



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

PERSONALITY BEHAVIOUR PREDICTION

A Project Report of Capstone Project - 3

Submitted by

**ROSHAN KUMAR CHAUDHARY
(1613101597 / 16SCSE101192)**

*in partial fulfillment for the award of the degree
of*

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

**Mr. B. BHARATHI KANNAN
Professor**

APRIL/MAY- 2020



SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING
BONAFIDE CERTIFICATE

Certified that this project report **“PERSONALITY BEHAVIOUR PREDICTION”** is the bonafide work of **“ROSHAN KUMAR CHAUDHARY(1613101696)”** who carried out the project work under my supervision.

SIGNATURE OF HEAD

Dr.Satyajee Srivastava, Assoc. Prof.,

PhD (CS)

Professor

School of Computing Science &

Engineering

SIGNATURE OF SUPERVISOR

Mr.B. Bharathi Kannan, Asst. Prof.,

Professor

School of Computing Science &

Engineering

TABLE OF CONTENTS

TITLE	1
CERTIFICATE	2
ACKNOWLEDGEMENT	4
ABSTRACT	5
LIST OF TABLES	6-8
CHAPTERS	9-51
CHAPTER 1(INTRODUCTION)	9-10
CHAPTER 2(LITERATURE SURVEY)	11-12
CHAPTER 3(SYSTEM DESIGN)	13-17
CHAPTER 4(REQUIREMENTS)	18-27
CHAPTER 5(IMPLEMENTATION)	27-46
CHAPTER 6(CONCLUSION)	48
CHAPTER 7(OUTPUT)	49-57
CHAPTER 8(REFERENCES)	52

ACKNOWLEDGMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to **Mr. B. Bharathi kannan** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I am extremely indebted to Mr. B. Bharathi kannan Department of Computer Science and Engineering, Galgotias University for his valuable suggestions and constant support throughout my project tenure. I would like to express our thanks to all faculty and Staff members of Department of Computer Science and Engineering, Galgotias University for their support in completing this project on time.

We also express gratitude towards our parents for their kind co-operation and encouragement which help me in completion of this project. My thanks and appreciations also go to our friends in developing the project and people who have willingly helped me out with their abilities.

Roshan kumar chaudhary
163101597
16SCSE101192

Abstract

Curiosity to predict personality, behavior and need for this is not as new as invent of social media. Personality prediction to better accuracy could be very useful for society. There are many papers and researches conducted on usefulness of the data for various purposes like in marketing, dating suggestions, organization development, personalized recommendations and health care to name a few.

With the introduction and extreme popularity of Online Social Networking Sites like Facebook, Twitter and LinkedIn numerous researches were conducted based on public data available, online social networking applications and social behavior towards friends and followers to predict the personality. Structured mining of the social media content can provide us the ability to predict some personality traits. This survey aims at providing researchers with an overview of various strategies used for studies and research concentrating on predicting user personality and behavior using online social networking site content. There positives, limitations are well summarized as reported in the literature. Finally, a brief discussion including open issues for further research in the area of social networking site based personality prediction preceding conclusion.

List of Tables

Areas of Facebook Research Identified in the Literature Review

Area of research	No. of articles	% of total	Associated research question
User Analysis	97	24%	Who is using Facebook and what they are doing on it.
Motivation behind use of Facebook	78	19%	Why they use Facebook?
Based on Social interaction	112	27%	How the Facebook is related to building the relations?
Related to privacy and information disclosures	75	18%	Why people are so intend to disclose their personal information despite the fact of security concern
Related to Identity	50	12%	How are people presenting themselves on Facebook?

“The Five Factors of the Five-Factor Model of Personality,”

Dimension	Sample items	Description	Examples of behaviours predicted by the trait
Openness to experience	“I have a vivid imagination”; “I have a rich vocabulary”; “I have excellent ideas.”	A general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience	Individuals who are highly open to experience tend to have distinctive and unconventional decorations in their home. They are also likely to have books on a wide variety of topics, a diverse music collection, and works of art on display.
Conscientiousness	“I am always prepared”; “I am exacting in my work”; “I follow a schedule.”	A tendency to show self-discipline, act dutifully, and aim for achievement	Individuals who are conscientious have a preference for planned rather than spontaneous behaviour.
Extraversion	“I am the life of the party”; “I feel comfortable around people”; “I talk to a lot of different people at parties.”	The tendency to experience positive emotions and to seek out stimulation and the company of others	Extraverts enjoy being with people. In groups they like to talk, assert themselves, and draw attention to themselves.
Agreeableness	“I am interested in people”; “I feel others’ emotions”; “I make people feel at ease.”	A tendency to be compassionate and cooperative rather than suspicious and antagonistic toward others; reflects individual differences in general concern for social harmony	Agreeable individuals value getting along with others. They are generally considerate, friendly, generous, helpful, and willing to compromise their interests with those of others.
Neuroticism	“I am not usually relaxed”; “I get upset easily”; “I am easily disturbed”	The tendency to experience negative emotions, such as anger, anxiety, or depression; sometimes called “emotional instability”	Those who score high in neuroticism are more likely to interpret ordinary situations as threatening and minor frustrations as hopelessly difficult. They may have trouble thinking clearly, making decisions, and coping effectively with stress.

Chapter 1. Introduction

(i) Overall View - Personality is a way person respond to a particular situation. It is combination of characteristics that make an individual unique. Assessment of personality over the past two decades in various researches has revealed that personality can be defined by five dimensions known as Big Five personality traits.

In general, study of personality considered as a psychology research based on the survey or questionnaire. But this limits the research data to less number of persons. Hence there is a need of something through which we can increase the number of peopleinvolved in survey and to make the process automated.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MOTIVATION

- Online social networking sites bring both benefits and dangers to society, which need to be carefully examined and analyzed.
- The universal popularity of Facebook/Twitter makes it a topic worthy of study and research. It can help us in depict the popularity and growth of Facebook with addition of various features in it.
- We can find that what type of personality use these Social sites more.
- If we do this analysis time to time then we can find what impact social networking brings on individuals personality .

SCOPE

- This system will help many firms to market their products and also helps them to target the right customers.
- Many users will get to know about their personality based on their personality traits.
- This system is useful for the social networking sites which help these sites to increase their rating scale.
- As the results are based on previous data stored in database, so the system will provide with appropriate result.
- There can be module where user will be provided with career guidance which matches his personality. For example: if a user has the ability to speak well and able to convince opposite person. So, this user will be good in marketing field.

Chapter 2. Literature Survey

Study of Social Media

Social media is giving people a platform to interact with each other. Popularity of social media sites like Facebook, Twitter, and LinkedIn has increased incredibly in last few years. The core of the social networking experience centres on user's ability to perform following action:

- Online space for self-expression by means of posting self-relevant information on individualized profile page.
- Connectivity/interaction with friends and followers.
- Network Expansion.

Online Social Networking sites provide web space to people and allow them to create an individualized public profile where they can express themselves and their thoughts using text, audio, pictures and video. They also provide notifications of people that we many know and those we should add as a friend using some friend recommendation system based on the principle known as homophile. We can also see other people's connections. These features help in Network expansion. Sharing of content on social sites comes with certain potential privacy risk. User can control it to certain degree with privacy settings provided by these sites.

Social media data can be leveraged in two type of analysis:

- Content based
- Linkage data

Content based analysis: Social networking sites like Facebook, twitter and LinkedIn have tremendous amount of content in the form of text, image, audio and video. This huge database can be used for various researches.

Linkage Data based analysis: Social Network can be analyzed with mapping and measuring of relationships between various entities. Analysis is often represented using diagram as shown in figure1. It is based on network structure.

Here Nodes represent actor, object, people or group. Edges represent relationship between those actors. This type of social network analysis is useful for the work related to organization development.

Combining both the approaches Linkage data and content based analysis provide input to wide range of applications including in prediction of personality traits.

Chapter 3. System Design

System Design

The most creative and challenging face of the system development is System Design. It provides the understanding and procedural details necessary for implementing the system recommended in the feasibility study. Design goes through the logical and physical stages of development.

In designing a new system, the system analyst must have a clear understanding of the objectives, which the design is aiming to fulfill. The first step is to determine how the output is to be produced and in what format. Second, input data and master files have to be designed to meet the requirements of the proposed output. The operational phases are handled through program construction and testing.

Design of a system can be defined as a process of applying various techniques and principles for the purpose of defining a device, a process or a system in sufficient detail to permit its physical realization. Thus system design is a solution to “how to” approach to the creation of a new system. This important phase provides the understanding and the procedural details necessary for implementing the system recommended in the feasibility study. The design step provides a data design, architectural design, and a procedural design.

Chapter 4. Requirements

System Requirements

SOFTWARE REQUIREMENTS

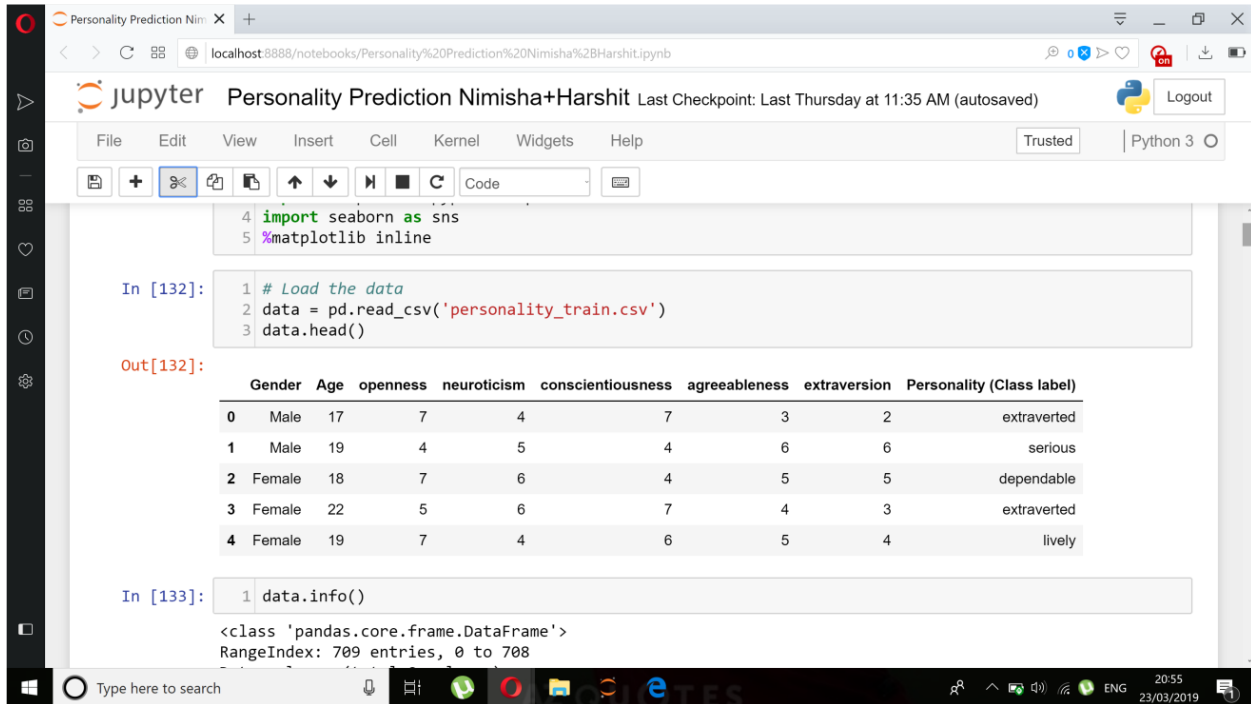
- Operating System : Windows.
- Server: Tomcat Apache.
- Platform: Jupyter
- Programming Language: Python
- Database: Oracle DB Server/MySQL.
- Web Technologies: HTML/CSS/JS/BOOTSTRAP

Hardware Requirements

- Minimum 500GB Hard Disk
- Processor: Pentium series and above
- RAM: 512MB

Chapter 5. Implementation

Part 1: Using csv file containing scores (BIG 5 MODEL)



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
4 import seaborn as sns
5 %matplotlib inline
```

In [132]:

```
1 # Load the data
2 data = pd.read_csv('personality_train.csv')
3 data.head()
```

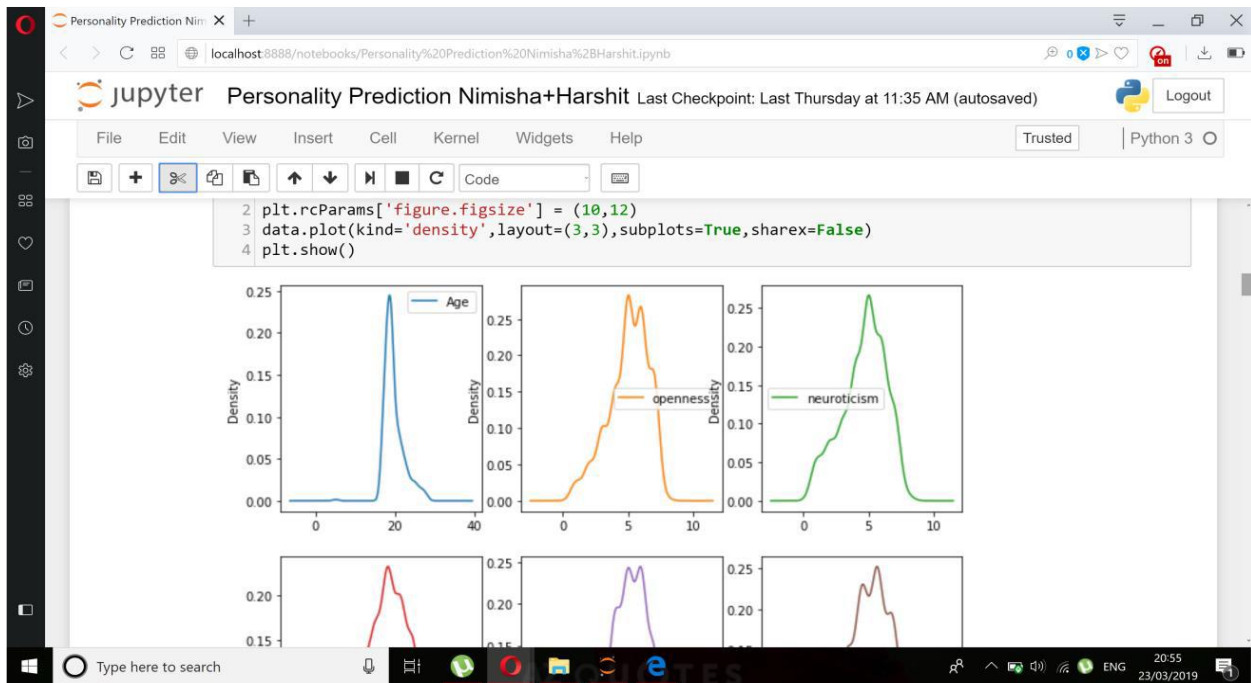
Out[132]:

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
0	Male	17	7	4	7	3	2	extraverted
1	Male	19	4	5	4	6	6	serious
2	Female	18	7	6	4	5	5	dependable
3	Female	22	5	6	7	4	3	extraverted
4	Female	19	7	4	6	5	4	lively

In [133]:

```
1 data.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 709 entries, 0 to 708



Personality Prediction Nimisha+Harshit Last Checkpoint: Last Thursday at 11:35 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [138]: 1 data['Personality (Class label)'] = le.fit_transform(data['Personality (Class label)']
2 data.head(4)
```

Out[138]:

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
0	2	17	7	4	7	3	2	1
1	2	19	4	5	4	6	6	4
2	1	18	7	6	4	5	5	0
3	1	22	5	6	7	4	3	1

```
In [140]: 1 one_hot = pd.get_dummies(data['Gender'])
2 data = data.drop('Gender',axis = 1)
3 # Join the encoded df
4 data = data.join(one_hot)
5 data.head(4)
```

Out[140]:

	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)	0	1	2
0	17	7	4	7	3	2	1			
1	19	4	5	4	6	6	4			
2	18	7	6	4	5	5	0			
3	22	5	6	7	4	3	1			

Type here to search

Personality Prediction Nimisha+Harshit Last Checkpoint: Last Thursday at 11:35 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
warnings.warn("variables are collinear.")
C:\Users\Abhishek\Anaconda3\lib\site-packages\sklearn\discriminant_analysis.py:388: UserWarning: Variables are collinear.
warnings.warn("Variables are collinear.")
```

```
In [154]: 1 from sklearn.metrics import confusion_matrix
2 from sklearn.metrics import accuracy_score
3 from sklearn.metrics import classification_report
4 for nam, mod in model:
5     print('method {}'.format(nam))
6
7     mod.fit(x_train, y_train)
8     predict = mod.predict(x_test)
9     print(accuracy_score(y_test, predict))
10    print(confusion_matrix(y_test, predict))
11    print(classification_report(y_test, predict))
```

method LR
0.19662921348314608

```
[[ 1 20  4  0  8]
 [ 5 13  2  2 13]
 [ 3 14  3  1 13]
 [ 3  9  4  4 12]]
```

Type here to search

Personality Prediction Nim X +

localhost:8888/notebooks/Personality%20Prediction%20Nimisha%20Harshit.ipynb

jupyter Personality Prediction Nimisha+Harshit Last Checkpoint: Last Thursday at 11:35 AM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [136]:

```
1 # making box plot
2 data.plot(kind='box',subplots=True,layout=(3,3),share=False)
3 plt.show()
```

Most of the data is following Gaussian Distribution

Age openness neuroticism

Type here to search

20:56 23/03/2019

Personality Prediction Nimisha+Harshit Last Checkpoint: 03/21/2019 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```

method LR
0.19662921348314608
[[ 1 19 4 0 9]
 [ 3 14 2 2 14]
 [ 1 15 3 1 14]
 [ 1 9 3 4 15]
 [ 2 17 7 5 13]]
method LDA
0.21910112359550563
[[ 1 19 4 0 9]
 [ 3 14 1 4 13]
 [ 3 15 3 1 12]
 [ 1 7 4 7 13]
 [ 1 15 5 9 14]]
method KNN
0.25280898876404495
[[12 8 4 3 6]
 [ 9 11 3 7 5]
 [ 7 8 10 3 6]
 [ 8 11 8 3 2]
 [10 9 13 3 9]]
method SVM
0.2640449438707247

```

Type here to search

10:41 30/03/2019

Part 2: k-NN and Logistic Regression Algorithm (Myers Briggs Algorithm)

Personality Prediction 2 Last Checkpoint: 11 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```

In [2]: 1 data = pd.read_csv('mbti_1.csv')
        2 data.head(3)

```

Out[2]:

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw ...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one ____ https://www.youtube.com/wat...

Words per count

```

In [3]: 1 data['words_per_comments'] = data['posts'].apply(lambda x:len(x.split())/50)
        2 data.head()

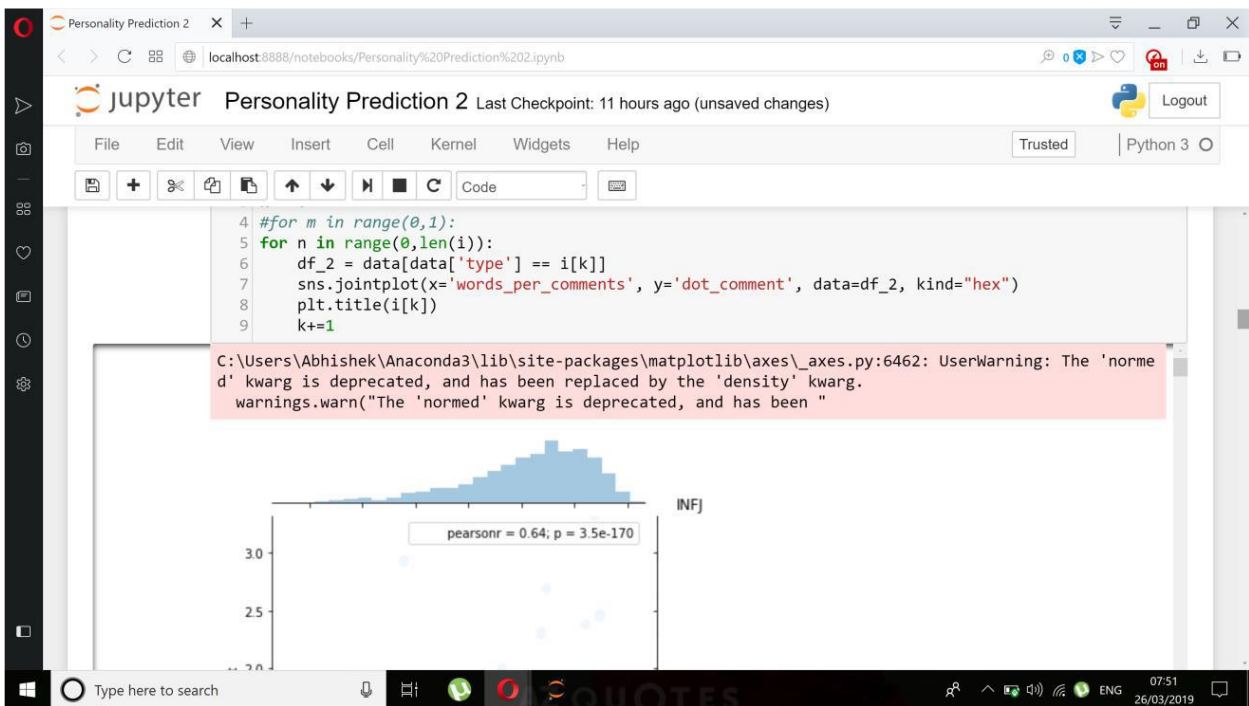
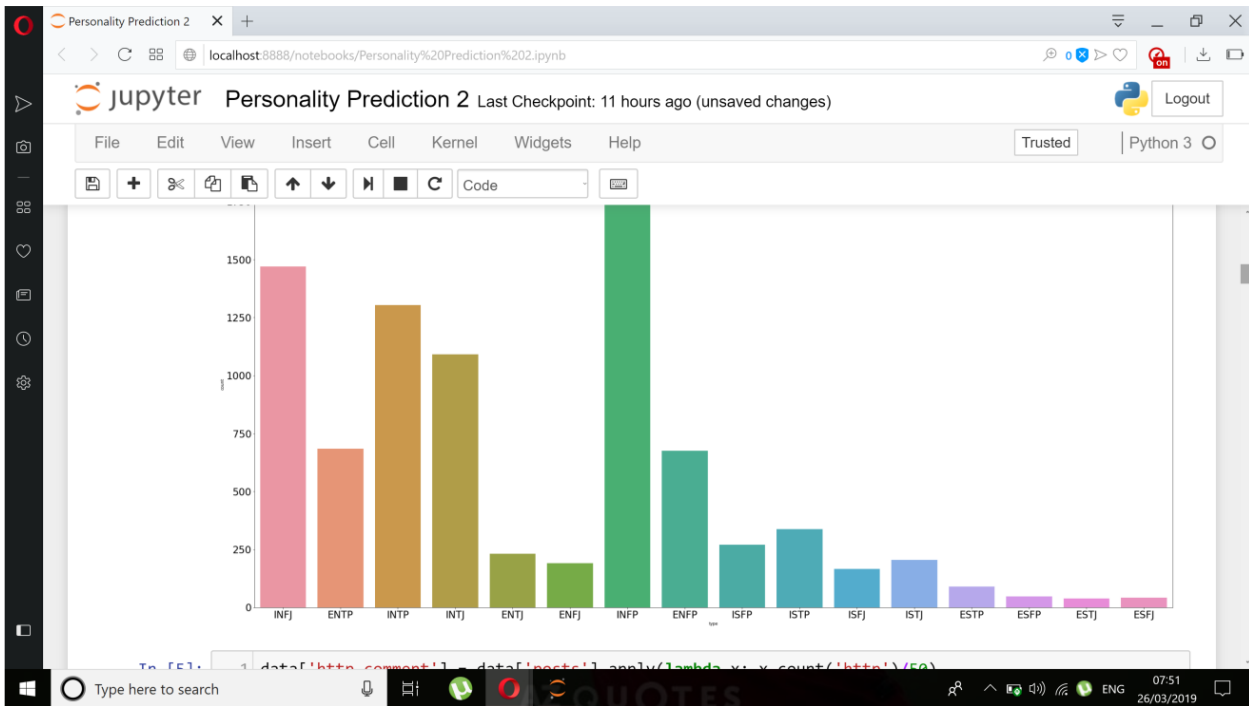
```

Out[3]:

	type	posts	words_per_comments
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw ...	11.12
1	ENTP	'I'm finding the lack of me in these posts ver...	23.40

Type here to search

07:51 26/03/2019



Personality Prediction 2 X +

localhost:8888/notebooks/Personality%20Prediction%202.ipynb

jupyter Personality Prediction 2 Last Checkpoint: 11 hours ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
20 print(PearArray)
21 print(TypeError)
22 plt.scatter(TypeError, PearArray)
```

In [16]:

```
1 map1 = {"I": 0, "E": 1}
2 map2 = {"N": 0, "S": 1}
3 map3 = {"T": 0, "F": 1}
4 map4 = {"J": 0, "P": 1}
5 df['I-E'] = df['type'].astype(str).str[0]
```

In [17]:

```
1 df.head()
```

Out[17]:

	type	posts	words_per_comments	http_comment	music_comment	question_comment	img_commer
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw ...	11.12	0.48	0.02	0.36	0.1
1	ENTP	'I'm finding the lack of me in these posts ver...	23.40	0.20	0.00	0.10	0.0
2	INTP	'Good one https://www.youtube.com/wat...	16.72	0.10	0.00	0.24	0.0

Type here to search 07:51 26/03/2019

Personality Prediction 2 X +

localhost:8888/notebooks/Personality%20Prediction%202.ipynb

jupyter Personality Prediction 2 Last Checkpoint: 11 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
predict = mod.predict(x_test)
print(accuracy_score(y_test, predict))
print(confusion_matrix(y_test, predict))
print(classification_report(y_test, predict))
#sns.heatmap(confusion_matrix(y_test, predict), annot=True, fmt='d', cmap="YLGNBu")
```

```
[ 0  0  0  0  0  0  0  0  4 11  1  1  0  0  0  0]
[ 0  0  0  0  0  0  0  0  7 19  0  6  0  0  0  0]]
```

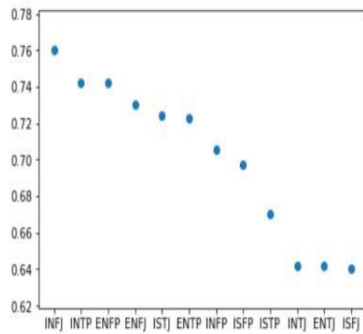
	precision	recall	f1-score	support
0	0.00	0.00	0.00	17
1	0.00	0.00	0.00	76
2	0.00	0.00	0.00	25
3	0.00	0.00	0.00	70
4	0.00	0.00	0.00	4
5	0.00	0.00	0.00	3
6	0.00	0.00	0.00	3
7	0.00	0.00	0.00	6
8	0.19	0.19	0.19	148
9	0.22	0.72	0.34	184
10	0.10	0.01	0.02	111
11	0.17	0.15	0.16	131

Type here to search

07:51 26/03/2019

0.7054114704760908
0.6417417099995272
0.7418360967765789
0.7423804864273493
0.6701370930896875
0.7223598412495943
0.64210438310988
0.7239129106930459
0.7305845421989159
0.760423088098855
0.6974642499076579
[0.760423088098855, 0.7423804864273493, 0.7418360967765789, 0.7305845421989159, 0.7239129106930459, 0.7223598412495943, 0.7054114704760908, 0.6974642499076579, 0.6701370930896875, 0.64210438310988, 0.6417417099995272, 0.6399820319539542]
['INFP', 'INTP', 'ENFP', 'ENFJ', 'ISTJ', 'ENTP', 'INFP', 'ISFP', 'ISTP', 'INTJ', 'ENTJ', 'ISFJ']

Out[10]: <matplotlib.collections.PathCollection at 0x1823ea0d358>



```
In [11]: li = []
l = ['posts', 'type', 'words_per_comments']
for i in data.columns:
```

Chapter 6. Conclusion

Social behavior in online social networking sites can be used to predict User's big five personality traits. Psychologist used to follow personality questionnaire approach. This process is costly and impractical at times. With the popularity of online social networks, researches envisaged to predict the personality automatically. Researches tried to assess the personality based on internet and social network site usage. However only some of the personality traits like Extraversion and emotional stability could be assessed using this approach. Through linkage and content based analysis of these online social networking sites data, researchers were able to predict personality traits quite accurately. Based on Facebook "Likes", Network Structure like number of friends and groups, Status update, Photo upload, Tags and then using various regression and machine learning

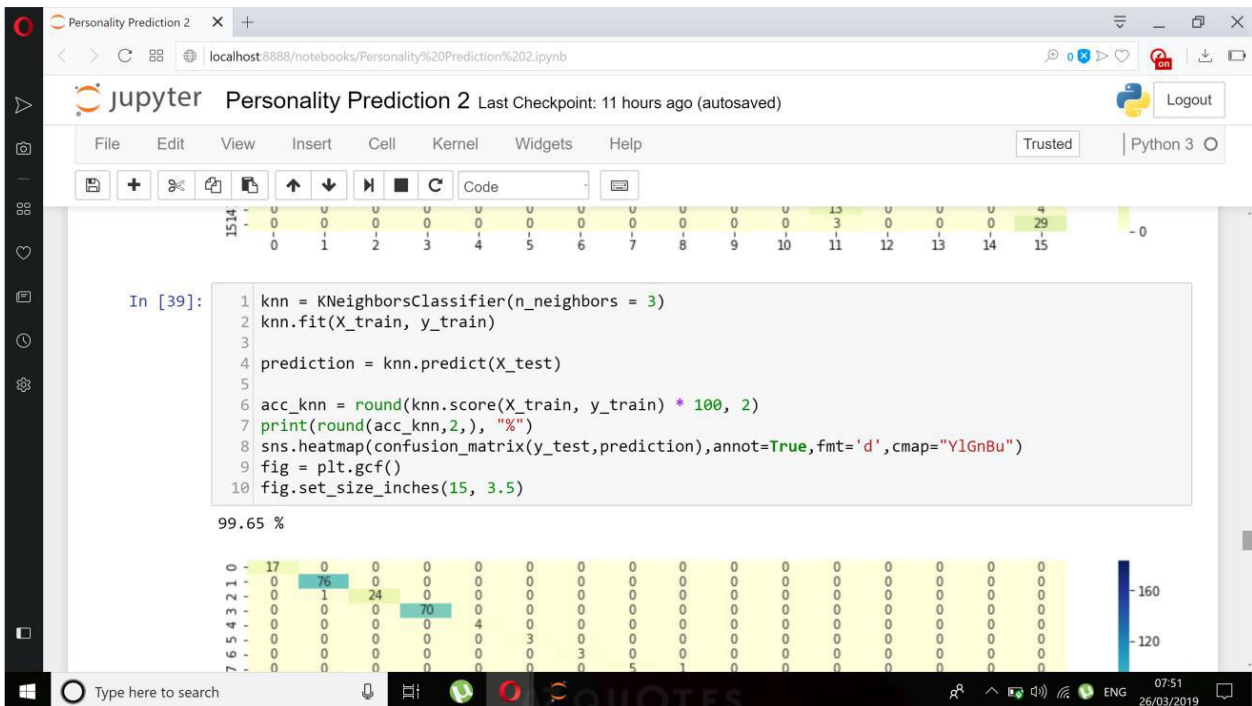
Chapter 7. Output

There were various approaches for classifying an individual into different personalities using various algorithms and models. We have basically used 2 models/theories to show the accuracy difference.

First we used Big 5 Model and then we used the Myers Brigg model.

Accuracy of Big 5 model was very less as compared to Myers Briggs because of the size of data set. Big 5 model works well with smaller data set.

We got approximate 98% accuracy when we used Myers Briggs model when we used kNN algorithm.



REFERENCES

- [1] S. Adali and J. Golbeck, "Predicting Personality with Social Behavior", Proc. the 2012 International Conference on Advances in Social Networks Analysis and Mining.
- [2] H.V. Zhao, W.S. Lin and K.J.R Liu, "Behavior modeling and forensics for multimedia social networks" Proc. Signal Processing Magazine, IEEE .
- [3] J. Sang, "Collective search and recommendation in social media", Proc. the 20th ACM international conference on Multimedia.
- [4] J.C. Prates, E. Fritzen, S.W.M. Sequeira, M. H. L.B. Braz and L.C.V.D. Andrade, "Contextual web searches in Facebook using learning materials and discussion messages", Computers in Human Behavior, Volume 29, Issue 2.
- [5] R. Srivastava, J. Feng, S. Roy, S. Yan and T. Sim, "Don't ask me what i'm like, just watch and listen", Proc. the 20th ACM international conference on Multimedia.
- [6] N. Albaloooshi, N. Mavaridis and N. Al-Qirim, "A survey on social networks and organization development", 2012 International Conference on Collaboration Technologies and System(CTS), IEEE.
- [7] J. Golbeck, C. Robles, M. Edmondson and K. Turner, "Predicting Personality from Twitter", Privacy, security, risk and trust , 2011 iee third international conference on and 2011 IEEE third international conference on social computing .

