# Abstract

Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. We are beginning to understand how some of this information can be utilized to improve the users’ experiences with interfaces and with one another.

In this paper, we are interested in the personality of users. Personality has been shown to be relevant to many types of interactions; it has been shown to be useful in predicting job satisfaction, professional and romantic relationship success, and even preference for different interfaces. Until now, to accurately gauge users’ personalities, they needed to take a personality test. This made it impractical to use personality analysis in many social media domains. In this paper, we present a method by which a user’s personality can be somewhat predicted through the publicly available information on their Twitter tweets. We will describe the type of data collected, our methods of analysis, and the machine learning techniques that allow us to successfully predict personality. We then discuss the implications this has for social media design, interface design, and broader domains.

**Table of Contents**

[Acknowledgements 2](#_Toc74818908)

[Abstract 3](#_Toc74818909)

[Chapter 1 Introduction 5](#_Toc74818910)

[1.1 Background of the Project 5](#_Toc74818911)

[Chapter 2 System Analysis 6](#_Toc74818912)

[2.1 Software and Hardware Requirement Specification 6](#_Toc74818913)

[2.2 Concepts Used 6](#_Toc74818914)

[2.3 Software Tools Used 6](#_Toc74818915)

[2.3.1 Jupyter 6](#_Toc74818916)

[2.3.2 Ms Excel 6](#_Toc74818917)

[Chapter 3 Project Objectives 7](#_Toc74818918)

[3.1 Personality Research and Social Media 7](#_Toc74818919)

[Chapter 4 Data used 8](#_Toc74818920)

[4.1 Extrovert data 8](#_Toc74818921)

[4.2 Introvert data 8](#_Toc74818922)

[4.3 Elon Musk's Tweets 9](#_Toc74818923)

[Chapter 5 Code and Running time screenshots 10](#_Toc74818924)

[5.1 Source Code 10](#_Toc74818925)

[5.2 Running Time Screenshots 12](#_Toc74818926)

[Chapter 6 Conclusion and Future Work 14](#_Toc74818927)

[6.1 Future Work 14](#_Toc74818928)

[6.2 Conclusion 14](#_Toc74818929)

[Chapter 7 References 15](#_Toc74818930)

# 

# Chapter 1

# Introduction

Social networking on the web has grown dramatically over the last decade. In January 2005, a survey of social networking websites estimated that among all sites on the web there were roughly 115 million members. Just over five years later, Twitter alone has exceeded 200 million members. In the process of creating social networking profiles, users reveal a lot about themselves both in what they share and how they say it. Through self-description, status updates, photos, and interests, much of a user’s personality comes out through their profile.

## 1.1 Background of the Project

For decades, psychology researchers have worked to understand personality in a systematic way. After extensive work to develop and validate a widely accepted personality model, researchers have shown connections between general personality traits and many types of behavior. Relationships have been discovered between personality and psychological disorders, job performance and satisfaction, and even romantic success.

We have a supervised this project to predict personality of a person using his twitter tweets and label him as an introvert or extrovert.

# Chapter 2

# System Analysis

In this, we will discuss and analyze about the developing process of Checking Personality of Different People including software requirement specification (SRS) and comparison between existing and proposed system.

## 2.1 Software and Hardware Requirement Specification

|  |  |
| --- | --- |
| Processor | Intel Core Processor or any other Processor with better performance |
| Operating System | Windows 7, Windows 10 |
| Memory | 4 GB RAM or more |
| Hard Disk Space | Minimum requirement is 5 - 10 MB |
| Software used | Jupyter, Ms Excel |
| Programming Language used | Python 3 or above installed in your system |

## 2.2 ****Concepts Used****

* Python Programming Language (Python 3 or above)
* Some basic machine learning algorithms

## 2.3 Software Tools Used

The whole project is done in back end only. The Console output is the front end generated from the compiler.

### 2.3.1 Jupyter

Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document.

### 2.3.2 Ms Excel

Microsoft Excel is the industry leading spreadsheet software program, a powerful data visualization and analysis tool. It has been a very widely applied spreadsheet for these platforms, especially since version 5 in 1993, and it has replaced Lotus 1-2-3 as the industry standard for spreadsheets. Excel forms part of the Microsoft Office suite of software.

# Chapter 3

# Project Objectives

This paper attempts to bridge the gap between social media and personality research by using the information people reveal in their tweets. Our core research question asks whether social media profiles can predict personality traits. If so, then there is an opportunity to integrate the many results on the implications of personality factors and behavior into the users’ online experiences and to use social media profiles as a source of information to better understand individuals. For example, the friend suggestion system could be tailored to a user based on whether they are more introverted or extraverted.

## 3.1 Personality Research and Social Media

It has been shown in that extroversion and conscientiousness positively correlate with the perceived ease of use of social media websites. extroversion was also shown to have a positive correlation with perceived usefulness of such sites. Not surprisingly, extroversion was also shown to correlate with the size of a user’s social network in several studies. There have also been mixed results for other personality traits. Work in showed that individuals with high agreeableness scores were selected more often as friends and that people tended to choose friends with similar agreeableness, extroversion, and openness scores. This was not repeated, but a correlation between openness and number of friends.

# Chapter 4

# Data used

## 4.1 Extrovert data

tonight,party,gonna,partying,rockin,town,poppin,club,messy,homies,bottles,champagne,rocking,drunken,booze,love,gotta,thunderstorms,love,miss,soo,guys,sooo,soooo,sooooo,missing,missed,soooooo,soooooooo,sooooooooo,youuu,privilege

night,tonight,girls,blast,forward,hanging,needed,bonfire,babysitting,sleepover,girly,partying,favorites,bestie,girlies love,loved,truely,freely,shown,dearly love,yo,adore,

xoxo,admire,extraordinary,absolutly,genuine,entitled,mentioned,growth

computer,windows,laptop,pc,drive,gb,mac,files,installed,virus,computers,system,software,install,xp it's,thing,sort,isn't,odd,suppose,strange,thinks, apparently, unusual, weird,telling, happening,fact,fairly computer,error,program,photoshop,server,properly,file,website,message,view,web,code,type,download,flash

computer,finally,internet,fixed,working,laptop,yay,works,comp,online,slow,computers,library,crashed,access

read,book,reading,pages,chapter,books,written,page,article,notes,chapters,write,story,history,newspaper

acid,body,brain,chemical,natural,oxygen,chemistry,liquid,process,energy,cells,science,reaction,levels,element

## 4.2 Introvert data

light,eyes,dark,soul,sky,sun,stars,darkness,moon,wind,earth,beauty,shadows,shadow,dreams

opinion,opinions,logic,based,political,fact,moral,beliefs,philosophy,argument,logical,facts,society,simply

human,beings,nature,spiritual,experience,compassion,sense,existence,reality,universe,humans,physical,divine,individual,humanity

i'm,letting,alive,i've,barely,holding,that's,breathing,moment,standing,falling,living,hanging,move,fine

i've,learned,lot,things,words,conclusion,forgotten,past,sum,recently,turned,learnt,noticed,lived,dealing

writing,write,read,story,poetry,poem,wrote,book,reading,stories,poems,written,creative,journal,writer

wait,till,can't,cnt,til,friday,tomorow,wed,nxt,cameron,tommorow,tmrw,wifey

great,lunch,nice,dinner,family,enjoyed,church,wonderful,afternoon,sunday,kids,evening,shopping,meeting,hubby

home,work,ready,hubby,bed,relax,dinner,shower,relaxing,heading,nap,early,chill,supper,snuggle happy,birthday,wishing,sister,years,wonderful,st,daughter,nephew,brother,son,turns,niece,special,celebrate

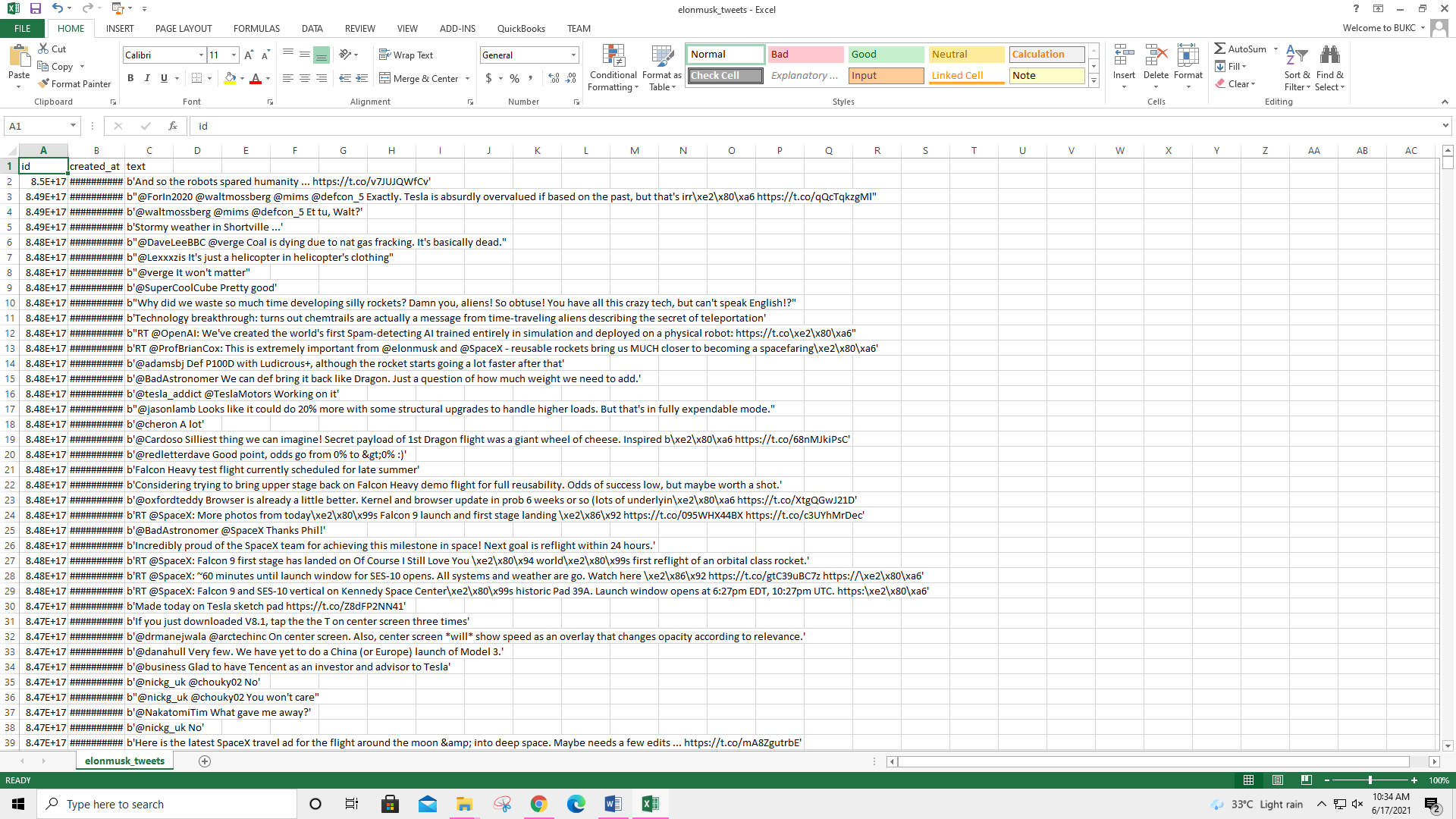
week,days,work,school,weeks,starting,till,yay,holiday,holidays,half,wednesday,schedule,stressful,hectic

week,weekend,great,forward,hope,fast,monday,end,coming,start,hoping,ready,sunday,flew

## 4.3 Elon Musk's Tweets

Source:

<https://www.kaggle.com/kulgen/elon-musks-tweets>



.

# Chapter 5

# Code and Running time screenshots

In this, we will share the source code and running time screenshots of the project to verify that the project is working properly.

## 5.1 Source Code

import pandas as pd

import numpy as np

from sklearn.tree import DecisionTreeClassifier

from sklearn.utils import shuffle

count=0

predicted1=0

total\_data=0

data= pd.read\_csv("C:\\Users\\user\\Desktop\\AI PRoject\\dataset\\elonmusk\_tweets.csv")

tweets=data.values[:,2]

print(tweets)

for i in range(0,len(tweets)):

print(f"index{i}",tweets[i])

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

from difflib import SequenceMatcher

import nltk

import csv

nltk.download('punkt')

nltk.download('stopwords')

labels=""

row=[]

stop\_words = set(stopwords.words('english'))

f1=open("C:\\Users\\user\\Desktop\\AI PRoject\\introvert.txt")

f2=open("C:\\Users\\user\\Desktop\\AI PRoject\\extrovert.txt")

file1=f1.read()

file2=f2.read()

print(file1)

print(file2)

for i in range(0,len(tweets)):

#tokennizing tweets

word\_token = word\_tokenize(tweets[i])

#print(word\_token)

#filtering sentences by removing stop words

filtered\_sentence = []

for w in word\_token:

if w not in stop\_words:

filtered\_sentence.append([w])

similarity\_intro=SequenceMatcher(None,file1,tweets[i]).ratio()

similarity\_extro=SequenceMatcher(None,file2,tweets[i]).ratio()

similarity\_intro\*=100

similarity\_extro\*=100

if similarity\_intro > similarity\_extro:

label="introvert"

elif similarity\_intro < similarity\_extro:

label="extrovert"

else:

label="neutral"

row.append([tweets[i],similarity\_intro,similarity\_extro,label])

with open('labels.csv', 'w', newline='') as file:

writer = csv.writer(file)

for i in range(0,len(row)):

writer.writerow(row[i])

import pandas as pd

data1=pd.read\_csv("C:\\Users\\user\\Desktop\\AI PRoject\\dataset\\labels.csv",names=["Tweets","Intro\_Ratio","Extro\_Ratio","Label"])

data1.head()

import seaborn as sns

from sklearn.feature\_extraction.text import CountVectorizer, TfidfVectorizer

from sklearn.naive\_bayes import MultinomialNB

from sklearn.model\_selection import train\_test\_split

import numpy as np

tweet\_get=data1["Tweets"]

label\_get=data1["Label"]

cv=TfidfVectorizer(min\_df=1,stop\_words='english')

#cv=CountVectorizer()

features\_train,features\_test,label\_train,label\_test = train\_test\_split(tweet\_get,label\_get,test\_size=0.2,random\_state=1)

features\_train\_cv=cv.fit\_transform(features\_train)

features\_toarray=features\_train\_cv.toarray()

print(features\_toarray)

print(cv.inverse\_transform(features\_toarray[0]))

mnb=MultinomialNB()

mnb.fit(features\_train\_cv,label\_train)

features\_test\_cv=cv.transform(features\_test)

#print(features\_test\_cv)

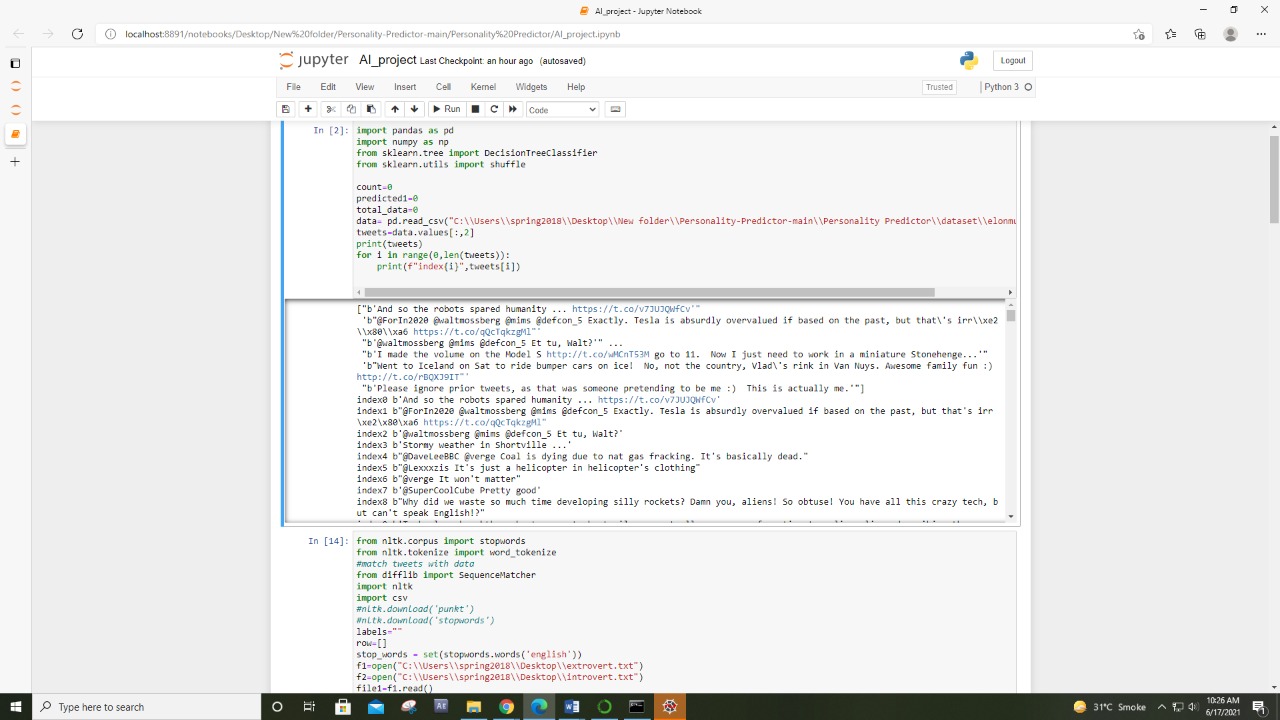
predict=mnb.predict(features\_test\_cv)

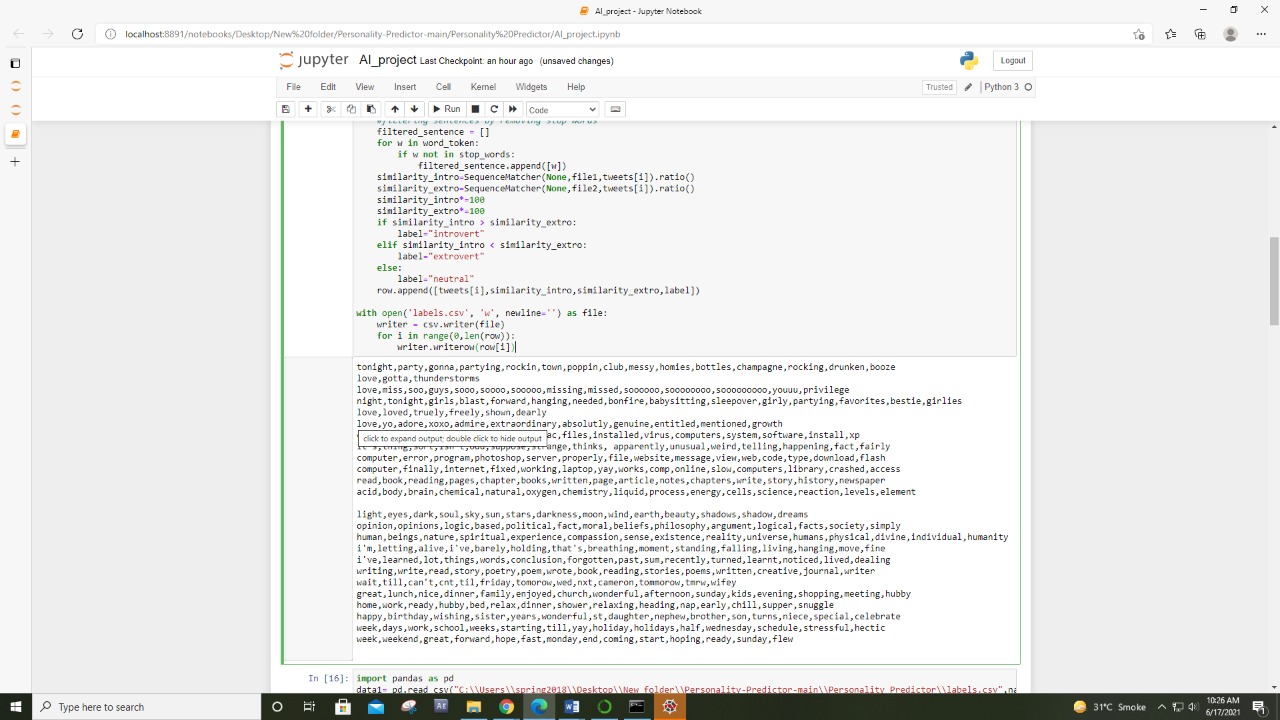
print(predict)

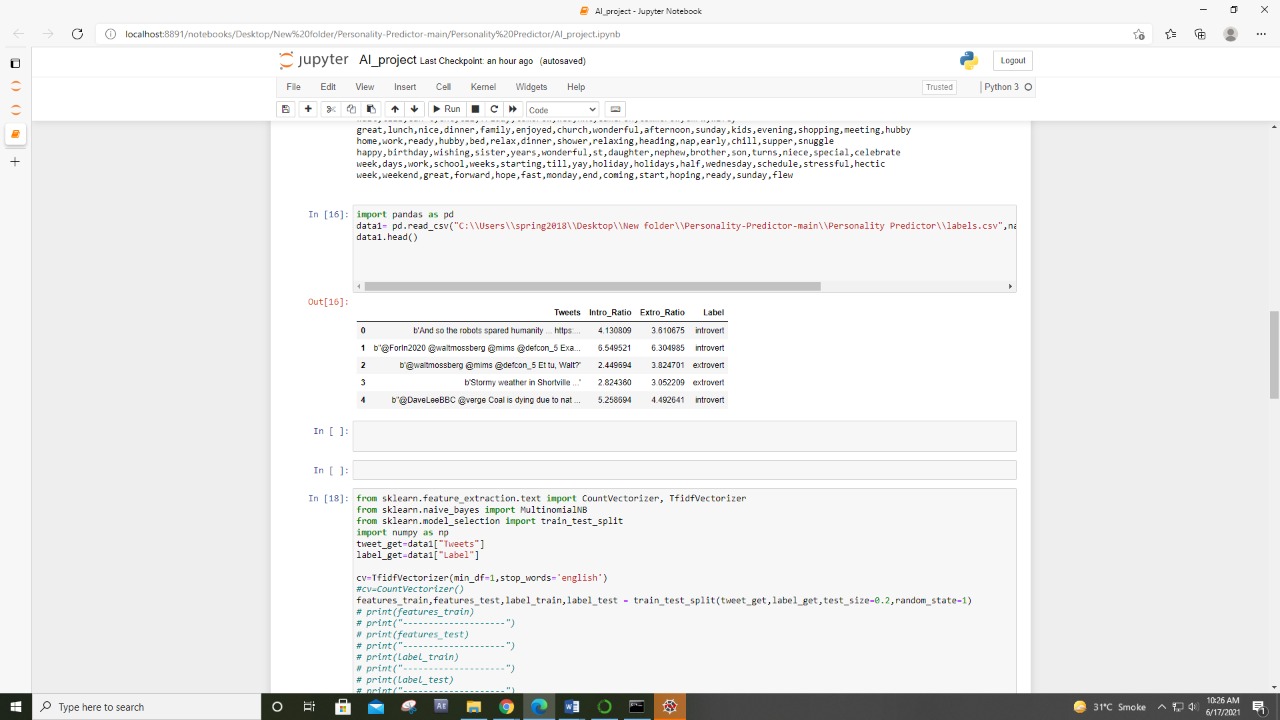
actual=np.array(label\_test)

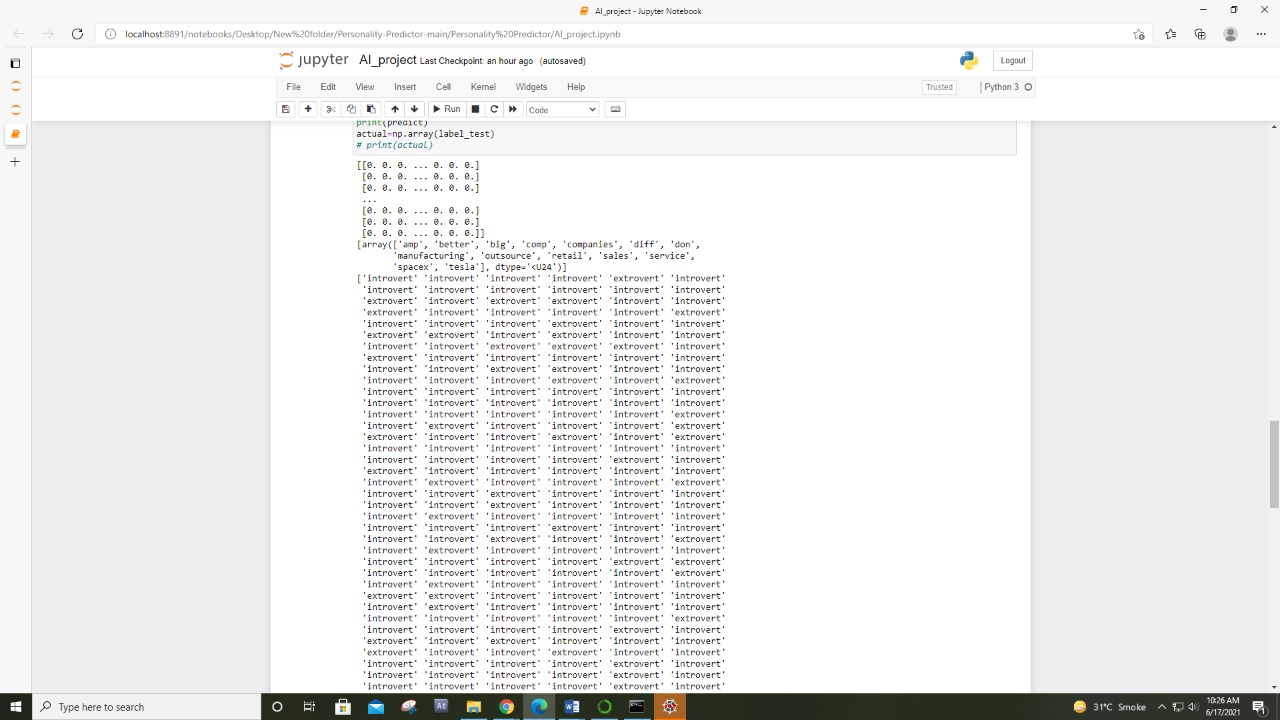
# print(actual)

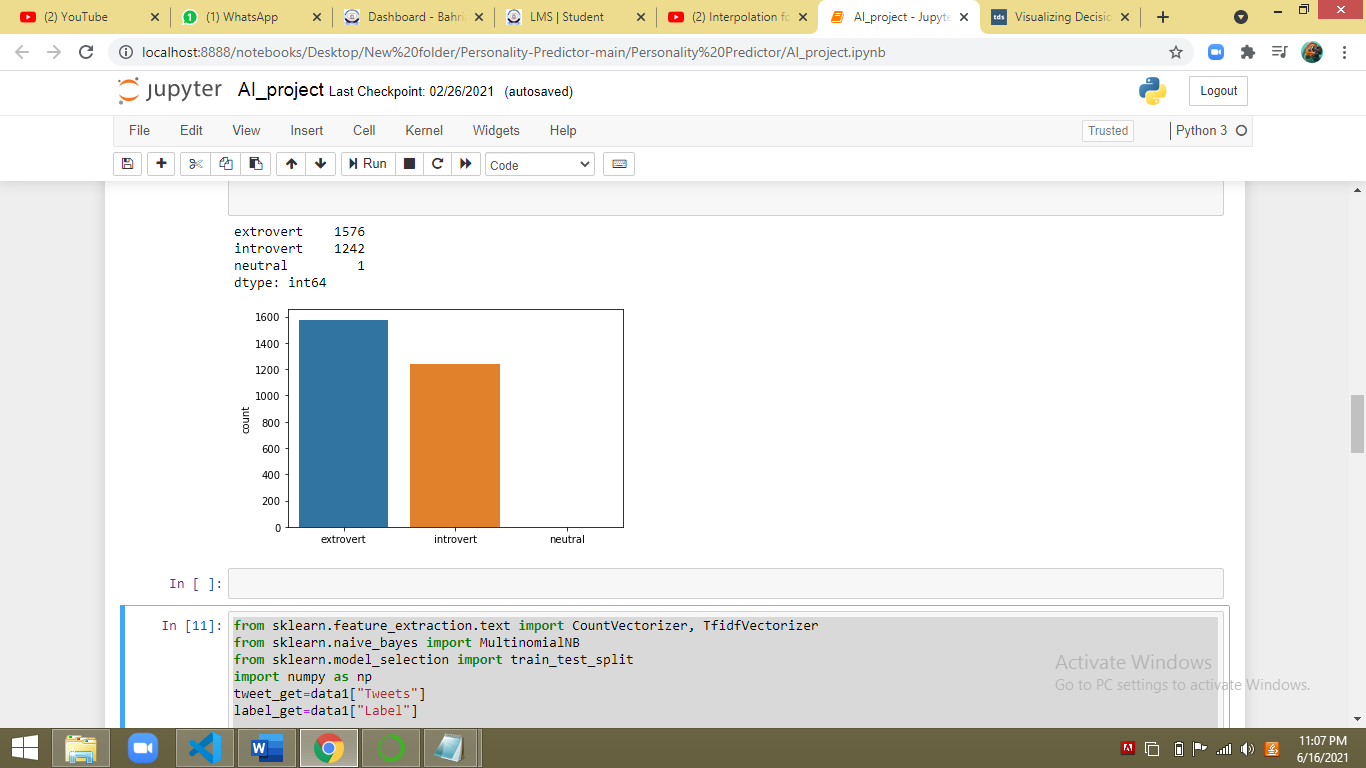
## 5.2 Running Time Screenshots











# Chapter 6

# Conclusion and Future Work

In this, we will discuss about the Future Work of this system along with Executive Summary.

## 6.1 Future Work

With the ability to guess a user’s personality traits, many opportunities are opened for personalizing interfaces and information. We discussed some of these opportunities for marketing and interface design above. However, there is much work to be pursued in this area.

One area that deserves attention is the connection between personality and the actual social network. We considered two structural features - number of friends and network density but we did not look at personality scores between friends. Understanding the connections between personality, tie strength, trust, and other related factors is an open space for research.

6.2 Conclusion

This prediction is almost 8-15% close to the original personality trait a person has. To predict the personality of a person, social media websites play a vital role. We have used the Twitter tweets to train our model and also to make the calculations of what type of a personality the person has more likely.

# Chapter 7

# References

Following are the links to the references which helped in understanding the Project and code:

* <https://www.demenzemedicinagenerale.net/pdf/2011%20%20Predicting%20Personality%20from%20Twitter.pdf>
* <https://www.skyfilabs.com/project-ideas/personality-prediction-using-machine-learning>