SENTIMENT ANALYSIS OF U.S. TELECOM COMPANIES

IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

SUBMITTED BY

ROHIN SANDHU

1907586

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

RAYAT BAHRA INSTITUTE OF ENGINEERING & NANOTECHNOLOGY

HOSHIARPUR –146104, PUNJAB (INDIA)

JULY, 2022

DECLARATION

I, **ROHIN SANDHU**, bearing Roll Number **1907586** , a student of B. Tech Computer Science and Engineering (CSE) under the department named Computer Science & Engineering, hereby declare that I own the full responsibility for the information, results, etc. provided in the PROJECT titled **"SENTIMENT ANALYSIS OF U.S. TELECOM COMPANIES"** submitted to "**Rayat Bahra Institute of Engineering & Nano-Technology, Hoshiarpur"** for the award of B. Tech (CSE) Degree. I have taken care in all respect to honour the intellectual property right and have acknowledged the contribution of others for using them in academic purpose and further declare that in case of any violation of an intellectual property right or copyright I, as a candidate, will be fully responsible for the same. My Supervisor should not be held responsible for full or partial violation of copyright or intellectual property right.

Name: Rohin Sandhu

Roll Number: 1907586

Date: 16 May, 2022

Place: Phagwara

ACKNOWLEDGEMENT

I express my sincere gratitude to my respected guide **Er. Rajvinder Kaur**, Assistant Professor, Rayat-Bahra Institute of Engineering & Nanotechnology, Hoshiarpur, for giving me stimulating guidance, continuous support and supervision throughout the course of present work.

I would like to place on record my deep sense of thankfulness to **Er. Prince Rana**, Assistant Professor & Head of the Department, Department CSE & IT, Rayat-Bahra Institute of Engineering & Nanotechnology, Hoshiarpur, for his generous guidance, help and useful suggestions.

I express my sincere gratitude to my respected teacher, **Dr. H.P.S. Dhami**, Director Principal, Rayat-Bahra Institute of Engineering & Nanotechnology, Hoshiarpur, for giving me stimulating guidance, continuous support and supervision throughout the course of present work.

I also wish to extend my thanks to all faculty members, Rayat-Bahra Institute of Engineering & Nanotechnology, Hoshiarpur, their insightful comments and constructive suggestions to improve the quality of this research work.

I am extremely thankful to **Dr. Chander Mohan**, Campus-Director, Rayat-Bahra Institute of Engineering & Nanotechnology, Hoshiarpur, for providing me infrastructural facilities to work in, without which this work would not have been possible.

ROHIN SANDHU

1907586

ABSTRACT

Predicting customer churn is a big challenge and a survival basic for Telecom operators. In a large and competitive market like U.S., it is very essential to gather real-time customer feedback as a health indicator. Social networks have evolved as a rich source of real-time sentiments and opinions of the general public. In this research, tweets for the Twitter handle of 7 major telecom brands in U.S.: AT&T, Avaya, CenturyLink, Spark Light, Sprint Corporation, U.S. Cellular & Verizon were extracted for six months to develop a prediction model for telecom subscriber churn prediction using the sentiment score. Pandas and TextBlob library of Python were used to assign polarities to user sentiments. Customer satisfaction represented by the overall monthly sentiment score has been used to predict customer churn. The predictions made by the model were validated using IBM SPSS and were within the acceptable limits. The results of the sentiment analysis based prediction model can be of great use for telecom operators to take timely actions for improving the future customer experience and avoiding customer churn.

TABLE OF CONTENT

LIST OF FIGURES

Overview of Technology Used For Making Project:

1.1 Twitter: Twitter is an online news and social networking site where people communicate in short messages called tweets. Tweeting is posting short messages for anyone who follows you on Twitter, withthe hope that your words are useful and interesting to someone in your audience. Another description of Twitter and tweeting might be microblogging. Some people use Twitter to discover interesting people and companies online, opting to follow their tweets.

Twitter's big appeal is how scan-friendly it is. You can track hundreds of engaging Twitter users and read their content with a glance, which is ideal for our modern attention-deficit world. Twitter employs apurposeful message size restriction to keep things scan-friendly: every microblog tweet entry is limited to 280 characters or less. This size cap promotes the focused and clever use of language, which makes tweets easy to scan, and challenging to write. This size restriction made Twitter a popular social tool.

Figure 1: Twitter's Logo

How Twitter Works: Twitter is easy to use as either broadcaster or a receiver. You join with a free account and Twitter name. Then you send broadcasts (tweets) daily, hourly, or as frequently as you like. Go to the What's Happening box next to your profile image, type 280 or fewer characters, and click Tweet**.** People who follow you, and potentially others who don't, will see your tweet. Encourage people you know to follow you and receive your tweets in their Twitter feeds. Let your friends know you are on Twitter to build up a following slowly. When people follow you, Twitter etiquette calls for you to follow them back. To receive Twitter feeds, find someone interesting (celebrities included) and press Follow to subscribe to their tweets. If their tweets aren't as interesting as you hoped, you can always unfollow them. Go to your account at Twitter.com day or night to read your Twitter feed, which is continually changing as people post. Check out Trending topics to see what's going on in the world. Twitter is that simple.

Why People Tweet: People send tweets for all sorts of reasons besides sharing their thoughts: vanity, attention, shameless self-promotion of their web pages, or pure boredom. The vast majority of tweeters microblog recreationally. It's a chance to shout out to the world and revel in how many people read their tweets. However, a growing number of Twitter users send out useful content, and that's the real value of Twitter. It provides a stream of quick updates from friends, family, scholars, news journalists, and experts. It empowers people to become amateur journalists of life, describing and sharing something that they found interesting about their day. Twitter has a lot of drivel, but at the same time, there is a base of useful news and knowledgeable content. You'll need to decide for yourself which content is worth following there.

Twitter as a Form of Amateur News Reporting: Among other things, Twitter is a way to learn about the world through another person's eyes. Tweets may come from people in Thailand as their cities become flooded. Your soldier cousin in Afghanistan might describe his war experiences; your traveling sister in Europe shares her daily discoveries, or a rugby friend could tweet from the Rugby World Cup. These micro-bloggers are all mini-journalists in their own way, and Twitter gives them a platform to send a constant stream of updates right from their laptops and smartphones.

Twitter as a Marketing Tool: Thousands of people advertise their recruiting services, consulting businesses, and retail stores by using Twitter, and it works. The modern internet-savvy user has grown tired of television advertisements. People prefer advertising that is fast, less intrusive, and can be turned on or off at will. Twitter is precisely that; when you learn how the nuances of tweeting work, you can get good advertising results by using Twitter.

Twitter as a Social Messaging Tool Yes, Twitter is social media, but it's more than instant messaging. Twitter is about discovering interesting people around the world. It can also be about building a following of people who are interested in you and your work or hobbies and then providing those followers with some knowledge value every day. Twitter works well with other social tools, including Instagram, Snapchat, and Messenger. For example, if you like a tweet and want to share it on your Instagram Story, tap the tweet, then tap the Share icon and choose Instagram Stories. The tweet will appear as part of your Instagram Story. (This feature is currently supported only on iOS).

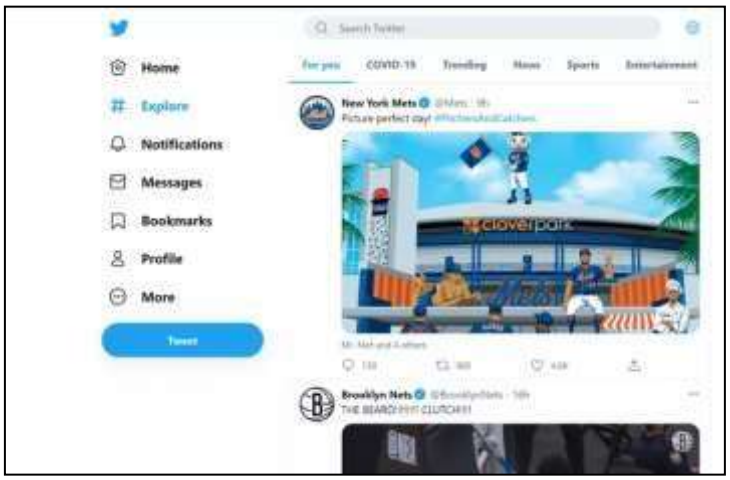


Figure 2: Twitter's Homepage

Why Celebrities Like Twitter: Twitter has become one of the most used social media platforms because it is both personal and rapid. Celebrities use Twitter to build a personal connection with their fans. Katy Perry, Ellen DeGeneres, and Dionne Warwick are some of the famous Twitter users. Their daily updates foster a sense of connectedness with their followers, which is powerful for advertising purposes and also compelling and motivating for the people who following the celebs.

Twitter Is Many Different Things Twitter is a blend of instant messaging, blogging, and texting, but with concise content and a broad audience. If you fancy yourself a bit of a writer with something to say, then Twitter is a channel worth exploring. If you don't like to write but are curious about a celebrity, a particular hobby topic, or even a long-lost cousin, then Twitter is one way to connect with that person or subject.

1.2 Python: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes simple approach very effective.

Figure 3: Python's Logo

Libraries in Python: A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer. As we don't need to write the same code again and again for different programs. Python libraries play a very vital role in fields of Machine Learning, Data Science, Data Visualization, etc.

Working of Python Library: As is stated above, a Python library is simply a collection of codes or modules of codes that we can use in a program for specific operations. We use libraries so that we don't need to write the code again in our program that is already available. But how it works. Actually, in the MS Windows environment, the library files have a DLL extension (Dynamic Load Libraries). When we link a library with our program and run that program, the linker automatically searches for that library. It extracts the functionalities of that library and interprets the program accordingly. That's how we use the methods of a library in our program. We will see further, how we bring in the libraries in our Python programs.

Use of Libraries in Python Program: As we write large-size programs in Python, we want to maintain the code's modularity. For the easy maintenance of the code, we split the code into different parts and we can use that code later ever we need it. In Python, *modules* play that part. Instead of using the same code in different programs and making the code complex, we define mostly used functions in modules and we can just simply import them in a program wherever there is a requirement. We don't need to write that code but still, we can use its functionality by importing its module. Multiple interrelated modules are stored in a library. And whenever we need to use a module, we import it from its library. In Python, it's a very simple job to do due to its easy syntax. We just need to use **import**.

1.2.1 RE In Python: Regular Expressions, also known as “regex” or “regexp”, are used to match strings of text such as particular characters, words, or patterns of characters. It means that we can match and extract any string pattern from the text with the help of regular expressions. Some applications of RE is listed below:

Form Validation: The most common use of regular expressions is form validation, i.e. email validation, password validation, phone number validation, and many other fields of the form.

Bank Account details: You must have noticed that every bank has an IFSC code for its different branches that starts with the name of the bank. The credit card number consists of 16 digits and the first few digits represent whether the card is Master, Visa, or Rupay. In all these cases, regex is used.

Data Mining: How can we forget the importance of regex in data mining? When the data is present in unstructured form, i.e. in text form, it needs to be converted to numbers to train the model.
Therefore, regex plays an important role in analyzing the data, find patterns in the data, and finally performing operations on the dataset.

NLP: NLP is a process through which a computer understands and generates human language. In NLP, regular expressions are used to remove the unnecessary words i.e., stop words from the text—thus helps in data cleaning. Regex is also used in analyzing the texts and thus helps in the prediction of the algorithm to process the data.

Social Media Platforms: Social Media Platforms such as Google, Facebook, Twitter provide several techniques to search, which are different and efficient from a normal search. Believe me, if you know these techniques, you can explore much more. All these techs use regex in the backend to process these searches.
You can think of various other applications of regular expressions wherever pattern matching is required.

Implementation in Python: Regex is provided by many programming languages, such as python, java, java script, etc. Although the concept is the same everywhere yet, you may find some differences in different languages. Now we will look into the various functions provided by python to implement regex along with its code. Python does not provide an inbuilt regex module. You need to install it using the pip command and then import it into your Python IDE. Then we stored some text in a variable named string.

Wild Card patterns: The smallest individual units through the regular expressions are formed are called wild-card patterns. The list of commonly used patterns are:

- **^ :** This wild card matches the characters at the beginning of a line.
- **$:** This wild card matches the characters at the end of the line.
- **. :** This wild card matches any character in the line.
- **s :** This wild card is used to match space in a string.
- **S :** This wild card matches non-whitespace characters.
- **d :** This wild card matches one digit.
- *** :** This wild card repeats any preceding character zero or more times.
- ***? :** This wild card also repeats any preceding character/characters zero or more times.
- **+ :**This wild card repeats any preceding character one or more times.
- **+? :** This wild card repeats any preceding character one or more times.
- **(:** This wild card represents the beginning of the string extraction.
- **) :** This wild card represents the end of the string extraction.

1.2.2 Pandas In Python: Pandas is defined as an open-source library that provides high-performance data manipulation in Python. The name of Pandas is derived from the word Panel Data, which means an Econometrics from Multidimensional data. It is used for data analysis in Python and developed by Wes McKinney in 2008. Data analysis requires lots of processing, such as restructuring, cleaning or merging, etc. There are different tools are available for fast data processing, such as Numpy, Scipy, Cython, and Panda. But we prefer Pandas because working with Pandas is fast, simple and more expressive than other tools. Pandas is built on top of the Numpy package, means Numpy is required for operating the Pandas.

Before Pandas, Python was capable for data preparation, but it only provided limited support for data analysis. So, Pandas came into the picture and enhanced the capabilities of data analysis. It can perform five significant steps required for processing and analysis of data irrespective of the origin of the data, i.e., load, manipulate, prepare, model, and analyze.

Key Features of Pandas: Pandas in Python are very powerful module that has many key features. Some key features of Pandas are listed below:

- It has a fast and efficient Data Frame object with the default and customized indexing.
- Used for reshaping and pivoting of the data sets.
- Group by data for aggregations and transformations.
- It is used for data alignment and integration of the missing data.
- Provide the functionality of Time Series.
- Process a variety of data sets in different formats like matrix data, tabular heterogeneous, time series.
- It integrates with the other libraries such as SciPy, and scikit-learn.
- Provides fast performance, and If you want to speed it, even more, you can use the Cython.

Benefits of Pandas: The benefits of pandas over using other language are as follows:
Data Representation: It represents the data in a form that is suited for data analysis through its DataFrame and Series. Clear code: The clear API of the Pandas allows you to focus on the core part of the code. So, it provides clear and concise code for the user.

Python Pandas Data Structure: The Pandas provides two data structures for processing the data, i.e., Series and Data Frame, which are discussed below:

Series: It is defined as a one-dimensional array that is capable of storing various data types. The row labels of series are called the index. We can easily convert the list, tuple, and dictionary into series using "series' method. A Series cannot contain multiple columns.

It has one parameter: Data: It can be any list, dictionary, or scalar value. Creating Series from Array: Before creating a Series, Firstly, we have to import the numpy module and then use array() function in the program.

Data Frame: It is a widely used data structure of pandas and works with a two-dimensional array with labeled axes (rows and columns). Data Frame is defined as a standard way to store data and has two different indexes, i.e., row index and column index.

It consists of the following properties: The columns can be heterogeneous types like int, bool, and so on. It can be seen as a dictionary of Series structure where both the rows and columns are indexed. It is denoted as "columns" in case of columns and "index" in case of rows.

1.2.3. NumPy In Python: NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed. NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.

Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created NumPy package by incorporating the features of Numarray into Numeric package.

Operations using NumPy: Using NumPy, a developer can perform the following operations −

- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.
- Operations related to linear algebra.

Replacement for MatLab: NumPy is often used along with packages like SciPy (Scientific Python) and Matplotlib (plotting library). This combination is widely used as a replacement for MatLab, a popular platform for technical computing. However, Python alternative to MatLab is now seen as a more modern and complete programming language. It is open-source, which is an added advantage of NumPy.

The most important object defined in NumPy is an N-dimensional array type called ndarray. It describes the collection of items of the same type. Items in the collection can be accessed using a zero-based index. Every item in a ndarray takes the same size as the block in the memory. Each element in ndarray is an object of the data-type object (called dtype). Any item extracted from ndarray object (by slicing) is represented by a Python object of one of array scalar types. The following diagram shows a relationship between ndarray, data-type object (dtype) and array scalar type −

An instance of ndarray class can be constructed by different array creation routines described later in the tutorial. The basic ndarray is created using an array function in NumPy as follows-

numpy.array: It creates a ndarray from any object exposing an array interface, or from any method that returns an array. numpy.array(object, dtype = None, copy = True, order = None, subok = False, ndmin = 0)

The ndarray object consists of a contiguous one-dimensional segment of computer memory, combined with an indexing scheme that maps each item to a location in the memory block. The memory block holds the elements in row-major order (C style) or a column-major order (FORTRAN or MatLab style). The above constructor takes the following parameters:

Sr.No.	Parameter & Description
1	**object** Any object exposing the array interface method returns an array or any sequence.
2	**dtype** The desired data type of array. By default, the object is copied
3	**order**C (row-major) or F (column-major) or A (any) (default)
4	**subok** By default, returned array forced to be a base class array.
5	**ndmin** Specifies minimum dimensions of the resultant array

1.2.4 TextBlob In Python: TextBlob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

Sentiment Analysis using Textblob: Sentiment Analysis can assist us in determining the mood and feelings of the general public as well as obtaining useful information about the setting. Sentiment Analysis is the process of assessing data and categorizing it according to the needs.

The polarity and subjectivity of a statement are returned by TextBlob. The range of polarity is [-1,1], with -1 indicating a negative sentiment and 1 indicating a positive sentiment. Negative words are used to change the polarity of a sentence. Semantic labels in TextBlob aid in fine-grained analysis. Emoticons, exclamation marks, and emojis, for example. subjectivity falls under the numeric range of [0,1]. The degree of personal opinion and factual information in a text is measured by subjectivity.

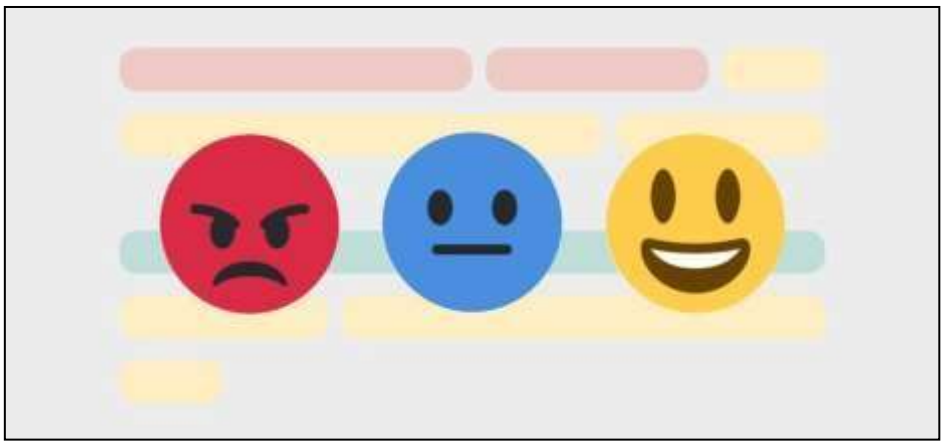


Figure 4: Sentiment Analysis using TextBlob

By providing an input sentence, the TextBlob's sentiment property returns a named tuple with polarity and subjectivity scores. The polarity score ranges from -1.0 to 1.0 and the subjectivity ranges from 0.0 to 1.0 where 0.0 is an objective statement and 1 is a subjective statement. It uses NLTK (Natural Language ToolKit) and the input contains a single sentence, The output of TextBlob is polarity and subjectivity. Polarity score lies between (-1 to 1) where -1 identifies the most negative words such as 'disgusting', 'awful', 'pathetic', and 1 identifies the most positive words like 'excellent', 'best'. Subjectivity score lies between (0 and 1), It shows the amount of personal opinion, If a sentence has high subjectivity i.e. close to 1, It resembles that the text contains more personal opinion than factual information. I was more concerned about the Polarity score as myobjective was not to identify factual information, so I skipped the subjectivity score in my project. TextBlob isa Lexicon-based sentiment analyzer It has some predefined rules or we can say word and weight dictionary, where it has some scores that help to calculate a sentence's polarity. That's why the Lexicon-based sentimentanalyzers are also called "Rule-based sentiment analyzers". Let's check some random sentences' polarity withTextBlob, The beauty of TextBlob is it has a very easy syntax.

1. It's a beautiful day. 2. This movie is badly directed. 3. The weather today is pleasant.

We get the polarity values as 0.85, -0.69, 0.73 respectively. In the above data, we have a negative sentence "This movie is badly directed" which has a polarity score of -0.69 which resembles one of the most negative sentences.

1.2.5 Tweepy In Python: Tweepy is a Python library for accessing the Twitter API. It is great for simple automation and creating twitter bots. Tweepy has many features. Some of them are listed below:

- Get tweets from our timeline.
- Creating and deleting Tweets.
- Follow and unfollow users.

1.2.6 CV In Python: OpenCV is a Python open-source library, which is used for computer vision in Artificial intelligence, Machine Learning, face recognition, etc. In OpenCV, the CV is an abbreviation form of a computer vision, which is defined as a field of study that helps computers to understand the content of the digital images such as photographs and videos.

OpenCV stands for Open Source Computer Vision Library, which is widely used for image recognition or identification. It was officially launched in 1999 by Intel. It was written in C/C++ in the early stage, but now it is commonly used in Python for the computer vision as well.

Working of Open CV: Human eyes provide lots of information based on what they see. Machines are facilitated with seeing everything, convert the vision into numbers and store in the memory. Here the question arises how computer convert images into numbers. So the answer is that the pixel value is used to convert images into numbers. A pixel is the smallest unit of a digital image or graphics that can be displayed and represented on a digital display device.

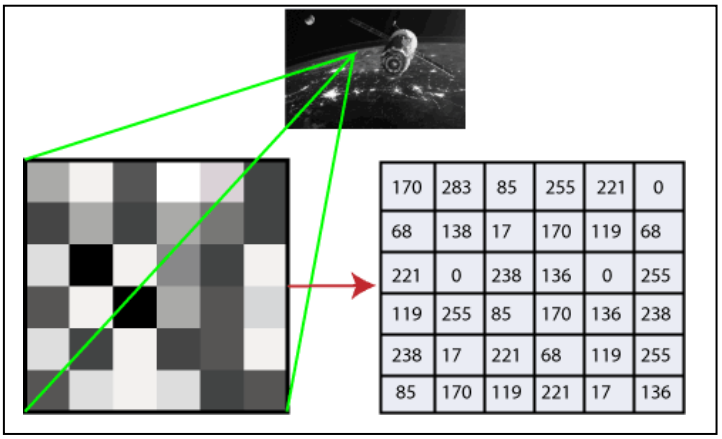


Figure 5: Working Of Open CV

The purpose of computer vision is to understand the content of the images. It extracts the description from the pictures, which may be an object, a text description, and three-dimension model, and so on. Computer vision allows the computer to perform the same kind of tasks as humans with the same efficiency. There are a two main task which are defined below:

Object Classification: In the object classification, we train a model on a dataset of particular objects, and the model classifies new objects as belonging to one or more of your training categories.

Object Identification: In the object identification, our model will identify a particular instance of an object

1.3 Microsoft Excel: Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications. Microsoft Excel is one of the most popular applications for data analysis. Equipped with built-in pivot tables, they are without a doubt the most sought-after analytic tool available. It is an all-in-one data management software that allows you to easily import, explore, clean, analyze, and visualize your data. In this article, we will discuss the various methods of data analysis in Excel.

Figure 6: Microsoft Excel's Logo

Data Cleaning: Excel Data Cleaning is a significant skill that all Business and Data Analysts must possess. In the current era of data analytics, everyone expects the accuracy and quality of data to be of the highest standards. A major part of Excel Data Cleaning involves the elimination of blank spaces, incorrect, and outdated information. Some simple steps can easily do the procedure of Data Cleaning in Excel by using Excel Power Query.

Remove Duplicates: There is a considerable probability that it might duplicate unintentionally the data without the user's knowledge. In such scenarios, you can eliminate the duplicate values.

Data Parsing from Text to Column: Sometimes, there is a possibility that one cell might have multiple data elements separated by a data delimiter like a comma. For example, consider that there is one column that stores address information.

Delete All Formatting: The formatting can be as simple as colouring your cells and aligning the text in the cells. It can be a logical condition applied to your cells using Excel's conditional formatting option from the home tab.

Spell Check: The feature of checking the spelling is available in MS Excel as well. To check the spellings of the words used in the spreadsheet, you can use the following method. Select the data cell, column, or sheet where you want to perform the spell check.

Data Visualization: It is the representation of data in a graphical format. It makes the data easier to understand. Excel is a spreadsheet that is used for data organization and data visualization as well. Excel provides various types of charts like Column charts, Bar charts, Pie charts, Linecharts, Area charts, Scatter charts, Surface charts, and much more.

Project Report Work:

2.1 Introduction: Sentiment analysis is contextual mining of text which identifies and extracts subjective malformation in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations.

Understanding people's emotions is essential for businesses since customers are able to express their thoughts and feelings more openly than ever before. By automatically analysing customer feedback, from survey responses to social media conversations, brands are able to listen attentively to their customers, and tailor products and services to meet their needs.

Example: Using sentiment analysis to automatically analyse 4,000+ reviews about your product could help you discover if customers are happy about your pricing plans and customer service.

Applications of Sentiment Analysis:

- Sentiment analysis helps businesses process huge amounts of data in an efficient & cost-effective way.
- Sentiment analysis models can identify critical issues in real time i.e. Real Time Analysis.
- 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.

2.2 Objectives: The objectives of Sentiment Analysis are:

- ✓ To mine the public opinion.
- ✓ To find the correlation between the moods of public.
- ✓ To help the investors to identify the intrinsic worth of a security even before investing in it.
- ✓ To Show it is indeed possible to have sense of where the market will go.

2.3 Methodology: A software/system development methodology in software engineering is a framework that is used to structure, plan, and control the process of developing an information system. Methodology for sentiment analysis of this project is as follows:

Data Collection: Consumers usually express their sentiments on public forums like the blogs, discussion boards, product reviews as well as on their private logs – Social network sites like Facebook and Twitter. Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and slang, making the data huge and disorganized. Manual analysis of sentiment data is virtually impossible. Therefore, special programming languages like 'Python' are used to process and analyse the data.

Text Preparation: Text preparation is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

Sentiment Detection: At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded.

Sentiment Classification: Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.

Presentation of Output: The main idea of sentiment analysis is to convert unstructured text into meaningful information. After the completion of analysis, the text results are displayed on graphs like pie chart, bar chart and line graphs.

2.3.1 DFD: DFD is the abbreviation for Data Flow Diagram. The flow of data of a system or a process is represented by DFD. It also gives insight into the inputs and outputs of each entity and the process itself. DFD for Sentiment analysis is listed below:

Figure 2: Data Flow Diagram

2.3.2 Entity Relationship Model: ER Model is used to model the logical view of the system from data perspective which consists of these components:

Entity: Entities are represented by means of rectangles. Rectangles are named with the entity set they represent.
Attributes: Attributes are the properties of entities. Attributes are represented by means of ellipses. Every ellipse represents one attribute and is directly connected to its entity (rectangle).
Relationship: Relationships are represented by diamond-shaped box. Name of the relationship is written inside the diamond-box. All the entities (rectangles) participating in a relationship, are connected to it by a line.
ER Diagram for Sentiment Analysis: Entity Relationship diagram for twitter sentiment analysis of this project is shown below:

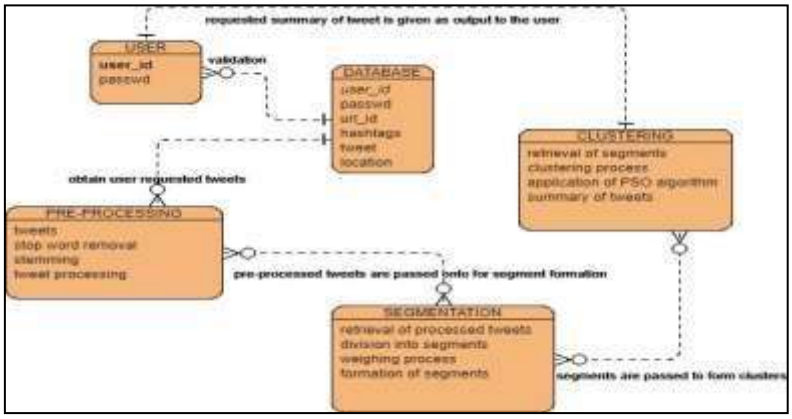


Figure 7: ER Diagram (Physical)

2.3.3 Use Case Diagram: A use case diagram is used to represent the dynamic behaviour of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system. Use Case Diagram for sentiment analysis is listed as follows:

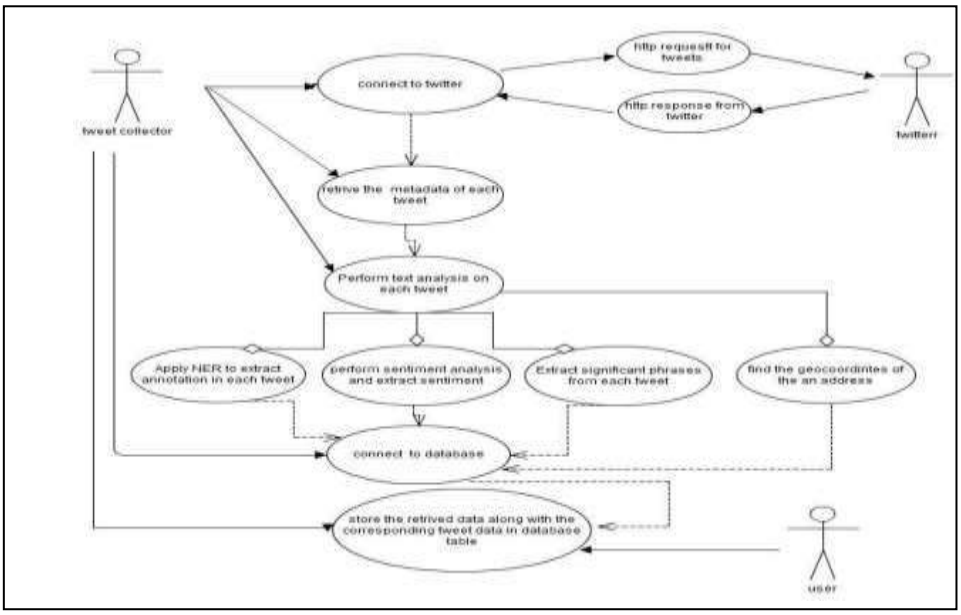


Figure 8: Use Case Diagram

2.3.4 Gantt Chart: Gantt charts help teams to plan work around deadlines and properly allocate resources.. They depict, among other things, the relationship between the start and end dates of tasks, milestones, and dependent tasks. Gantt chart for sentiment analysis of this project is as follows:

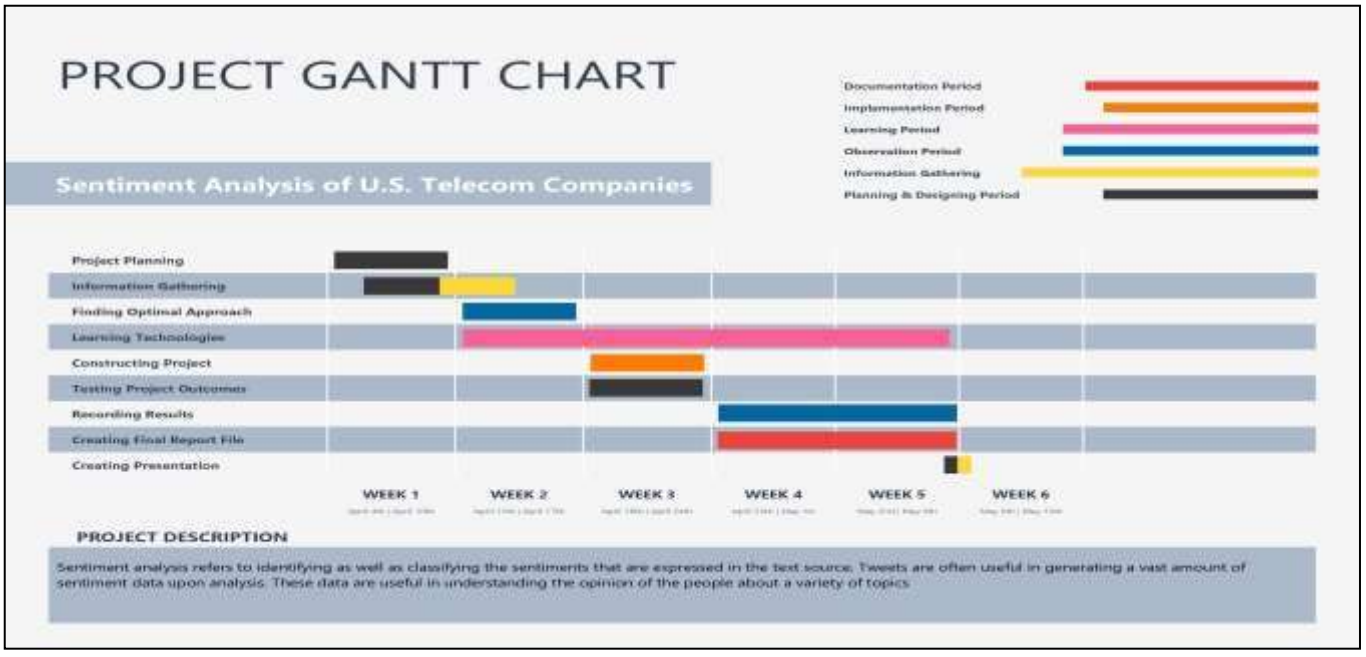


Figure 9: Gantt Chart

2.3.5 Project/Test Report: A test report is an organized summary of testing objectives, activities, and results. It is created and used to help stakeholders (product manager, analysts, testing team, and developers) understand product quality and decide whether a product, feature, or a defect resolution is on track for release. Test report for sentiment analysis of this project is as follows:

2.3.5.1 Telecom Companies: Followings are the U.S. Telecom Companies with having the Stock under New York Stock Exchange (NYSE).

1. AT & T *2. Avaya* *3. Century-Link* *4.Spark Light*

5. Sprint Corp. *6. U.S. Cellular* *7. Verizon*

2.3.5.2 Process of Sentiment Analysis:

Step 1: Twitter Developer Account

To have access to the Twitter API, you'll need to login the Twitter Developer website and create an application.

Followings are the steps to generate your Keys and Tokens:

Create a Twitter Developer account.

Create an app with your project name.

Go to the Keys and Tokens tab and generate your Keys and Tokens.

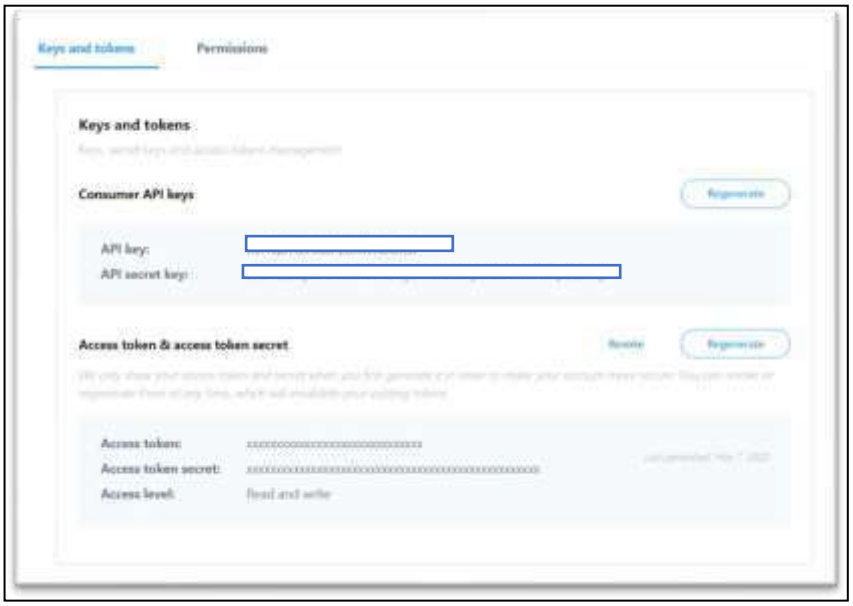


Figure 10: Twitter Access Keys & Tokens

Step 2: Data Mining

Write a Python Script to mine the tweets using the Twitter Credentials against the desired hashtags.

Download the results in a '.csv' file.

Open the files using 'Microsoft Excel'.

Figure 11: Tweets Mined Using Python

Step 3: Data Pre-Processing

Followings are the steps for text pre-processing to be followed before analyzing the sentiment score for each tweet.

Remove Duplicates

Go to the database stored in Microsoft Excel.

Click on the Data tab placed at the top of Microsoft Excel.

Select the Remove Duplicates option.

Select Ok.

You'll get the duplicate values eliminated from the database.

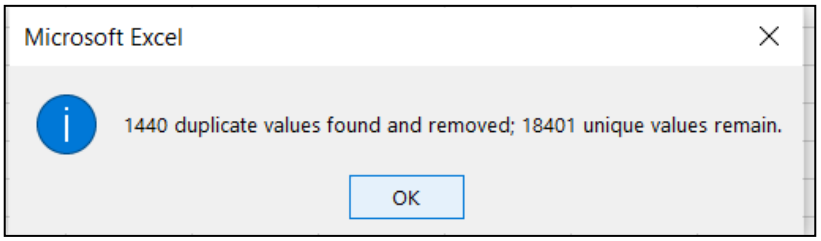


Figure 12: Removal of Duplicate Values

Text Pre-processing:

Write a Python script to remove un-necessary information like emoji, uniform resource locators (URL) and non-English characters using the python libraries i.e. 'TextBlob'.

	clean_tweet
2	Ahhh haven chance requirements with eyes until read wrong been https iLBrfwK
3	
4	Seokjin took director announced entire crew that they could them they didn break
5	atleast threaten kill good couple times tell done pull https NSjEIh
6	either what gotta
7	weren told cthis beauty salon teacher least once bitch
8	taeil saying johnny exposed calves cracking himself couldn stop laughing joke Kick
9	just excited about white dutifully show primary course election
10	Schumer should face censure expulsion threats directed Justices Gorsuch Kavanaugh Chief Roberts correctly
11	Every DEMOCRAT must asked agree with Chuck Schumer that Supreme Court Justices Gorsuch Kavanaugh must
12	Ready Commune this Saturday East Quadrangle forget bring your friends your
13	restaurant that happily used Resy since opened doesn take American Express Wonder that last
14	anniversary John Candy passing cooked small tribute comedic genius Canadian hero
15	Most young voters couldn find their voter info system They showed with affidavit printed
16	that think will happen really think https MZITFhCCPS
17	Having preference fine What being argued here that preference doesn have
18	talk today featured https wdyDVvAiYN ReTweet chance FREE https beXk
19	this holding hands rest your life want walk hand hand forever Dongpyo
20	Guys vote IDOL CHAMP they triple crown this time beat record That need
21	wish some would understand that these closures cancellations aren because they think everyone going
22	only struggling with think been there guess Chesterfield maybe
23	just bill Gates walked right into that pedi https bqcMQP
24	Marlborough diffrent kettle fish deny fool hardy braver https ePmJsi
25	Good morning Please watch this video shouldn home with random meet bars https beYGE
26	only problem with void season that signing death warrant clubs wouldn allowe https gFKkuVAEQv
27	take responsibility nFiled bankruptcy protection over times nSee
28	This stunning description what China about Covid what probably have state capacity

Figure 13: Cleaned Tweets using Python

Step 4: Analyse Sentiment

Python NLTK and TextBlob libraries were employed to perform sentiment analysis on the mined dataset. The sentiment function returns a tuple of properties as follows:

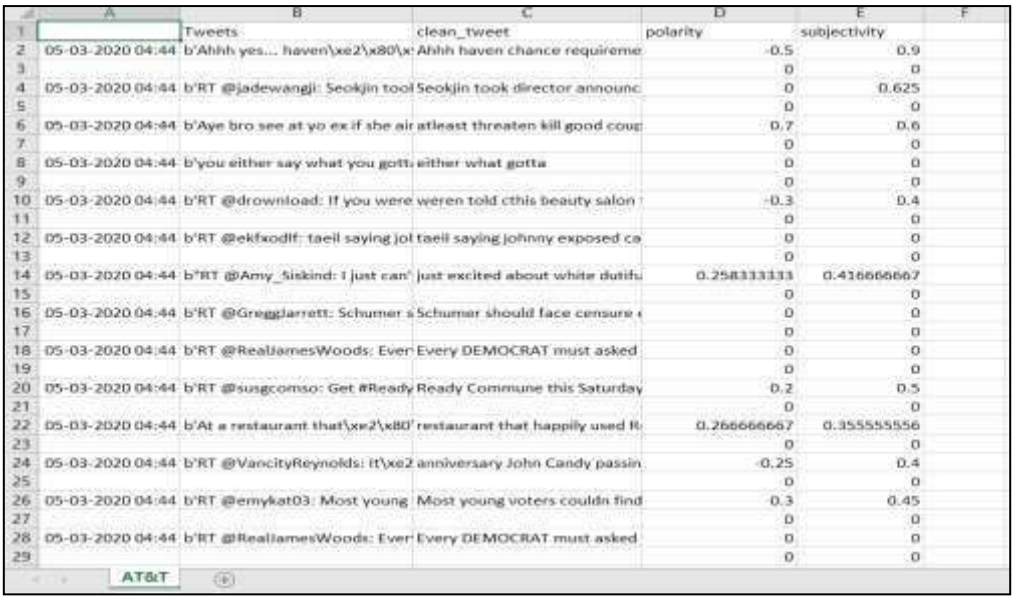
Polarity: Polarity is a floating-point number that lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.

Subjectivity: The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.

Sentiment Property	Range
Polarity	-1 to +1
Subjectivity	0.0 to 1.0

Figure 14: Range for Sentiment Properties

Write a python script for analyzing the sentiments using libraries and download the results in a '.csv' file.



	A	B	C	D	E	F
1		Tweets	clean_tweet	polarity	subjectivity	
2	05-03-2020 04:44	b'Ahhh yes... haven\xe2\x80\x	Ahhh haven chance requireme	-0.5	0.9	
3				0	0	
4	05-03-2020 04:44	b'RT @jadewangji: Seokjin tool	Seokjin took director announc	0	0.625	
5				0	0	
6	05-03-2020 04:44	b'Aye bro see at yo ex if she air	atleast threaten kill good cou	0.7	0.6	
7				0	0	
8	05-03-2020 04:44	b'you either say what you gott	either what gotta	0	0	
9				0	0	
10	05-03-2020 04:44	b'RT @download: If you were	weren told cthis beauty salon	-0.3	0.4	
11				0	0	
12	05-03-2020 04:44	b'RT @ekfxodlf: taeil saying jol	taeil saying johnny exposed ca	0	0	
13				0	0	
14	05-03-2020 04:44	b'RT @Amy_Siskind: I just can'	just excited about white dutifu	0.258333333	0.416666667	
15				0	0	
16	05-03-2020 04:44	b'RT @Gregglarrett: Schumer s	Schumer should face censure	0	0	
17				0	0	
18	05-03-2020 04:44	b'RT @RealJamesWoods: Ever	Every DEMOCRAT must asked	0	0	
19				0	0	
20	05-03-2020 04:44	b'RT @susgcomso: Get #Ready	Ready Commune this Saturday	0.2	0.5	
21				0	0	
22	05-03-2020 04:44	b'At a restaurant that\xe2\x80'	restaurant that happily used fi	0.266666667	0.355555556	
23				0	0	
24	05-03-2020 04:44	b'RT @VancityReynolds: It\xe2	anniversary John Candy passin	-0.25	0.4	
25				0	0	
26	05-03-2020 04:44	b'RT @emykat03: Most young	Most young voters couldn find	0.3	0.45	
27				0	0	
28	05-03-2020 04:44	b'RT @RealJamesWoods: Ever	Every DEMOCRAT must asked	0	0	
29				0	0	

AT&T

Figure 15: Results of Sentiment Analysis

Analyze the results by calculating the sentiment score of each company.

Sentiment Score = Polarity*Subjectivity

Compare the total sentiment score for each selected company.

Sr. No.	Telecom Company	Sentiment Score
1	AT&T	4.32%
2	Avaya	4.32%
3	CenturyLink	6.30%
4	Sparklight	4.68%
5	Sprint Corp.	2.28%
6	U.S. Cellular	5.34%
7	Verizon	5.06%

Figure 16: Resulted Sentiment Score

Step 5: Visualization

Create an Excel Dashboard representing the collected no. of tweets, cleaned tweets and the comparison for results of Sentiment Score.

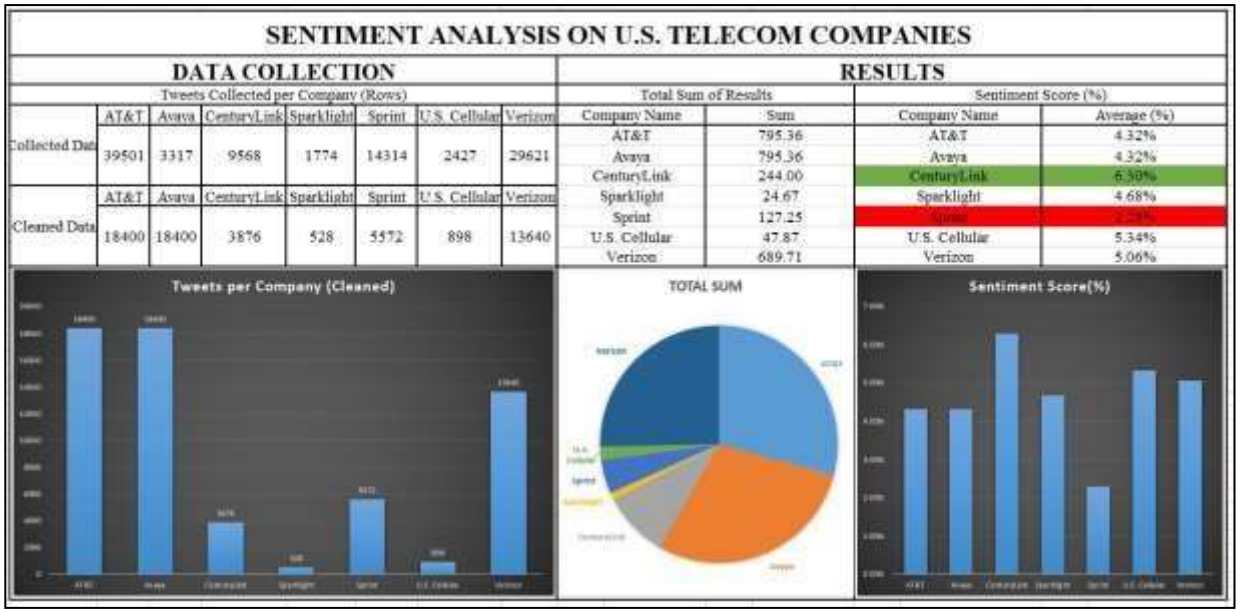


Figure 17: Excel Dashboard Representing Results

2.4 Hardware/Software Requirements:

a) Hardware Requirements: Followings are the minimum hardware requirements that are needed for the discussed project:

- Processor: Intel ® Core ™ i5-6440HQ CPU @ 2.60GHz 2.59GHz
- Installed Memory (RAM): 8.00 GB (7.64 GB Usable)
- System type: 64-bit Operating System

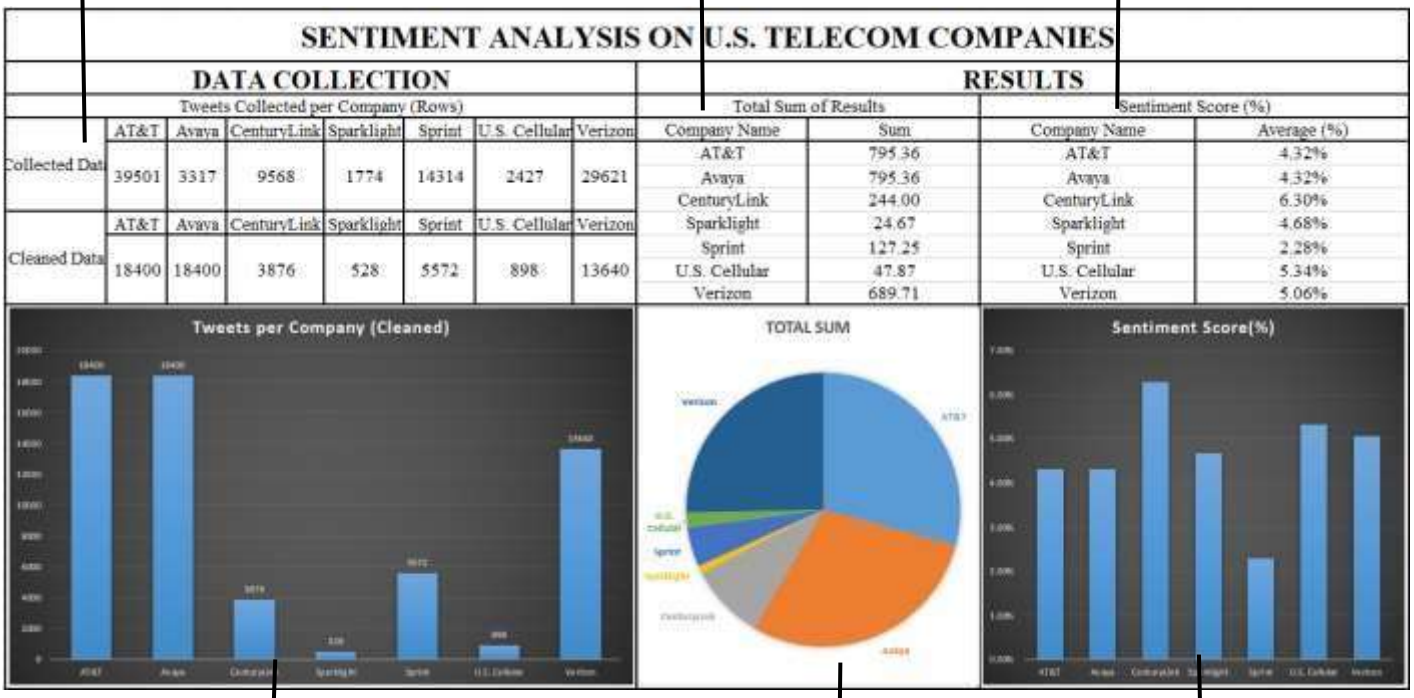
b) Software Requirements: Followings are the minimum software requirements that are needed for the discussed project:

- Back-end Data Service Provider: Twitter 9.4.10
- Operating System: Windows 10
- Front-end Tool: Microsoft Excel 2019, PyCharm 2021.2

User Guide:

3.1 Screen Shot with Labelling:

1. Collected & Cleaned Data **2. Total Sum of Results** **3. Sentiment Score (%)**



SENTIMENT ANALYSIS ON U.S. TELECOM COMPANIES

DATA COLLECTION

Tweets Collected per Company (Rows)

	AT&T	Avaya	CenturyLink	Sparklight	Sprint	U.S. Cellular	Verizon
Collected Data	39501	3317	9568	1774	14314	2427	29621
Cleaned Data	AT&T	Avaya	CenturyLink	Sparklight	Sprint	U.S. Cellular	Verizon
	18400	18400	3876	528	5572	898	13640

RESULTS

Total Sum of Results

Company Name	Sum
AT&T	795.36
Avaya	795.36
CenturyLink	244.00
Sparklight	24.67
Sprint	127.25
U.S. Cellular	47.87
Verizon	689.71

Sentiment Score (%)

Company Name	Average (%)
AT&T	4.32%
Avaya	4.32%
CenturyLink	6.30%
Sparklight	4.68%
Sprint	2.28%
U.S. Cellular	5.34%
Verizon	5.06%

4. Cleaned Tweets Per Company (Bar Graph) **5. Total Sum (Pie Chart)** **6. Sentiment Score (%) (Bar Graph)**

Conclusion & Future Scope:

4.1 Conclusion: On analysing sentiments from people's real-time reviews, and by visualizing the whole results, we get the conclusion that U.S. Telecom Company i.e. CenturyLink is most liked by the users as it gets the maximum positive reviews i.e. average (%) of the Sentiment score (from Dashboard).

Based on the Sentiment Analysis results for each selected company, we can conclude the following:

- ✓ **Most Liked Company:** CenturyLink
- ✓ **Least Liked Company:** Sprint Corp.

4.2 Future Scope: The Twitter sentiment score obtained from the dataset communities which is the online word of mouth representing the wisdom of the crowds accurately predicts the popularity and success of the Telecom operators. The success of Telecom operators is reflected in the correlation with the month-wise subscriber addition data. A positive sentiment score of a company is an indicator of the brand preference of the public and a negative sentiment score indicates customer dissatisfaction or inclination towards any other company which is better suited to their requirements. Data mining and sentiment analysis techniques can be used by managers to take timely actions to predict and prevent such customer churn.

References:

- **Sentiment Analysis: Mining Opinions, Sentiments, and Emotions ~ Bing Liu**
- **https://www.youtube.com/Code-With-Harry/Python-For-Absolute-Beginners/**
- **https://www.youtube.com/Code-With-Harry/Big-Data/**
- **https://www.youtube.com/Code-With-Harry/Machine-Learning/**
- (PDF) Research of Sentiment Analysis Tools (researchgate.net)