PERSONALITY SYSTEM USING MACHINE LEARNING ALGORITHMS

A CAPSTONE PROJECT REPORT

Submitted in partial fulfillment of the requirement for the award of the Degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

by

Sangam Mahajan (18BCE7005)

Under the Guidance of

DR. Sunil Kumar Singh

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
VIT-AP UNIVERSITY
AMARAVATI- 522237

JANUARY 2022

CERTIFICATE

This is to certify that the Capstone Project work titled "**Personality System Using Machine Learning Algorithms**" that is being submitted by **Sangam Mahajan (18BCE7005)** is in partial fulfillment of the requirements for the award of Bachelor of Technology, is a record of bonafide work done under my guidance. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for award of any degree or diploma and the same is certified.

Dr. Sunil Kumar Singh
Guide

The thesis is satisfactory / ~~unsatisfactory~~

Approved by

PROGRAM CHAIR
B. Tech. CSE

DEAN
School Of Computer Science and Engineering

ACKNOWLEDGEMENTS

In effectively finishing this venture, many individuals have helped me. I might want to thank every one of the people who are connected with this venture. Primarily, I would say thanks to God for having the option to finish this undertaking with progress. Then, at that point, I will thank my Guide (Dr. Sunil Kumar Singh) and all my educators, under whose direction I gained some useful knowledge about this undertaking. His ideas and bearings have helped in the finishing of this task.

At long last, I might want to thank my folks and companions who have assisted me with their important ideas and direction and have been exceptionally useful in different phases of venture finishing.

ABSTRACT

In this paper an efficient system is presented which can automatically predict and create personality systems. Information Science and Artificial Intelligence are changing the world through specialized changes. We can notice many AI applications in everyday lives, except perhaps the best utilization of AI is to group people dependent on their character attributes. Each individual on this advertising efforts' viability by focusing on explicit individuals. Such character based correspondences are exceptionally compelling in expanding the notoriety and allure of items and administrations. It expanded use, consumer loyalty, and more extensive acknowledgment among clients.

The utilization of interpersonal organizations is expanding quickly. Different information is shared broadly through online media, for example Facebook. Data about clients and what they communicated through notices are such significant resources for research in the field of conduct learning and human character. Comparable explorations have been led in this field and it develops consistently till now. This review endeavors to fabricate a framework that can foresee an individual's character dependent on Facebook client data. Character model utilized in this examination is Big Five Model Personality. While other past explorers utilized more established AI calculation in building their models, this exploration attempts to execute a few profound learning structures to see the examination by doing a far reaching investigation strategy through the precision result. The outcomes prevailed to beat the exactness of past comparable examinations with the normal precision of 84.17%.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

The Big Five character attributes, otherwise called the five-factor model (FFM) and the OCEAN model, is a scientific categorization, or gathering, for character qualities. At the point when factor examination (a measurable procedure) is applied to character study information, a few words used to depict parts of character are regularly applied to a similar individual. For instance, somebody portrayed as principled is bound to be depicted as "arranged all of the time" rather than "muddled". This hypothesis depends along these lines on the relationship between words however not on neuropsychological examinations. This hypothesis utilizes descriptors of normal language and consequently recommends five wide aspects ordinarily used to portray the human character and mind.

Web-based media has turned into the most generally utilized correspondence and connection device between individuals over the past not many years. Direct association between individuals is diminishing as individuals will generally convey by implication through cell phones. In this way, it is very hard to perceive an individual's character. Be that as it may, what's written in web-based media might assist us with getting the data required as individuals invest a lot of energy really taking a look at web-based media and communicating their sentiments and musings through situations with, and refreshes. Facebook has the biggest clients arriving at 1.8 billion clients with around 800 million clients going through around 40 minutes daily utilizing it. Facebook clients for the most part express their sentiments and assessments through announcements or remarks. In spite of the fact that Facebook is right now more broadly used to share photographs and recordings, this exploration centers around clients' etymological viewpoint which is their announcements. Studies in the area of brain science showed that there is a relationship among character and the etymological conduct of an individual. This connection can be viably examined and shown utilizing normal language handling approaches. Along these lines, the objective of this exploration is to assemble a forecast framework that can naturally anticipate client character dependent on their exercises in Facebook.

There are a few character models utilized in foreseeing character, like Big Five Personality, MBTI (MyersBriggs Type Indicator) or Disk (Dominance Influence Steadiness Conscientiousness). In any case, after some contemplations and a writing survey process, Big Five Personality is utilized in this concentration as it is the most well known and exact in telling somebody's character characteristics. Attributes in this model consist of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

1.1 Objectives

The following are the objectives of this project:

- Personalization of online commercial missions prompts more income and navigate rates.
- Character qualities are firmly connected with a singular's conduct and inclinations. Thus the combination of a character based methodology has principally expanded the Recommender System's engaging quality.
- Character based transformations are likewise used to give customized perceptions and could even propose better music proposals.
- Character attributes could likewise address the "cool beginning" issue utilizing deductively approved and generally stable idle elements of an individual dependent on customized frameworks.
- An inside and out comprehension of how character expectation is helpful.
- How might AI calculations anticipate your character depends on your conduct on Facebook or other online media channels.
- Steps to execute your own character indicator.
- How do nations like Japan utilize computerized reasoning methods to track down the ideal counterpart for couples to tackle a higher separation rate.

1.2 Background and Literature Survey

Contextual investigation: Japan's AI-empowered dating administration

Since the beginning of the century, Japan saw a diminishing pattern in the rate of birth. It is a direct result of decreasing the quantity of yearly relationships from 800,000 of every 2000 to 600,000 out of 2019. These days, it is too hard to even consider tracking down the ideal mate, even in this COVID-19 time when nearly everything has become on the web and virtual.
Thus to help Japan's declining rate of birth and individuals observe their everlasting adoration, Japan's administration is utilizing Artificial Intelligence and Machine Learning with the goal that they could get hitched and start their families.

In any case, Japan's Cabinet accepts that current dating administrations have not progressed enough in tracking down the ideal pair. They have depended on inclinations like age, pay, and instructive level filled by the clients. Subsequently the Japanese Government looked for Artificial Intelligence's assistance to observe the ideal match dependent on more secret examples.

The new AI and ML-based dating frameworks have shown amazing outcomes by zeroing in on people's esteems and characters. Henceforth utilizing this more customized approach rather than simply utilizing age, pay, schooling level, and the matched pair has a higher likelihood of getting hitched. The public authority is likewise paying 66% of the better than ever AI dating frameworks' working expenses to help such administrations. Right now, Japan's Cabinet Office is additionally searching for endorsement of two billion yen for the new and progressed AI-empowered dating administration in the spending plan.

The use of Machine Learning techniques in mental examination is normal to increment pointedly soon. Personalization is the way to organizations extending also offering client situated administrations. Likewise, personalization offers better choices and gives better choices to people dependent on their character.

AI has incredible potential in deciding character attributes, which can be additionally utilized for self-checking and organizations to recruit representatives in light of their character standards.

1.3 Organization of the Report

The remaining chapters of the project report are described as follows:

- Chapter 2 contains the proposed system, methodology, and software details.
- Chapter 3 gives the cost involved in the implementation of the project.
- Chapter 4 discusses the results obtained after the project was implemented.
- Chapter 5 concludes the report.
- Chapter 6 consists of codes.
- Chapter 7 gives references.

CHAPTER 2

The PERSONALITY SYSTEM

This Chapter describes the proposed system, working methodology, software and hardware details.

2.1 Proposed System

The following block diagram (figure 1) shows the system architecture of this project.

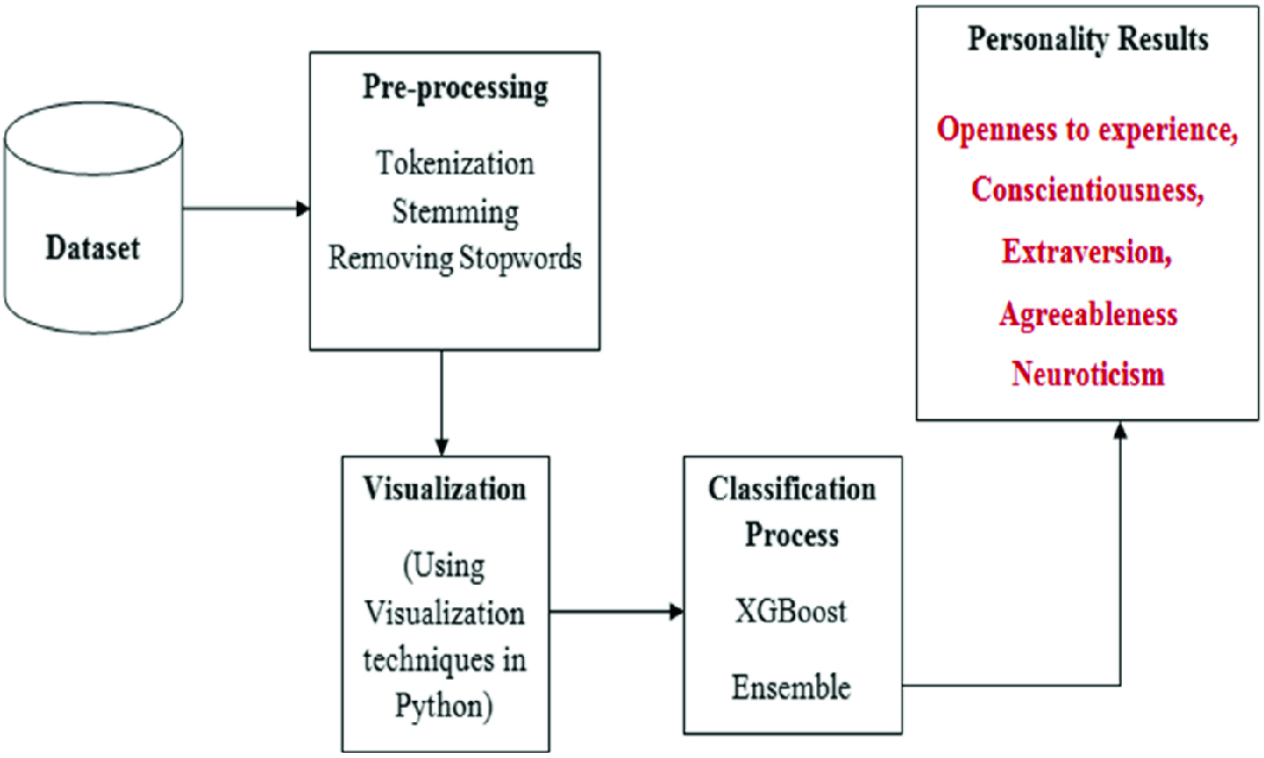


Figure 1 System Block Diagram

2.2 Working Methodology

Character Prediction and Machine Learning

With the accessibility of high-layered and fine-grained information about human conduct, it turns out to be too convenient to even consider exploring and notice human conduct. Utilizing portable detecting studies, information gathered from our everyday exercises have definitely adjusted how analysts perform research and attempt character evaluations. AI models are a help to specialists and are utilized to advance exceptionally complex connections and assess their generalizability and heartiness utilizing the resampling technique.

It can possibly change examination and evaluation in character brain research. Calculations can deal with huge datasets, including a large number of traits, without surrendering to collinearity issues. Also, ML calculations are exceptionally productive in perceiving designs in datasets that people can't see.

The utilization of these ML models can prompt better, more evenhanded, and computerized character appraisals. Individuals collaborate and communicate their preferences, contemplations, sentiments, and conclusions via online media, catching their character qualities. AI models have been effectively utilizing a wide scope of information to anticipate people's Big Five (OCEAN) character qualities. Different administered AI calculations like Naïve Bayes and Support Vector Machines are generally utilized among ventures to foresee character characteristics. Besides, as of late, specialists have begun to apply solo learning techniques to recognize other mental builds in computerized information.

2.3 Standards

Personality Predictions Based on User Behavior on Facebook

In recent years, social media such as Facebook, Twitter, Instagram, Linkedin have become some of the most popular destinations for internet users. Social network activities provide an excellent platform for researchers to study and understand someone's online behaviors, preferences, and personality. Different personalities are related to forming different social relations and interaction behaviors on status preferences.

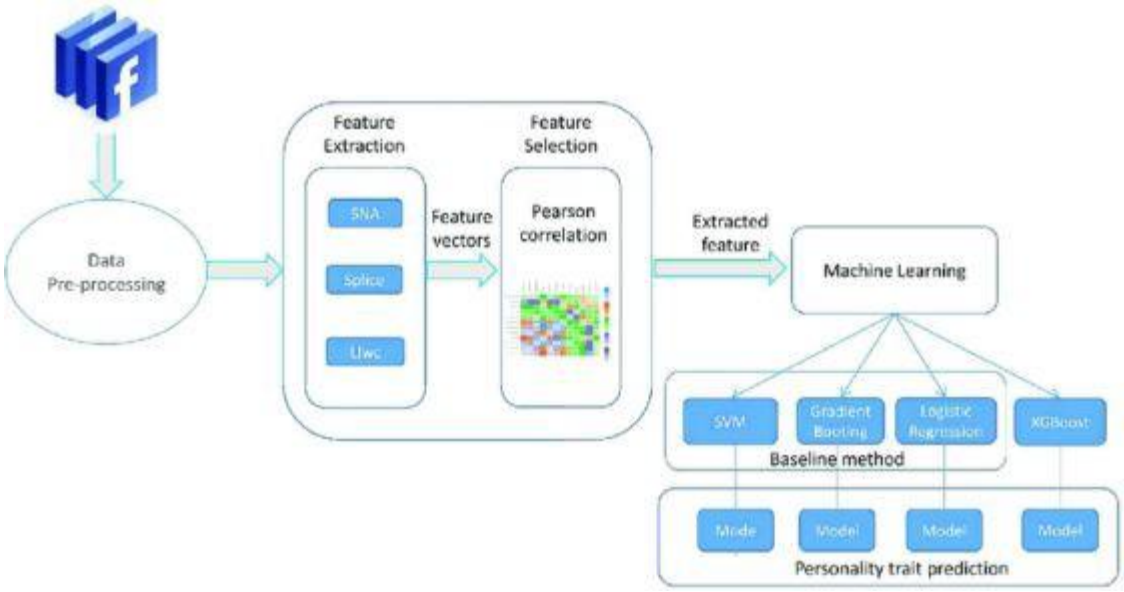


Figure 2: Personality Predictions Based on User Behavior on the Facebook Social Media Platform

With the development of social networks, it has become convenient to determine users' personalities based on their social activities. The figure below describes how to predict the personality traits of Facebook users based on different features and measures of the Big Five model.

2.4 Software and Implementation Details

- **Big five personality traits**

The Big Five personality traits was the model to comprehend the relationship between personality and academic behaviors. This model was defined by several independent sets of researchers who used factor analysis of verbal descriptors of human behavior. These researchers began by studying relationships between a large number of verbal descriptors related to personality traits. They reduced the lists of these descriptors by 5–10 fold and then used factor analysis to group the remaining traits (using data mostly based upon people's estimations, in self-report questionnaires and peer ratings) in order to find the underlying factors of personality.

Big Five Personality Trait model is a notable model dependent on mental speculations and used to gauge character. This model gives an outline of the general character of the individual. It is otherwise called the OCEAN model.

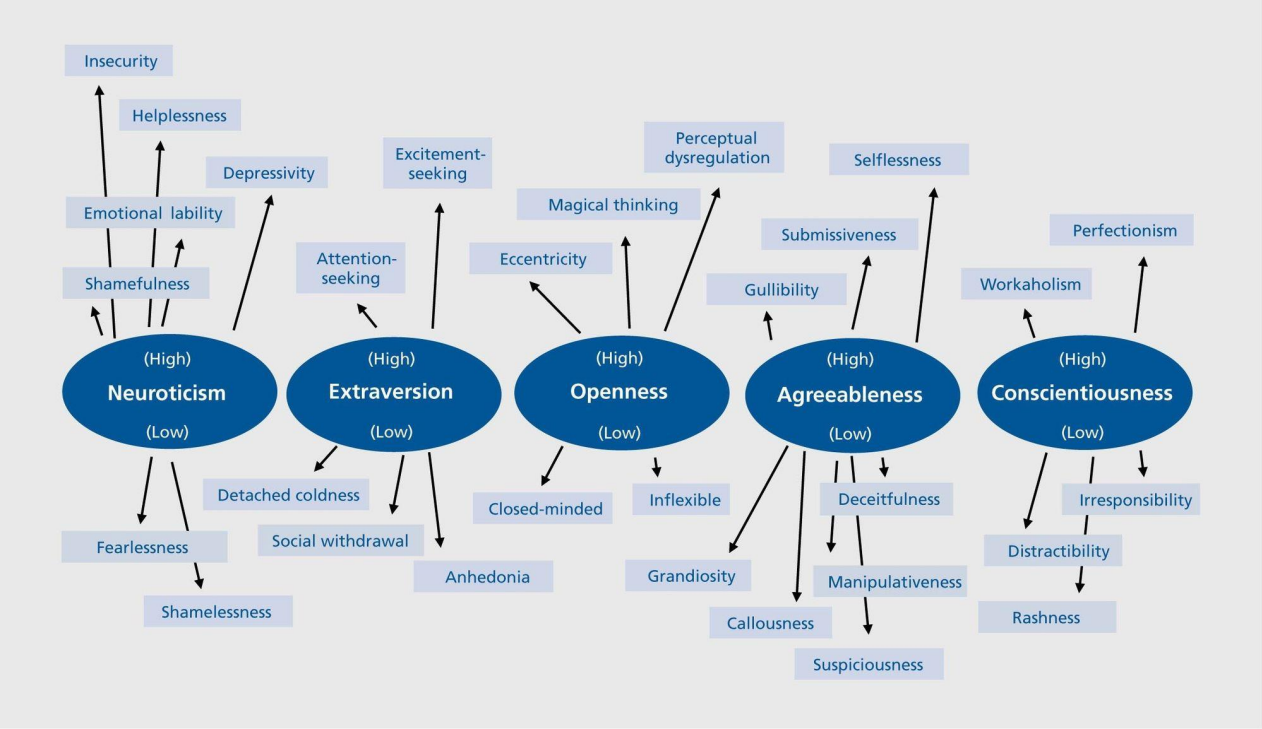


Figure 3 : Personality Traits

Openness: Insight, creative mind, awareness, mindfulness, and interest are the huge qualities of this characteristic. Individuals with such attributes have dynamic characters and a wide scope of interests. They are continually able to investigate the world, are interested with regards to different things, and study others. Such individuals will generally be more inventive.

Conscientiousness: Conscientiousness is utilized to depict the watchfulness, discipline, thought, and industriousness of the individual. Individuals with this quality are profoundly objective and have excellent heartbeat control with the most significant levels of mindfulness. They have excellent authoritative capacities and are inconceivably aware of cutoff times.

Extroversion: Extrovert individuals are sincerely expressive and state their thoughts on others. This quality portrays how individuals are great with their social abilities. Such individuals are typically friendly, energetic, and volatile. They like to make a move and get results with no doubt.

Agreeableness: Agreeableness examinations individual conduct dependent on liberality. Individuals with such characteristics are profoundly agreeable, loving, kind, and are staggeringly trustable. They are obliging, liberal, and chivalrous.

Neuroticism: Neuroticism portrays an individual's passionate precariousness, glumness, and how an individual has mind-set swings. Individuals with undeniable degrees of neuroticism are effectively peevish.

- **Big Five Personality Test Implementation**

This segment performs Big Five Personality Test forecasts utilizing a dataset consisting of 1,015,342 poll answers gathered online by Open Psychometrics. How about we take a gander at how the dataset shows up in reality. The Number of members = 1015341.

```
data_raw=pd.read_csv("data-final.csv",sep='\t')
data = data_raw.copy()
pd.options.display.max_columns = 150
data.drop(data.columns[50:107], axis=1, inplace=True)
data.drop(data.columns[51:], axis=1, inplace=True)
print('Number of participants: ', len(data))
data.head()
```

	EXT1	EXT2	EXT3	EXT4	EXT5	...	OPN7	OPN8	OPN9	OPN10	Clusters
0	0.8	0.2	1.0	0.4	1.0	...	1.0	0.6	0.8	1.0	3
1	0.6	1.0	0.6	0.8	0.6	...	0.8	0.4	1.0	0.6	4
2	0.4	0.6	0.8	0.8	0.6	...	1.0	0.6	0.8	0.8	4
3	0.4	0.4	0.4	0.6	0.8	...	0.8	0.8	0.6	0.6	0
4	0.6	0.6	0.6	0.6	1.0	...	1.0	0.6	1.0	1.0	3
5	0.6	0.6	0.8	0.4	0.8	...	1.0	0.8	1.0	0.4	3
6	0.8	0.6	0.8	0.6	0.6	...	1.0	1.0	0.8	0.8	2
7	0.6	0.2	1.0	0.4	1.0	...	0.8	0.6	0.8	1.0	3
8	0.4	0.4	0.6	0.6	0.8	...	1.0	0.4	1.0	1.0	2
9	0.2	1.0	0.6	1.0	0.4	...	0.8	0.6	0.6	0.6	0

Figure 4: Head of the dataset

Dataset: Consists of 110 columns

- **Dataset Description**

The Big Five person characteristics, in any case called the five-factor model (FFM) and the OCEAN model, is a logical order, or assembling, for character attributes. Exactly when factor assessment (a quantifiable method) is applied to character outline data, a couple of words used to portray portions of character are habitually applied to a comparable person. For example, someone portrayed as principled will undoubtedly be depicted as "organized constantly" rather than "chaotic". This theory relies hence upon the connection between words anyway not on neuropsychological investigations. This theory uses descriptors of ordinary language and hence prescribes five sweeping perspectives for the most part used to depict the human person and psyche.

For every character characteristic, certain inquiries are posed, and members need to pick between 1 to 5. The scale was marked between 1=Disagree, 3=Neutral, 5=Agree. Here EST relates to the Extroversion quality, AGR compares to Agreeable Personality, and so forth.

- **Visualization**

We should investigate how inquiries for every character attribute are appropriated. Here we are showing the recurrence conveyance of inquiries for Extroversion and Conscientious Personality.

Conscientious Personality

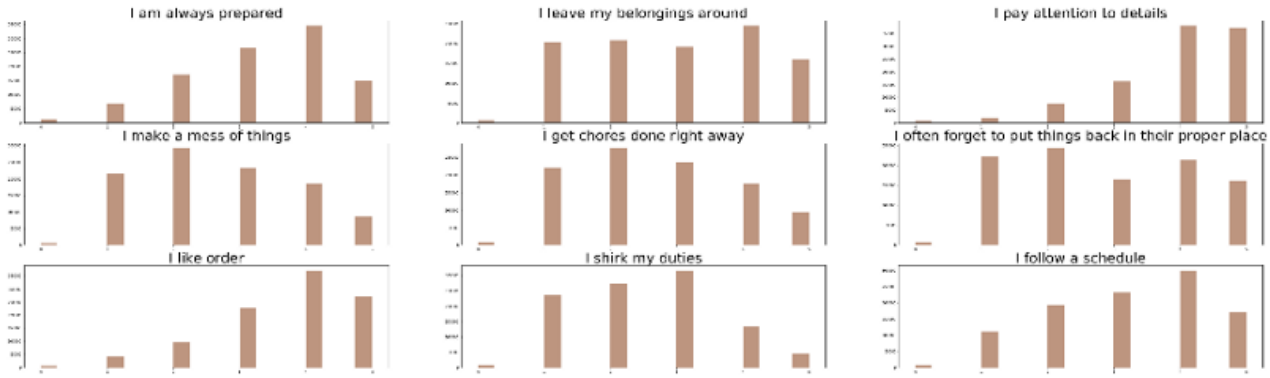


Figure 5: Conscientious trait questionnaire

Extroversion Personality

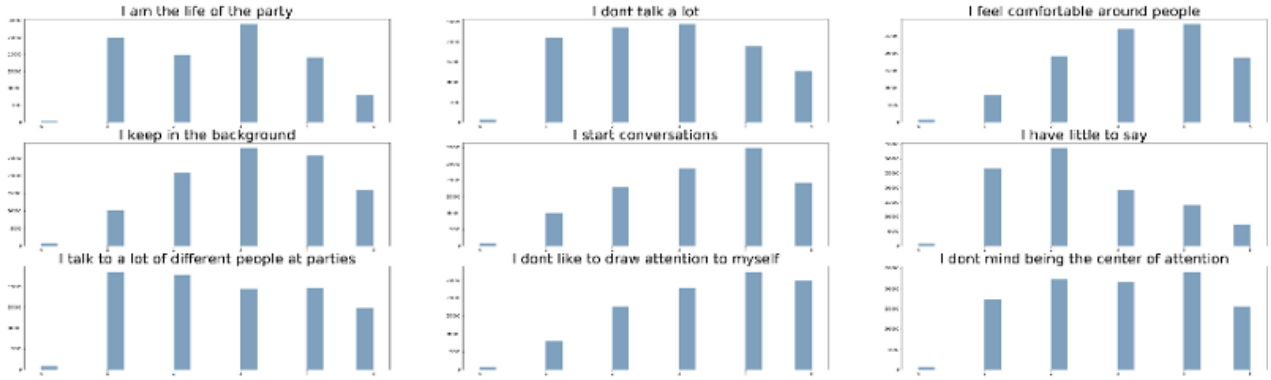


Figure 6: Extroversion trait questionnaire

- **Data Preprocessing and Clustering**

All information in this type of English went through the preprocessing stage before it very well may be handled. Pre-preprocessing steps consist of eliminating URLs, images, names, spaces, bringing down case, stemming, and eliminating stop words. While, information in Bahasa went through an extra preprocessing process; it was substitution of shoptalk words or nonstandard words which was physically led.

We scaled the information utilizing MinMaxScaler to scale between 0–1. Scaling helps in improving the model's presentation and producing better outcomes.

- **Model Building**

We have our information in the ideal arrangement. So how about we make a stride ahead and take care of business by framing five groups where each bunch relates to every character train from the OCEAN model. For this issue, we are utilizing the **K-means** bunching calculation.

❖ **ML - Clustering K-Means Algorithm**

K-Means Clustering is an unaided learning calculation that is utilized to take care of the grouping issues in AI or information science. In this subject, we will realize what K-means grouping calculation is, how the calculation works, alongside the Python execution of k-implies bunching.

The k-means bunching calculation is an information mining and AI apparatus used to group perceptions into gatherings of related perceptions with next to no earlier information on those connections. By examining, the calculation endeavors to show in which class, or bunch, the information has a place with, with the quantity of groups being characterized by the worth k.

The k-means calculation is one of the most straightforward grouping strategies and it is generally utilized in clinical imaging, biometrics, and related fields. The benefit of k-means bunching is that it tells about your information (utilizing its unaided structure) rather than you training the calculation about the information toward the beginning (utilizing the administered type of the calculation).

❖ Working Of K-means

The k-means calculation is a transformative calculation that acquires its name from its strategy for activity. The calculation bunches perceptions into k gatherings, where k is given as an info boundary. It then, at that point, allots every perception to bunches dependent on the perception's vicinity to the mean of the group. The bunch's mean is then recomputed and the interaction starts once more. This is the way the calculation works:

1. The calculation subjectively chooses k focuses as the underlying group communities (the means).
2. Each point in the dataset is relegated to the shut group, in light of the Euclidean distance between each point and each bunch place.
3. Each bunch community is recomputed as the normal of the places in that group.
4. Stages 2 and 3 rehash until the bunches join. Combination might be characterized diversely relying on the execution, yet it ordinarily means that either no perceptions change groups when stages 2 and 3 are rehashed, or that the progressions don't have a material effect in the meaning of the bunches.

❖ Choosing Number of clusters

One of the principal benefits to k-means grouping is the way that you should indicate the quantity of bunches as a contribution to the calculation. As planned, the calculation isn't equipped for deciding the fitting number of bunches and relies on the client to recognize this ahead of time.

For instance, assuming you had a gathering of individuals that are to be grouped dependent on paired sexual orientation way of life as male or female, calling the k-means calculation utilizing the info k=3 would constrain individuals into three bunches when just two, or a contribution of k=2, would give a more normal fit.

Additionally, assuming a gathering of people was handily bunched dependent on home state and you called the k-means calculation with the info k=10, the outcomes may be too summed up to be in any way powerful. In this project we have made 10 clusters and 100 batches.

❖ Cluster Distribution

```
from sklearn.cluster import MiniBatchKMeans
kmeans=MiniBatchKMeans(n_clusters=10,random_state=0,batch_size=100, ax_iter=100).fit(X)

len(kmeans.cluster_centers_)
        10

one = kmeans.cluster_centers_[0]
two = kmeans.cluster_centers_[1]
three =kmeans.cluster_centers_[2]
four = kmeans.cluster_centers_[3]
five =kmeans.cluster_centers_[4]
six = kmeans.cluster_centers_[5]
seven = kmeans.cluster_centers_[6]
eight = kmeans.cluster_centers_[7]
nine= kmeans.cluster_centers_[8]
ten = kmeans.cluster_centers_[9]
one
array([3.51875306, 1.8936703 , 4.45234737, 2.11922512, 4.35885553,
```

1.56325379, 3.98647132, 2.76422671, 3.81405256, 2.3716646 ,
2.05604329, 3.97875709, 2.92127457, 3.4824271 , 1.8897844 ,
1.67228923, 1.87329092, 1.47554762, 1.86143174, 1.5434501 ,
1.79666676, 4.49912207, 1.70769408, 4.29999136, 1.7910538 ,
3.85187531, 1.50729685, 4.16565441, 4.14265565, 4.31282922,
3.87432716, 2.42716099, 4.31288679, 1.68691172, 3.30911603,
2.07210501, 3.97875709, 1.76252842, 3.69278374, 4.00077718,
4.0497395 , 1.58055324, 4.18154342, 1.64966466, 4.2821738 ,
1.53360583, 4.44892202, 3.33657638, 4.18433551, 4.42462796])

- **PCA - Principal component analysis**

Principal Component Analysis (PCA) is a measurable technique that utilizes a symmetrical change that changes a bunch of related factors over to a bunch of uncorrelated factors. PCA is the most broadly involved device in exploratory information examination and in AI for prescient models. In addition, PCA is an unaided measurable method used to inspect the interrelations among a bunch of factors. It is otherwise called an overall element investigation where relapse decides a line of best fit.

PCA works by considering the change of each characteristic in light of the fact that the high property shows the great split between the classes, and consequently it lessens the dimensionality. A few true uses of PCA are picture handling, film proposal framework, upgrading the power assignment in different correspondence channels. It is an element extraction strategy, so it contains the significant factors and drops the most non-significant variable.

The PCA calculation depends on a few numerical ideas, for example, Change and Covariance Eigenvalues and Eigen factors

A few normal terms utilized in PCA calculation:

Dimensionality: It is the quantity of highlights or factors present in the given dataset. All the more effectively, it is the quantity of segments present in the dataset.

Connection: It means how emphatically two factors are connected with one another. For example, on the off chance that one changes, the other variable additionally gets changed. The connection esteem goes from - 1 to +1. Here, - 1 happens assuming factors are contrarily relative to one another, and +1 shows that factors are straightforwardly corresponding to one another.

Symmetrical: It characterizes that factors are not associated to one another, and thus the connection between the pair of factors is zero.

Eigenvectors: If there is a square framework M, and a non-zero vector v is given. Then, at that point, v will be an eigenvector assuming Av is the scalar variety of v.

Covariance Matrix: A network containing the covariance between the pair of factors is known as the Covariance Matrix.

Principal Components in PCA

As portrayed over, the changed new highlights or the result of PCA are the Principal Components. The quantity of these PCs are either equivalent to or not exactly the first highlights present in the dataset. A few properties of these principal parts are given beneath:

- The principal part should be the straight blend of the first elements.
- These parts are symmetrical, i.eThe relationship between a couple of factors is zero.
- The significance of every part diminishes when going from 1 to n, it means the 1 PC has the most significance, and then PC will have the least significance.

Steps for PCA calculation

- **Getting the dataset**

First and foremost, we want to take the information dataset and partition it into two subparts X and Y, where X is the preparation set, and Y is the approval set.

- **Addressing information into a construction**

Presently we will address our dataset into a construction. For example, we will address the two-layered lattice of free factor X. Here each line relates to the information things, and the segment compares to the Features. The quantity of segments is the components of the dataset.

- **Normalizing the information**

In this progression, we will normalize our dataset. For example, in a specific section, the elements with high fluctuation are more significant contrasted with the highlights with lower change.

Assuming that the significance of elements is free of the change of the component, then, at that point, we will separate every information thing in a section with the standard deviation of the segment. Here we will name the lattice as Z.

- **Ascertaining the Covariance of Z**

To ascertain the covariance of Z, we will take the network Z, and will translate it. After translation, we will increase it by Z. The result framework will be the Covariance network of Z.

- **Ascertaining the EigenValues and EigenVectors**

Presently we really want to compute the eigenvalues and eigenvectors for the resultant covariance framework Z. Eigenvectors or the covariance framework are the bearings of the tomahawks with high data. Furthermore the coefficients of these eigenvectors are characterized as the eigenvalues.

- **Arranging the EigenVectors**

In this progression, we will take every one of the eigenvalues and will sort them in diminishing requests, which means from biggest to littlest. Also all the while sort the eigenvectors in like manner in network P of eigenvalues. The resultant grid will be named as P*.

- **Ascertaining the new elements Or Principal Components**

Here we will work out the new elements. To do this, we will increase the P* framework to the Z. In the resultant network Z*Every perception is the direct blend of unique highlights. Every section of the Z* The grid is autonomous of one another.

- **Eliminate less or immaterial elements from the new dataset.**

The new list of capabilities has happened, so we will choose here what to keep and what to eliminate. It means, we will just keep the significant or significant highlights in the new dataset, and irrelevant elements will be eliminated out.

- **Utilizations of Principal Component Analysis**

PCA is fundamentally utilized as the dimensionality decrease method in different AI applications like PC vision, picture pressure, and so on

It can likewise be utilized for tracking down secret examples assuming information has high aspects. A few fields where PCA is utilized are Finance, information mining, Psychology, and so forth.

- **Module Needed**

```
import seaborn as sns
plt.figure(figsize=(10,10))
sns.scatterplot(data=df_pca, x='PCA1', y='PCA2', hue='Clusters'
plt.title('Personality Clusters after PCA');
```

- **Performance Evaluation**

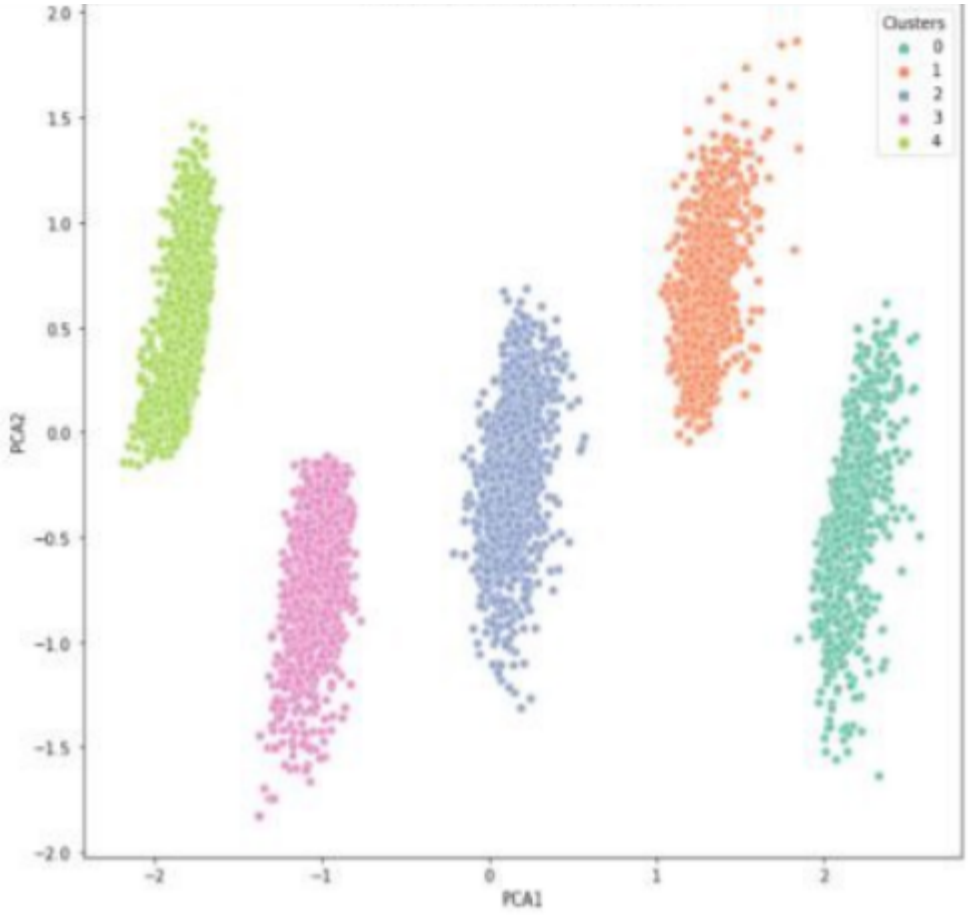
For assessing the model exhibition, recreation blunder is utilized. Indeed, PCA is utilized to extend the focus into the low layered space. The first focuses are remade by extending the low layered portrayals back into the high layered space.

The distance between the reproductions and unique focuses is contrarily connected with how well the model catches the design present in the information. Likewise, reproduction mistakes can likewise be utilized to register the R2 score, estimating the presentation. Thus along these lines, we can get the Big Five Personality Traits Prediction.

CHAPTER 3

RESULTS AND DISCUSSIONS

For assessing the model exhibition, reconstruction error is utilized. Indeed, PCA is utilized to extend the focus into the low layered space. The first focuses are remade by extending the low layered portrayals back into the high layered space. The distance between the reconstructions and unique focuses is contrarily connected with how well the model catches the design present in the information. Additionally, reconstruction error can likewise be utilized to figure the R2 score, estimating the exhibition. Subsequently along these lines, we can get the Big Five Personality Traits Prediction.



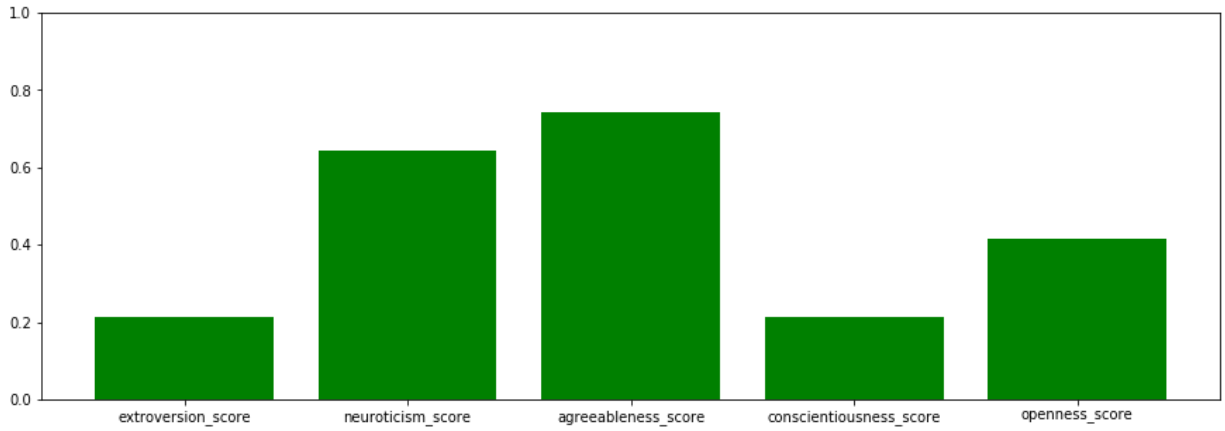


Figure : Personality 1

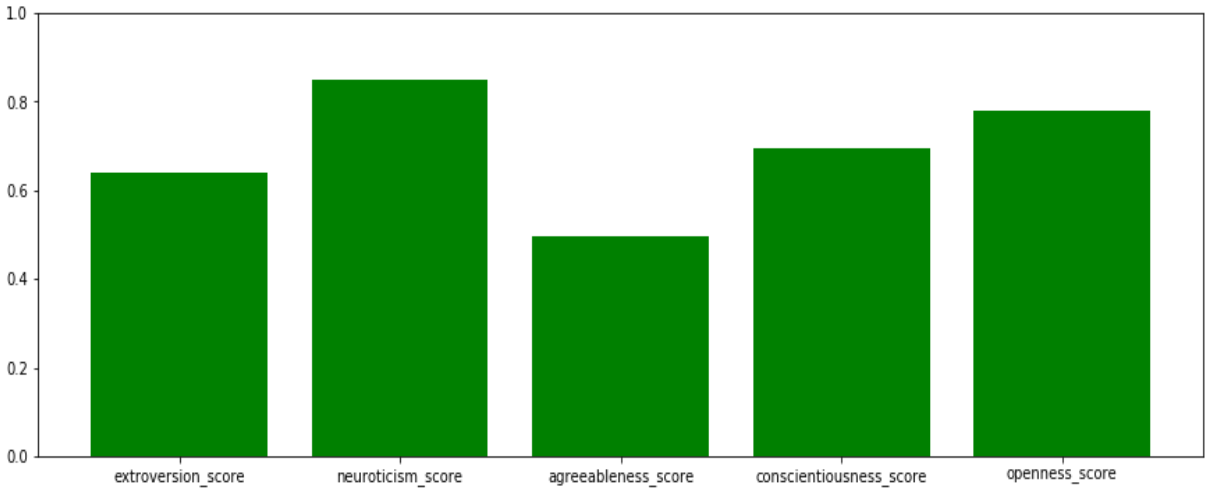


Figure : Personality 2

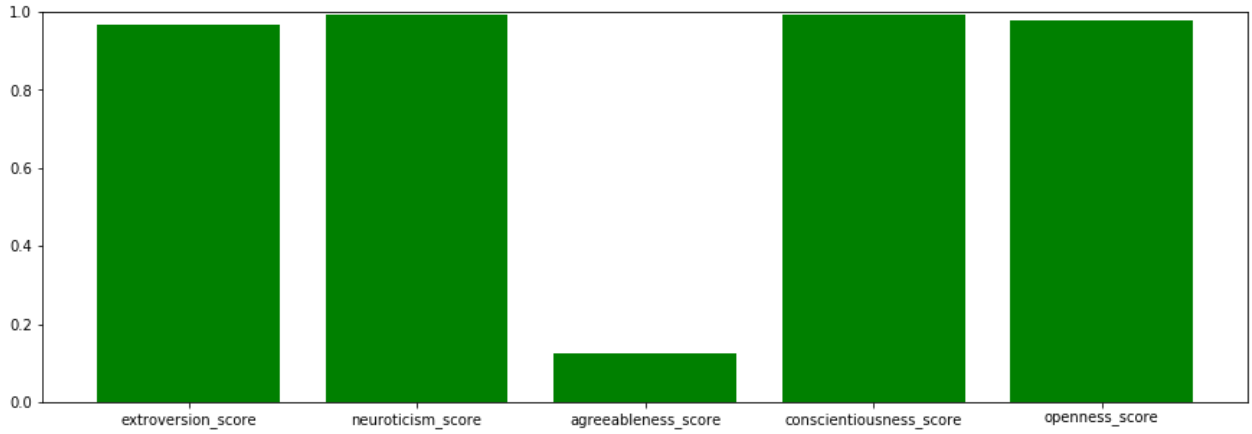


Figure : Personality 3

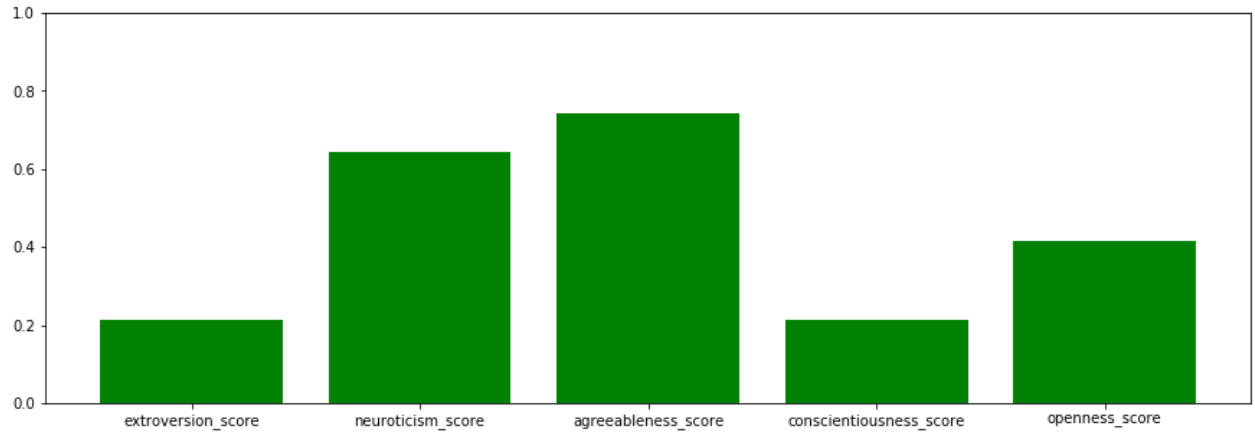


Figure : Personality 4

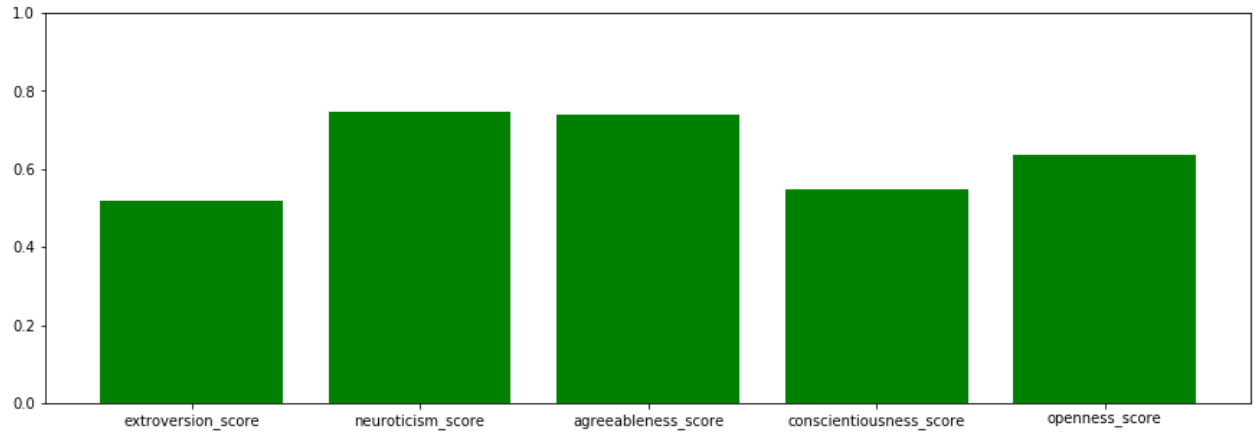


Figure : Personality 5

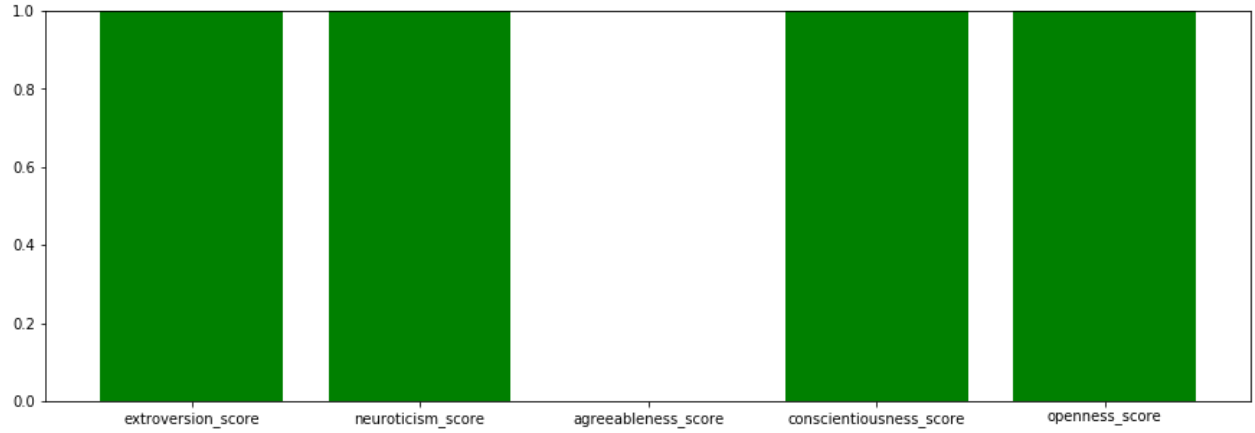


Figure : Personality 6

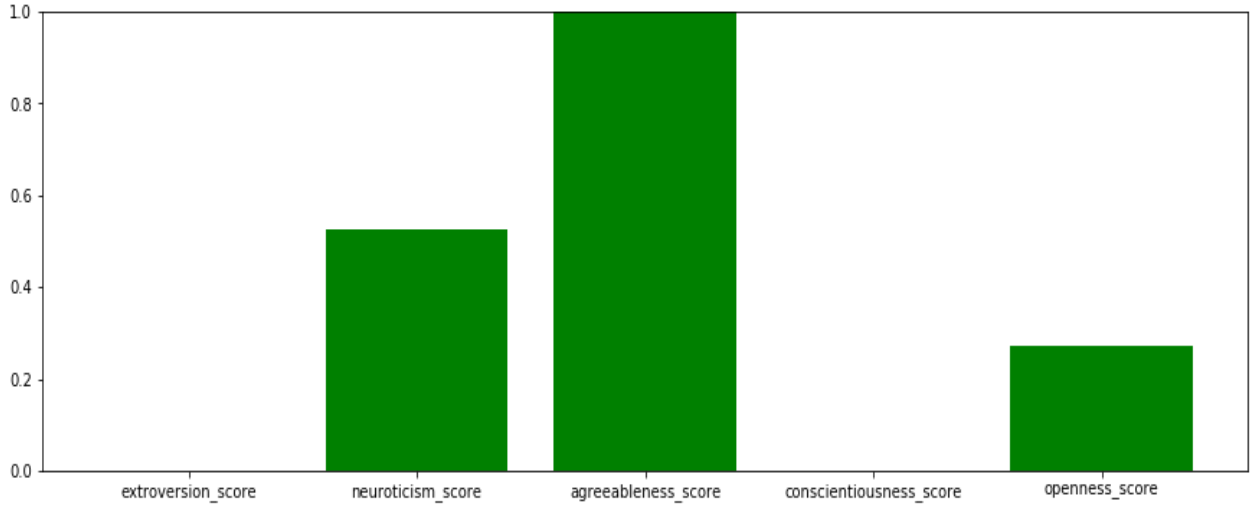


Figure : Personality 7

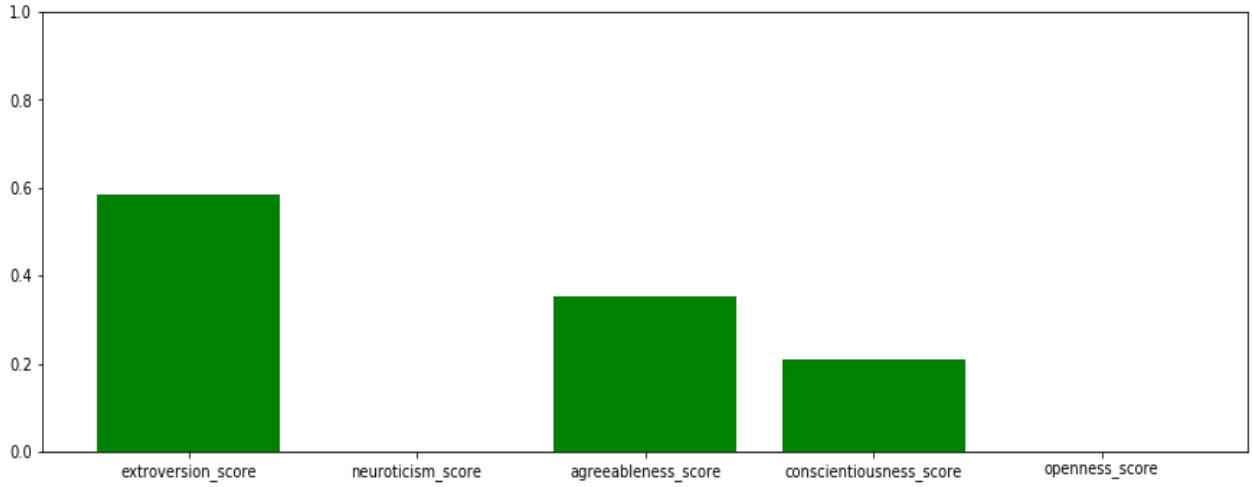


Figure : Personality 8

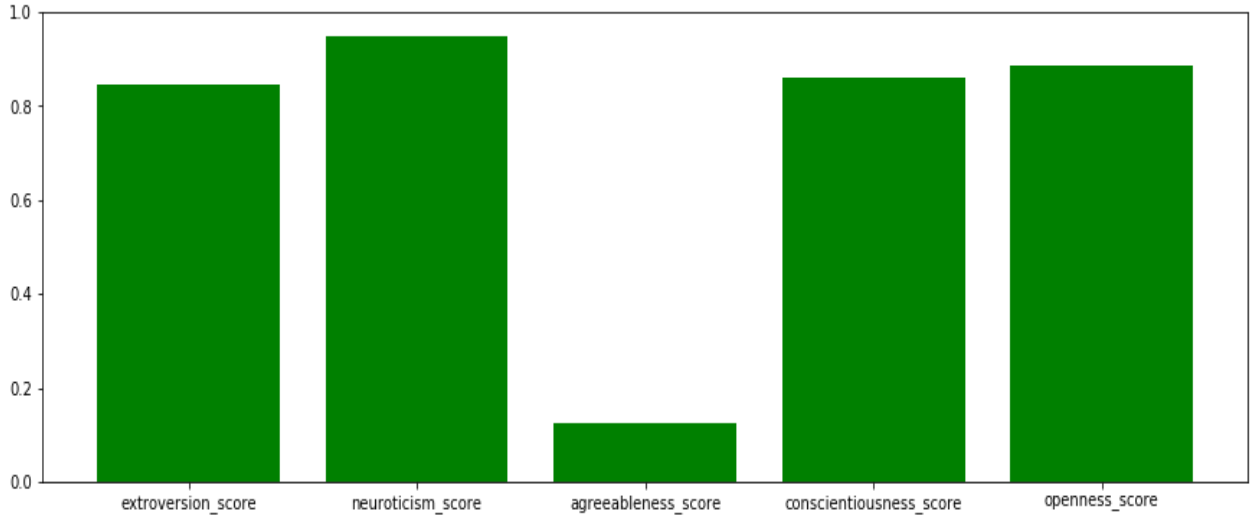


Figure : Personality 9

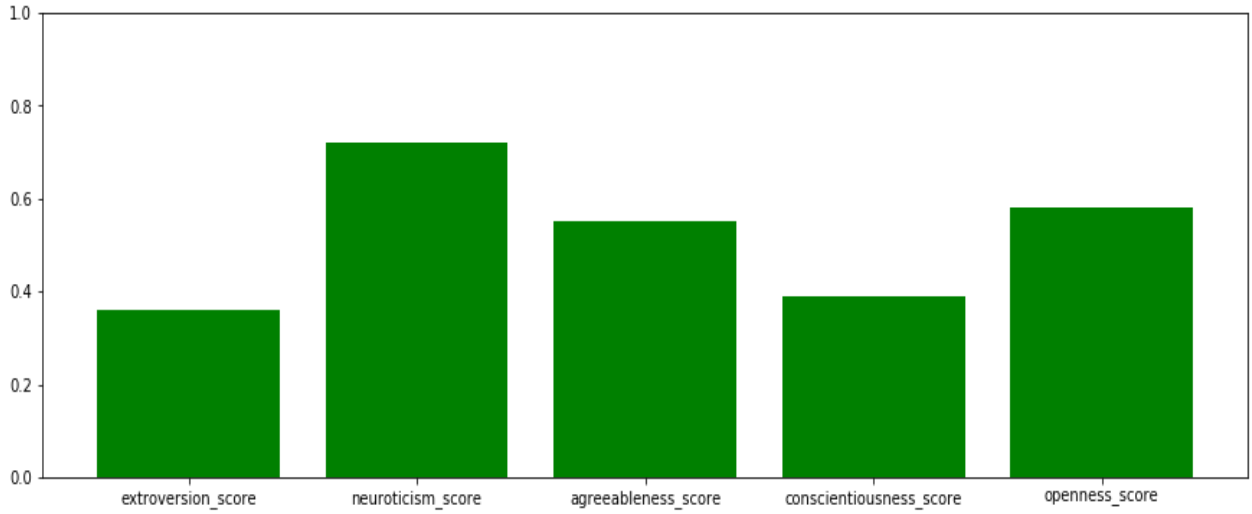


Figure : Personality 10

And, these are the results that we'll get after a pyplot for our personality system.

```
import numpy as np
import matplotlib.pyplot as plt

plt.figure(figsize=(15,5))
plt.ylim(0, 1)

plt.bar(list(normalized_all_types_scores['one'].keys()),
normalized_all_types_scores['one'].values(), color='g')
plt.show()
```

CHAPTER 4

CONCLUSION AND FUTURE WORK

The use of Machine Learning techniques in mental examination is relied upon to increment pointedly soon. Personalization is the way to organizations extending and offering client arranged administrations. Additionally, personalization offers better choices and gives better choices to people dependent on their character. AI has extraordinary potential in deciding character characteristics, which can be additionally utilized for self-checking and organizations to enlist representatives dependent on their character measures.

Japan's Cabinet accepts that current dating administrations have not progressed enough in tracking down the ideal pair. They have depended on inclinations like age, pay, and instructive level filled by the clients. Thus the Japanese Government looked for Artificial Intelligence's assistance to observe the ideal match dependent on more secret examples. The new AI and ML-based dating frameworks have shown brilliant outcomes by zeroing in on people's esteems and characters. Consequently utilizing this more customized approach rather than only utilizing age, pay, instruction level, and the matched pair has a higher likelihood of getting hitched. The public authority is likewise paying 66% of the better than ever AI dating frameworks' working expenses to help such administrations. Presently, Japan's Cabinet Office is likewise searching for endorsement of two billion yen for the new and progressed AI-empowered dating administration in the financial plan.

CHAPTER 5

APPENDIX

Python Code

```
import pandas as pd

df = pd.read_csv('data-final.csv', delimiter="\t")
df

columns = df.columns

for column in columns:
    print(column)

# **************

import numpy as np

X = df[df.columns[0:50]]

pd.set_option('display.max_columns', None)

X

X = X.fillna(3.0)

from sklearn.cluster import MiniBatchKMeans
```

```
kmeans = MiniBatchKMeans(n_clusters=10, random_state=0, batch_size=100, max_iter=100).fit(X)

# **********************

len(kmeans.cluster_centers_)

one = kmeans.cluster_centers_[0]

two = kmeans.cluster_centers_[1]

three =kmeans.cluster_centers_[2]

four = kmeans.cluster_centers_[3]

five =kmeans.cluster_centers_[4]

six = kmeans.cluster_centers_[5]

seven = kmeans.cluster_centers_[6]

eight = kmeans.cluster_centers_[7]

nine= kmeans.cluster_centers_[8]

ten = kmeans.cluster_centers_[9]

one

# ********************
```

```
one_scores = {}

one_scores['extroversion_score'] =  one[0] - one[1] + one[2] - one[3] + one[4] - one[5] + one[6] - one[7] + one[8] - one[9]

one_scores['neuroticism_score'] =  one[0] - one[1] + one[2] - one[3] + one[4] + one[5] + one[6] + one[7] + one[8] + one[9]

one_scores['agreeableness_score'] =  -one[0] + one[1] - one[2] + one[3] - one[4] - one[5] + one[6] - one[7] + one[8] + one[9]

one_scores['conscientiousness_score'] =  one[0] - one[1] + one[2] - one[3] + one[4] - one[5] + one[6] - one[7] + one[8] + one[9]

one_scores['openness_score'] =  one[0] - one[1] + one[2] - one[3] + one[4] - one[5] + one[6] + one[7] + one[8] + one[9]



one_scores

# ********************

all_types = {'one':one, 'two': two, 'three' :three, 'four':four, 'five':five, 'six': six, 'seven': seven, 'eight': eight,
            'nine': nine, 'ten': ten}

all_types_scores ={}

for name, personality_type in all_types.items():
    personality_trait = {}
```

```
        personality_trait['extroversion_score'] = personality_type[0] - personality_type[1] +personality_type[2] - personality_type[3] + personality_type[4] - personality_type[5] +personality_type[6] - personality_type[7] + personality_type[8] -personality_type[9]

        personality_trait['neuroticism_score'] = personality_type[0] - personality_type[1] + personality_type[2] -personality_type[3] + personality_type[4] + personality_type[5] + personality_type[6] + personality_type[7] + personality_type[8] + personality_type[9]

        personality_trait['agreeableness_score'] = -personality_type[0] +personality_type[1] - personality_type[2] + personality_type[3] - personality_type[4] - personality_type[5] + personality_type[6] - personality_type[7] + personality_type[8] + personality_type[9]

        personality_trait['conscientiousness_score'] = personality_type[0] - personality_type[1] + personality_type[2] -personality_type[3] +personality_type[4] - personality_type[5] +personality_type[6] -personality_type[7] + personality_type[8] + personality_type[9]

        personality_trait['openness_score'] = personality_type[0] -personality_type[1] + personality_type[2] - personality_type[3] + personality_type[4] - personality_type[5] +personality_type[6] + personality_type[7] + personality_type[8] + personality_type[9]

    all_types_scores[name] = personality_trait


all_types_scores

# **********

all_extroversion = []
all_neuroticism =[]
all_agreeableness =[]
all_conscientiousness =[]
```

```
all_openness =[]

for personality_type, personality_trait in all_types_scores.items():

    all_extroversion.append(personality_trait['extroversion_score'])
    all_neuroticism.append(personality_trait['neuroticism_score'])
    all_agreeableness.append(personality_trait['agreeableness_score'])
    all_conscientiousness.append(personality_trait['conscientiousness_score'])
    all_openness.append(personality_trait['openness_score'])



all_extroversion_normalized =
(all_extroversion-min(all_extroversion))/(max(all_extroversion)-min(all_extroversion))

all_neuroticism_normalized =
(all_neuroticism-min(all_neuroticism))/(max(all_neuroticism)-min(all_neuroticism))

all_agreeableness_normalized =
(all_agreeableness-min(all_agreeableness))/(max(all_agreeableness)-min(all_agreeableness))

all_conscientiousness_normalized =
(all_conscientiousness-min(all_conscientiousness))/(max(all_conscientiousness)-min(all_conscie
ntiousness))

all_openness_normalized =
(all_openness-min(all_openness))/(max(all_openness)-min(all_openness))

all_extroversion_normalized

counter = 0
```

```
normalized_all_types_scores ={}

for personality_type, personality_trait in all_types_scores.items():
    normalized_personality_trait ={}
    normalized_personality_trait['extroversion_score'] = all_extroversion_normalized[counter]
    normalized_personality_trait['neuroticism_score'] = all_neuroticism_normalized[counter]
    normalized_personality_trait['agreeableness_score'] = all_agreeableness_normalized[counter]
                                        normalized_personality_trait['conscientiousness_score']          =
all_conscientiousness_normalized[counter]
    normalized_personality_trait['openness_score'] = all_openness_normalized[counter]

    normalized_all_types_scores[personality_type] = normalized_personality_trait

    counter+=1

normalized_all_types_scores

# **********************

import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10,10))
sns.scatterplot(data=df_pca, x='PCA1', y='PCA2', hue='Clusters';
plt.title('Personality Clusters after PCA');

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
```

```
plt.bar(list(normalized_all_types_scores['one'].keys()),
normalized_all_types_scores['one'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['two'].keys()),
normalized_all_types_scores['two'].values(), color='g')
plt.show()


plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['three'].keys()),
normalized_all_types_scores['three'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['four'].keys()),
normalized_all_types_scores['four'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['five'].keys()),
normalized_all_types_scores['five'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
```

```
plt.bar(list(normalized_all_types_scores['six'].keys()), normalized_all_types_scores['six'].values(),
color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['seven'].keys()),
normalized_all_types_scores['seven'].values(), color='g')
plt.show()


plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['eight'].keys()),
normalized_all_types_scores['eight'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['nine'].keys()),
normalized_all_types_scores['nine'].values(), color='g')
plt.show()

plt.figure(figsize=(15,5))
plt.ylim(0, 1)
plt.bar(list(normalized_all_types_scores['ten'].keys()),
normalized_all_types_scores['ten'].values(), color='g')
plt.show()
```

REFERENCES

[1] Suyash Namdeo, Ravish Raj, “ Personality Prediction using Machine Learning ”, enjoyalgoriths.com.

[2] Sam Costello, “ K-means Clustering ”, lifewire.com.

[3] Bojan Tunguz, “Big five Personality trait test”, kaggle.com.

[4] “A guide to Principal component Analysis (PCA)”, keebola.com.

[5] Machine learning || Principal Component Analysis (PCA), geeksforgeeks.com.

[6] Lloyd’s algorithm, “ Clustering in k-means using Python ”, The Data Science Lab, thedatasciencelab.wordpress.com.

[7] Supriya Singh, “ Japanese firm adopts AI dating app to help firm find love amid pandemic”, japanesetoday.com.

[8] WION Web Team, Tokyo, Japan “Technically perfect: Japan to fund AI matchmaker to increase birth rate”, wionews.com.

BIODATA

Name : Sangam Mahajan
Mobile Number : +91-6283501227
E-mail : sangam.18bce7005@vitap.ac.in
Permanent Address: 18/R, BBMB Block, Nangal,
Punjab, Pin code- 140124