



**L**OVELY  
**P**ROFESSIONAL  
**U**NIVERSITY

**SIX WEEKS SUMMER TRAINING REPORT**

On

**DATA SCIENCE**

Submitted by

**Arjun Rao Rachakonda**

**11903860**

**Computer Science & Engineering (CSE)**

Under the Guidance of

**Naman Agarwal**

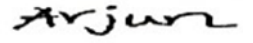
**School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

(June-July, 2021)

## DECLARATION

I hereby declare that I have completed my six weeks summer training at **Skill Vertex, Bengaluru** from 7<sup>th</sup> June, 2021 to 30<sup>th</sup> July, 2021 under the guidance of **Naman Agarwal**. I have declare that I have worked with full dedication during these six weeks of training and my learning outcomes fulfil the requirements of training for the award of degree of Bachelor of Technology (B.Tech), Lovely Professional University, Phagwara.



Arjun Rao Rachakonda

Registration No.: 11903860

Date: 30<sup>th</sup> July, 2021

## **ACKNOWLEDGEMENT**

It is a matter of great pleasure to present this report on development of “Sentiment Analysis Model”. I’m grateful to Lovely Professional University and Skill Vertex for providing me with this great opportunity to develop an application for the project.

I am also grateful to our Mentor Naman Agarwal for his valuable lecture’s on Data Science and suggestions.

### **About the Project**

Analysis of public information from social media could yield interesting results and insights into the world of public opinions about almost any product, service, or personality. Social network data is one of the most effective and accurate indicators of public sentiment. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result, there has been an eruption of interest in people to mine these vast resources of data for opinions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments, and subjectivity of text. In this project report we will be discussing a methodology which allows utilization and interpretation of twitter data to determine public opinions. Developing a program for sentiment analysis is an approach to be used to computationally measure customers' perceptions. This paper reports on the design of a sentiment analysis, extracting and training a vast number of tweets. Results classify customers' perspective via tweets into positive and negative, which is represented in a pie chart, bar diagram, scatter plot using matplotlib.



**ARJUN RAO RACHAKONDA**

**DATA SCIENCE INTERNSHIP**

**CERTIFICATE OF COMPLETION**

**FROM 7<sup>TH</sup> JULY 2021 TO 7<sup>TH</sup> AUGUST 2021**

Over the period of July 2021, intern has successfully completed the program with:

- Fine adaptability
- Effective communication
- Strong presentation skills
- Excellent efficiency

Student UIN - RA1711338



Certificate ID- 2000338

  
Shivam Sharma  
Vice President  
Human Resources

## **CERTIFICATE OF TRAINING COMPLETION**

PRESENTED TO

**ARJUN RAO RACHAKONDA**

FOR SUCCESSFULLY COMPLETING DATA SCIENCE PROGRAM

FROM 7<sup>th</sup> June 2021 to 7<sup>th</sup> July 2021

During the course we found student to be consistent & determined candidate.



Student UIN- RA1711338



Certificate No. - 2000338



**Shivam Sharma**  
VICE PRESIDENT  
HUMAN RESOURCE

# Table of Contents

<b>S.No</b>	<b>Content</b>	<b>Page No.</b>
<b>1.</b>	Introduction	6
<b>2.</b>	Technology Learnt	7
<b>3.</b>	Reason for choosing this technology	9
<b>4.</b>	Profile of the Problem	11
<b>5.</b>	Existing System	14
<b>6.</b>	Problem Analysis <ul style="list-style-type: none"><li>• Product definition</li><li>• Feasibility Analysis</li></ul>	18
<b>7.</b>	Software Requirement Analysis	22
<b>8.</b>	Design <ul style="list-style-type: none"><li>• Tables and their relationships</li><li>• Flowcharts/Pseudo code</li></ul>	23
<b>9.</b>	Implementation	25
<b>10.</b>	Learning Outcome from training/technology learnt	28
<b>11.</b>	Gantt chart	29
<b>12.</b>	Project Legacy <ul style="list-style-type: none"><li>• Technical and Managerial lessons learnt</li></ul>	30
<b>13.</b>	Bibliography	31

# INTRODUCTION

**Skill vertex** is a self-funded start up and a sophisticated Edu-tech start up that provides courses and placement programs to the Graduating Students. It is an e-learning platform with a holistic environment to upskill students for Industry. Recognizing Skill Vertex as the avenue to provide mentor-based training and internship programs directly from the oracles of the industry.

The Company has a vision to empower and upskill the youth of the future with absolute requisites and stance to seize better employment opportunities and a mission to endow our mentees with skills in compliance with the emerging market requirements to become industry ready.

Skill Vertex is an innovative organization with a goal to impart the aspiring learners with rigorous training and appropriate exposure for a promising future. they believe that pragmatic abilities weigh more over the educational degree and strive forward to provide an existence where youth can find their energy and transform it into vocation. With a vision of a world loaded with opportunities and conceivable outcomes, they help youth to build a completely certain and assured foundation to stake guarantee on their skills for a brilliant future that awaits.

Skill Vertex offers many Programs Such as:

- Training & Internship Programs
- Crash Courses
- Placement Assurance Programs

Skill Vertex provide Training & Internship opportunities for Students Studying in different Streams such as: Computer Science, Management, Civil, Mechanical Engineering, Electronics & Communication.

Skill Vertex provides e-learning platform for many other Courses Such as: C Programming, C++ Programming, Python, Java Programming.

Skill Vertex Offers few Placement assurance programs such as Digital Marketing, Full Stack Web Development.

# TECHNOLOGY LEARNT

## **Data Science**

Data Science is a blend of mathematics, business acumen, tools, algorithms and machine learning techniques all of which help us in finding out the hidden insights or patterns from raw data which can be of major use in the formation of big business decisions.

- Natural language processing (NLP)
- Understanding clean data
- Big data analytics
- Gained knowledge of different data science models
- Data filtration through python-based algorithms

## **OpenCV**

OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision. The library is cross-platform and free for use under the open-source Apache 2 License, OpenCV features GPU acceleration for real-time operations.

OpenCV is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human. When it is integrated with various libraries, such as NumPy which is a highly optimized library for numerical operations, then the number of weapons increases in your Arsenal i.e.; whatever operations one can do in NumPy can be combined with OpenCV.

- Face Detection Using Haar Cascade.
- Edge Detection
- Corner Detection
- Feature Matching
- Template Matching
- Background Reduction
- Foreground Extraction
- Motion Detection

## **Sentiment Analysis**

Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. However, analysis of social media streams is usually restricted to just basic sentiment analysis and count based metrics. This is related to just scratching the surface and missing out on those high value insights that are waiting to be discovered.

With the recent advances in deep learning, the ability of algorithms to analyse text has improved considerably. Creative use of advanced artificial intelligence techniques can be an effective tool for doing in-depth research. I believe it is important to classify incoming customer conversation about a brand based on following lines:

- Key aspects of a brand's product and service that customers care about.
- Users' underlying intentions and reactions concerning those aspects.

These basic concepts when used in combination, become a very important tool for analysing millions of brand conversations with human level accuracy.

### **Text Classifier:**

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative or neutral.

Similarly, In the Project I have used this text classification tool that analyses the comments on tweets by a specific user in Twitter by their username and then it displays a pie chart with underlying sentiment % of positive, negative, Weakly Positive, Strongly Positive, Weakly Negative, Strongly Negative or neutral comments on the tweets after Analysis.



## **REASON FOR CHOOSING THIS TECHNOLOGY**

Data Science has become a revolutionary technology that everyone seems to talk about, While I wish to become Data Scientist. So, I Choose this technology.

Data Science is the study of data. It is about extracting, analysing, visualizing, managing, and storing data to create insights. These insights help the companies to make powerful data-driven decisions. Data Science requires the usage of both unstructured and structured data. It is a multidisciplinary field that has its roots in statistics, maths, and computer science. It is one of the most highly sought-after jobs due to the abundance of data science position and a lucrative pay-scale.

The field of Data Science is massive and has its own fair share of advantages.

### **1. It's in Demand**

Data Science is greatly in demand. Prospective job seekers have numerous opportunities. It is the fastest growing job on Linked in and is predicted to create 11.5 million jobs by 2026. This makes Data Science a highly employable job sector.

### **2. Abundance of Positions**

There are very few people who have the required skill set to become a complete Data Scientist. This makes Data Science less saturated as compared with other IT sectors. Therefore, Data Science is a vastly abundant field and has a lot of opportunities. The field of Data Science is high in demand but low in supply of Data Scientists.

### **3. A Highly Paid Career**

Data Science is one of the most highly paid jobs. According to Glassdoor, Data Scientists make an average of \$116,100 per year. This makes Data Science a highly lucrative career option.

### **4. Data Science is Versatile**

There are numerous applications of Data Science. It is widely used in health care, banking, consultancy services, and e-commerce industries. Data Science is a very versatile field. Therefore, I would get the opportunity to work in various fields.

## **5. Data Science Makes Data Better**

Companies require skilled Data Scientists to process and analyse their data. They not only analyse the data but also improve its quality. Therefore, Data Science deals with enriching data and making it better for their company.

## **6. Data Scientists are Highly Prestigious**

Data Scientists allow companies to make smarter business decisions. Companies rely on Data Scientists and use their expertise to provide better results to their clients. This gives Data Scientists an important position in the company.

## **7. No More Boring Tasks**

Data Science has helped various industries to automate redundant tasks. Companies are using historical data to train machines in order to perform repetitive tasks. This has simplified the arduous jobs undertaken by humans before.

## **8. Data Science Makes Products Smarter**

Data Science involves the usage of Machine Learning which has enabled industries to create better products tailored specifically for customer experiences. For example, Recommendation Systems used by e-commerce websites provide personalized insights to users based on their historical purchases. This has enabled computers to understand human behaviour and make data driven decisions.

## **9. Data Science can Save Lives**

Healthcare sector has been greatly improved because of Data Science. With the advent of machine learning, it has been made easier to detect early-stage tumours. Also, many other health-care industries are using Data Science to help clients.

## PROFILE OF THE PROBLEM

**Sentiment analysis** (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative, or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback and understand customer needs.

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically analysing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customer's needs.

For example, using sentiment analysis to automatically analyse 4,000+ reviews about product could help us discover if customers are happy about pricing plans and customer service.

If we want to gauge brand sentiment on social media, in real time and over time, so we can detect disgruntled customers immediately and respond as soon as possible.

### **Types of Sentiment Analysis**

Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc), urgency (urgent, not urgent) and even intentions (interested or not interested).

Depending on how we want to interpret customer feedback and queries, we can define and tailor our categories to meet our sentiment analysis needs.

## **1. Fine-grained Sentiment Analysis**

polarity precision is important to analyse, we can consider expanding our polarity categories to include:

- Positive
- Weakly Positive
- Strongly Positive
- Neutral
- Negative
- Weakly Negative
- Strongly Negative

## **2. Emotion detection**

This type of sentiment analysis aims to detect emotions, like happiness, frustration, anger, sadness, and so on. Many emotion detection systems use lexicons (i.e. lists of words and the emotions they convey) or complex machine learning algorithms.

One of the downsides of using lexicons is that people express emotions in different ways. Some words that typically express anger, like bad or kill might also express happiness.

## **3. Aspect-based Sentiment Analysis**

Usually, when analysing sentiments of texts, let's say product reviews, we want to know which particular aspects or features people are mentioning in a positive, neutral, or negative way. That's where aspect-based sentiment analysis can help, for example ; "The battery life of this camera is too short", an aspect-based classifier would be able to determine that the sentence expresses a negative opinion about the feature battery life.

## **4. Multilingual sentiment analysis**

Multilingual sentiment analysis can be difficult. It involves a lot of pre-processing and resources. Most of these resources are available online (e.g. sentiment lexicons), while others need to be created (e.g. translated corpora or noise detection algorithms), but we need to know how to code to use them.

Alternatively, we could detect language in texts automatically with Monkey Learns language classifier, then train a custom sentiment analysis model to classify texts in the language of your choice.

## **Why is Sentiment Analysis Important?**

Sentiment analysis is extremely important because it allows businesses to understand the sentiment of their customers towards their brand. By automatically sorting the sentiment behind social media conversations, reviews, and more, businesses can make better and more informed decisions.

It's estimated that 90% of the world's data is unstructured, in other words it's unorganized. Huge volumes of unstructured business data are created every day: emails, support tickets, chats, social media conversations, surveys, articles, documents, etc). But it's hard to analyse for sentiment in a timely and efficient manner.

### **The overall benefits of sentiment analysis include:**

#### **Sorting Data at Scale**

We Cannot manually sort through thousands of tweets, customer support conversations, or surveys? There's just too much business data to process manually. Sentiment analysis helps businesses process huge amounts of data in an efficient and cost-effective way.

#### **Real-Time Analysis**

Sentiment analysis can identify critical issues in real-time, Sentiment analysis models can help us immediately identify different kinds of situations, so we can take action right away.

#### **Consistent criteria**

It's estimated that people only agree around 60-65% of the time when determining the sentiment of a particular text. Tagging text by sentiment is highly subjective, influenced by personal experiences, thoughts, and beliefs. By using a centralized sentiment analysis system, companies can apply the same criteria to all of their data, helping them improve accuracy and gain better insights.

## EXISTING SYSTEM

There are different algorithms we can implement in sentiment analysis models, depending on how much data we need to analyse, and how accurate our model is, we use below three Models.

- **Rule-based:**

These systems automatically perform sentiment analysis based on a set of manually crafted rules.

- **Automatic:**

Systems rely on machine learning techniques to learn from data.

- **Hybrid:**

Systems combine both rule-based and automatic approaches.

### 1. Rule-based Approaches

Usually, a rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion.

These rules may include various NLP techniques developed in computational linguistics, such as:

- Stemming
- Tokenization
- Part-of-speech tagging
- Parsing.

Lexicons (i.e. lists of words and expressions).

#### **Basic example of how a rule-based system works:**

1. Defines two lists of polarized words (e.g. negative words such as bad, worst, ugly, etc and positive words such as good, best, beautiful, etc).
2. Counts the number of positive and negative words that appear in each text.

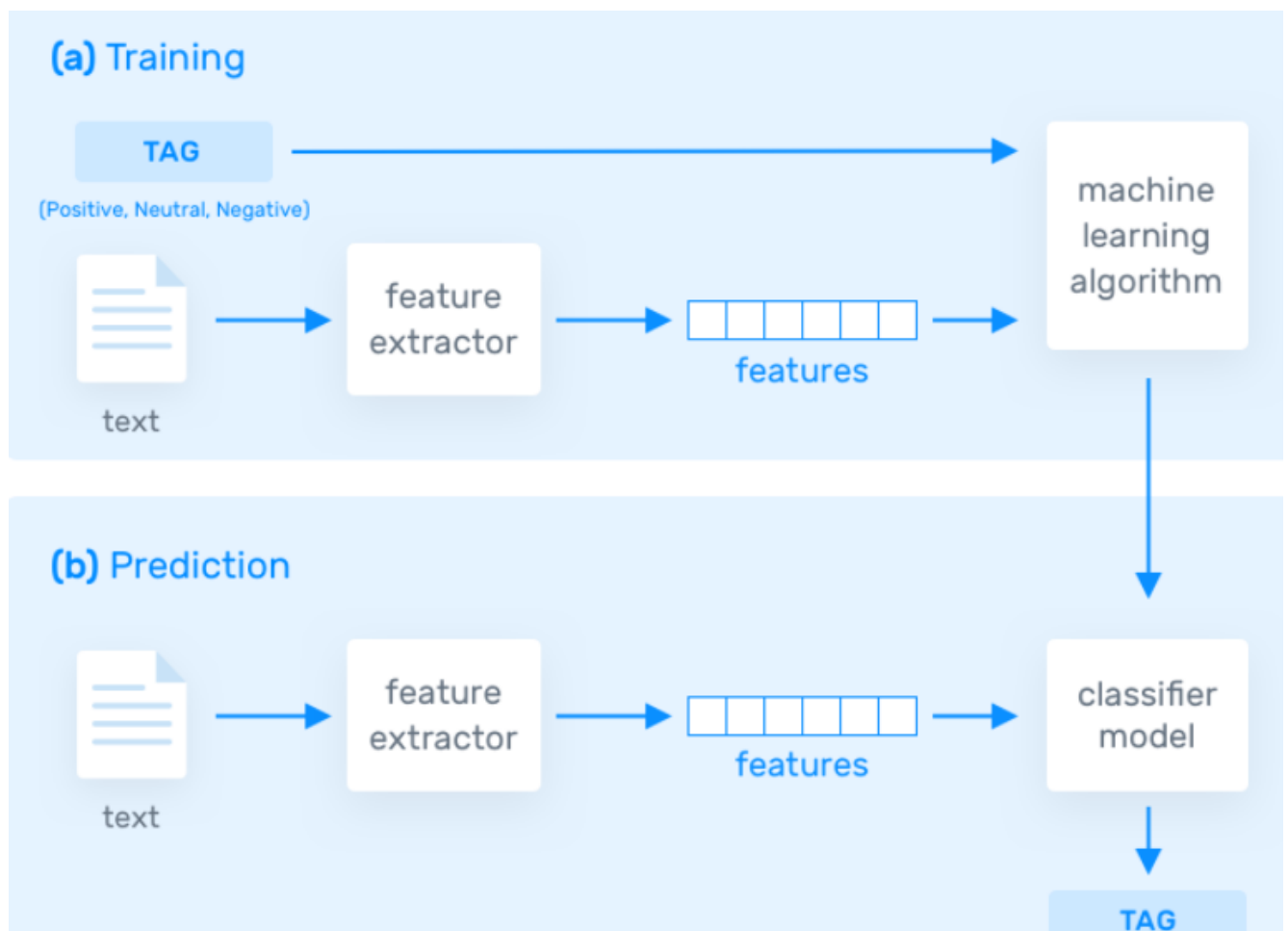
3. If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment.

Rule-based systems are very naive since they don't consider how words are combined in a sequence. Of course, more advanced processing techniques can be used, and new rules added to support new expressions and vocabulary. However, adding new rules may affect previous results, and the whole system can get very complex. Since rule-based systems often require fine-tuning and maintenance, they'll also need regular investments.

## 2. Automatic Approaches

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modelled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

A machine learning classifier can be implemented:



## **The Training and Prediction Processes**

In the training process our model learns to associate a particular input (i.e. a text) to the corresponding output (tag) based on the test samples used for training. The feature extractor transfers the text input into a feature vector. Pairs of feature vectors and tags (e.g. positive, negative, or neutral) are fed into the machine learning algorithm to generate a model.

In the prediction process the feature extractor is used to transform unseen text inputs into feature vectors. These feature vectors are then fed into the model, which generates predicted tags (positive, negative, or neutral).

### **Feature Extraction from Text**

The first step in a machine learning text classifier is to transform the text extraction or text vectorization, and the classical approach has been bag-of-words or bag-of-n grams with their frequency.

More recently, new feature extraction techniques have been applied based on word embeddings (also known as word vectors). This kind of representations makes it possible for words with similar meaning to have a similar representation, which can improve the performance of classifiers.



## Classification Algorithms

The classification step usually involves a statistical model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks:

- **Naïve Bayes:**

A family of probabilistic algorithms that uses Bayes's Theorem to predict the category of a text.

- **Linear Regression:**

A very well-known algorithm in statistics used to predict some value (Y) given a set of features (X).

- **Support Vector Machines:**

A non-probabilistic model which uses a representation of text examples as points in a multidimensional space. Examples of different categories (sentiments) are mapped to distinct regions within that space. Then, new texts are assigned a category based on similarities with existing texts and the regions they're mapped to.

- **Deep Learning:**

A diverse set of algorithms that attempt to mimic the human brain, by employing artificial neural networks to process data.

### 3. Hybrid Approaches

Hybrid systems combine the desirable elements of rule-based and automatic techniques into one system. One huge benefit of these systems is that results are often more accurate.

## PROBLEM ANALYSIS

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis, which is also known as opinion mining, studies people's sentiments towards certain entities. Internet is a resourceful place with respect to sentiment information. From a user's perspective, people can post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. For instance, Twitter currently has three different versions of APIs available, namely the REST API, the Search API, and the Streaming API. With the REST API, developers are able to gather status data and user information; the Search API allows developers to query specific Twitter content, whereas the Streaming API is able to collect Twitter content in real time. Moreover, developers can mix those APIs to create their own applications. Hence, sentiment analysis seems having a strong fundament with the support of massive online data.

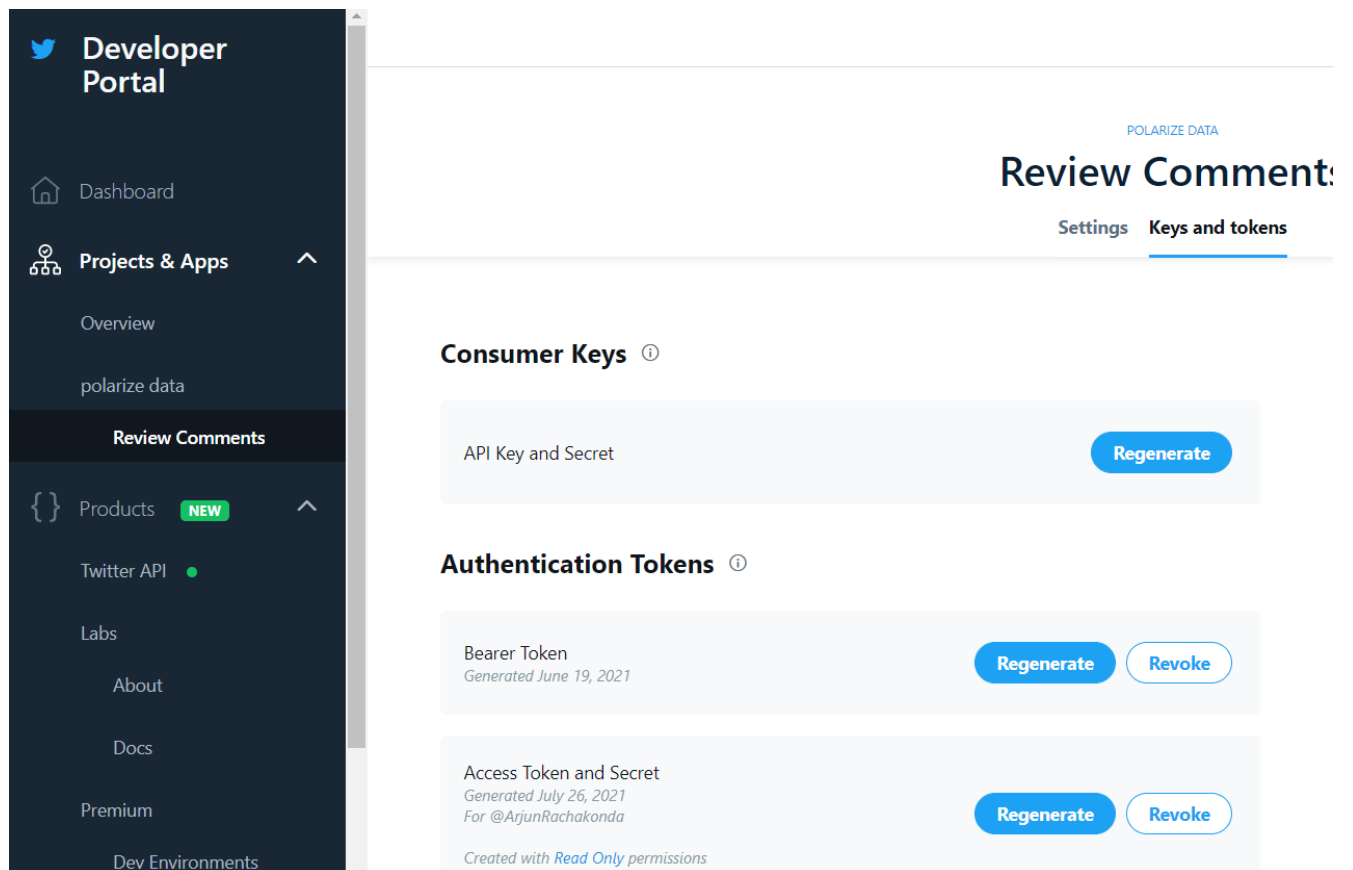
However, those types of online data have several flaws that potentially hinder the process of sentiment analysis. The first flaw is that since people can freely post their own content, the quality of their opinions cannot be guaranteed. For example, instead of sharing topic-related opinions, online spammers post spam on forums. Some spam is meaningless at all, while others have irrelevant opinions also known as fake opinions. The second flaw is that ground truth of such online data is not always available. A ground truth is more like a tag of a certain opinion, indicating whether the opinion is positive, negative, or neutral. The Stanford Sentiment 140 Tweet Corpus is one of the datasets that has ground truth and is also public available. The corpus contains 1.6 million machine-tagged Twitter messages.

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of micro blogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times, these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment.

## PRODUCT DEFINITION

My project model analyses tweets by the people using **Rule-based approach**. To do this, I have analysed tweets from Twitter. Tweets are a reliable source of information mainly because people tweet about anything and everything, they do include buying new products and reviewing them. Besides, all tweets contain hash tags which make identifying relevant tweets a simple task. Several research works have already been done on twitter data.

Most of which mainly demonstrates how useful this information is to predict various outcomes. My current research deals with outcome prediction and explores localized outcomes.



To Collect data from Twitter we must have a twitter account then we must apply to become a Twitter Developer after few steps and verification we can access data from twitter using API's

We can collect data using the Twitter public API which allows developers to extract tweets from twitter programmatically using “tweepy” in python. The collected data, because of the random and casual nature of tweeting, need to be filtered to remove unnecessary information using Natural Language Processing (NLP). Filtering out these and other problematic tweets such as redundant ones, and ones with no proper sentences was done next.

As the pre-processing phase was done in certain extent it was possible to guarantee that analysing these filtered tweets will give reliable results. Twitter does not provide the gender as a query parameter, so it is not possible to obtain the gender of a user from his or her tweets. It turned out that twitter does not ask for user gender while opening an account so that information is seemingly unavailable.

### **Statement of the problem**

- **Phrase Level Sentiment Analysis in Twitter:**

Given a message containing a marked instance of a word or a phrase, determine whether that instance is positive, negative, or neutral in that context.

- **Sentence Level Sentiment Analysis in Twitter:**

Given a message, decide whether the message is of positive, negative, or neutral sentiment. For messages conveying both a positive and negative sentiment, whichever is the stronger sentiment should be chosen, then store then in a Data Frame.

### **Objectives**

The objectives of this project are:

- To implement an algorithm for automatic classification of text into positive and negative.
- Sentiment Analysis to determine the attitude of the mass is positive, negative, or neutral towards the subject of interest.
- Graphical representation of the sentiment in form of Pie-Chart, Bar Diagram and Scatter Plot.

### **Scope of project**

This project will be helpful to the companies, political parties as well as to the common people. It will be helpful to political party for reviewing about the program that they are going to do or the program that they have performed. Similarly, companies also can get review about their new product on newly released hardware or software. Also, the movie maker can take review on the currently running movie. By analysing the tweets analyser can get result on how positive or negative or neutral are peoples about it.

### **System Overview**

This application which is used to analyse the tweets. We will be performing sentiment analysis in tweets and determine where it is positive, negative, or neutral. This application can be used by any

organization office to review their works or by political leaders or by any others company to review about their products or brands.

### **System Features**

The main feature of my application is that it helps to determine the opinion about the peoples on products, government work, politics or any other by analysing the tweets. The computed or analysed data will be represented in various diagram such as Pie-chart, Bar graph and Scatter Plot.

## **FEASIBILITY ANALYSIS**

My Application is technically feasible since all the required tools are easily available. Python can be easily handled. Although all tools seem to be easily available there are challenges too.

Proposed project is beneficial only if it can be turned into information systems that will meet the operating requirements. The proposed was to make a simplified application. It is simpler to operate and can be used in any webpages. It is free and not costly to operate.

## SOFTWARE REQUIREMENT ANALYSIS

After the extensive analysis of the problems in the system, I'm familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements.

These requirements are listed below:

### **1. Functional Requirements**

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users.

Based on this, the functional requirements that the system must require are as follows:

- System should be able to process new tweets stored in data frame after retrieval.
- System should be able to analyse data and classify each tweet polarity.

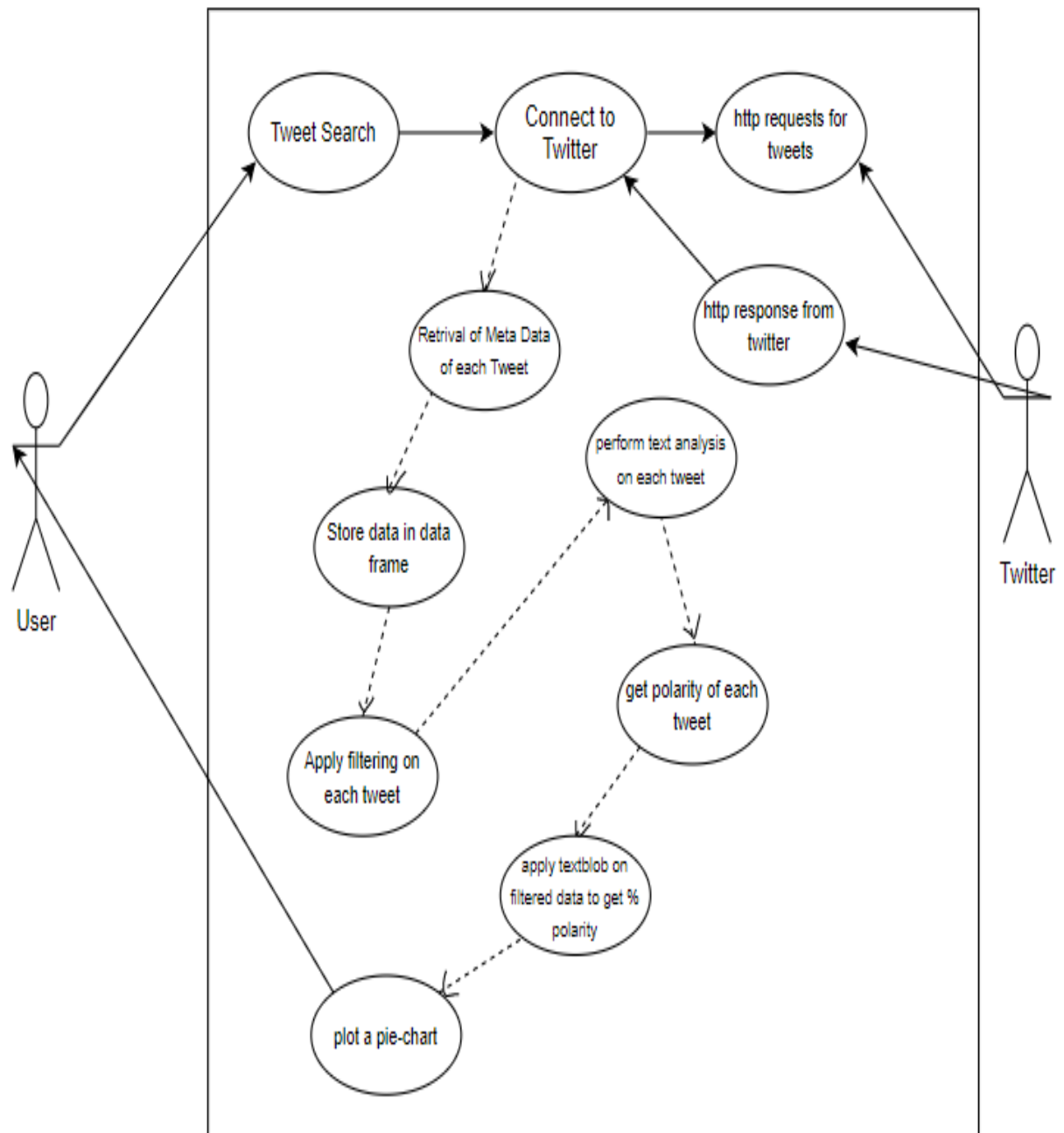
### **2. Non-Functional Requirements**

Non-functional requirements are a description of features, characteristics, and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services.

Based on these the non-functional requirements are as follows:

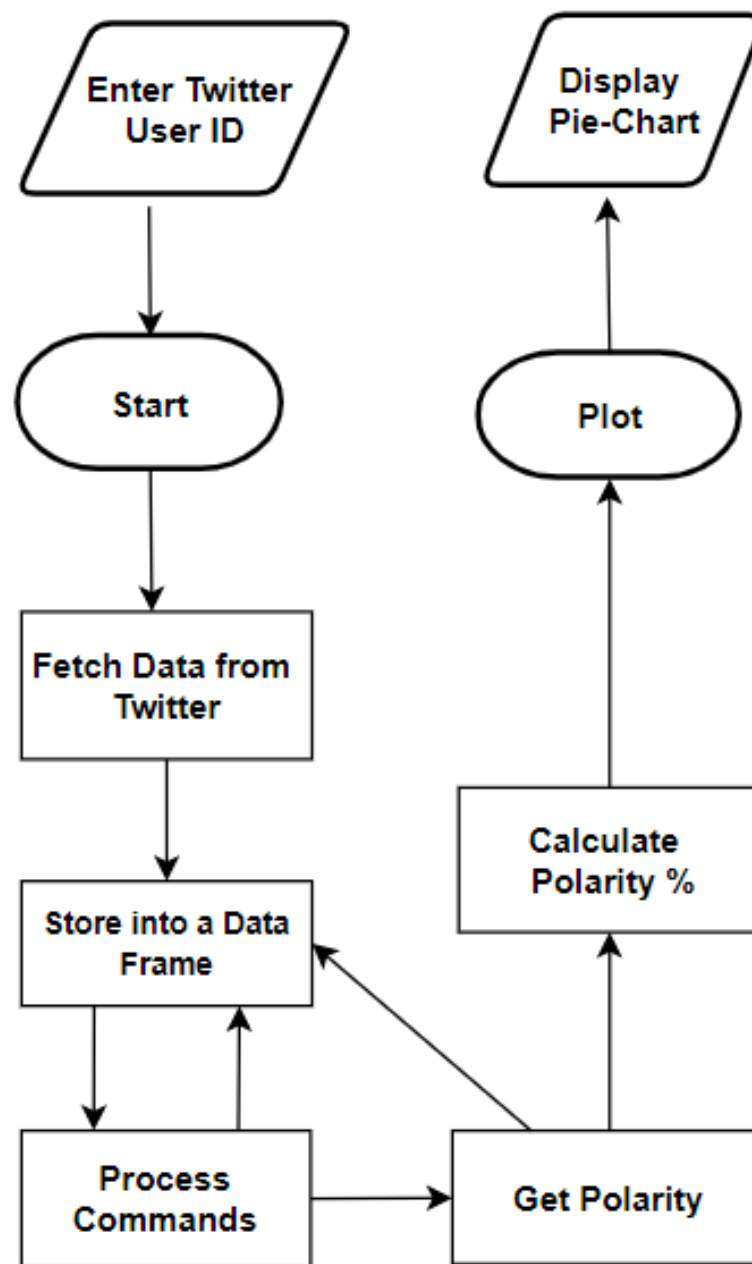
- User friendly.
- System should provide better accuracy.
- To perform with efficient throughput and response time.

## DESIGN



Use Case Diagram

## Flow Chart





## IMPLEMENTATION

Code for the Project was written in Python which is a widely used high-level, general-purpose, interpreted, dynamic programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than possible in languages such as C or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale.

### NLTK

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. NLTK has been called “a wonderful tool for teaching, and working in, computational linguistics using Python,” and “an amazing library to play with natural language.” NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analysing linguistic structure, and more.

Algorithms implement in this sentiment analysis model is Rule-based which automatically perform sentiment analysis based on set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion, rules may include various NLP techniques.

### Natural language processing (NLP)

Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages. This approach utilizes the publicly available library of SentiWordNet, which provides a sentiment polarity values for every term occurring in the document. In this lexical resource each term t occurring in WordNet is associated to three numerical scores describing the objective, positive and negative polarities of the term, respectively. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. WordNet is also freely and publicly available for download. WordNet’s structure makes it a useful tool for computational linguistics and natural language processing. It groups words together based on their meanings. Synet is nothing but a set of one or more Synonyms. This approach uses Semantics to understand the language.

Major tasks in NLP that helps in extracting sentiment from a sentence:

- Extracting part of the sentence that reflects the sentiment.
- Understanding the structure of the sentence.
- Different tools which help process the textual data.

Basically, Positive and Negative scores got from SentiWordNet according to its part-of-speech tag and then by counting the total positive and negative scores we determine the sentiment polarity based on which class (i.e. either positive or negative) has received the highest score.

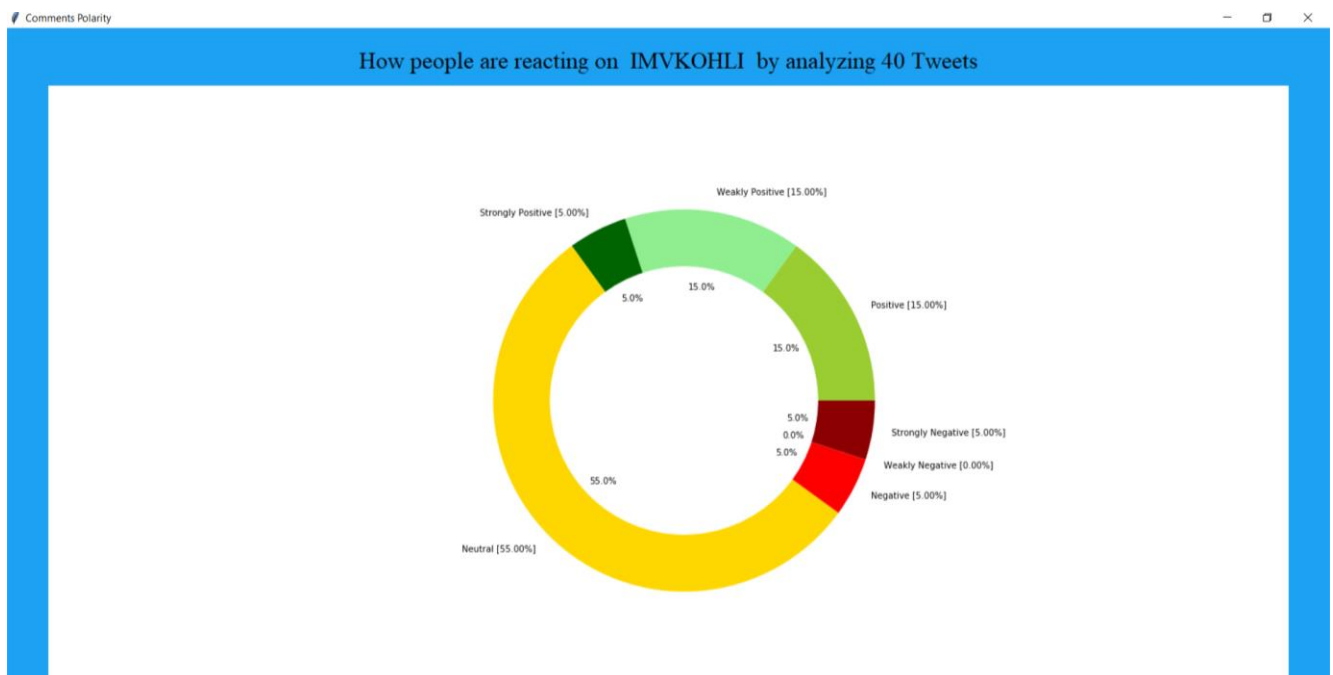
## Matplotlib

“matplotlib” is a charting library written in Python, offering an easy way of adding interactive charts to tkinter window.

## Result

After facing a number of errors, successful elimination of those error I have completed project with continuous effort. At the end of the project the results can be summarized as:

- A user-friendly application.
- No expertise is required for using the application.
- Organizations can use the application to visualize product or brand review graphically.



**Pie-chart Representation**

Filtered Tweets		polarity
1	medal india back india celebrate could even enjoy much deserved stay celebrate	Weakly Positive
2	virat say ipus thanks appreciating motivating student	Weakly Positive
3	team india captain appreciated ipu achievement tokyo olympics virat say ipus	Weakly Positive
4	wish ipu sends student indian cricket team also look forward encouraging new cric	Weakly Positive
5	india sri lanka ha postponed due kunal pandye found corona positive play tomorrow	Weakly Positive
6	kudos nearly indian olympian tokyoolympics vertos say	Weakly Positive
7	praising youngster successful olympics virat say ipus	Strongly Positive
8	vk vohra superb make adidas logo nike customize bro	Strongly Positive
9	myself getting unnecessary hate today b c ipu post	Strongly Negative
10	manita idolo reach level hate	Strongly Negative
11	could better compliment coming praising athlete virat say ipus	Positive
12	thanks ipu training brilliant student amazing compliment king virat say i	Positive
13	could better compliment coming praising athlete virat say ipus	Positive
14	vk vohra fine bro	Positive
15	think jersey outstanding team india playe	Positive
16	vohra ok bro follow dm	Positive
17	compliment king kohli encouraging virat say ipus	Neutral
18	ha back ipu student dream moment ipu get appreciation mr kohli virat say	Neutral
19	vk vohra make one upcoming wc jersey	Neutral
20	shounak support till last breathe	Neutral
21	backing ipu university serve boost university train nurture i	Neutral
22	appreciate work virat say ipus	Neutral
23	compliment king kohli encouraging virat say ipus	Neutral
24	vohra tell made copied	Neutral
25	trace decoded	Neutral
26	k like	Neutral
27	well got paid crore	Neutral
28	manita meanwhile virat	Neutral
29	virat kohli team india get affected second ti sri lanka india ha postpo	Neutral
30	many engineer unfortunately ipu provide sport	Neutral
31	big news	Neutral

## Filtered Tweets and their respective Polarity Stored in a Data Frame

## LIMITATION AND FUTURE ENHANCEMENT

### Limitation

The system we designed is used to determine the opinion of the people based on twitter data. somehow, I completed project and was able to determine only positivity, negativity, Neutral of tweet. currently the application could analyse only 200 live tweets. This may not give proper value and results. The results are not much accurate.

### Future Enhancement

- Analysing sentiments on emoji / smiley.
- Determining neutrality.
- Potential improvement can be made to our data collection and analysis method.
- Future research can be done with possible improvement such as more refined data and more accurate algorithm.

## CONCLUSION

I have completed project using python as language, for output presentation. I'm able to determine the positivity and negativity of each tweet. Based on those tweets we represented them in a diagram like Pie-chart. All the diagrams related to outcome are shown above. designed system is user friendly. All displaying results are displayed in tkinter window.

## **LEARNING OUTCOME FROM TRAINING/ TECHNOLOGY LEARNT**

- Can be able to understand and implement machine learning models.
- Can be able to analyse and classify the large data.
- Can be able to understand and implement various types of Recommendation Systems.
- Can be able to implement various data visualization techniques.
- Can be able use OpenCV and make many computer vision models such as Motion Detection, Face Detection, Edge Detection, Corner Detection, Background reduction, Foreground Extraction ... etc.
- Can be able understand and implement various Classification and Regression Models.

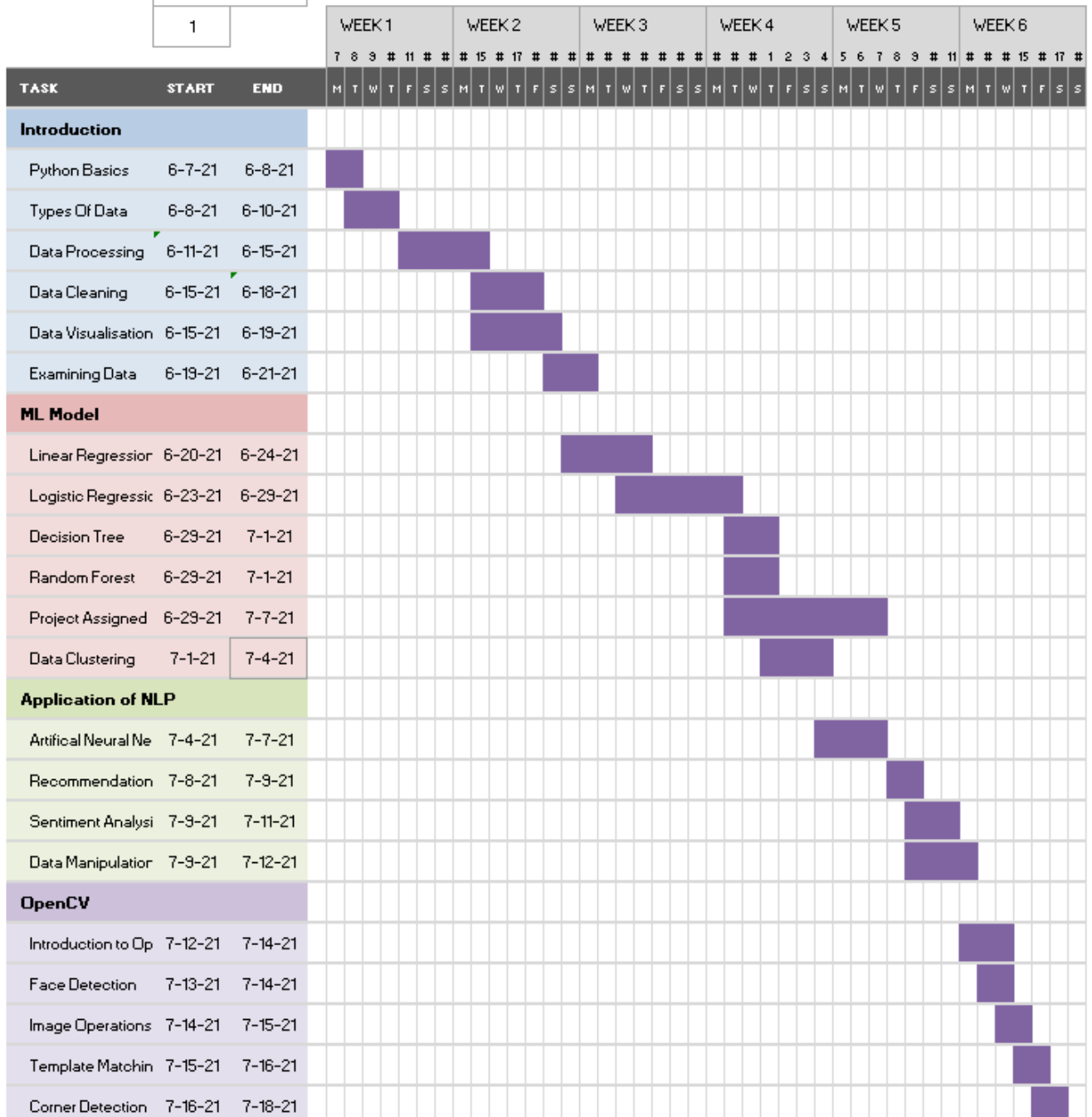
## GANTT CHART

## Data Science

### Skill Vertex

Mon, 6-7-2021

1



## PROJECT LEGACY

Sentiment Analyzer for Twitter has a number of applications:

**Business:** Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customers' feelings towards products or brand, how people respond to their campaigns or product launches and why consumers are not buying certain products.

**Politics:** In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analysing election results.

**Public Actions:** Twitter Sentiment Analysis also is used for monitoring and analysing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere.

## BIBLIOGRAPHY

1. Kim S-M, Hovy E (2004) Determining the sentiment of opinions In: Proceedings of the 20th international conference on Computational Linguistics, page 1367. Association for Computational Linguistics, Stroudsburg, PA, USA.
2. Liu B (2010) Sentiment analysis and subjectivity In: Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca. Liu B, Hu M, Cheng J (2005) Opinion observer: Analyzing and comparing opinions on the web In: Proceedings of the 14th International Conference on World Wide Web, WWW '05, 342–351. ACM, New York, NY, USA.
3. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining In: Proceedings of the Seventh conference on International Language Resources and Evaluation. European Languages Resources Association, Valletta, Malta.
4. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04.. Association for Computational Linguistics, Stroudsburg, PA, USA.
5. Pang B, Lee L (2008) Opinion mining and sentiment analysis
6. Twitter API's: <https://developer.twitter.com>
7. Sentiment Analysis: [https://en.wikipedia.org/wiki/Sentiment\\_analysis](https://en.wikipedia.org/wiki/Sentiment_analysis)
8. Sentiment Analysis: A Definitive Guide → [https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20\(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs](https://monkeylearn.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20(or%20opinion%20mining,feedback%2C%20and%20understand%20customer%20needs)
9. How to display a dataframe in tkinter: <https://stackoverflow.com/questions/44798950/how-to-display-a-dataframe-in-tkinter>