***

Technical Project Report: Personality Prediction System

Author: Syed Tayyab Raza Tabassam
Date: November 14, 2025

***

1. Abstract

This report documents the development of a **Personality Prediction System** using machine learning techniques. The project's primary objective was to accurately classify individuals into distinct personality categories based on a given dataset of behavioral and demographic features. The methodology involved extensive data preprocessing, feature engineering, and the application of an advanced ensemble learning algorithm. The final model achieved a high performance metric, demonstrating its effectiveness as a reliable prediction tool for personality profiling.

***

2. Introduction

Understanding and predicting human personality traits is a critical area in behavioral science and artificial intelligence. Such a system has applications ranging from personalized user experiences to optimizing team composition in professional environments. The goal of this project is to implement a robust machine learning model capable of predicting an individual's personality type from a set of provided features. The foundational dataset, which captures various aspects of the subjects, serves as the basis for this classification task. The initial project brief highlighted the need to classify individuals into groups, which is addressed through a supervised learning approach.

***

3. Methodology

The project followed a standard **supervised machine learning pipeline**. The initial phase involved **Data Preprocessing**, where the raw data was cleaned, missing values were imputed, and categorical features were encoded (e.g., using one-hot encoding) to make them

suitable for model consumption. Key features were selected and scaled using **Z-score normalization (StandardScaler)** to ensure features contribute equally to the model training process. The dataset was then partitioned into training and testing sets ($\text{X\_train, X\_test, y\_train, y\_test}$) using the `train_test_split` function from `sklearn`.

The core of the methodology centered on **Ensemble Learning**, specifically utilizing the `XGBoostClassifier`. Extreme Gradient Boosting (XGBoost) was chosen for its proven performance in classification tasks due to its ability to handle complex non-linear relationships and its inherent efficiency. The model was trained on the processed training data and subsequently validated on the unseen test set to measure its generalization capability.

***

4. Packages Used

This project utilized the following primary Python libraries for data manipulation, machine learning implementation, and visualization:

matplotlib, **numpy**, **pandas**, **sklearn**, **xgboost**

***

5. Results and Discussion

The trained model was evaluated on the test set, yielding promising results in the personality classification task. The primary performance indicator, **Accuracy**, was utilized alongside other metrics like Precision, Recall, and the F1-Score, derived from the Classification Report.

The XGBoost model achieved a final test **Accuracy of 0.7719**, indicating strong predictive power. This accuracy suggests that the model is correctly classifying approximately **77.19%** of the test subjects into their respective personality groups.

Specific classification report and final metrics were observed, showing high scores across the board for key classes (e.g., 'Introvert' and 'Extrovert'), confirming the model's reliability and balanced performance.

***

Conclusion

The Personality Prediction System successfully leverages ensemble machine learning to accurately classify personality types. The robust methodology, centered on feature

engineering and XGBoost, yielded a model with strong generalization capabilities, demonstrating the value of predictive analytics in behavioral science. Future work can focus on model interpretability and deployment.