A Project/Dissertation Review-1 Report

on

PERSONALITY PREDICTION MODEL

Submitted in partial fulfillment of the requirement

for the award of the degree of

B.TECH IN COMPUTER SCIENCE & ENGINEERING

Under The Supervision of :
Designation:

Submitted By

Name of Student/s	Enrollment/Admission No.
Harshit Agarwal	19021011461/19SCSE1010272
Rachna Jha	19021011937 / 19SCSE1010800

GALGOTIAS UNIVERSITY OF COMPUTER SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE ANDENGINEERING GALGOTIAS UNIVERSITY,GREATER NOIDA
INDIA OCTOBER,2021

Abstract

The project is based on identifying the personality of an individual using machine learning algorithms and big 5 models. The personality of a human plays a major role in his personal and professional life. Nowadays, many organizations have also started shortlisting the candidates based on their personality as this increases the efficiency of the work because the person is working in what he is good at rather than what he is forced to do.

The **Big Five model** is also known as the **Five-Factor Model (FFM**) and **OCEAN model** was developed in the early 1980s according to many psychological theories. When statistical analysis is applied to personality survey data, some words used to describe the person and these words give a summary of the overall character or personality of the person accurately.

In our project, we have tried to combine both personality predictions using ML algorithms like K-Mean and Logistic regression to predict the personality of a person and used the term frequency algorithm to get the skill in which the person is good at. Users can easily identify his personality and his technical skill from this model or system.

List of Figures

Acronyms

B.Tech	Bachelor of Technology
BCA	Bachelor of computer Application
M.Tech	Masters of technology
MCA	Master of computer Application
Bsc. (cs)	Bachelor of science in computer science
Msc.(cs)	Master of science in computer science
SCSE	School of computer science and Engineering

Table of Contents

CHAPTER-1
Introduction

1.1 Introduction

The personality of a human plays a major role in his personal and professional life. Nowadays, many organizations have also started shortlisting the candidates based on their personality as this increases the efficiency of the work because the person is working in what he is good at rather than what he is forced to do.

The **Big Five model** is also called the **Five-Factor Model** and **OCEAN model** .It was developed in the early 1980's according to many psychological theories. When statistical analysis is applied to personality survey data, some words used to describe the person and these words give a summary of the overall character or personality/ traits of the person accurately.

- **Open to Experience:** It involves various dimensions, like attentiveness, imagination, sensitivity, preference to curiosity and variety
- **Conscientiousness**: This trait is used to describe the diligence and carefulness of the person. It is the quality that describes how well mannered and efficient a person is.
- **Extraversion:** It is the trait that describes how the best candidates can interact with people that is how good are his/her social skills.
- **Agreeableness:** It is a quality that analyses the individual behavior based on the generosity, sympathy, cooperativeness and ability to adjust with people.
- **Neuroticism:** This trait usually describes a person to have mood swings and has extreme expressive power.

Application:

- **Can be used to predict a person's personality with an accuracy of**

85.81%

- **Used to identify the right candidate to the right candidate based on his personality and skill**
- **Can be used to match marital profiles**

1.2 Formulation of Problem

Nowadays everyone is concerned about human behaviour because it can cause many errors and problems related to other domains like medical error,social media trolling ,etc .

The main role of this project is to know about the traits of the individual without wasting too much time analyzing the data by ourselves using prediction models . problem such as physiological conditions of individuals,If the person is good for the required job or not by using the big 5 personality traits via machine learning

Another aspect of human features that is also related to;personality type,behavior,cognitive and emotional factors is their problem solving ability also related to the individual performance which in turn can affect human error rates and so can cause medical errors in hospitals.

1.2.1 Tool and Technology Used :

- Tools:
  - ☐ Programming Language:python
  - ☐ Python libraries
- Sklearn
- Numpy
- Pandas
- Tensorflow
- Matplotlib

- Technology : Machine learning
- K-MEANS(Unsupervised learning)
- Logistic regression

CHAPTER-2
LiteratureSurvey/Project Design

LiteratureSurvey :

previous researches on SNS mostly focus on topological characteristics (Kwak, 2007),web community mining(Kevin, 2010) and so on. From these meaningful results, virtual world is a facsimile version of the nature society which follows most sociological principles such as Six Degrees of Separation andRule of150 (Yaguang, 2009). It is also found that onlineusers tend to join with each other to form some small communities. Meanwhile, the growing user demand in SNS world triggers the taking off for techniques of characterized recommendation (Jie, 2011) and in-formation retrieval (Christopher,2010) recent year.Junco Reynol (Reynol, 2011) researched on relationship
between Facebook use and student engagement and found that Facebook use was negatively predictive of engagement scale score and positively predictive of time spent on SNS. However, these works were based on user's statistic information such as common friendcount, familiar shared resources, time spent on SNS or information checked frequency which considers user's SNS usage instead of her inner preferences and personality.Personality is one of the hottest topics in Psychology.According to Big Five personality traits theory, personality can be divided into five different dimensions which are openness, conscientiousness, extraversion,agreeableness and neuroticism. Berkeley PersonalityLab (Berkeley, 2012), focusing on personality,self perception, and individual differences in emotion regulation, designed a Big Five Inventory which is widely used around the world. It contains 44 questions with high validity and reliability and can give back a quantized personality score with five dimensions.Until now, researches that combine personality and SNS together have a few bases (Shaoqi, 2011). EmilyS. Orr discussed the influence of shyness on the use of SNS in undergraduate samples in 2009. He discovered that shyness was significantly positively correlated with the time spent on SNS and negativelycorrelated with the number of "friends" (Sisic, 2009).Meanwhile, Teresa Correa analyzed the intersection of users' personality and social media (Correa,2010)and found that openness and extraversion had positive relation to using experience of social media while neuroticism was a negative predictor. However, these works could only give the association relation between personality and behavior instead of a quantization of personality metrics.Samuel D. Gosling (Gosling, 2011) experimented on the

manifestations of personality inSNS. In this research, a mapping between personality and SNS behavior is announced. Theyexamined the personality with self-reported Facebook usage and observable profile information and finally gave the correlation factor between personality and online behavior. They designed 11features, friends count, weekly usage and 9 other functions using frequency. However, their features are all based on statistical characteristics with-out any inner properties of the user.
The data collections are based on self-reported usage and observable profile information which will need a large amount of manual operation. Therefore, experiment objectivity will get a discount.

Project Design:

Our system comes from a combination of ml from computer science and big five inventory

Steps to proceed with the projects.

1 We will download 1 million responses for 50 personality questions for big 5 personality tests.
2 feed the data into machine learning algorithms. 3 normalise the data .
4 plot the results.

3.2 DATA SET UNDERTAKEN:

For training and testing data we will be using big 5 personality tests

This data was collected (2016-2018) through an interactive on-linepersonality test. The personality test was constructed with the "Big-Five Factor Markers" fromthe IPIP.
https://ipip.ori.org/newBigFive5broadKey.htm

WORK FLOW CHART