* **Managing a crisis better:** Brands listen to people and monitor all their opinions, identify what happened and why, review and respond accordingly mainly on their social media channels.
* **Providing better product analytics:** Companies use sentiment analysis to keep an eye on what is working for them and what is not.
* **Improving customer support:** Being able to track that ONE bad customer review or response that can cost a business again and again.
* **Keeping an eye on your competition:** What the customers are thinking about your brand when compared to the brand you are competing with can also help the companies improve.
* **Assuring the quality of products:** Mostly companies come to know about the errors or faults in their products after launching them in the market. Sentiment analysis helps them in detecting those errors based on actual user experience.
* **In politics:** We can easily determine what are the views of the people regarding certain situation or what are they happy or angry for.

1. **Conclusion**

This draft or research paper introduced us to the field of sentiment analysis or opinion mining along with various other aspects of it. First, we came to know what **Sentiment Analysis** actually is with a proper introduction in the beginning of this draft and what are the major steps involved in the process of opinion mining. Later on, we came to know the actual need of processes like this in our daily life especially in the cases of businesses and marketing. Through that portion it was clear to us that the need for sentiment analysis is real and it can have a huge impact on various number of fields like industries, businesses and even brands. Now that we are dealt with the introduction and need part of the project, covering the working part of this process was also important. As we know that, this particular research paper deals with the working of **Sentiment analysis** of **Twitter posts** using **Python** and discussing about the tools required for it was just as necessary. The whole process is explained in detail in this draft so that the people can actually implement the procedure by themselves, obviously on shorter scale.

As the basics were covered properly, the prominent benefits as well as challenges were also to be mentioned. It is very simple to deduce that this technique is pretty challenging technically as well as practically, although it has some benefits too. That being said, we all are aware of the fact that each and every industry and business wants to know how well their products are performing in the market and how the consumers are perceiving their products as well as services and how it is different than those of their competitors. The same thing can be said about the customers who are using the products and services because every consumer wants to know what they are about to invest in by knowing the opinions of existing users. Just because of these benefits, technical challenges and practical needs, the field will be as important as of now in the coming future.

Talking about what we have done so far, we need to investigate more on the topic and I believe that these in-depth investigations will eventually allow us to build such systems which are integrated. Now, these integrated systems will try to solve all the collective sub-problems together so as to help us reach to a solution of each individual sub-problem. We just have to be optimistic that this whole situation will meet its conclusion in the coming future and this process will have a huge number of applications.

Now coming back to the main aim of this project which was to discuss how we can analyse the sentiments of various twitter posts using twitter API with the working based on python language. Sentiment analysis can be done in a number of ways but, in this paper, I have particularly focused on the method using python, tweepy, textblob and twitter. I’ve discussed in detail from **‘what is sentiment analysis’** to **‘why do we need it’** and which fields can implement the process and use this type of method in today’s cloud-based architecture and obviously, how it is changing our day to day life.

1. **References**

<https://developer.twitter.com/en/apps>

<https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

<http://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf>

<https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis>

<https://textblob.readthedocs.io/en/dev/_modules/textblob/en/sentiments.html>

<https://monkeylearn.com/sentiment-analysis/>

<https://www.wonderflow.co/blog/sentiment-analysis-examples>

<https://www.growthaccelerationpartners.com/tech/sentiment-analysis/>

<https://www.quora.com/What-are-the-applications-of-sentiment-analysis-Why-is-it-in-so-much-discussion-and-demand>