Industrial Training cum Summer Internship along with Project

**SCRAPPING DATA FROM TWITTER USING PYTHON**  
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTFOR THE AWARD OF THE DEGREE OF

**BACHELOR OF TECHNOLOGY**

(Electronics & Communication Engineering)



Date of Internship: From 10.12.2021 to 10.01.2022

**SUBMITTED BY: UNDER THE GUIDANCE OF:**

MOHAMMED FAYAZ PANTECH E-LEARNING

20671A0420

ELECTRONICS AND COMMUNICATIONS ENGINEERING

JB INSTITUTE OF ENGINEERING AND TECHNOLOGY

# BONAFIED CERTIFICATE

This is to certify that the project report entitled, **“SCRAPPING DATA FROM TWITTER USING PYTHON** “that is being submitted by **‘MOHAMMED FAYAZ (20671A0420)’** in a partial fulfilment of the requirements for the degree of Bachelor of Technology in Electronics and Communication Engineering of the Jawaharlal Nehru Technological University, Hyderabad, during the Academic year 2021-2022, is a bonafied record of work carried out under the guidance and supervision.

## Internship Guide: Head of The Department

Mrs.K.Deepa Rao Dr.Towheed Sultana

Asst.Professor Professor



This is to certify that Mr. / Ms. Mohammed Fayaz, Roll Number – (20671A0420), who is pursuing Electronics and Communication Engineering Department at JB Institute of Engineering and Technology (Autonomous), has successfully completed his/her Virtual Internship training program at, Pantech E Learning Pvt Ltd on (“**Scraping data from twitter**”) and has submitted the report.

During the internship period, the candidate has shown keen interest and commitment towards learning and his/her performance was good.

Period of Internship: From 10.12.2021 to 10.01.2022



(Branch Manager)

Pantech E Learning Pvt Ltd

4th Floor, Delta Chambers,

Behind Chennai Shopping Mall, Ameerpet, Hyderabad, Telangana – 500 016 Phone: +91 040-40077960. | hr@pantechmail.com

#### JB INSTITUTE OF ENGINEERING AND TECHNOLOGY

**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the Project entitled “**SCRAPPING DATA FROM TWITTER USING PYTHON**” by “**MOHAMMED FAYAZ (20671A0420)**” in partial fulfillment of requirements for the award of degree of B.Tech. (Electronics and Communication Engineering) submitted to the Department of Electronics and Communication Engineering at JB I N S T I T U T E O F ENGINEERING AND TECHNOLOGY is the authentic record of my own work carried out during a period from 10.12.2021 to 10.01.2021 The matter presented in this project has not been submitted by me or anybody else in any other University / Institute for the award of B.Tech Degree.

**Signature of the Student**

**ACKNOWLEDGEMENT**

I express our sincere gratitude and deep sense of respect to our beloved principal, **Dr. P.C.**

**Krishnamacharya principal, J.B. Institute of Engineering & Technology** for making us available all the required assistance and for his support and inspiration to carry Project work in the institute.

Our sincere thanks to **Dr. Towheed Sultana**, Professor and Head of Department, Department of Electronics and Communication Engineering for her valuable suggestions and advices during our completion of this project.

I express our sincere thanks to the management of J.B. Institute of Engineering & Technology, for giving us this opportunity to work in their esteemed organization.

**ABSTRACT**

Twitter is a popular social networking website where users posts and interact with messages known as “tweets”. This serves as a mean for individuals to express their thoughts or feelings about different subjects. Various different parties such as consumers and marketers have done sentiment analysis on such tweets to gather insights into products or to conduct market analysis. Furthermore, with the recent advancements in machine learning algorithms, we are able improve the accuracy of our sentiment analysis predictions.

In this report, we will attempt to conduct sentiment analysis on “tweets” using various different machine learning algorithms. We attempt to classify the polarity of the tweet where it is either positive or negative. If the tweet has both positive and negative elements, the more dominant

sentiment should be picked as the final label. We use the dataset from Kaggle which was crawled and labelled positive/negative. The data provided comes with emoticons, usernames and hashtags which are required to be processed and converted into a standard form. We also need to extract useful features from the text such unigrams and bigrams which is a form of representation of the “tweet”. We use various machine learning algorithms to conduct sentiment analysis using the extracted features. However, just relying on individual models did not give a high accuracy so we pick the top few models to generate a model.

#### TABLE OF CONTENTS

#### INTRODUCTION-------------------------------------------------------------1

#### LITERATURE SURVEY ----------------------------------------------------2

#### PROPOSED SYSTEM--------------------------------------------------------3

#### WHAT IS PYTHON----------------------------------------------------------4

#### PYTHON FEATURES---------------------------------------------------------5

#### HOW TO EXECUTE PYTHON---------------------------------------------7

#### BLOCK DIAGRAM-----------------------------------------------------------14

#### SCRAPPING DATA FROM TWITTER USING PYTHON ---------15

#### OUTPUT-------------------------------------------------------------------------21

#### TOOLS USED-------------------------------------------------------------------22

#### ADVANTAGES-----------------------------------------------------------------22

#### APPLICATIOSN---------------------------------------------------------------22

#### LIMITATIONS-----------------------------------------------------------------22

#### FUTURE SCOPE--------------------------------------------------------------23

#### REFERENCE------------------------------------------------------------------24

1. **INTRODUCTION**

With the huge amount of increase in the web technologies, the no of people expressing their views and the opinion via web are increasing. This information is useful for everyone like businesses, governments and individuals . with 500+ million tweets per day , twitter is becoming a major source of information. Twitter is a microblogging site, which is popularly known for its short messages known as tweets. It has a limit of 140 characters. Twitter has a user base of 240+ million active users and hence it is a useful source of information. The users often discuss their personal views on various subjects and also on current affairs via tweets. Out of all popular social medias like Facebook , Twitter, Google+, and Myspace we choose Twitter because of the reasons like :

* Twitter contains vast number of text posts and it grows day by day. The collected corpus can be arbitrarily large.
* Twitter’s audience varies from regular users to celebrities, Politicians , company representatives, and even country’s president. Therefore it is possible to collect text posts of users from different social and interests groups.
* Tweets are small in length and thus less ambiguous and are unbiased in nature.

Using social media, models are built for classifying “tweets” into positive , negative, and neutral classes .The models are build for two classification tasks : a 3-way classification of already separated phrases in a tweet into positive, negative , and neutral classes and another 3 way classifications of entire message into positive , negative and neutral classes.

This paper is experimented with baseline model and feature based model. An incremental analysis is done to the features. It is also experimented with a combination of models: combining baseline and feature based model. The baseline model is done to the phrase based classification task which achieves an accuracy of 62.24% and is 29% more than the chance baseline. The feature based model uses features and achieves an accuracy of 77.86%.

These combinations achieves an accuracy of 77.90% which outperforms the baseline by 16%. For message based classification task the baseline model comes out with 51% of accuracy which is 18% more than the chance baseline. The feature based model uses the features with the accuracy of 57.43% . The combination achieves 58.00% of accuracy which outperforms the baseline by 7%.

**Existing System:-**

* The biggest reason to adopt CNN in image analysis and classification is due to CNN can extract an area of features from global information, and it is able to consider the relationship among these features. The above solution can achieve a higher accuracy in analysis and classification.
* For natural language processing, texts data features also can be extracted piece by piece and to consider the relationship among these features, but without the consideration of context or whole sentence, the sentiment might be understood wrong.

**Proposed System:-**

* The raw twitter data is given as input to the system. The unstructured voluminous input data can be obtained from various product and twitter data for and external sources Algorithms are SVM algorithm and Naïve bayes classification.
* To explore Big Data, proposed system analyzed several challenges at the data, model, and system levels. To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data.
* At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values. In other situations, privacy concerns, noise, and errors can be introduced into the data, to produce altered data copies.

# 4.What is Python

**Python** is an object-oriented, high level language, interpreted, dynamic and multipurpose programming language.

Python is *easy to learn* yet powerful and versatile scripting language which makes it attractive for Application Development.

Python's syntax and *dynamic typing* with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas.

Python supports *multiple programming pattern*, including object oriented programming, imperative and functional programming or procedural styles.

Python is not intended to work on special area such as web programming. That is why it is known as *multipurpose* because it can be used with web, enterprise, 3D CAD etc.

We don't need to use data types to declare variable because it is *dynamically typed* so we can write a=10 to declare an integer value in a variable.

Python makes the development and debugging *fast* because there is no compilation step included in python development and edit-test-debug cycle is very fast.

# 5.Python Features

##### Easy to Use:

Python is easy to very easy to use and high level language. Thus it is programmer-friendly language.

##### Expressive Language:

Python language is more expressive. The sense of expressive is the code is easily understandable.

##### Interpreted Language:

Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.

##### Cross-platform language:

Python can run equally on different platforms such as Windows, Linux, Unix

, Macintosh etc. Thus, Python is a portable language.

##### Free and Open Source:

Python language is freely available(www.python.org).The source-code is also available. Therefore it is open source.

##### Object-Oriented language:

Python supports object oriented language. Concept of classes and objects comes into existence.

##### Extensible:

It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in your python code.

##### Large Standard Library:

Python has a large and broad library.

##### GUI Programming:

Graphical user interfaces can be developed using Python.

##### Integrated:

It can be easily integrated with languages like C, C++, JAVA etc.

**6.How to execute python**

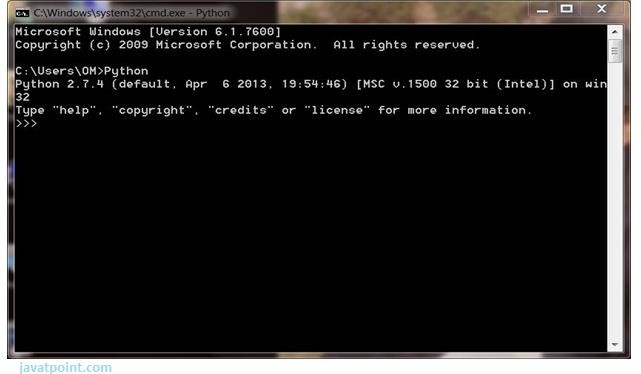
There are three different ways of working in Python:

1. Interactive Mode:

You can enter python in the command prompt and start working with Python.

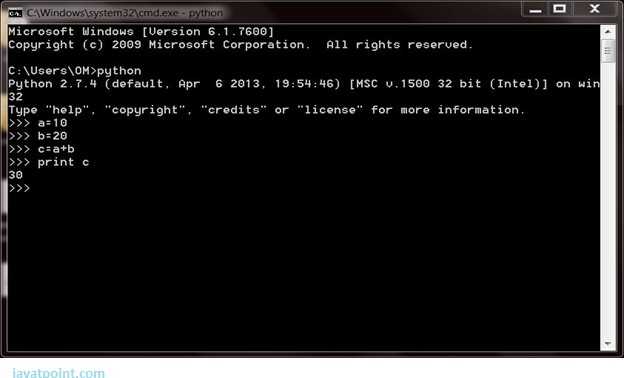


Press Enter key and the Command Prompt will appear like:



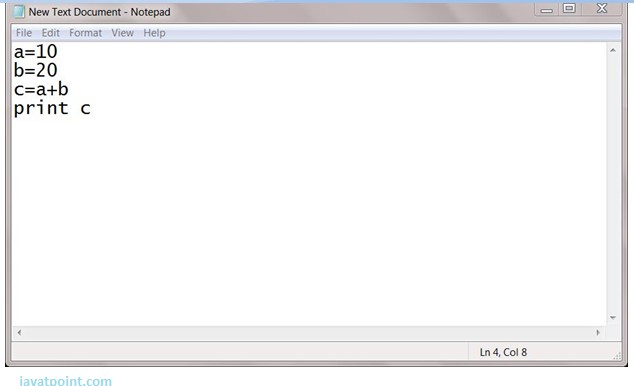
Now you can execute your Python commands.

#### Eg:

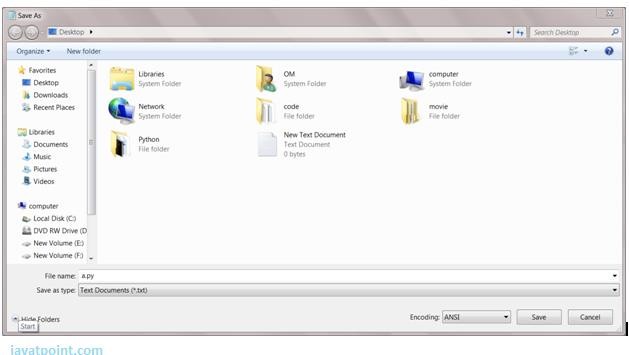


1. Script Mode:

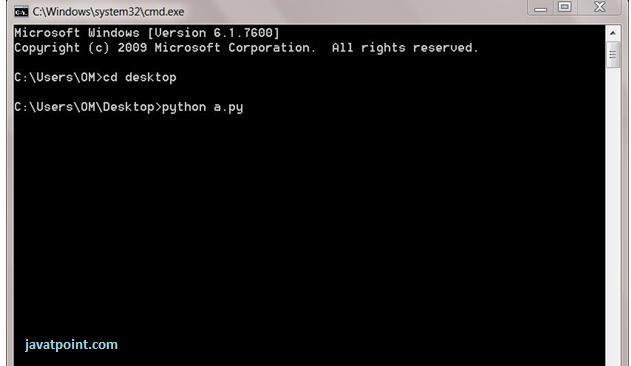
Using Script Mode , you can write your Python code in a separate file using any editor of your Operating System.



Save it by .py extension.



Now open Command prompt and execute it by :

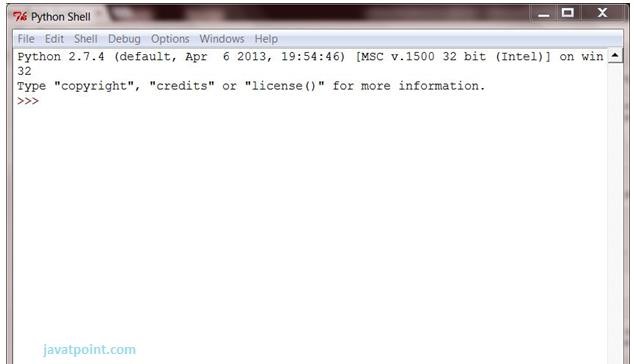


NOTE: Path in the command prompt should be where you have saved your file. In the above case file should be saved at desktop.

1. Using IDE: (Integrated Development Environment)

You can execute your Python code using a Graphical User Interface (GUI). All you need to do is:

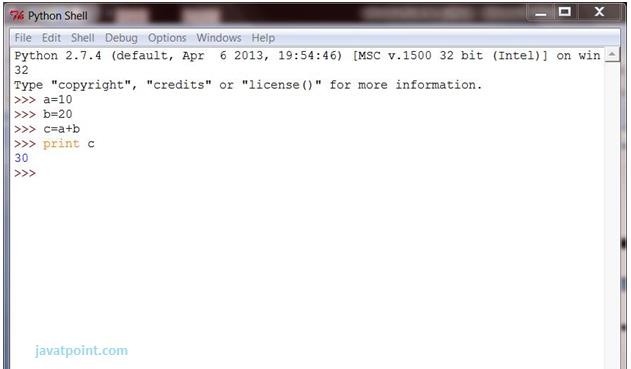
Click on Start button -> All Programs -> Python -> IDLE(Python GUI)



You can use both Interactive as well as Script mode in IDE.

#### Using Interactive mode:

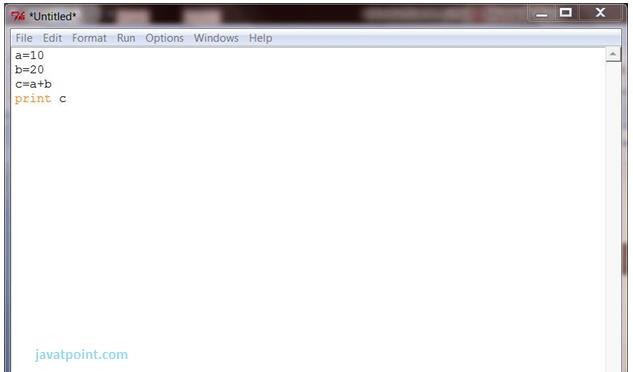
Execute your Python code on the Python prompt and it will display result simultaneously.



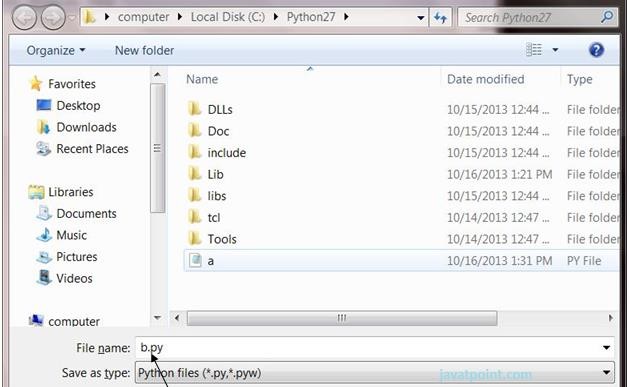
#### Using Script Mode:

* 1. Click on Start button -> All Programs -> Python -> IDLE(Python GUI)
  2. Python Shell will be opened. Now click on File -> New Window.

A new Editor will be opened . Write your Python code here.

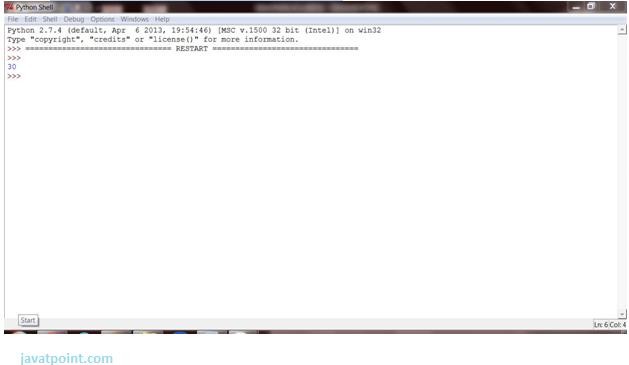


Click on file -> save as



Run then code by clicking on Run in the Menu bar. Run -> Run Module

Result will be displayed on a new Python shell as:



**CODE:-**

{"nbformat":4,"nbformat\_minor":0,"metadata":{"colab":{"name":"twitter sentiment analysis .ipynb","provenance":[]},"kernelspec":{"display\_name":"Python 3","language":"python","name":"python3"},"language\_info":{"codemirror\_mode":{"name":"ipython","version":3},"file\_extension":".py","mimetype":"text/x-python","name":"python","nbconvert\_exporter":"python","pygments\_lexer":"ipython3","version":"3.7.3"}},"cells":[{"cell\_type":"code","metadata":{"id":"KJT1d4LnHNBs","executionInfo":{"status":"ok","timestamp":1644928819754,"user\_tz":-330,"elapsed":477,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["import numpy as np\n","import pandas as pd\n","import matplotlib.pyplot as plt\n","from sklearn.metrics import confusion\_matrix\n"],"execution\_count":8,"outputs":[]},{"cell\_type":"code","metadata":{"id":"mjzL9-QQHNBy","colab":{"base\_uri":"https://localhost:8080/","height":310},"executionInfo":{"status":"error","timestamp":1644928842133,"user\_tz":-330,"elapsed":512,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}},"outputId":"53da57ca-bb22-41e8-9f7b-0edb7b913fda"},"source":["data = pd.read\_csv('twitter.csv')"],"execution\_count":11,"outputs":[{"output\_type":"error","ename":"FileNotFoundError","evalue":"ignored","traceback":["\u001b[0;31m---------------------------------------------------------------------------\u001b[0m","\u001b[0;31mFileNotFoundError\u001b[0m Traceback (most recent call last)","\u001b[0;32m<ipython-input-11-9edd0f3daf75>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m()\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mdata\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mpd\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mread\_csv\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m'twitter.csv'\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/util/\_decorators.py\u001b[0m in \u001b[0;36mwrapper\u001b[0;34m(\*args, \*\*kwargs)\u001b[0m\n\u001b[1;32m 309\u001b[0m \u001b[0mstacklevel\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mstacklevel\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 310\u001b[0m )\n\u001b[0;32m--> 311\u001b[0;31m \u001b[0;32mreturn\u001b[0m \u001b[0mfunc\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m\*\u001b[0m\u001b[0margs\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\*\*\u001b[0m\u001b[0mkwargs\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 312\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 313\u001b[0m \u001b[0;32mreturn\u001b[0m \u001b[0mwrapper\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/readers.py\u001b[0m in \u001b[0;36mread\_csv\u001b[0;34m(filepath\_or\_buffer, sep, delimiter, header, names, index\_col, usecols, squeeze, prefix, mangle\_dupe\_cols, dtype, engine, converters, true\_values, false\_values, skipinitialspace, skiprows, skipfooter, nrows, na\_values, keep\_default\_na, na\_filter, verbose, skip\_blank\_lines, parse\_dates, infer\_datetime\_format, keep\_date\_col, date\_parser, dayfirst, cache\_dates, iterator, chunksize, compression, thousands, decimal, lineterminator, quotechar, quoting, doublequote, escapechar, comment, encoding, encoding\_errors, dialect, error\_bad\_lines, warn\_bad\_lines, on\_bad\_lines, delim\_whitespace, low\_memory, memory\_map, float\_precision, storage\_options)\u001b[0m\n\u001b[1;32m 584\u001b[0m \u001b[0mkwds\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mupdate\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mkwds\_defaults\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 585\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m--> 586\u001b[0;31m \u001b[0;32mreturn\u001b[0m \u001b[0m\_read\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mfilepath\_or\_buffer\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0mkwds\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 587\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 588\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/readers.py\u001b[0m in \u001b[0;36m\_read\u001b[0;34m(filepath\_or\_buffer, kwds)\u001b[0m\n\u001b[1;32m 480\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 481\u001b[0m \u001b[0;31m# Create the parser.\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m--> 482\u001b[0;31m \u001b[0mparser\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mTextFileReader\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mfilepath\_or\_buffer\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\*\*\u001b[0m\u001b[0mkwds\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 483\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 484\u001b[0m \u001b[0;32mif\u001b[0m \u001b[0mchunksize\u001b[0m \u001b[0;32mor\u001b[0m \u001b[0miterator\u001b[0m\u001b[0;34m:\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/readers.py\u001b[0m in \u001b[0;36m\_\_init\_\_\u001b[0;34m(self, f, engine, \*\*kwds)\u001b[0m\n\u001b[1;32m 809\u001b[0m \u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0moptions\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m\"has\_index\_names\"\u001b[0m\u001b[0;34m]\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mkwds\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m\"has\_index\_names\"\u001b[0m\u001b[0;34m]\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 810\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m--> 811\u001b[0;31m \u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0m\_engine\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0m\_make\_engine\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mengine\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 812\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 813\u001b[0m \u001b[0;32mdef\u001b[0m \u001b[0mclose\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mself\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m:\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/readers.py\u001b[0m in \u001b[0;36m\_make\_engine\u001b[0;34m(self, engine)\u001b[0m\n\u001b[1;32m 1038\u001b[0m )\n\u001b[1;32m 1039\u001b[0m \u001b[0;31m# error: Too many arguments for \"ParserBase\"\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m-> 1040\u001b[0;31m \u001b[0;32mreturn\u001b[0m \u001b[0mmapping\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0mengine\u001b[0m\u001b[0;34m]\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mf\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\*\*\u001b[0m\u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0moptions\u001b[0m\u001b[0;34m)\u001b[0m \u001b[0;31m# type: ignore[call-arg]\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 1041\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 1042\u001b[0m \u001b[0;32mdef\u001b[0m \u001b[0m\_failover\_to\_python\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mself\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m:\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/c\_parser\_wrapper.py\u001b[0m in \u001b[0;36m\_\_init\_\_\u001b[0;34m(self, src, \*\*kwds)\u001b[0m\n\u001b[1;32m 49\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 50\u001b[0m \u001b[0;31m# open handles\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m---> 51\u001b[0;31m \u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0m\_open\_handles\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0msrc\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0mkwds\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 52\u001b[0m \u001b[0;32massert\u001b[0m \u001b[0mself\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mhandles\u001b[0m \u001b[0;32mis\u001b[0m \u001b[0;32mnot\u001b[0m \u001b[0;32mNone\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 53\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/parsers/base\_parser.py\u001b[0m in \u001b[0;36m\_open\_handles\u001b[0;34m(self, src, kwds)\u001b[0m\n\u001b[1;32m 227\u001b[0m \u001b[0mmemory\_map\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mkwds\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mget\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m\"memory\_map\"\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;32mFalse\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 228\u001b[0m \u001b[0mstorage\_options\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mkwds\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mget\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m\"storage\_options\"\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;32mNone\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m--> 229\u001b[0;31m \u001b[0merrors\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mkwds\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mget\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m\"encoding\_errors\"\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\"strict\"\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 230\u001b[0m )\n\u001b[1;32m 231\u001b[0m \u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;32m/usr/local/lib/python3.7/dist-packages/pandas/io/common.py\u001b[0m in \u001b[0;36mget\_handle\u001b[0;34m(path\_or\_buf, mode, encoding, compression, memory\_map, is\_text, errors, storage\_options)\u001b[0m\n\u001b[1;32m 705\u001b[0m \u001b[0mencoding\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0mioargs\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mencoding\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[1;32m 706\u001b[0m \u001b[0merrors\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0merrors\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m--> 707\u001b[0;31m \u001b[0mnewline\u001b[0m\u001b[0;34m=\u001b[0m\u001b[0;34m\"\"\u001b[0m\u001b[0;34m,\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m\u001b[1;32m 708\u001b[0m )\n\u001b[1;32m 709\u001b[0m \u001b[0;32melse\u001b[0m\u001b[0;34m:\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n","\u001b[0;31mFileNotFoundError\u001b[0m: [Errno 2] No such file or directory: 'twitter.csv'"]}]},{"cell\_type":"code","metadata":{"colab":{"base\_uri":"https://localhost:8080/","height":165},"id":"emg4TL1LHNB0","outputId":"c6d787f3-a0f7-4b42-ae53-3fc52f63a6d9","executionInfo":{"status":"error","timestamp":1644928873362,"user\_tz":-330,"elapsed":500,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["data.head()"],"execution\_count":13,"outputs":[{"output\_type":"error","ename":"NameError","evalue":"ignored","traceback":["\u001b[0;31m---------------------------------------------------------------------------\u001b[0m","\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)","\u001b[0;32m<ipython-input-13-304fa4ce4ebd>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m()\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mdata\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mhead\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m","\u001b[0;31mNameError\u001b[0m: name 'data' is not defined"]}]},{"cell\_type":"code","metadata":{"id":"IozoOYWKHNB2","executionInfo":{"status":"ok","timestamp":1644928769233,"user\_tz":-330,"elapsed":505,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["import re \n","def remove\_pattern(input\_txt,pattern):\n"," r = re.findall(pattern,input\_txt)\n"," for i in r:\n"," input\_txt = re.sub(i,'',input\_txt)#sub(characters we want to keep, removed character replaced by space, string to work on)\n"," \n"," return input\_txt\n"],"execution\_count":3,"outputs":[]},{"cell\_type":"code","metadata":{"id":"tHNkFL1rHNB3","colab":{"base\_uri":"https://localhost:8080/","height":165},"executionInfo":{"status":"error","timestamp":1644928780756,"user\_tz":-330,"elapsed":525,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}},"outputId":"5ed3491c-56f7-4664-d985-c65126919d5f"},"source":["data['tidy\_tweet'] = np.vectorize(remove\_pattern)(data['tweet'], \"@[\\w]\*\")"],"execution\_count":5,"outputs":[{"output\_type":"error","ename":"NameError","evalue":"ignored","traceback":["\u001b[0;31m---------------------------------------------------------------------------\u001b[0m","\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)","\u001b[0;32m<ipython-input-5-4dde69953b8f>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m()\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tidy\_tweet'\u001b[0m\u001b[0;34m]\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mnp\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mvectorize\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mremove\_pattern\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tweet'\u001b[0m\u001b[0;34m]\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\"@[\\w]\*\"\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m","\u001b[0;31mNameError\u001b[0m: name 'data' is not defined"]}]},{"cell\_type":"code","metadata":{"id":"kwHoEQjsHNB4","colab":{"base\_uri":"https://localhost:8080/","height":182},"executionInfo":{"status":"error","timestamp":1644928788955,"user\_tz":-330,"elapsed":594,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}},"outputId":"ee8cc029-38cf-441f-c1d6-0ee996ce95fc"},"source":["import string\n","data['tidy\_tweet'] = data['tidy\_tweet'].str.replace(\"[^a-zA-Z#]\", \" \")"],"execution\_count":6,"outputs":[{"output\_type":"error","ename":"NameError","evalue":"ignored","traceback":["\u001b[0;31m---------------------------------------------------------------------------\u001b[0m","\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)","\u001b[0;32m<ipython-input-6-97494f983134>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m()\u001b[0m\n\u001b[1;32m 1\u001b[0m \u001b[0;32mimport\u001b[0m \u001b[0mstring\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0;32m----> 2\u001b[0;31m \u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tidy\_tweet'\u001b[0m\u001b[0;34m]\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tidy\_tweet'\u001b[0m\u001b[0;34m]\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mstr\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mreplace\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m\"[^a-zA-Z#]\"\u001b[0m\u001b[0;34m,\u001b[0m \u001b[0;34m\" \"\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m","\u001b[0;31mNameError\u001b[0m: name 'data' is not defined"]}]},{"cell\_type":"code","metadata":{"id":"xS7Zni1xHNB4","colab":{"base\_uri":"https://localhost:8080/","height":165},"executionInfo":{"status":"error","timestamp":1644928799270,"user\_tz":-330,"elapsed":514,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}},"outputId":"f9293398-89fc-491a-c91a-6bb5852ababf"},"source":["data['tidy\_tweet'] = data['tidy\_tweet'].apply(lambda x: ' '.join([w for w in x.split() if len(w)>3 ]))"],"execution\_count":7,"outputs":[{"output\_type":"error","ename":"NameError","evalue":"ignored","traceback":["\u001b[0;31m---------------------------------------------------------------------------\u001b[0m","\u001b[0;31mNameError\u001b[0m Traceback (most recent call last)","\u001b[0;32m<ipython-input-7-72c31d398655>\u001b[0m in \u001b[0;36m<module>\u001b[0;34m()\u001b[0m\n\u001b[0;32m----> 1\u001b[0;31m \u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tidy\_tweet'\u001b[0m\u001b[0;34m]\u001b[0m \u001b[0;34m=\u001b[0m \u001b[0mdata\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0;34m'tidy\_tweet'\u001b[0m\u001b[0;34m]\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mapply\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;32mlambda\u001b[0m \u001b[0mx\u001b[0m\u001b[0;34m:\u001b[0m \u001b[0;34m' '\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0mjoin\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m[\u001b[0m\u001b[0mw\u001b[0m \u001b[0;32mfor\u001b[0m \u001b[0mw\u001b[0m \u001b[0;32min\u001b[0m \u001b[0mx\u001b[0m\u001b[0;34m.\u001b[0m\u001b[0msplit\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0;34m)\u001b[0m \u001b[0;32mif\u001b[0m \u001b[0mlen\u001b[0m\u001b[0;34m(\u001b[0m\u001b[0mw\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m>\u001b[0m\u001b[0;36m3\u001b[0m \u001b[0;34m]\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m)\u001b[0m\u001b[0;34m\u001b[0m\u001b[0;34m\u001b[0m\u001b[0m\n\u001b[0m","\u001b[0;31mNameError\u001b[0m: name 'data' is not defined"]}]},{"cell\_type":"code","metadata":{"id":"QxHFSUs8HNB5","executionInfo":{"elapsed":56,"status":"aborted","timestamp":1644928592196,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["data.head()"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"5lV7PU9OHNB6","executionInfo":{"status":"aborted","timestamp":1644928592198,"user\_tz":-330,"elapsed":57,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["tokenized\_tweet = data['tidy\_tweet'].apply(lambda x: x.split())"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"IVLSOTBDHNB7","executionInfo":{"elapsed":58,"status":"aborted","timestamp":1644928592200,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["tokenized\_tweet.head()"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"\_T66zZbaHNB8","executionInfo":{"elapsed":62,"status":"aborted","timestamp":1644928592205,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["import nltk\n","from nltk.stem.porter import \*\n","stemmer = PorterStemmer()\n","tokenized\_tweet = tokenized\_tweet.apply(lambda x : [stemmer.stem(i) for i in x])\n","tokenized\_tweet.head()"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"2Se2Ov\_oHNB8","executionInfo":{"status":"aborted","timestamp":1644928592207,"user\_tz":-330,"elapsed":64,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["for i in range(len(tokenized\_tweet)):\n"," tokenized\_tweet[i] = ' '.join(tokenized\_tweet[i])\n","\n","data['tidy\_tweet'] = tokenized\_tweet"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"O0lEXA2fHNB9","executionInfo":{"status":"aborted","timestamp":1644928592208,"user\_tz":-330,"elapsed":64,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["all\_words = ' '.join([text for text in data['tidy\_tweet']])\n"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"Qp9C5R3NHNB9","executionInfo":{"status":"aborted","timestamp":1644928592210,"user\_tz":-330,"elapsed":66,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["normal\_words =' '.join([text for text in data['tidy\_tweet'][data['label'] == 0]])"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"qcHtR-5YHNB-","executionInfo":{"status":"aborted","timestamp":1644928592211,"user\_tz":-330,"elapsed":66,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["negative\_words = ' '.join([text for text in data['tidy\_tweet'][data['label'] == 1]])"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"pXy8ZW90HNB-","executionInfo":{"status":"aborted","timestamp":1644928592212,"user\_tz":-330,"elapsed":67,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["def hashtag\_extract(x):\n"," hashtags = []\n"," # Loop over the words in the tweet\n"," for i in x:\n"," ht = re.findall(r\"#(\\w+)\", i)\n"," hashtags.append(ht)\n","\n"," return hashtags"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"vnX3DZLMHNB\_","executionInfo":{"status":"aborted","timestamp":1644928592213,"user\_tz":-330,"elapsed":67,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["HT\_regular = hashtag\_extract(data['tidy\_tweet'][data['label'] == 0])\n","\n","\n","HT\_negative = hashtag\_extract(data['tidy\_tweet'][data['label'] == 1])\n","\n","# unnesting list\n","HT\_regular = sum(HT\_regular,[])\n","HT\_negative = sum(HT\_negative,[])"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"tQPm5ufoHNCA","executionInfo":{"status":"aborted","timestamp":1644928592214,"user\_tz":-330,"elapsed":68,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["# Bag of words\n","from sklearn.feature\_extraction.text import CountVectorizer\n","bow\_vectorizer = CountVectorizer(token\_pattern=r\"(?u)\\b\\w+\\b\",stop\_words=None, ngram\_range=(2,2), analyzer='word')\n","# bag-of-words feature matrix\n","bow = bow\_vectorizer.fit\_transform(data['tidy\_tweet'])"],"execution\_count":null,"outputs":[]},{"cell\_type":"markdown","metadata":{"id":"LllKupJVh0GI"},"source":[""]},{"cell\_type":"code","metadata":{"id":"bYShTIB6HNCA","executionInfo":{"status":"aborted","timestamp":1644928592216,"user\_tz":-330,"elapsed":70,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["#TF-IDF\n","from sklearn.feature\_extraction.text import TfidfVectorizer\n","tfidf\_vectorizer = TfidfVectorizer(token\_pattern=r\"(?u)\\b\\w+\\b\", stop\_words=None, ngram\_range=(2,2), analyzer='word')\n","# TF-IDF feature matrix\n","tfidf = tfidf\_vectorizer.fit\_transform(data['tidy\_tweet'])"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"OaaceNa\_HNCB","executionInfo":{"elapsed":71,"status":"aborted","timestamp":1644928592218,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["bow[1]"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"qIt9iY\_HHNCC","executionInfo":{"elapsed":74,"status":"aborted","timestamp":1644928592221,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["data.head()"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"KDYgHJm9HNCC","executionInfo":{"elapsed":75,"status":"aborted","timestamp":1644928592222,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"},"user\_tz":-330}},"source":["#On Bag of Words Model\n","from sklearn.linear\_model import LogisticRegression\n","from sklearn.model\_selection import train\_test\_split\n","from sklearn.metrics import f1\_score\n","\n","\n","# splitting data into training and validation set\n","xtrain\_bow, xvalid\_bow, ytrain, yvalid = train\_test\_split(bow, data['label'], random\_state=42, test\_size=0.1)\n","\n","lreg = LogisticRegression()\n","lreg.fit(xtrain\_bow, ytrain) # training the model\n","\n","prediction = lreg.predict\_proba(xvalid\_bow) # predicting on the validation set\n","lreg.score(xvalid\_bow,yvalid)\n","#prediction\_int = prediction[:,1] >= 0.3 # if prediction is greater than or equal to 0.3 than 1 else 0\n","#prediction\_int = prediction\_int.astype(np.int)\n","\n","#f1\_score(yvalid, prediction\_int) # calculating f1 score"],"execution\_count":null,"outputs":[]},{"cell\_type":"markdown","metadata":{"id":"sZgTKfeAHNCD"},"source":["# Random Forest"]},{"cell\_type":"code","metadata":{"id":"uhHDzk5ZHNCE","executionInfo":{"status":"aborted","timestamp":1644928592223,"user\_tz":-330,"elapsed":75,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["from sklearn.ensemble import RandomForestClassifier"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"QSBiasdCHNCE","executionInfo":{"status":"aborted","timestamp":1644928592225,"user\_tz":-330,"elapsed":76,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["random=RandomForestClassifier()"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"YHzI-nMDHNCE","executionInfo":{"status":"aborted","timestamp":1644928592227,"user\_tz":-330,"elapsed":78,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["random.fit(xtrain\_bow,ytrain)"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"V4dh84gwHNCF","executionInfo":{"status":"aborted","timestamp":1644928592228,"user\_tz":-330,"elapsed":79,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["y\_pred = random.predict(xvalid\_bow)"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"iGrkoHjgHNCF","executionInfo":{"status":"aborted","timestamp":1644928592230,"user\_tz":-330,"elapsed":81,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["random.score(xvalid\_bow,yvalid)"],"execution\_count":null,"outputs":[]},{"cell\_type":"markdown","metadata":{"id":"E1dBdx78HNCG"},"source":["# SVM"]},{"cell\_type":"code","metadata":{"id":"NDo0sSxVHNCG","executionInfo":{"status":"aborted","timestamp":1644928592231,"user\_tz":-330,"elapsed":81,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["from sklearn import svm\n","s=svm.SVC()\n","s.fit(xtrain\_bow,ytrain)"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"dq-neQ3lHNCG","executionInfo":{"status":"aborted","timestamp":1644928592233,"user\_tz":-330,"elapsed":83,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["y\_pred =s.predict(xvalid\_bow)"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"o8VXOewqHNCH","executionInfo":{"status":"aborted","timestamp":1644928592234,"user\_tz":-330,"elapsed":84,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":["print (s.score(xvalid\_bow,yvalid))"],"execution\_count":null,"outputs":[]},{"cell\_type":"code","metadata":{"id":"6bUqeEIPHNCH","executionInfo":{"status":"aborted","timestamp":1644928592236,"user\_tz":-330,"elapsed":85,"user":{"displayName":"pantech e learning","photoUrl":"https://lh3.googleusercontent.com/a/default-user=s64","userId":"02085171050783995221"}}},"source":[""],"execution\_count":null,"outputs":[]}]}

**10.Tool used:**

* Python 3.10.0
* t\_movie()
* Theater()
* Timing()
* Movie()
* Center()
* City()

### 11.Advantages:

* Easy and user-friendly
* Visually appealing design

### 12.Applications:

* Manage the information of customer
* Integration of all record of seats
* It tracks all the information of customer,payment,shows etc

### 13.Limitations:

* Less flexibility:only pre-programmed instructions can be executed
* System functionality cannot be adjusted or adapted

## 15.REFERENCES:

* + 1. ["The ineluctable middlemen"](http://www.economist.com/node/21560866). *The Economist*. 25 August 2012. Retrieved 29 August 2012.
    2. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-2) [How a Computer Outage Can Take Down a Whole Airline](https://www.wired.com/2016/08/computer-outage-can-take-whole-airline/)
    3. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-3) [American Airlines and others carriers were left helpless after](https://www.businessinsider.com.au/airlines-left-crippled-with-delays-after-sabre-system-outage-2021-5?r=US&IR=T) [a system outage crippled operations, causing delays](https://www.businessinsider.com.au/airlines-left-crippled-with-delays-after-sabre-system-outage-2021-5?r=US&IR=T)
    4. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-4) [A computer-system outage grounded Southwest Airlines](https://www.businessinsider.com.au/southwest-airlines-flight-delays-caused-by-computer-outage-2021-6?r=US&IR=T) [flights, causing delays for the second day in a row](https://www.businessinsider.com.au/southwest-airlines-flight-delays-caused-by-computer-outage-2021-6?r=US&IR=T)
    5. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-5) [Travellers still facing delays as Virgin and Rex airlines try to](https://www.abc.net.au/news/2021-05-21/virgin-and-rex-computer-outage-impacting-flights/100157452) [recover after global IT outage](https://www.abc.net.au/news/2021-05-21/virgin-and-rex-computer-outage-impacting-flights/100157452)
    6. ^ [Jump up to:***a***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-mars1_6-0) [***b***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-mars1_6-1) [【Hitachi and Japanese National Railways】MARS-1](http://museum.ipsj.or.jp/en/computer/dawn/0030.html), [Information Processing Society of Japan](https://en.wikipedia.org/wiki/Information_Processing_Society_of_Japan)
    7. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-dawn_7-0) [Early Computers: Brief History](http://museum.ipsj.or.jp/en/computer/dawn/history.html), [Information Processing](https://en.wikipedia.org/wiki/Information_Processing_Society_of_Japan) [Society of Japan](https://en.wikipedia.org/wiki/Information_Processing_Society_of_Japan)
    8. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-8) R. Blair Smith, OH 34. Oral history interview by Robina Mapstone, May 1980. Charles Babbage Institute, University of Minnesota,

Minneapolis. <http://www.cbi.umn.edu/oh/display.phtml?id=9>

Archived 2002-08-16 at the [Wayback Machine](https://en.wikipedia.org/wiki/Wayback_Machine)

* + 1. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-9) Strauss, Michael (2010), [Value Creation in Travel](https://www.amazon.com/dp/0557612462) [Distribution](https://www.amazon.com/dp/0557612462)
    2. [**^**](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-10) "American Airlines - Direct Connect".

Directconnect.aa.com. Archived from [the original](http://directconnect.aa.com/Default.aspx) on 2012-08- 31. Retrieved 2012-11-08.

* + 1. ^ [Jump up](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-0)

[to:***a***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-0)[***b***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-1)[***c***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-2)[***d***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-3) [***e***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-4)[***f***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-5) [***g***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-6)[***h***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-7)[***i***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-8)[***j***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-9) [***k***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-10)[***l***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-11) [***m***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-12)[***n***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-13) [***o***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-14)[***p***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-15)[***q***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-16) [***r***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-17)[***s***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-18)[***t***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-19) [***u***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-20)[***v***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-21)[***w***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-22)[***x***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-23) [***y***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-24)[***z***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-25) [***aa***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-26)[***ab***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-27) [***ac***](https://en.wikipedia.org/wiki/Computer_reservation_system#cite_ref-CH-Aviation_11-28)["InteliSys amelia](https://www.ch-aviation.com/portal/airlines/it/system/AMEL) [RES"](https://www.ch-aviation.com/portal/airlines/it/system/AMEL). *Ch-aviation*. Retrieved 4 November 2020.