

## **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 outside world. This project is helpful where we have data related to personal behaviour. This personal behaviour data can be useful for identifying person based on his/her personality traits. The personality characteristics will be already stored in 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 persons 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.

**Key Words:** Personality, Behaviour, Logistic regression, Decision tree, Support Vector Machine, Big Five personality Traits.

# Contents

<b>ABSTRACT</b>	5
<b>LIST OF SYMBOLS</b>	8
<b>LIST OF FIGURES</b>	9
<b>LIST OF TABLES</b>	10
<b>LIST OF ABBREVIATIONS</b>	11
<b>CHAPTER 1: INTRODUCTION</b>	12
1.1 INTRODUCTION	12
1.1.1 BIG FIVE PERSONALITY TRAITS	13
1.1.2 MACHINE LEARNING	16
1.1.3 DATA PREPROCESSING	18
1.1.4 DATA MINING	20
1.1.5 PYTHON	22
1.2 MOTIVATION OF THE WORK	26
1.3 PROBLEM STATEMENT	27
<b>CHAPTER 2: LITERATURE SURVEY</b>	28
2.1. Novel approaches to automated personality classification:	28
2.2. Educational Game (Detecting personality of players in an educational game):	28
2.3. A System for Personality and Happiness Detection:	29
2.4. Using Twitter Content to Predict Psychopathy:	29
2.5 An Examination of Online learning effectiveness using data mining	30
<b>CHAPTER 3: PROPOSED METHODOLOGY</b>	31
3.1 INTRODUCTION	31
3.2 SYSTEM ARCHITECTURE	32
<b>CHAPTER 4: EXPERIMENTAL ANALYSIS AND RESULTS</b>	33
4.1 MODULE DIVISION	33
4.1.1 Data Collection	33
4.1.2 Attribute Selection	34
4.1.3 Pre-Processing of Data	34
4.1.4 Prediction of Personality Classification	34
4.2 ALGORITHM	35
4.3 SAMPLE CODE	45
4.4 RESULTS	62
4.4.1 INPUT IMAGES	62

## LIST OF FIGURES

FIGURE NO.	TITLE
1	Big Five personality trait
2	Machine Learning
3	Splitting of data set
4	Data mining
5	System Architecture
6	Logistic Regression
7	Decision Tree
8	Support Vector Machine

# **CHAPTER 1: INTRODUCTION**

## **1.1 INTRODUCTION**

Personality identification of a human being by their nature an old technique. Earlier these were done manually by spending lot of time to predict the nature of the person. Data mining is primarily used today by companies with a strong consumer focus- retail, financial, communication, and marketing organizations. Methods used to analyse the data include surveys, interviews, questionnaires, classroom activities, shopping website data, social network data about the user experiences and problems they are facing. But these traditional methods are time consuming and very limited in scale. Our Proposed system will provide information about the personality of the user. Based on the personality traits provided by the user, System will match the personality traits with the data stored in database. System will automatically classify the user's personality and will match the pattern with the stored data. System will examine the data stored in database and will match the personality traits of the user with the data in database. Then system will detect the personality of the user. Based on the personality traits of the user, 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 recruitment process, career counselling, health counselling, etc. Predicting personality by analysing the behaviour of the person is an old technique. This manual method of personality prediction required a lot of time and resources. Analysing personality based on one's nature was a tedious task and a lot of human effort would be required to do such analysis. Also, this manual analysis did not give accurate results while analysing the personality of a user from their nature and behaviour. Since analysis was done manually, it affects the accuracy of the results as humans prone to be prejudice and generally see the things accordingly.

### **1.1.1 BIG FIVE PERSONALITY TRAITS**

The Big Five Personality traits are the five dimensions or the domains of personality that can be used to analyse or predict the personality of a user. The Big Five Personality Model is the most widely accepted and researched model for predicting the personality of a user. The Big Five Personality traits are found in a variety of people of different ages, locations and cultures. The Big Five Personality results are very accurate and predict the true personality of a user to a large extent.

The Big Five Factors are:

1. Openness to Experience or Imagination Capability
2. Agreeableness
3. Extraversion
4. Neuroticism or Emotional Stability
5. Conscientiousness

- The Big Five personality traits extraversion (also often spelled extroversion), agreeableness, openness, conscientiousness, and neuroticism.
- Each trait represents a continuum. Individuals can fall anywhere on the continuum for each trait.
- The Big Five remain relatively stable throughout most of one's lifetime.
- They are influenced significantly by both genes and the environment, with an estimated heritability of 50%.
- They are also known to predict certain important life outcomes such as education and health.

#### **1. Openness to Experience:**

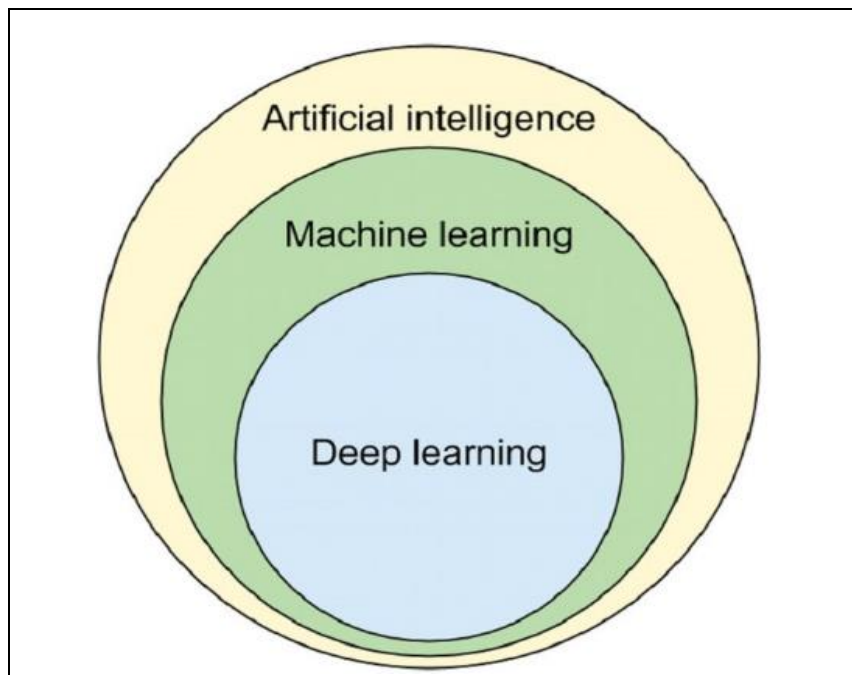
Openness to experience refers to one's willingness to try new things as well as engage in imaginative and intellectual activities. It includes the ability to "think outside of the box."

This trait features characteristics such as imagination and insight. People who are high in this trait also tend to have a broad range of interests. They are curious about the world and other people and eager to learn new things and enjoy new experiences.

People who are high in this trait tend to be more adventurous and creative. People low in this trait are often much more traditional and may struggle with abstract thinking.

## 1.1.2 MACHINE LEARNING

**Machine learning (ML)** is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed task



**Fig(2). Deep learning is a subset of machine learning which is a subset of artificial intelligence.**

In addition to an informed, working definition of machine learning (ML), we detail the challenges and limitations of getting machines to ‘think,’ some of the issues being tackled today in deep learning (the frontier of machine learning), and key takeaways for developing machine learning applications for business use-cases. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

All of these things mean it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results – even on a very

large scale. And by building precise models, an organization has a better chance of identifying profitable opportunities – or avoiding unknown risks.

Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict.

- **Supervised learning:**

In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.

Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression etc.

- **Unsupervised learning:**

This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.

Examples of Unsupervised Learning: Apriori algorithm, K-means.

- **Semi-supervised learning:**

This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.

- **Reinforcement learning:**

Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to

complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

Example of Reinforcement Learning: Markov Decision Process.

### 1.1.3 DATA PREPROCESSING

There are seven significant steps in data pre-processing in Machine Learning:

#### 1. Acquire the dataset

To build and develop Machine Learning models, you must first acquire the relevant dataset. This dataset will be comprised of data gathered from multiple and disparate sources which are then combined in a proper format to form a dataset. Dataset formats differ according to use cases

#### 2. Import all the crucial libraries

Since Python is the most extensively used and also the most preferred library by Data Scientists around the world, we'll show you how to import Python libraries for data preprocessing in Machine Learning. Read more about [Python libraries for Data Science here](#). The predefined Python libraries can perform specific data preprocessing jobs. The three core Python libraries used for this data preprocessing in Machine Learning are:

- **NumPy**  
NumPy is the fundamental package for scientific calculation in Python. Hence, it is used for inserting any type of mathematical operation in the code. Using NumPy, you can also add large multidimensional arrays and matrices in your code.
- **Pandas**  
Pandas is an excellent open-source Python library for data manipulation and analysis. It is extensively used for importing and managing the datasets. It packs in high-performance, easy-to-use data structures and data analysis tools for Python.
- **Matplotlib**  
Matplotlib is a Python 2D plotting library that is used to plot any type of charts in Python. It can deliver publication-quality figures in numerous hard copy formats and interactive environments across platforms (IPython shells, Jupyter notebook, web application servers, etc.).

#### 3. Import the dataset

In this step, you need to import the dataset/s that you have gathered for the ML project at hand. However, before you can import the dataset/s, you must set the current directory as the working



Training set denotes the subset of a dataset that is used for training the machine learning model. Here, you are already aware of the output. A test set, on the other hand, is the subset of the dataset that is used for testing the machine learning model. The ML model uses the test set to predict outcomes.

## **7. Feature scaling**

Feature scaling marks the end of the **data preprocessing in Machine Learning**. It is a method to standardize the independent variables of a dataset within a specific range. In other words, feature scaling limits the range of variables so that you can compare them on common grounds.

### **1.1.4 DATA MINING**

The process of extracting information to identify patterns, trends, and useful data that would allow the business to take the data-driven decision from huge sets of data is called Data Mining.

In other words, we can say that Data Mining is the process of investigating hidden patterns of information to various perspectives for categorization into useful data, which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cost-cutting and generating revenue.

Data mining is the act of automatically searching for large stores of information to find trends and patterns that go beyond simple analysis procedures. Data mining utilizes complex mathematical algorithms for data segments and evaluates the probability of future events. Data Mining is also called Knowledge Discovery of Data (KDD).

Data mining works in conjunction with predictive analysis, a branch of statistical science that uses complex algorithms designed to work with a special group of problems. The predictive analysis first identifies patterns in huge amounts of data, which data mining generalizes for predictions and forecasts. Data mining serves a unique purpose, which is to recognize patterns in datasets for a set of problems that belong to a specific domain.

**Fig(3).Data Mining involves of Statistics, Artificial Intelligence, Machine Learning**



We can also define data mining as a technique of investigation patterns of data that belong to particular perspectives. This helps us in categorizing that data into useful information. This useful information is then accumulated and assembled to either be stored in database servers, like data warehouses, or used in data mining algorithms and analysis to help in decision making. Moreover, it can be used for revenue generation and cost-cutting amongst other purposes.

### **Advantages of Data Mining**

- The Data Mining technique enables organizations to obtain knowledge-based data.
- Data mining enables organizations to make lucrative modifications in operation and production.
- Compared with other statistical data applications, data mining is a cost-efficient.
- Data Mining helps the decision-making process of an organization.
- It Facilitates the automated discovery of hidden patterns as well as the prediction of trends and behaviours.
- It can be induced in the new system as well as the existing platforms.
- It is a quick process that makes it easy for new users to analyze enormous amounts of data in a short time.

### **Data Mining Applications**

- Data Mining in Healthcare
- Data Mining in Market Basket Analysis
- Data mining in Education
- Data Mining in Manufacturing Engineering
- Data Mining in CRM (Customer Relationship Management)
- Data Mining in Fraud detection
- Data Mining in Lie Detection
- Data Mining Financial Banking