AJEENKYA
D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY

School of
Engineering

PERSONALITY CLASSIFICATION

Madhavi Kad
2020-B-31102002A
BCA Data Science (DS)

Sakshi Panchal
2020-B-08022003
BCA Data Science (DS)

Salomi Pawar
2020-B-01122002
BCA Data Science (DS)

Personality Classification

Thesis submitted in partial fulfilment

of the requirements of the degree of

Bachelor of Computer Application

in

Data Science (DS)

by

Madhavi Kad

2020-B-31102002A

Sakshi Panchal

2020-B-08022003

Salomi Pawar

2020-B-01122002

Under the Supervision of

Dr. Bhargavi Dalal

April 2023

School of Engineering

Ajeenkya DY Patil University, Pune

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY
School of Engineering

April 24, 2023

CERTIFICATE

This is to certify that the dissertation entitled **"Personality Classification"** is a bona-fide work of "**Madhavi kad (URN No.: 2020-B-31102002A) and Sakshi Panchal (URN No.: 2020-B-0802003) and Salomi Pawar(URN No.: 2020-B-01122002)**" submitted to the School of Engineering, Ajeenkya D Y Patil University, Pune in partial fulfilment of the requirement for the award of the degree of **"*Bachelor of Technology in Data Science (DS)*"**.

Dr. Bhargavi Dalal
Supervisor

Internal-Examiner/s

External Examiner

Dr. Bhargavi Dalal
Head-School of Engineering

May 12, 2023

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY
School of Engineering

Dr. Bhargavi Dalal
Assistant Professor

Supervisor's Certificate

This is to certify that the dissertation entitled **"Personality Classification"** is a bona-fide work of "**Madhavi Kad (URN No.: 2020-B-31102002A) and Sakshi Panchal (URN No.: 2020-B-08022002) and Salomi Pawar(URN No.: 2020-B-01122002)**", is a record of original work carried out by him/her under my supervision and guidance in partial fulfillment of the requirements of the degree "***Bachelor of Technology in Data Science (DS)***", **Ajeenkya D Y Patil University, Pune, Maharashtra-412105**. Neither this dissertation nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

Dr. Bhargavi Dalal
Supervisor

May 12, 2023

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY
School of Engineering

Declaration of Originality

We, ***Madhavi Kad (URN No.: 2020-B-31102002A) and Salomi Pawar (URN No.: 2020-B-01122002) and Sakshi Panchal (URN No.: 2020-B-08022003)***, hereby declare that this dissertation entitled "***PERSONALITY CLASSIFICATION"*** presents my original work carried out as a bachelor student of School of Engineering, Ajeenkya D Y Patil University, Pune, Maharashtra. To the best of my knowledge, this dissertation contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of Ajeenkya D Y Patil University, Pune or any other institution. Any contribution made to this research by others, with whom I have worked at Ajeenkya D Y Patil University, Pune or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I am fully aware that in case of any non-compliance detected in future, the Academic Council of Ajeenkya D Y Patil University, Pune may withdraw the degree awarded to me on the basis of the present dissertation.

Date:

Place: Lohegaon, Pune

_____

Madhavi Kad

_____

Sakshi Panchal

_____

Salomi Pawar

May 12, 2023

Acknowledgement

I remain immensely obliged to **Dr. Bhargavi Dalal**, for providing me with the idea of this topic, and for his invaluable support in garnering resources for me either by way of information or computers also his guidance and supervision which made this Internship/Project happen.

I would like to say that it has indeed been a fulfilling experience for working out this Internship/Project.

Madhavi Kad
Sakshi Panchal
Salomi Pawar

ABSTRACT

Personality of a person decides whether he can play the role of leader, influence people around, mastering communication skills, do collaborative work, able to do negotiation is business and handle stress. This project deals with the areas wherever it determines the characteristics of someone based on the frequent patterns observed. Personality classification refers to the psychological classification of different types of individuals. The analysis is done using vast set of data in dataset and is being compared with the user input. In this project, the classification of personalities will be done on the basis of these specific characteristics; conscientiousness, openness, extroversion, agreeableness, neuroticism. Researchers have utilized social media data for auto predicting personality. However, it is confusing and complex to mine the social media data as the data can be noisy. The paper proposes machine learning techniques using Random Forest, Logistic Regression, Decision Tree, Support Vector Machine, KNN. The process of implementation and obtaining of the result will include certain steps like- Data collection, Attribute selection, Preprocessing of data, Prediction of personality. The type of personality classification and prediction can be used in certain fields like business intelligence, marketing and psychology. Research in prediction and analysis of human being is in great demand these days. Predicting the personality of candidates by this system has made things simple in varied fields like recruitment procedure, medical counselling and likewise. Personality prediction using the questionnaire helps to find out the behavioural features of the individuals taking the survey.

Key terms—personality classification and prediction, Random forest, SVM, decision tree, logistic regression, KNN, frequent patterns.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

SVM – SUPPORT VECTOR MACHINE

KNN- KN NEAREST NEIGHBOUR

DM- DATA MINING

FORTRAN- FORMULA TRANSLATION

BSD- BERKELY SOURCE DISTRIBUTION

NLTK- NATURAL LANGUAGE PROCESSING TOOLKIT

LAN- LOCAL AREA NETWORK

UML- UNIFIED MODELING LANGUAGE

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Personality is defined as the characteristics that make a person unique in terms of their way of thoughts, emotions, behaviours, habits and interests while influencing how one make decisions in life. Personality classification is one of the most studied topics in recent years. Personality is a combination of an individual's characteristics features that determines how he/she will behave in various situations. Personality influences every choice a person makes, from books to clothing to music and movies. Personality also influences its interaction with the outside world and its environment.

Personality can be used in the hiring process, career counselling, health counseling, etc. It can also be used as an additional feature.The routine method of self-estimation is time consuming and limited in scale. Also, this manual breakdown couldn't provide precise conclusions while analysing the personality of a user from their nature and conduct. Since the analysis is suited manually, it affects the exactness of the conclusions because people are biased. Data mining techniques are therefore used to study and analyse data and then identify any hidden patterns or information from a large dataset. This process is used to create my user profile, which is used to train models to predict the future behaviour of other users. Motivational influences and human behaviour are the best predictors in personality that will predict an individual's work performance. People's experiences which are emotionally important with circumstances, can also be influenced by personality. This approach reflects a person's character.

Based on the above mentioned parameters they defined the personality of a person as a set of attributes that describes a likelihood on the uniqueness of behaviour, feeling and thoughts of the person. These attributes of a person change through time and situations. In a easy term, we can admire personality as a blend of characteristics and standards that built an candidates's one in many character.

There are many different personalities used to define a person. Here we are using these big five personality traits openness, conscientiousness, neuroticism, extraversion and agreeableness. People currently deliver their thoughts and feelings through social media platforms. The posts can be in so numerous ways, similar as using an image, URL link, and music. People's personality also can be fetched using social media. The personality of people has proved to be useful in predicting job satisfaction, professional and romantic relationship success.

In the process of recruiting the rightful candidates, companies currently tend to examine the candidates' social media profiles to know the personality of the applicants for a selected job. The big five come from the statistical analysis of responses to personality details. Using a technique called factor analysis researchers can look at the responses of people to hundreds of personality items and ask the question" what is the best was to summarize an individual?".

This has been done with numerous samples from all over the world and the overall conclusion is that, while there seem to be limitless personality variables, five stand out from the rest in accordance to explaining a lot of a individual answers to questions about their personality:

extraversion, neuroticism, agreeableness, conscientiousness and openness to experience. The big five aren't associated with any particular test, a variety of measures have been developed to measure them. This test uses the Big- Five Factor Markers from the International Personality Item Pool, evolved by Goldberg(1992).

9.4 PROJECT BACKGROUND

Our Big Five Personality Traits:

• Openness: As the word suggests, This quality features characteristics such as openness and imagination and curiosity. Openness is a quality that includes imagination and insight. The world, other people, and the desire to learn and experience new things are especially high for this personality trait. This leads to you having a wide range of interests and being more adventurous when it comes to making decisions.

Creativity also plays a large role in the openness trait; this leads to a greater comfort zone when it comes to abstract and lateral thinking.

Think of a person who always orders the most exotic thing on the menu, goes to different places, and has interests that you would never have thought of... that is someone who has the traits of high openness. Anyone with this trait tends to be seen with more traditional approaches to life and may struggle when it comes to solving problems outside of their comfort zone of knowledge.

• Conscientiousness: Conscientiousness talks about a high amount of thoughtfulness, a goal-oriented attitude and good decision-makers. Conscientiousness is a trait that includes a high level of consideration, good impulse control, and goal-oriented behavior. This organized and structured approach is often found in people who work in science and even highly retail finance where detail orientation and organization are required as a skill set.

A highly conscientious person will regularly plan ahead and analyze their own behavior to see how it affects others. Project management teams and HR departments regularly have highly conscientious people on their teams to help balance structural roles within the overall team development.

A good example of a conscientious person would be someone you know who always plans ahead for the next time you meet – and in the meantime you keep in touch regularly to check in on how you're doing. They like to organize themselves around certain dates and events and focus on you when you meet.

People with low conscientiousness tend to dislike structure and plans, procrastinate on important tasks, and also fail to complete tasks.

• Extraversion: Extraversion also means extroversion is identified by excitement, talkativeness and assertiveness. Extraversion (sometimes referred to as extroversion) is a trait that many have encountered in their lives. He is easily identifiable and widely recognized as "someone who gets energized in the company of others".

This, in addition to other traits that include talkativeness, assertiveness, and a great deal of emotional expressiveness, has made extroverted people widely recognizable over many years of social interaction.

We all have that one friend or family member - or several - who aren't exactly wallflowers in social interaction. They thrive on being the center of attention, love meeting new people, and somehow tend to have the greatest friends and acquaintances you've ever known.

The opposite, of course, is someone else in our lives that we may know, an introvert. They prefer solitude and have less energy in social situations. Being the center of attention or having a conversation can be quite challenging.

Extroverts tend to have very public roles including areas such as sales, marketing, teaching and politics. Extroverts, seen as leaders, will lead rather than stand in a crowd and be seen doing nothing.

• Agreeableness: Agreeableness refers to features such as trust, affection and social behaviour of an individual. People who display high agreeableness will show signs of trust, altruism, kindness, and affection. Highly agreeable people tend to have highly prosocial behavior, which means they are more inclined to help other people.

Sharing, comforting, and cooperation are qualities that lend themselves to very agreeable personality types. Empathy towards others is commonly thought of as another form of friendliness, although the term doesn't quite fit.

The opposite of agreeableness is unpleasantness, but it manifests itself in behavioral traits that are socially unpleasant. Manipulation and nastiness towards others, lack of care or compassion, lack of concern for others and their problems are quite common.

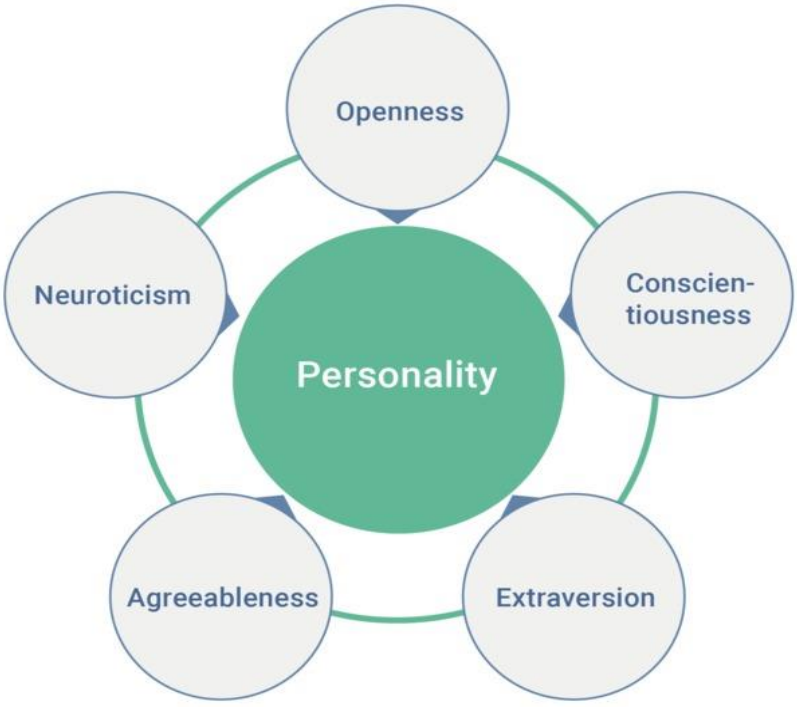
Pleasant people tend to seek careers in areas where they can help the most.

• Neuroticism: Neuroticism includes attributes like sadness, moodiness and sudden burst of emotions. As these five dimensions cover almost all avenues needed to know someone it is the right method that forms the basis of a person’s overall personality. Neuroticism is defined by various melancholic moods or feelings, and emotional instability. Often mistaken for antisocial behavior or, worse, a larger psychological problem, neuroticism is a physical and emotional reply to stress and perceived threats in someone's daily life.

Individuals who exhibit high levels of neuroticism will experience sudden mood changes, anxiety, and irritability. Some individuals who experience sudden changes I their behavioural characteristics from a day-to-day perspective may be highly neurotic and they may respond to high extent of stress in their work and personal lives.

Anxiety, which plays a major role in the composition of neuroticism, is about an individual's ability to deal with stress and real risk. People who suffer from neuroticism think about many situations and have difficulty relaxing even in their own space.

Of course, those who are lower on the neurotic level will show a more stable and emotionally resilient attitude towards stress and situations. Low neurotic people also rarely feel sad or depressed, taking the time to focus on the present moment and not engage in mental arithmetic about possible stressors.



CHAPTER 2

EXISTING AND PROPOSED SYSTEM

2.1 EXISTING SYSTEM

In the existing system, an intelligent personality prediction system is designed, which predicts the personality of candidates based on Twitter data. The current method uses machine learning to mine user characteristics and learn patterns from large amounts of personal behavioral data.

This system can automatically evaluate the personality traits of candidates by processing various attributes and eliminate the time-consuming process required by the conventional approach.

In the past few years, many studies have used various machine learning algorithms to predict personality types. One of the first studies on the Predicting Personality System from Facebook users was developed in 2017 by Tandera.

The goal of this study was to build a prediction system that can automatically predict the personality of users based on their Facebook activities. They also analyzed the accuracy of traditional machine learning and deep learning algorithm on personality prediction by implementing the Big Five personality models.

In 2018, Giulio Carducci also conducted a study on Computing Personality Traits from Tweets using Word Embedding and Supervised Learning. The researcher used a supervised learning approach to compute personality traits from an individual's historical tweets.

They developed three machine learning algorithms namely Support Vector Machine (SVM), LASSO and Logistic Regression to predict the Big Five personality model. Mohammad Hossein Amirhosseini conducted a study in 2020 on a machine learning approach to predict personality type based on the Myers–Briggs Type Indicator .

The study developed a new machine learning method to automate the process of meta program detection and personality type prediction based on the MBTI personality type indicator . The Natural Language Processing Toolkit (NLTK) and XGBoost, which is based on the Gradient Boosting library in Python, are used to implement the machine learning algorithms.

2.2 PROPOSED SYSTEM

- The system to be developed here is an Chat facility.
- It is a centralized system.
- It is Client Server system with centralized database server.
- Al local dients are connected to the centralized server via LAN.
- There is a two way communication between different clients and server.
- This chat application can be used for group discussion.
- It allows users to find other logged in users.

2.3 REQUIREMENTS

2.3.1 System Requirements:

The software system shall be compatible with Windows 10, macOS, and Linux operating systems.

The system shall require a minimum of 4GB of RAM and 1GB of free disk space.

The system shall support a minimum screen resolution of 1280x720.

CHAPTER 3

OBJECTIVE

3.1 OBJECTIVE:

The main objective of this project is to overview the data mining algorithms which are used to predict the personality of the user.

In this paper, we focus on an online dataset and then personality would be predicted accordingly based on the Big Five Personality traits. In this way, we can filter candidates applying for certain positions in the organization. Therefore, it would save the resources of the organization and they would also solicit only those applicants which would be most suitable for the job.

i. To analyze the challenges that individuals' face in career decision-making process.

ii. To analyze the relationship between personality and career.

iii. To review the techniques that have been used in career assessment.

iv. To develop an automated personality classification system that can match given personality traits with a given vocation

v. To test the developed system.

3.2 SURVEY

As part of the development process for the web-based disease prediction system, a survey was conducted to gather insights into existing approaches and technologies used in similar projects. The survey aimed to explore the use of machine learning algorithms for disease prediction, web-based interfaces for user interaction, and database management techniques. The survey findings helped in identifying the most suitable algorithms, frameworks, and tools for the project, such as SVM, KNN, Random Forest, Flask, and SQL-Alchemy.

By conducting the survey, a comprehensive understanding of the current state-of-the-art in disease prediction systems and the associated data flow was obtained. This knowledge served as a foundation for designing an effective and robust system architecture that integrates machine learning algorithms, web-based interfaces, and database management.

The survey results also informed the selection of appropriate tools and technologies, ensuring that the developed system aligns with industry best practices and meets the project's objectives.

3.3 PROBLEM STATEMENT

The world of work is very complex. It is faced with recurrent changes that have made career paths unpredictable, multi-decisional, and unstable (Mitchell, Levin, & Krumboltz, 1999). In the world of work today, there is a need to empower individuals to become autonomous decision-makers.

For successful career development, this empowerment should aim to help individuals to acquire decision-making skills.

The ability to make a career choice while faced with many alternatives is further complicated by the unpredictable work changes (Schwartz, 2004). The many alternatives present an overload of choice that require lots of effort and expertise that the subject may not have.

This means that subjects are hardly likely to benefit from the large pool of alternatives (Schwartz,2004).

To make good use of the alternatives, deliberating individuals have the option to carry out an in depth information gathering. With substantial quality information, they are then likely to do a detailed exploration of any promising alternatives after weighing their pros and cons.

3.4 ANALYSIS

In analyzing the data collected from the survey, it is clear that there is a significant demand for more environmentally friendly products. A majority of the respondents stated that they would be willing to pay more for products that are sustainable and have a lower environmental impact. This shows a growing concern for the environment and a desire to make more conscious purchasing decisions.

Furthermore, the data also revealed that many consumers are unsure about what makes a product environmentally friendly.

This highlights the need for better education and awareness regarding sustainable products and their benefits. Companies can play a crucial role in providing this education and promoting sustainable products to their customers.

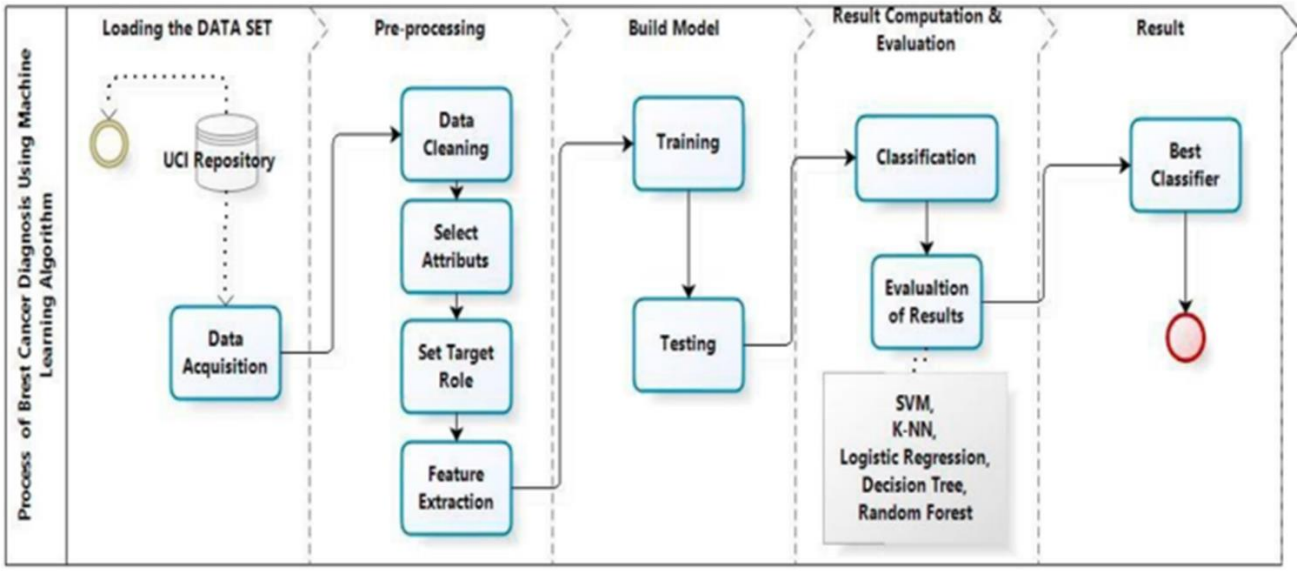
By doing so, they can not only meet the demands of consumers but also contribute to a more sustainable future.

CHAPTER 4

METHODOLOGY

4.1 SYSTEM ARCHITECTURE

Using different and machine learning algorithms like KNN, Random forest, Linear 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. 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 give suggestions to those whose personality matches like friend suggestion.



4.2 LIBRARIES

i.). NumPy:-

NumPy is a software library of python which supports scientific computing like large, multi-dimensional arrays and matrice. NumPy has many contributors and open source software interfaces.

It also has the following things:-

- An array object with powerful N-dimensions.
- Functions of broadcasting.
- C/C++ and FORTRAN code can be integrated by tools.

NumPy useful for mathematical computation like linear algebra, Fourier transform, etc. In numPy data type and size can be effectively defined. Speedily integration with a large form of databases can be allowed by NumPy. It is also licensed under BSD license but with some restrictions.

ii. Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

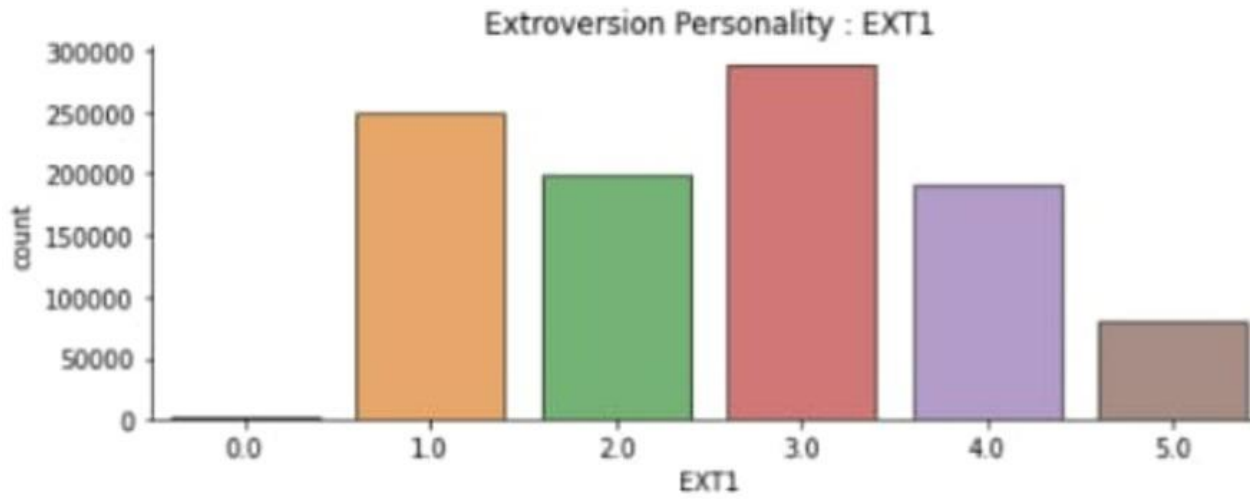
The name “Pandas” has a reference to both “Panel Data”, and “Python Data Analysis” and was created by Wes McKinney in 2008.

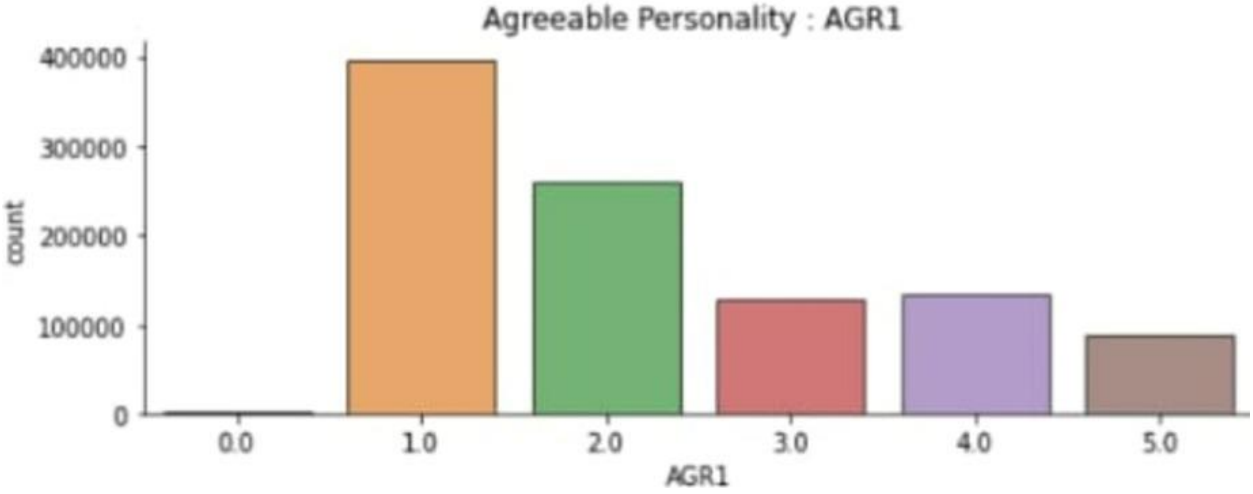
iii. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

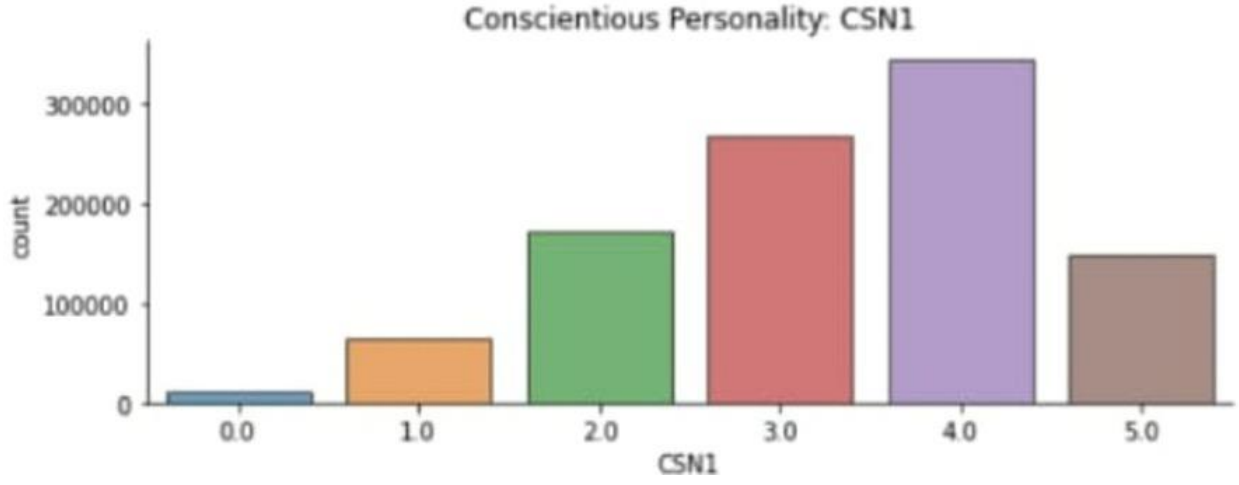
Create publication quality plots.
Make interactive figures that can zoom, pan, update.
Customize visual style and layout.
Export to many file formats.
Embed in JupyterLab and Graphical User Interfaces.
Use a rich array of third-party packages built on Matplotlib.

CHAPTER 5

ALGORITHM


Extroversion Personality : EXT1
300000
250000
200000
150000
100000
50000
0
count
0.0
1.0
2.0
3.0
4.0
5.0
EXT1


Agreeable Personality : AGR1
400000
300000
200000
100000
0
count
0.0
1.0
2.0
3.0
4.0
5.0
AGR1


Conscientious Personality: CSN1
300000
200000
100000
0
count
0.0
1.0
2.0
3.0
4.0
5.0
CSN1

Step1:
The first step is to import libraries from python. Like pandas, numpy, matplotlib, seaborn, scikit learn, certain stats models and linear models. This can be done through writing simple commands in jupyter notebook.

Step2:
Upload dataset in csv form in the jupyter notebook or any IDLE that you are using. This dataset involves the records of certain statements that are being given as survey questions.

Step3:
Read the dataset- This step involves the extraction of information from dataset and briefing a summary out of it.

Step4:
Finding the null values of the dataset so as to increase the accuracy of further algorithms. If the missing values are less we fill it with 0 and if there are a lot of null values in a certain column we will delete that column to fulfill the above mentioned reason.

Step5:
Visualization and analysis-In this step we analyze the big five personalitites with the help of various plots and graphs available in the python matplotlib and seaborn libraries. Here we derive the conclusion of the entire dataset on the basis of the big five personality traits.

Step 6:
Classification and prediction- We then classify the dataset and predict the personalities on the given data and on the foundation of these traits like agreeableness, neuroticism etc.,.

CHAPTER 6
IMPLEMENTATION

6.1 Data Collection:

In this project, the dataset for disease prediction is collected from Kaggle, a popular platform for data science competitions and datasets. The dataset contains medical information for many patients, including their age, gender, medical history, and other relevant features.

The use of the Kaggle dataset allows for a comprehensive analysis of the disease prediction problem and provides a large amount of data for the training and testing of the machine learning models. The dataset has been used by many researchers and data scientists for similar projects, ensuring that the results obtained are reliable and accurate.

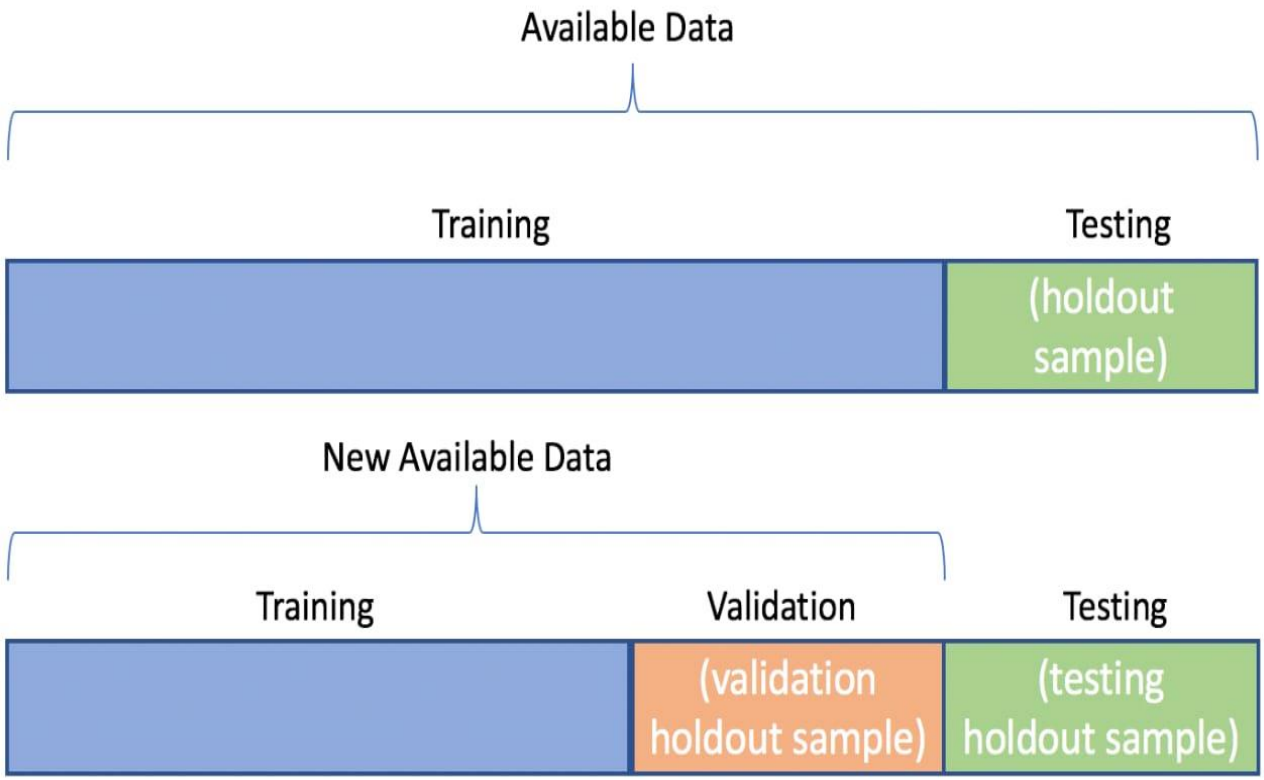
6.2. Data Preparation:

Data pre-processing is a critical step in any machine learning project. In this project, the dataset for disease prediction is pre-processed to handle missing values, outliers, and categorical variables. The dataset is first cleaned to remove any irrelevant or duplicate data points. (e main use of data cleaning is to ignore the missing value and compute the value in a filled format [3].

Next, missing values in the dataset are handled by either imputing values using mean or median or by removing the entire row if a significant number of values are missing. Outliers, which can skew the training process of machine learning models, are identified, and removed using statistical techniques like Z-score or IQR (Interquartile Range).

Feature selection is also an important step in data pre-processing, as it helps to identify the most relevant and significant features that contribute to the accuracy of the model. In this project, feature selection is done using various techniques such as correlation matrix, recursive feature elimination, and principal component analysis (PCA).

Correlation matrix is used to identify the correlation between different features in the dataset. Features that are highly correlated with each other are removed to avoid overfitting. Recursive feature elimination (RFE) is a backward selection process that recursively removes the least significant features until the optimal number of features is reached. PCA is a dimensionality reduction technique that transforms the dataset into a lower-dimensional space while preserving most of the variability in the data.



6.3. Feature Selection:

The feature selection process using two methods; UFS and RFECV (Recursive Feature Elimination with Cross-Validation). Resulted in two different sets of features. The resulted subset of features has been used in the training of Logistic Regression, SVM, and Decision Tree algorithms. The selected features for all three algorithms were different because recursive feature elimination with cross-validation can automatically eliminate less predictive features iteratively depending on the model. Shows the number of selected features for binary and multi-class models.

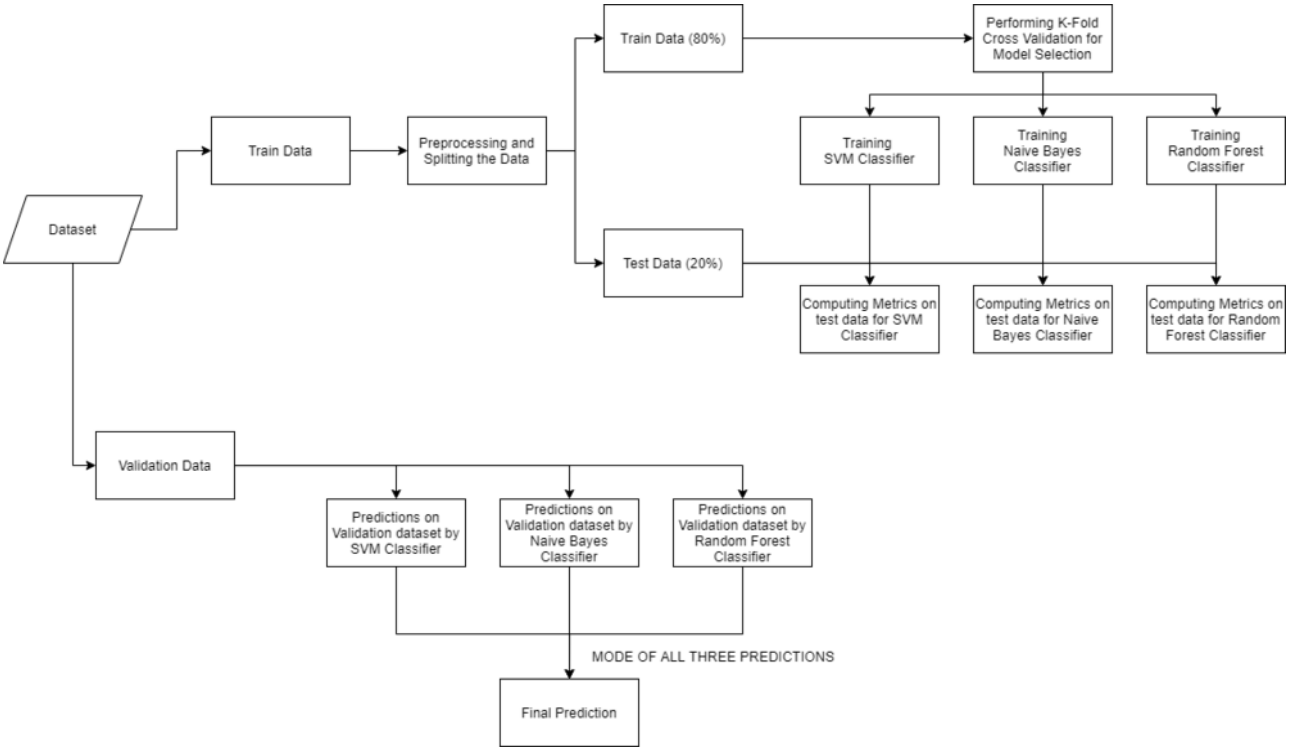
6.4. Model Preparation:

The proposed work aims to develop a multiple disease prediction system using machine learning algorithms such as logistic regression, support vector machine, and decision tree. The dataset will be collected from reliable sources and pre-processed to handle missing values and feature selection will be done and dataset is divided into training data, testing data, and validation data. Where various classifiers are then compared, and the best suited classifier for the dataset is then chosen and applied to obtain the most significant predictors.

The machine learning models will be trained on the pre-processed dataset to predict the likelihood of various diseases. Performance metrics such as accuracy, precision, and recall will be used to evaluate the models, and the best performing model will be selected for deployment

CHAPTER 7

UML DIAGRAM



The UML diagram for the machine learning model data flow can include the following components and their interactions:

Data Pre-processing:

The data pre-processing component is responsible for cleaning and transforming the input data before feeding it to the machine learning model.

It may involve steps such as handling missing values, outlier detection and removal, feature scaling, and encoding categorical variables.

The pre-processed data is then passed to the feature selection component.

Feature Selection:

The feature selection component aims to identify the most relevant features from the pre-processed data that contribute to the prediction task.

It can utilize techniques such as statistical tests, correlation analysis, or feature importance scores from the machine learning model.

The selected features are then passed to the model training component.

Model Training:

The model training component involves training the machine learning model on the selected features and corresponding target labels.

It can include various algorithms such as Random Forest, XGBoost, or Gradient Boosting, as mentioned earlier.

The trained model is then ready for prediction.

Prediction:

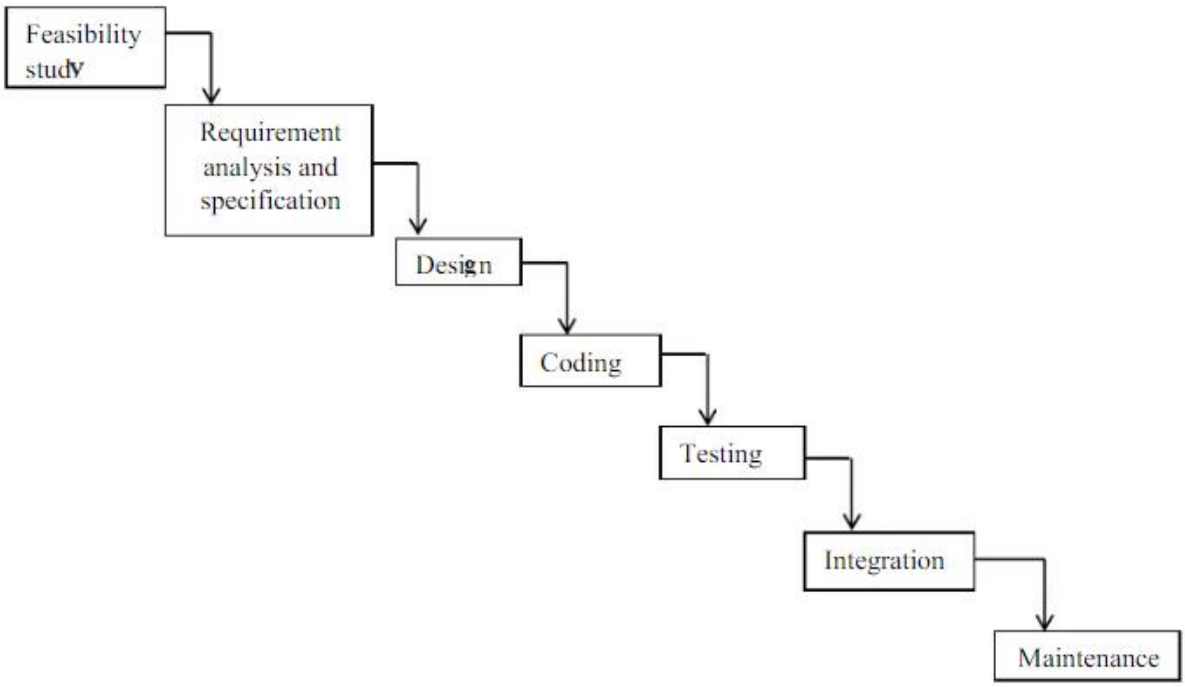
The prediction component takes new input data from the user interface or external sources and applies the pre-processing steps performed during training.

The pre-processed data is fed into the trained machine learning model for prediction.

The output of the prediction component is the predicted disease or the probability of occurrence, which is then displayed to the user through the user interface.

Note: The UML diagram can further include interactions with the database, user authentication, and other components specific to the overall system architecture**.**

CHAPTER 8
SDLC DIAGRAM



Every activity has a life cycle and the software development process is no exception. Even if you are not aware of the SDLC, you still have to follow it unconsciously. However, if a software professional is aware of the SDLC, they can execute the project in a very controlled manner.

One of the big benefits of this awareness is that hot-blooded developers don't start executing (coding) directly, which can really lead to a project running out of control. Second, it helps customers and software professionals avoid confusion by anticipating problems and issues in advance.

In short, the SDLC defines the different phases of the software life cycle. But before we try to understand what SDLC is all about. We need to get a broader view of the start and end of the SDLC.

Any project started if it doesn't have a beginning and an end, it's already in trouble. It's like when you go on a drive, you should know where to start and where to end, otherwise you'll be moving endlessly. Below is a figure that shows a typical flow in an SDLC that has five main models.

• Waterfall – big bang and phasing model.
• Iterative – spiral and incremental model.

Iterative model
The iterative model was introduced due to the problems faced by the waterfall model. An iterative waterfall model is used in system development. The system is developed in increments, with each increment adding some functionality to the system until the entire system is fully implemented. The advantage of this approach is that it will lead to better testing because testing each increment is easier than testing the entire system as a whole. In addition, this approach provided us with important feedback that was very useful in the implementation of the system.

CHAPTER 9
HARDWARE AND SOFTWARE REQUIREMENTS

9.1 SOFTWARE REQUIREMENTS

ANACONDA

Anaconda is primarily developed to support data science and machine learning tasks. It focuses on the distribution of the R and Python programming languages and aims to simplify data management and deployment of said languages.
It offers all the necessary packages related to data science at once. Despite this, programmers still choose this program for its quick installation and ease of use.
Installing the app is simple as it only requires following the instructions of the setup wizard.
Upon completion, the app will provide you with more than 1,500 packages in its distribution. In it you will find Anaconda Navigator, which is a graphical alternative to the command line interface. This makes it easy for users to run applications and manage packages and environments without using command line commands.
You will also find the Conda package, which is a virtual environment manager. This feature installs dependencies quickly along with frequent updates.
It also makes it easy to create and load at the same speed and even allows for easy environment switching.
Conda differs from Python's PIP in that it checks for dependency requirements before installing. More importantly, it provides warnings if dependencies already exist.
Anaconda is a data science platform made on the Python programming language. This open source development tool acts as an all-in-one data management tool, creating an environment that facilitates access to large amounts of data. If you and your team need to secure, interpret, scale and store important data, this application can help.
However, users must be aware that Anaconda is aimed at a very narrow audience. It focuses on a large amount of data, so it is not suitable for small projects. Competitions like Dev C++ are a better choice when you're working to produce smaller amounts of data.

EXCEL

Excel is typically used to organize data and perform financial analysis. It is used across all business functions and in companies from small to large.
The main uses of Excel include:
Data entry
Data management
Bookkeeping
Financial analysis
Charts and graphs
Programming
Organization of time
Task management
Financial modeling
Customer Relationship Management (CRM)
Almost everything that needs to be organized!
Excel is defined as a "data" managing tool, the data that is most often managed is financial. At CFI, we would define Excel as the best financial software. While there are other pieces of financial software that are tailored to perform specific tasks, Excel's strongest point is its robustness and openness. Excel models are as powerful as an analyst wants them to be.

Py-Charm

PyCharm is an integrated development environment (IDE) used for programming in Python. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems, and supports the development of web applications with Django. PyCharm is developed by the Czech company JetBrains.

It is multi-platform, works on Microsoft Windows, MacOS and Linux. PyCharm has a Professional Edition, released under a proprietary license, and a Community Edition, released under the Apache License. PyCharm Community Edition is less extensive than Professional Edition.

Coding help and analysis with code completion, syntax and error highlighting, linter integration and hotfixes

Project and code navigation: dedicated project views, file structure view and quick jump between files, classes, methods and usages

Python code refactoring: including renaming, method extraction, variable introduction, constant introduction, pull up, move down and more

Support for web frameworks: Django, web2py and Flask

Integrated Python debugger

Integrated unit testing with line-by-line coverage

Google App Engine Python development

Version control integration: unified user interface for Mercurial, Git, Subversion, Perforce and CVS with changelists and merge

Scientific tools integration: integrates with IPython Notebook, has an interactive Python console and supports Anaconda as well as several scientific packages including Matplotlib and NumPy.

HARDWARE REQUIRED

- Processor - Intel Core I7 9th Generation
- Speed - 3 GHz
- RAM - 8 GB RAM
- Device - VOSTRO

CHAPTER 10
COMPARATIVE STUDY

SR.NO	Publication And Year.	Features	Drawbacks	Technology used
1.	IEEE Access 2020	Automatic personality recognition	Uneven Data Format	Logistic Regression, Random Forest
2.	IJISRT 2022	Predicts personality labels	Less accurate	KNN, SVM
3.	IJACSA 2020	Predicts individual work performance	Classifies employment data only	Random forest classifier
4.	IJRAR 2020	Multi-label classifiers	Time consuming	SVM, Decision tree, Naive Bayes
5.	IRJETS 2022	Text based personality prediction	overfitting problem	KNN , Logistic Regression
6.	GCET 2021	High Efficiency of algorithms.	Time Consuming	Naïve Bayes Classifier , SVM
7.	CEID, Greece 2021	Automated candidate ranking	Deteriorating accuracy	Decision Tree, linear regression

CHAPTER 11
FUTURE SCOPE AND CONCLUSION

11.1 CONCLUSION

In this paper, we have used various Machine Learning Algorithms such as Logistic Regression, Naive Bayes, Random Forest, SVM and KNN for Personality prediction using CV Analysis. Using pyre sparser, spaCy and Phrase Matcher we were able to predict the personalities of various candidates. The results indicate Random Forest has the maximum accuracy of 0.71 however due to lack of available data the accuracy is much lesser than it was anticipated. The proposed system can be used by various companies in order to streamline the recruitment process by considering the personality of potential candidates. Future work can also be done to increase the efficiency and performance of the proposed system in order to predict personality using CV analysis more accurately. The predictive analysis of personality assessments provides an overview of a candidate's behavioural tendencies and that allows recruiters to really understand if a candidate will, in fact, be a top performer and if he will fit the culture of the company. It helps people learn about their personality types and attributes. It reveals the candidate's personality traits, such as work preferences, motivation, strengths and weaknesses and attitude. Personality tests can reveal whether or not the candidate fits the position and the team in more than just capability and skills. It displays whether they are capable of thinking on their feet, how they approach problem solving, and whether they display leadership skills when necessary or under pressure.

REFERENCES

A Demetriou, L. Kyriakides, and C. Avraamidou, The missing link in the relations between intelligenceand personality in Journal of Research in Personality,vol. 37 issue 6,December 2003, pp 547- 581.

M. Kalghatgi, M Ramannavar, and Dr. N. S. Sidnal, Neural Network approach to personality prediction based on the Big-Five Model in IJIRAE, vol2 issue 8, August 2015, pp 56-63.

A..Robey, K. Shukla, K. Agarwal, K. Joshi, Professor S. Joshi Personality prediction system through CV Analysis, in IRJET vol 6, issue 02, February2019.

J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem Resume Ranking using NLP and Machine Learning, unpublished

Md Tanzim Reza, and Md. Sakib Zaman, Analyzing CV/Resume using natural language processing and machine learning, unpublished.

https://www.ijrar.org/papers/IJRAR19H1202.pdf

https://www.google.com/search?q=future+scope+and+conclusion+for+personality+classification&oq=future+scope+and+conclusion+for+personality+classification&aqs=chrome.0.69i59.5557j0j1&sourceid=chrome&ie=UTF-8#imgrc=APgvrcKhl4N7dM

https://www.thomas.co/resources/type/hr-guides/what-are-big-5-personality-traits