



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.524

Volume 14, Issue 1, January 2025

Personality Prediction using ML

Pratik Shinde¹, Tushar Malode², Abhishek Khemnar³, Pruthviraj Dhale⁴, Prof. Rupali Patil⁵

Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, India¹

Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, India²

Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, India³

Department of Information Technology, Sandip Institute of Technology and Research Centre, Nashik, India⁴

Assistant Professor, Department of Information Technology, Sandip Institute of Technology and Research Centre,
Nashik, India⁵

ABSTRACT: Data Science and AI are revolutionizing the planet through technical transformations. One of the most significant uses of machine learning is the classification of people based on their personality features. We can see many applications of machine learning in our daily lives. Every human on the world has distinct personality type. This is made by the availability of high dimensional data. Using the Big Five personality traits, we created a system for predicting personality. t. Daily numerous students are writing competitive exams that mainly focuses on his/her Personality. The main motive of these exams is to check the student's personality and skills. 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.

KEY WORDS: personality prediction, Machine Learning, K-means Clustering, Gaussian Mixture model Ocean Model

I. INTRODUCTION

The big five personality traits, often mentioned as OCEAN. OCEAN stand for Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. These five personality traits represent broad domains of human behaviour and account for differences in both personality and decision making. Today, the model is employed by HR practitioners to measure potential employees and marketers to understand the audiences of their products We are deploying OCEAN Model in this project for developing the algorithm

Openness to experience: Also called as intellect or imagination, this personality trait represents the willingness to try new things and think out of the box. This trait includes insightfulness, originality, and curiosity.

Conscientiousness: The desire to be careful, diligent and regulate immediate gratification with self-discipline. This trait includes ambition, discipline, consistency and reliability.

Extroversion: A state where an individual draw energy from others and seek social connections or interaction, as opposed to being alone. This trait includes being outgoing, energetic, and confident.

Agreeableness: The measure of how an individual interacts with other individuals, characterized by the degree of compassion and cooperation. This trait includes tactfulness, kindness, and loyalty.

Neuroticism: A tendency for negative personality traits, emotional instability, and self-destructive thinking. This trait includes pessimism, anxiety, insecurity and feature .



II. LITERATURE SURVEY

- 1. Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi (2019)** – Personality Prediction System through CV Analysis: A machine learning system analyses CVs using NLP and the OCEAN model to predict personality traits. Challenges include diversity in CV formats and data preprocessing.
- 2. Devesh Agarwal, M. Karthikeyan (2022)** – Personality Prediction Using Machine Learning: Machine learning (Decision Trees, Random Forest) is used to predict personality traits from user data, with high accuracy, especially with Random Forest. Useful for recruitment.
- 3. Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat (2021)** – Personality Prediction Via CV Analysis using Machine Learning: CV text is analysed using NLP and algorithms like Naive Bayes and SVM. TF-IDF feature extraction improves accuracy but is hindered by unstructured CVs.
- 4. M. Kalghatgi, M.Ramannavar, Dr. N. S. Sidnal (2015)** – Neural Network Approach to Personality Prediction Based on the Big-Five Model: Neural networks outperform traditional ML methods in predicting personality traits using the Big-Five model, offering better accuracy and generalization.
- 5. Zubeda, M. Shaheen, G. Narsayya Godavari, S. Naseem – Resume Ranking Using NLP and Machine Learning:** NLP and machine learning rank resumes based on content. Though not focused on personality prediction, the study shows the importance of structured data in recruitment.
- 6. Md Tanzim Reza, Md. Sakib Zaman** – Analyzing CV/Resume Using NLP and Machine Learning: A system using NLP and ML to extract skills and personality traits from CVs for job compatibility. Preprocessing and feature extraction are key for accuracy.
- 7. Sudhir Bagade, Jayashree Rout, Pooja Yede** – Personality Evaluation and CV Analysis Using Machine Learning: ML and NLP are used to predict personality traits from CVs. Challenges include feature extraction and standardizing data for better prediction.

III. PROPOSED MODEL

One of the major challenges for the project will be the collection of input datasets for the algorithm. For conducting the test, we are using K – Mean Clustering Algorithm. The dataset for testing the algorithm is collected from the

participant. This is done by giving a questionnaire on personality classification. Then, the collected information is fed to the personality classification algorithm i.e., K-Mean Clustering Algorithm. Finally, the algorithm evaluates the data on the basis of the big five personality traits and displays the result. To give you an idea, here is a diagrammatical representation of the whole process we will follow for the conduction of the project.

3.1 Methodology

The OCEAN model, also known as the Big Five personality traits, is a widely accepted framework for understanding human personality. It comprises five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Given the rise of machine learning and artificial intelligence, these technologies can be used to predict personality traits based on data such as social media activity, text analysis, behavioral patterns, and more.

1. Data Collection: Collect data from Kaggle.com, Ensure that data collection complies with data privacy laws and obtain user consent where necessary.
2. Data Preprocessing: Remove noise, redundant information, and irrelevant data. Extract meaningful features from text data. Standardize data formats, normalize scales, and handle missing values.
3. Feature Selection: Use domain knowledge and statistical methods to identify features that correlate with the OCEAN traits. Create new features that can help with personality prediction (e.g. selecting more appropriate questions).
4. Model Selection: Consider machine learning algorithms suitable for this task, such as K mean clustering, Gaussian Mixture Model algorithm. Use ensemble learning techniques to combine multiple models for improved accuracy and robustness.
5. Model Training and Validation: Use a labeled dataset (if available) to train models. Labeled datasets contain known personality traits for a set of individuals. Implement cross-validation techniques to ensure the models robustness and avoid overfitting. Optimize model parameters to improve performance.
6. Model Evaluation: Evaluate model performance using appropriate metrics such as accuracy, precision, recall, F1-score, or others relevant to the application. Ensure the model is interpretable to understand why it makes certain predictions. This architecture outlines a comprehensive approach to building a system for personality prediction based on the OCEAN model using machine learning and AI. The focus is on accuracy, ethical considerations, and respecting user privacy throughout the process.

3.2 Flowchart

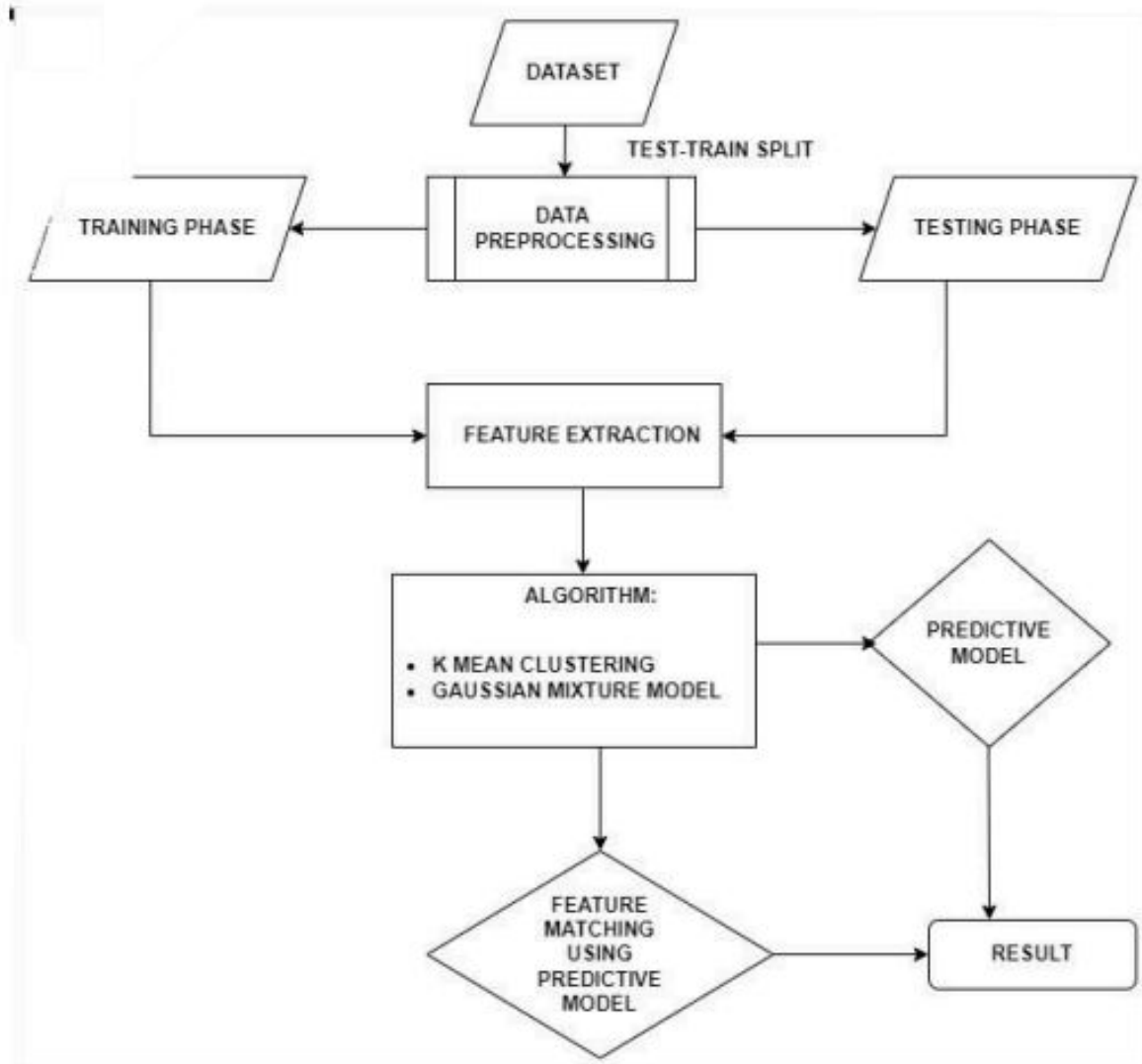
A **flowchart** is a **diagrammatic representation of a process** that visually illustrates the **sequence of steps** involved in solving a problem or executing a task. It uses different symbols to represent various actions, decisions, and processes.

A **flowchart** is a powerful tool to visually represent processes, making it easier to understand and analyze workflows, algorithms, and system operations. It is widely used in software development, business process modeling, and problem-solving.

1. Dataset:-The system starts with a dataset containing personality-related information. This dataset is essential for training and testing the predictive model.

2. Test-Train Split. The dataset is divided into two parts:

- **Training Set** – Used to train the machine learning model.



FLOWCHART :- PERSONALITY PREDICTION

- **Testing Set** – Used to evaluate the model’s performance.

2. Data Preprocessing:- Raw data is cleaned, normalized, and formatted to remove inconsistencies. This step ensures that the data is suitable for feature extraction and model training.

3. Training Phase & Testing Phase

- **Training Phase:** The machine learning model learns from the processed data.
- **Testing Phase:** The trained model is evaluated on unseen data to check accuracy.

4. Feature Extraction:- Important attributes are selected from the dataset to improve prediction accuracy. Extracted features play a crucial role in clustering and classification.

5. Algorithm Selection:-Two clustering techniques are used for personality classification:

- **K-Means Clustering** – A partition-based clustering algorithm that groups similar data points.
- **Gaussian Mixture Model (GMM)** – A probabilistic clustering approach that models data distribution.

6. **Predictive Model:-** The trained model analyzes new inputs based on learned patterns. The predictive model applies the best-suited algorithm for classification.

7. **Feature Matching Using Predictive Model:-** User input is compared with the trained model to find personality traits. The system ensures that the prediction is accurate based on feature similarity.

8. **Result Generation:-** The final personality prediction is displayed as the output. The user receives the predicted personality type based on the trained model's analysis.

DFD Diagram:

Data objects represented by labeled arrows and transformation are represented by circles also called as bubbles. DFD is presented in a hierarchical fashion i.e. the first data flow model represents the system as a whole. Subsequent DFD refine the context diagram (level 0 DFD), providing increasing details with each subsequent level. The DFD enables the software engineer to develop models of the information domain & functional domain at the same time. As the DFD is refined into greater levels of details, the analyst perform an implicit functional decomposition of the system. At the same time, the DFD refinement results in a corresponding refinement of the data as it moves through the process that embody the applications. A context-level DFD for the system the primary external entities produce information for use by the system and consume information generated by the system. The labeled arrow represents data objects or object hierarchy. The context diagram is the most abstract data flow representation of a system. It represents the entire system as a single bubble and. The various external entities with which the system interacts and the data flows occurring between the system and the external entities are also represented. The name context diagram is well justified because it represents the context in which the system is to exist i.e. the external entities (users) that would interact with the system and specific data items they would be receiving from the system.

DLF Diagrams:

A **Data Flow Diagram (DFD)** is a graphical tool used to represent the flow of data within a system. It helps visualize how data is processed, stored, and transmitted between different components of a system. Here are the key elements of a DFD:

- **Processes:** Indicate activities or functions that transform data.
- **Data Flows:** Arrows that represent the movement of data between processes, data stores, and external entities.
- **Data Stores:** Places where data is held for later use.
- **External Entities:** Sources or destinations of data that exist outside the system.

1. **DFD level 0:-** The given diagram represents a Level 0 Data Flow Diagram (DFD) for an application. Level 0 DFD is also known as a Context Diagram, which provides a high-level overview of the system and its interactions with external entities.

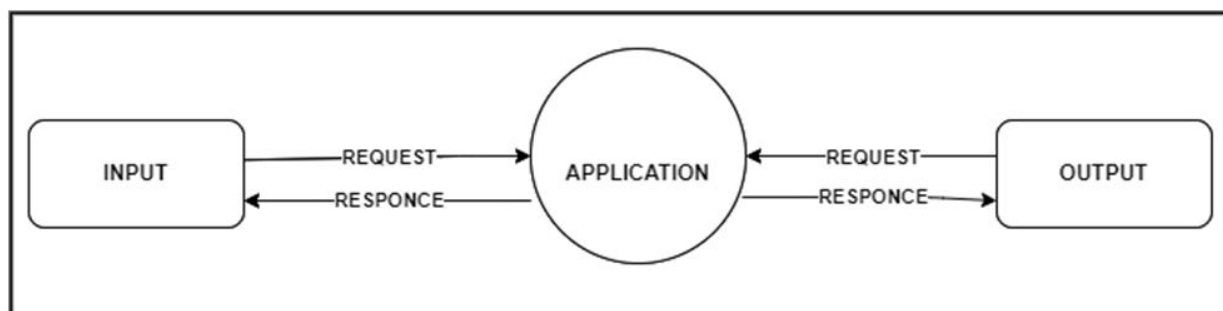


Fig: DFD level 0

Explanation of the Diagram:

Entities (External Users):

- **Admin:** Represents the administrator who interacts with the application, sending requests and receiving responses.
- **User:** Represents an end-user who interacts with the application by making requests and receiving responses.

Process:

- **Application (System):** The central process (represented by a circle) that handles requests from both Admin and Users. It processes these requests and provides appropriate responses.

Data Flow:

- **Request (Input):** Both the Admin and User send requests to the Application for different operations.
- **Response (Output):** The Application processes the request and sends back a response to the Admin and User accordingly.

Data Store or Output Process:

- **Prediction:** The application interacts with a Prediction module, possibly performing some computations or predictions based on the input data. The results may be stored or sent back to the user.

2. **DFD level 1 :-** The given diagram represents a Level 1 Data Flow Diagram (DFD) for an application. Level 1 DFD expands on the Level 0 (Context Diagram) by breaking down the main process into more detailed sub-processes and showing how data flows between them.

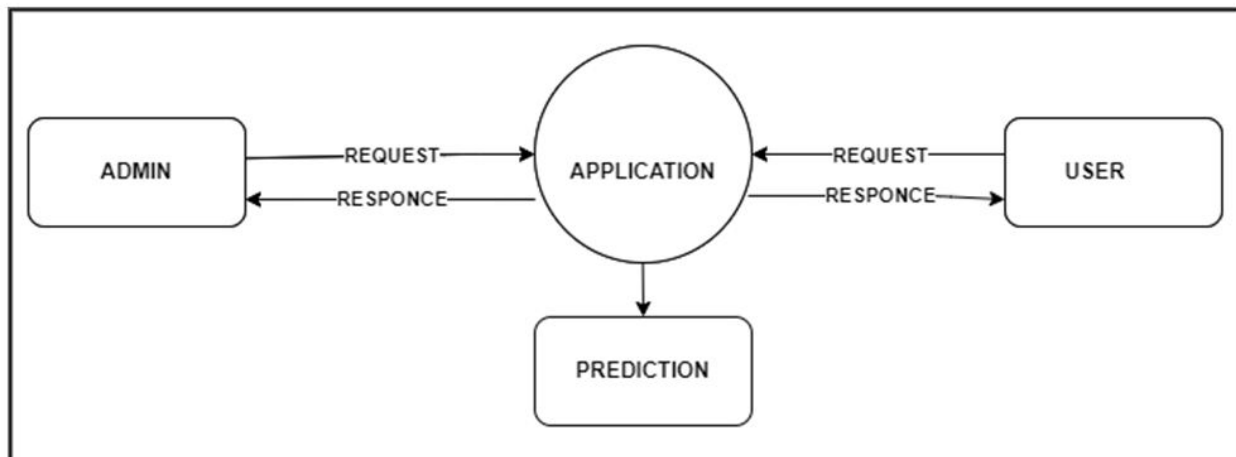


Fig: DFD level 1

Explanation of the Diagram in DFD Level 1 Context:

External Entities :

- **Admin:** The administrator who interacts with the application, sending requests and receiving responses. Admin users typically have higher privileges, such as managing data, users, and configurations.
- **User:** The general end-user who sends requests and receives responses from the application. Users interact with the system for specific functionalities like data retrieval, submission, or analysis.

Main Process:

- **Application:** This is the central process that handles requests from both the Admin and User. It processes their inputs and generates appropriate responses.

Data Flow (Arrows Representing Communication):

- Admin → Application: The admin sends a Request to the application.
- Application → Admin: The application processes the request and sends a Response back to the admin.
- User → Application: The user also sends a Request to the application.
- Application → User: The application processes the user's request and sends a Response back.

Sub-process or Data Store:

- Prediction Module: This module is used for generating predictions based on the data provided. The application sends relevant data to this module, which then processes it and returns the results.
- The output of the Prediction process may either be used within the application or sent back to the Admin/User as part of their response.

3) DFD level 2 : The given image represents a Level 2 Data Flow Diagram (DFD) for a Personality Prediction System. Here's a breakdown of the key components and their interactions:

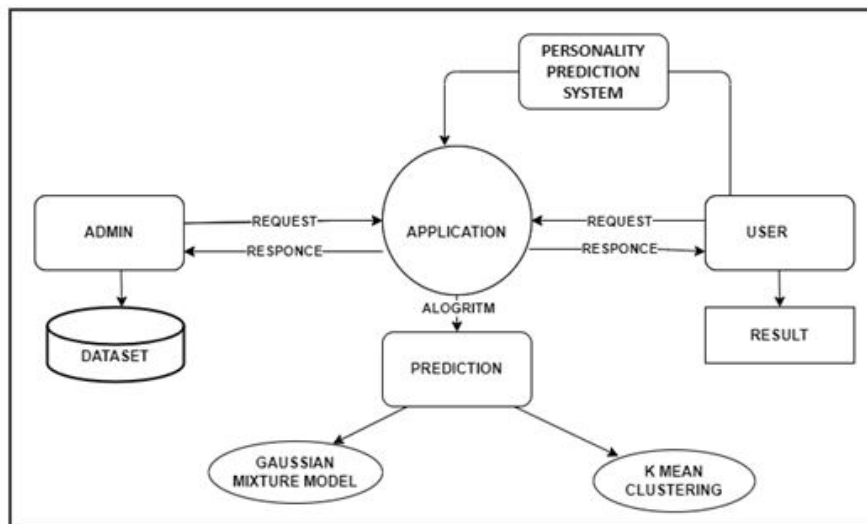


Fig: DFD level2

Explanation of the Diagram in DFD Level 1 Context:

Entities and Processes:

1. Admin

- Interacts with the system by sending a request to the **Application**.
- Manages and updates the **Dataset** used for personality prediction.
- Receives responses from the **Application** regarding dataset-related operations.

2. User

- Sends a request to the **Application** to predict personality traits.
- Receives the prediction results after processing.

3. Application (Main Processing Unit)

- Acts as the core system handling requests from both Admin and User.
- Communicates with the Personality Prediction System.
- Uses predictive algorithms to generate results.

4. Prediction Module

- Implements two algorithms for prediction:
 1. **Gaussian Mixture Model (GMM)**: A probabilistic clustering approach.
 2. **K-Means Clustering**: A machine learning technique to classify users into personality categories.

5. Dataset

- Contains stored personality data used for training and evaluation.
- Admin interacts with it to update or retrieve relevant information.

6. Result

- The final output after processing the user request.
- Delivered to the user based on the prediction from the application.

IV. CONCLUSION

Personality prediction using machine learning and AI based on the OCEAN model holds great potential. By analyzing large datasets, these algorithms can accurately predict personality traits such as openness, conscientiousness, extraversion, agreeableness, and neuroticism. This technology has applications in psychology, human resources, marketing, and mental health, allowing for personalized insights and recommendations. However, it is essential to address ethical concerns, such as biases in data and privacy issues, to ensure fair and responsible use of these predictions. Despite its promise, careful attention is required to avoid inaccuracies and protect individuals' data.

REFERENCES

- [1] Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi (2019). Personality Prediction System through CV Analysis, IRJET vol 06 issue 02, e-ISSN-2395-0056, p-ISSN-2395-0072.
- [2] Devesh Agarwal, Mr. M. Karthikeyan (2022). PERSONALITY PREDICTION USING MACHINE LEARNING, IRJMETS Vol:04 /Issue:04/ April-2022, e-ISSN-2582-5208.
- [3] Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat (2021). Personality Prediction Via CV Analysis using Machine Learning, IJERT Sep 2021 Vol 10 Issue 09, ISSN – 2278-0181.
- [4] M. Kalghatgi, M Ramannavar, and Dr. N. S. Signal, “Neural Network approach to personality prediction based on the Big-Five Model” in IJIRAE, vol2 issue 8, August 2015, pp 56-63.
- [5] J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem “Resume Ranking using NLP and Machine Learning”, unpublished.
- [6] Md Tanzim Reza, and Md. Sakib Zaman, “Analyzing CV/Resume using natural language processing and machine learning”, unpublished.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details