A Mini Project Report on

Personality Trait Prediction

T.E. - I.T Engineering

Submitted By

Rohin Ambati	**20104068**
Karan Maurya	**20104070**
Rupesh Mishra	**20104100**
Atharva Ankalwar	**20104098**

Under The Guidance Of

Prof. Yaminee Patil

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year: 2022-23

CERTIFICATE

This to certify that the Mini Project report on Loan Prediction System has been submitted by **Rohin Ambati (20104068)**, **Karan Maurya (20104070)**, **Rupesh Mishra (20104100), Atharva Ankalwar (20104098)** who are Bonafide students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2022-2023** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Yaminee Patil
Guide

Dr. Kiran Deshpande
HoD, Information Technology

Dr. Uttam D. Kolekar
Principal

External Examiner(s)
1.
2.

Place: A.P. Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Yaminee Patil**. Expressing gratitude to our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher **Ms. Charul Singh** who gave us her valuable suggestions and ideas when we needed them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

Personality is useful for recognizing how people lead, influence, communicate, collaborate, negotiate business, and manage stress. Personality is one of the important main features that determines how people interact with the outside world. This project is helpful where we have data related to personal behavior. This personal behavior data can be useful for identifying a person based on his/her personality traits. The personality characteristics will already be stored in the database. Later when user enters his personality characteristics his personality is examined in database and system will detect the personality of user, it is based on Big Five Personality Traits Personality is one feature that determines how people interact with the outside world. This data can be helpful to classify people using Automated personality classification (APC). This learning can now be used to classify/predict user personality based on past classifications. This system is useful to social networks as well as various ad selling online networks to classify user personality and sell more relevant ads. This system will be helpful for organizations as well as other agencies who would be recruiting applicants based on their personality rather than their technical knowledge. In this project, we propose a system which analyses the personality of an applicant.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

BIG FIVE PERSONALITY TRAITS:

The Big Five Personality results are very accurate and predict the true personality of a user to a large extent. They are as follows:

Openness to Experience: Openness to experience refers to one's willingness to try new things as well as engage in imaginative and intellectual activities. It includes the ability to "think outside of the box.

Conscientiousness: Conscientiousness describes a person's ability to regulate their impulse control in order to engage in goal-directed behaviors. It measures elements such as control, inhibition, and persistency of behavior.

Agreeableness: This personality dimension includes attributes such as trust, altruism, kindness, affection, and other prosocial behaviors. People who are high in agreeableness tend to be more cooperative while those low in this trait tend to be more competitive and sometimes even manipulative.

Extraversion: Extraversion (or extroversion) is characterized by excitability, sociability, talkativeness, assertiveness, and high amounts of emotional expressiveness.1 People who are high in extraversion are outgoing and tend to gain energy in social situations. Being around other people helps them feel energized and excited.

Neuroticism: Neuroticism is a trait characterized by sadness, moodiness, and emotional instability. Individuals who are high in this trait tend to experience mood swings, anxiety, irritability, and sadness. Those low in this trait tend to be more stable and emotionally resilient.

1.1 Purpose:

The system will examine the data stored in database and will match the personality traits of the user with the data in database. Then the system will detect the personality of the user. Based on the personality traits of the user, the system will provide other features that are relevant to the user's personality. Personality can also affect his/her interaction with the outside world and his/her environment. Personality can also be used as an additional feature during the recruitment process, career counselling, health counselling, etc. Predicting personality by analyzing the behavior of the person is an old technique. This manual method of personality prediction requires a lot of time and resources.

1.2 Problem Statement:

Now a days personality assessment has become the most used test to hire many employees. Classifying the personality of the user based on the big five personality traits using data mining. We need a strong model that predicts the personality of people based on their imagination and thoughts. The goal of this project is to build a model that predicts the personality of the people. The main aim of the proposed system is to predict the personality of the user by the answers given by the user. The project is aimed to developing software which will be helpful in identifying the personality of the person. It uses the concept of machine learning algorithms. After classification is done the user type personality will be displayed.

1.3 Objectives:

- To shortlist the right candidate according to the skills and experience.
- To eliminate the manual traditional recruitment process.
- To automate the recruitment process thereby saving cost and time considerably.
- To enhance human behaviour research and decision-making process.
- To uplift marketing strategies which could target specific people.

1.4 Scope

This project, we discuss about how the personality is identified using different classification algorithms. Here we study the relationship between a user and his/her personality. In this we used Decision Trees because it gives best accuracy around 86.53% while compared to other algorithms that are used previously like naive Bayes , SVM , Logistic regression is fast and give accurate results compared to other algorithms. Personality system is used in E-commerce sites, in Competitive exams , Psychometric tests , matrimonial sites , Government sectors like army, navy, Air force . Thus, the personality is automatically classified by the system after user attempts the survey by the data set provided in the back end . Personality analysis and prediction is more in recent times so further in future more personality traits can be added. Further any improvement can be made using the data set and algorithms to improve accuracy and can be helpful for career guidance module, if user has good speaking and convincing skills. This project discusses about personality Analysis and Prediction.

CHAPTER 2

LITERATURE REVIEW

Sr.no	Title	Author(s)	Year	Outcomes	Methodology	Result
1	"Predicting personality traits with social media"	Youyou, Wu, and Kosinki	2015	Achieved high accuracies from Facebook likes, with correlation of 0.56 for extraversion and 0.50 for Openness.	Analyzed social media Data (Facebook like) using machine learning algorithms to predict OCEAN model.	Found Facebook likes can be used to predict personality traits.
2	"Personality Trait Prediction using Twitter Data and Deep Learning"	Kumar Rastogi, and Varshney	2018	Achieved high accuracies from Twitter data, with correlation of 0.68 for extraversion and 0.64 for Openness.	Applied deep learning techniques – long short term memory networks to predict OCEAN model based on Twitter data.	Discovered that deep learning techniques can be used predict personality traits from twitter data.
3	"Predicting Big Five Personality Traits from Linguistic Data"	Tausczik and Pennebaker	2010	Found that certain linguistic Features were consistently associated with certain personality traits, while other features were more variable.	Conducted meta-analysis of studies that used linguistic data to predict big Five personality traits (OCEAN).	Learned a new category linguistics that can be used to determine traits.

CHAPTER 3

PROPOSED SYSTEM

To overcome the problems of the existing system an Automated personality classification system is proposed which uses some data mining techniques and machine learning algorithms are used to classify the personalities of different users. And by using different algorithms like Big Five Personality Model, Logistic regression, Decision Tree, and Support Vector Machine. By identifying the past data and their patterns it is easy to identify the personality by applying new techniques, so it overcomes the existing system. In this proposed system, the graph will give percentages of this Automated Personality System. Here it gives probabilistic values from 0 to 1. An Automated Personality Classification System is designed in which every applicant/user is given a separate username and password if registered or else user must get registered before taking the survey. Each applicant logins the survey /test using his /her username and password and takes the survey. It consists of 50 questions of each trait and aptitude test of 10 questions, and the user can take the survey so that it determines the Big five personality traits.

After taking the survey user can see the result of his/her personality. Thus, we can see the graph after getting result and then it gives suggestions to those whose personality matches i.e., like friend suggestion. By this, it is useful for many sectors like interview, recruitment process, government sectors, psychometric tests, and once a user results are suitable then can get into any organization which is based on personality type jobs. In this system we are finding the personality of each user and the average of each big five personality trait average out of 10 and finally getting the personality type of the user. Based up on the answers given by the user in personality test the type of the personality is predicted. By that personality type predicted user can easily apply for the job and can know the personality predicted. In the same way students can also know their personality and can participate in the competitive exams.

3.1 Features and Functionality

- Provides detailed analysis of the user's personality.
- Helps users to analyse their soft skills and target the weak areas.
- They can upload their respective CV or Resume.
- The GUI is simple, attractive and user-friendly.
- User can fill the questionnaire and submit responses.
- Personality characters can be stored in database.
- This system can be used in many business sectors that may require expert candidate.
- This system will reduce workload of the human resource department.
- This system will help the human resource department to select right candidate for job profile which in turn provide expert workforce for the organization.
- Admin or the concern person can easily shortlist a candidate based on their online test marks and can select an appropriate candidate for desired job profile.

CHAPTER 4

REQUIREMENT ANALYSIS

Attribute Selection Attribute of dataset are property of dataset which are used for system and for personality many attributes are like heart gender of the person, age of the person ,Big five traits like Openness, Neuroticism, Extraversion, Agreeableness, Consciousness(value 1 -10). The importance of feature selection can best be recognized when you are dealing with a dataset that contains a vast number of features. This type of dataset is often referred to as a high dimensional dataset. Now, with this high dimensionality, comes a lot of problems such as - this high dimensionality will significantly increase the training time of your machine learning model, it can make your model very complicated which in turn may lead to Overfitting. Pre-Processing of Data Pre-processing needed for achieving best result from the machine learning algorithms. In this, we gathered dataset, and it was pre-processed before it is sent to training stage. Sampling is a very common method for selecting a subset of the dataset that we are analysing. In most cases, working with the complete dataset can turn out to be too expensive considering the memory.

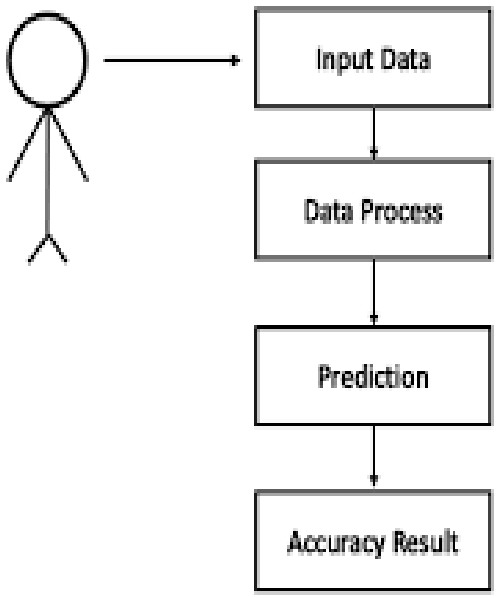
Using a sampling algorithm can help us reduce the size of the dataset to a point where we can use a better, but more expensive, machine learning algorithm. When we talk about data, we usually think of some large datasets with huge number of rows and columns. While that is a likely scenario, it is not always the case — data could be in so many different forms: Structured Tables, Images, Audio files, Videos etc. Machines don't understand free text, image or video data as it is, they understand 1s and 0s. So we pre-process the data. Prediction of Personality Classification In this, system we used machine learning algorithms is performed and whichever algorithm is used which it gives best accuracy for personality prediction. By applying all these modules, the personality is predicted, and the final result is personality of the user by using the training and testing dataset the personality of the user is classified.

CHAPTER 5

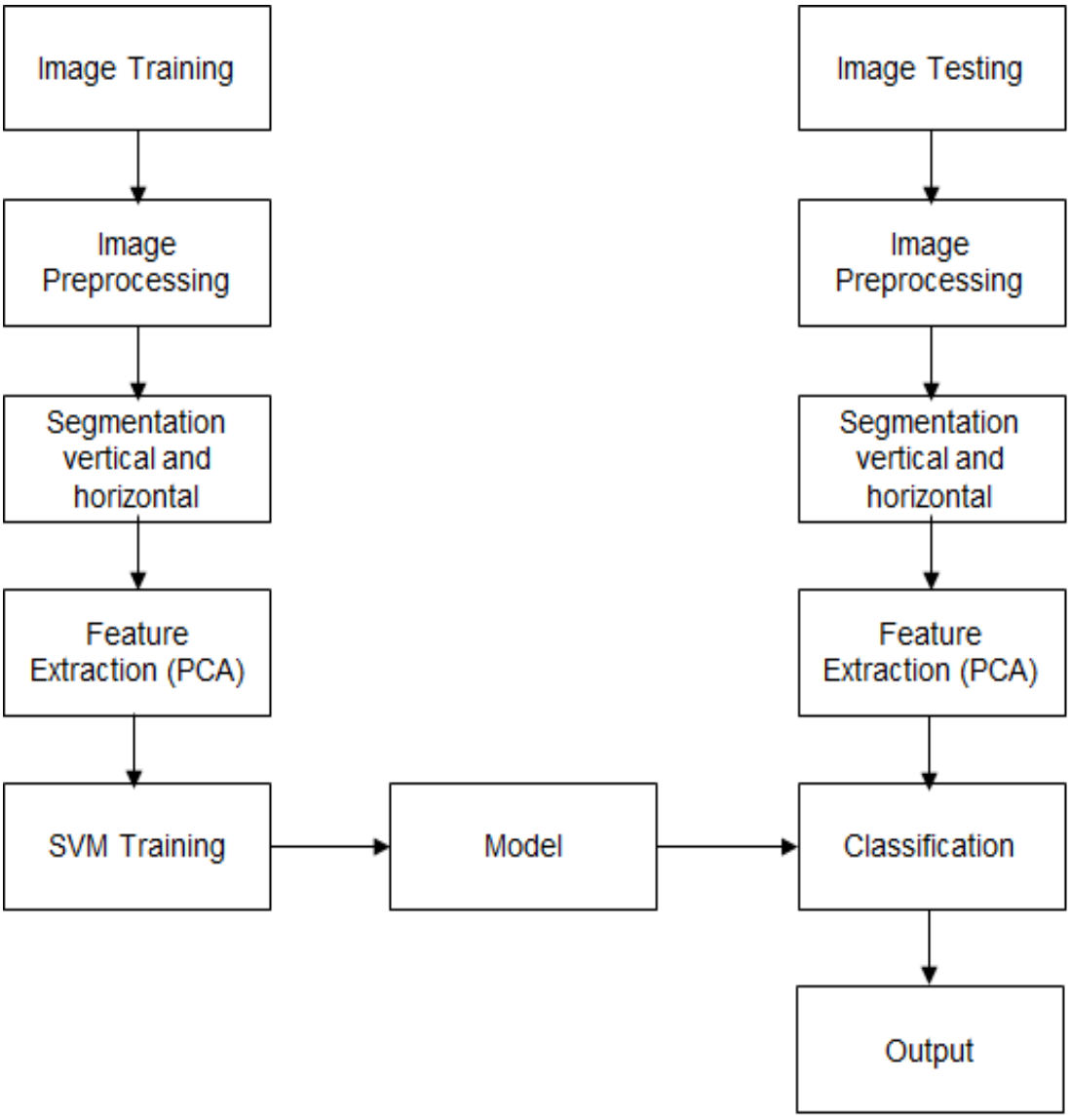
PROJECT DESIGN

In this phase, a logical system is built which fulfils the given requirements. Design phase of software development deals with transforming the client's requirements into a logically working system. Normally, design is performed in the following in the following two steps: Primary Design Phase: In this phase, the system is designed at block level. The blocks are created on the basis of analysis done in the problem identification phase. Different blocks are created for different functions emphasis is put on minimizing the information flow between blocks. Thus, all activities which require more interaction are kept in one block. Secondary Design Phase: In the secondary phase the detailed design of every block is performed. The general tasks involved in the design process are the following: 1. Design various blocks for overall system processes. 2. Design smaller, compact and workable modules in each block. 3. Design various database structures. 4. Specify details of programs to achieve desired functionality. 5. Design the form of inputs, and outputs of the system. 6. Perform documentation of the design. 7. System reviews.

5.1 Use Case Diagram:



5.2 DFD (Data Flow Diagram):



5.3 System Architecture:

The system architecture gives an overview of the working of the system. The working of the system starts with the collection of data from the database and here we divide the data into training data and testing data, selecting the attributes. And then pre-processing the required data is done so that it removes duplicate data and the error data. Firstly, user have to login and then write the personality test there are 50 questions each trait consists of 10 questions, user must answer those 50 questions, based up on the answers the algorithms are applied, and the model is trained using the training data .Here we are big five personality traits and then classify the personality type. Accuracy is measured by testing the system using testing data. So, after that Personality is predicted.

MODULE DIVISION Module Division is the process of dividing collection of source files required in the project into discrete units of functionality. Each module can be independently built, tested and debugged. Below are the modules which are divided in our project. 1.Data collection 2.Attribute selection 3.Pre-processing of data 4.Prediction of personality.

Data Collection First step for prediction system is data collection and deciding about the training and testing dataset. In this project we have imported dataset from Kaggle website which includes 70% of training dataset and 30% of testing dataset. Data collection is defined as the procedure of collecting, measuring, and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach of data collection is different for different fields of study, depending on the required information.

TRAINING DATASET: In a dataset, a training set is implemented to build up a model, while a test (or validation) set is to validate the model built. Here, you have the complete training dataset. You can extract features and train to fit a model and so on.

TESTING DATASET: Here, once the model is obtained, you can predict using the model obtained on the training set. Some data may be used in a confirmatory way, typically to verify that a given set of input to a given function produces some expected result. Other data may be used to challenge the ability of the program to respond to unusual, extreme, exceptional, or unexpected input.

CHAPTER 6

TECHNICAL SPECIFICATION

Development Environment: VS Code

VS Code also known as Visual Studio Code is a source code editor made by Microsoft for Windows, Linux, MacOS. It has various features such as Debugging, Syntax highlighting, extension, intelligent code completion.

Frontend: Python

Python is a multi-paradigm programming language. Object-oriented programming and structured programming are fully supported, and many of their features support functional programming and aspect-oriented programming (including metaprogramming and metaobjects).

Operating System (OS): Windows

Windows is a graphical operating system developed by Microsoft. It allows users to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

CHAPTER 7

PROJECT SCHEDULING

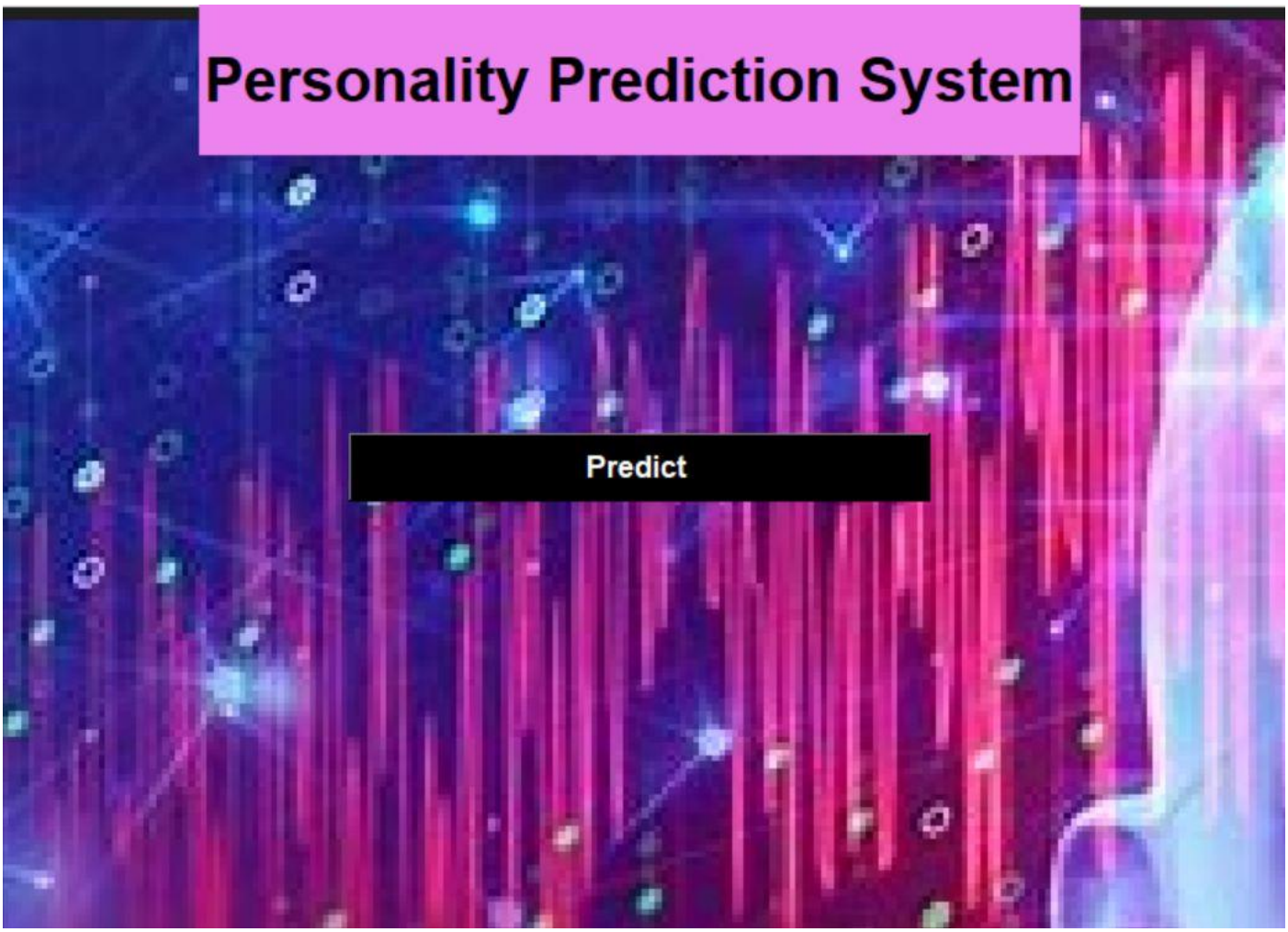
Sr. No	Group Member	Time duration	Work to be done
1	Rohin Ambati, Karan Maurya	1st week of July	Implementing 1st module/ functionality
		2nd week of July	Testing 1st module
2	Rupesh Mishra	3rd week of August	Implementing 2nd module/ functionality
3	Atharva Ankalwar	1st week of September	Implementing 3rd module/ functionality

CHAPTER 8

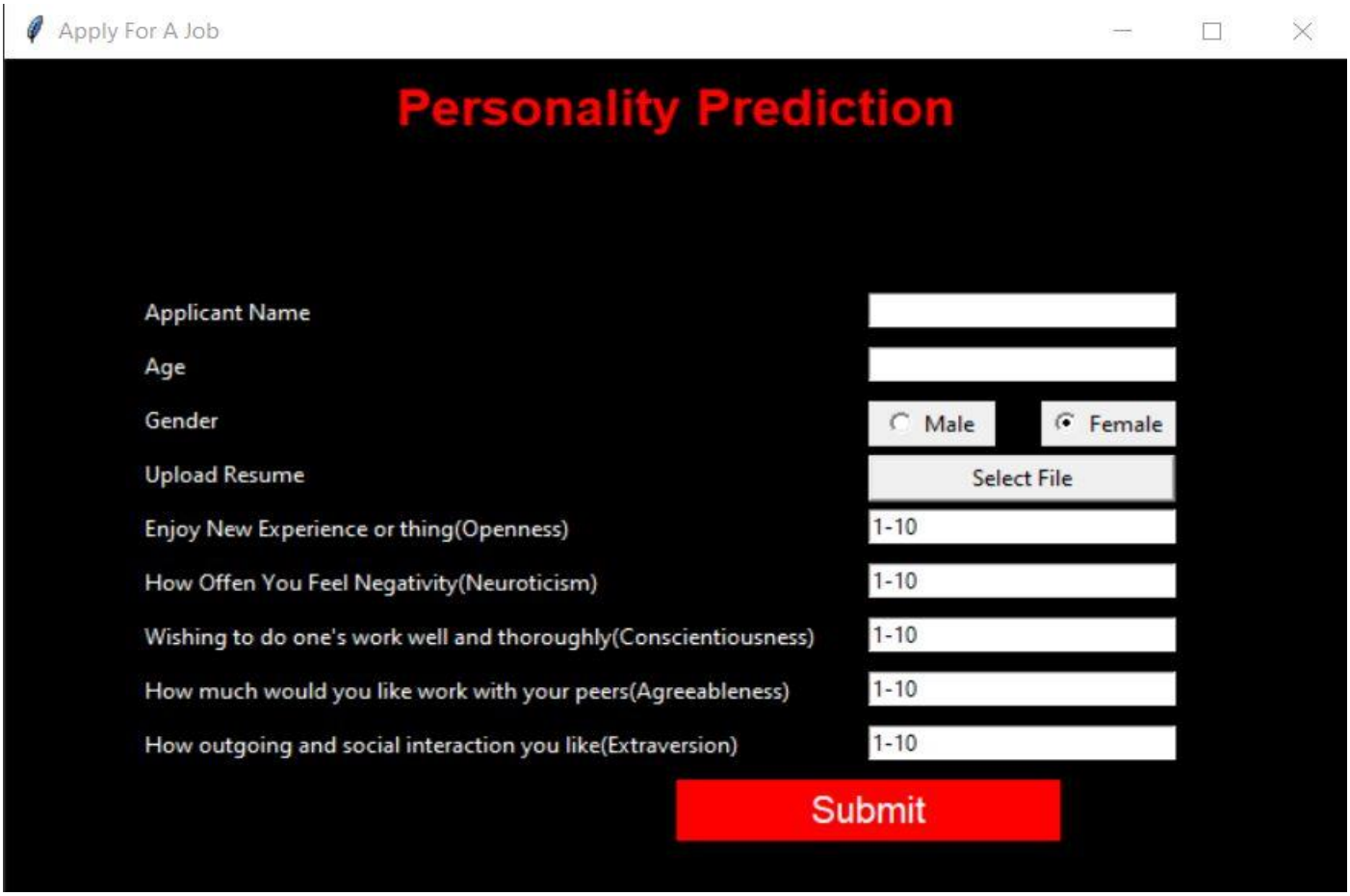
IMPLEMENTATION

Project Outputs:

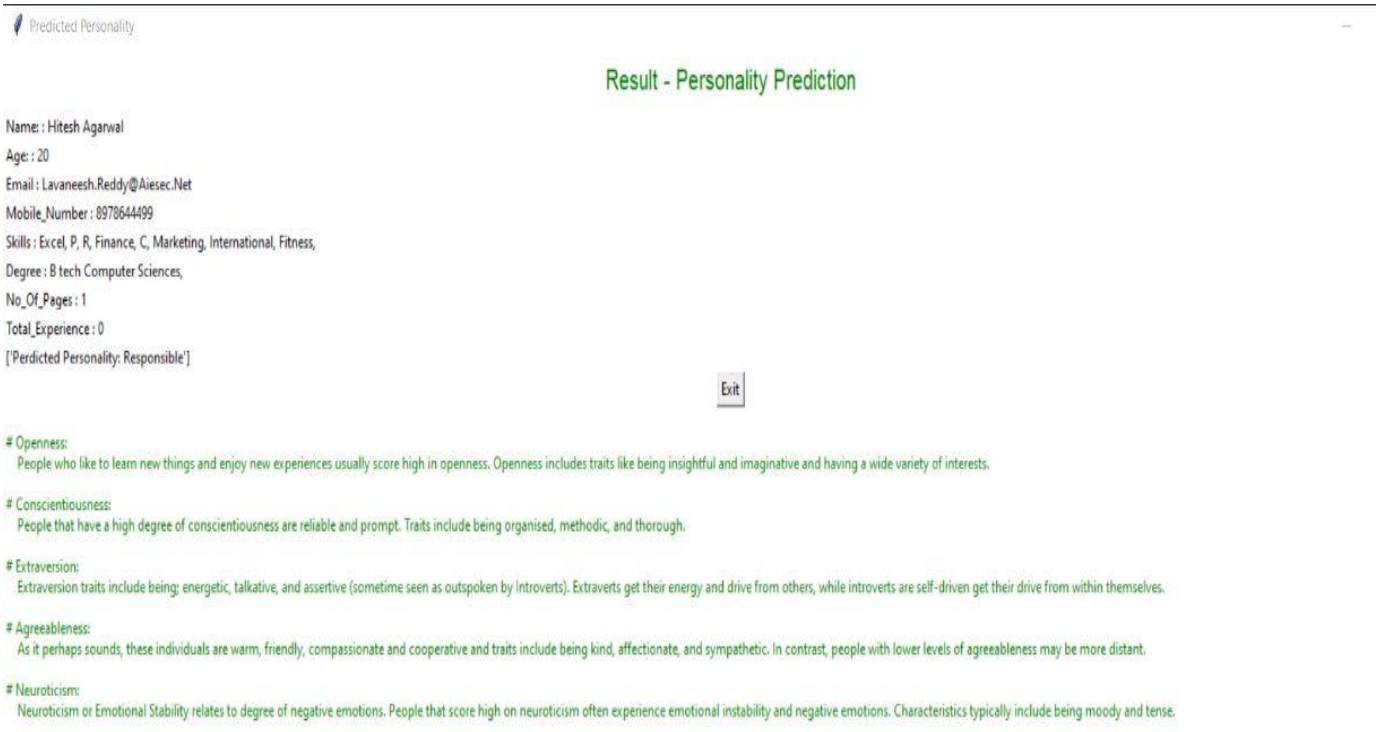
Home Page:



Main Page:



Result:



CHAPTER 9

RESULT AND DISCUSSION

Personality prediction systems are often based on the analysis of an individual's behavior, emotions, and traits, using various techniques such as self-reported questionnaires, behavioral observations, and physiological measures. These systems aim to provide insights into an individual's personality traits, such as openness, conscientiousness, extraversion, agreeableness, and neuroticism, and how they may affect their behavior and decision-making.

CV analysis, on the other hand, involves the analysis of an individual's resume or CV, to assess their skills, experience, education, and qualifications, and how they match the requirements of a specific job or role. CV analysis can provide valuable insights into an individual's professional strengths and weaknesses and their potential fit for a particular job or organization.

Combining personality prediction systems with CV analysis can provide organizations with a more comprehensive understanding of an individual's potential fit for a particular role, as it considers both their personality traits and professional qualifications. However, it is important to note that personality prediction systems are not perfect, and there is still ongoing debate around the reliability and validity of these systems.

Thus, in result, while personality prediction systems and CV analysis can provide valuable insights into an individual's personality and professional potential, they should be used as one part of a larger, holistic approach to recruitment and hiring, which also considers factors such as cultural fit, work experience, and references.

CHAPTER 10

CONCLUSION AND FUTURE SCOPE

Conclusion:-

In this project, we discussed how the personality is identified using different classification algorithms. Here we study the relationship between user and his/her personality. In this we used logistic regression because it gives best accuracy while comparing to other algorithms that are used previously like naive Bayes , SVM , Logistic regression is fast and gives accurate results compared to other algorithms. This project helps to write the personality test and check the personality of the person. From the personality classification, the person can view the type of personality and can improve the personality based upon the results.

Future Scope:-

We can modify the existing system and attach it to a questionnaire so that the personality score and the aptitude score will be calculated at the same time, thereby reducing the workload. This Personality prediction system can be extended further to other domains like Telecom, Healthcare, E-commerce and public sector jobs. Personality analysis and prediction is more in recent times so further in future more personality traits can be added. Further any improvement can be done using the data set and algorithms to improve the accuracy and can be helpful for career guidance module, if user has good speaking and convincing skills.

REFERENCES:

[1] Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C. Graesser,Using Data Mining Techniques to Detect the Personality of Players in an Educational Game

[2] R. Wald,T. M. Khoshgoftaar,A. Napolitano Using Twitter Content to Predict Psychopathy

[3]YagoSaez , Carlos Navarro , Asuncion Mochon and Pedro Isasi, A system for personality and happiness detection.

[4] https://www.enjoyalgorithms.com/blog/personality-prediction-using-ml

[5] Golbeck, J., Robles, C., and Turner, K. 2011a. Predicting Personality with Social Media. In Proc of the 2011 annual conference extended abstracts on Human factors in computing systems.

[6] DURGESH K.SRIVASTAVA, LEKHA BHAMBHU, “DATA Classification using Support Vector Machine,” Journal of Theoretical and Applied Information Technology

[7] YILUN WANG, “Understanding Personality through social media,” International of computer Science stand ford University.

[8] . MAHESH KINI, SAROJA DEVI, PRASHANT G DESAI, NIRANJAN CHIPLUNKAR,”Text Mining approach to classify Technical Research Document using naive Bayes”, International journal of Advanced Research in computer and communication engineering.

[9]. Cantandir, I. Fernandez-Tobia] z, A. Belllogin, "Relating personality types with user preferences in multiple entertainment domains," EMPIRE 1st Workshop on Emotions and Personality in Personalized Services, 2013. [10]. P.T. Costa, R.R. McCrae(1992).Revised NEO personality inventory (NEO-PIR) and NEO five-factor inventory (NEOFFI),Psychological Assessment Resource. IEEE conference templates.