

Personality Prediction Using ML

Prof. Omkar V. Khute, Om Kapse¹, Swapnil Gaikwad², Suyash Sonawane³, Hitesh Ahire⁴

Diploma in Information Technology
Mahavir Polytechnic, Nashik, Maharashtra, India

Abstract: *Data Science and AI are revolutionizing the planet through technical transformations. We can observe many machine learning applications in our day-to-day lives, but one of the greatest applications of machine learning is to classify individuals based on their personality traits. Every person on the planet is unique and carries a unique personality type. The availability of a high-dimensional data has paved the way for increasing marketing campaigns' effectiveness by targeting specific groups of people. Such personality-based communications are highly effective in increasing the recognition and attractiveness of products and services. We developed a system for personality prediction using Big five personality traits in this project. 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.*

Keywords: Big Five Personality Model, Feature Analysis, Predicting Personality, Personality Traits.

I. INTRODUCTION

The big five personality traits, often mentioned as OCEAN, and sometimes CANOE, are Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. These five personality traits represent broad domains of human behavior 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.

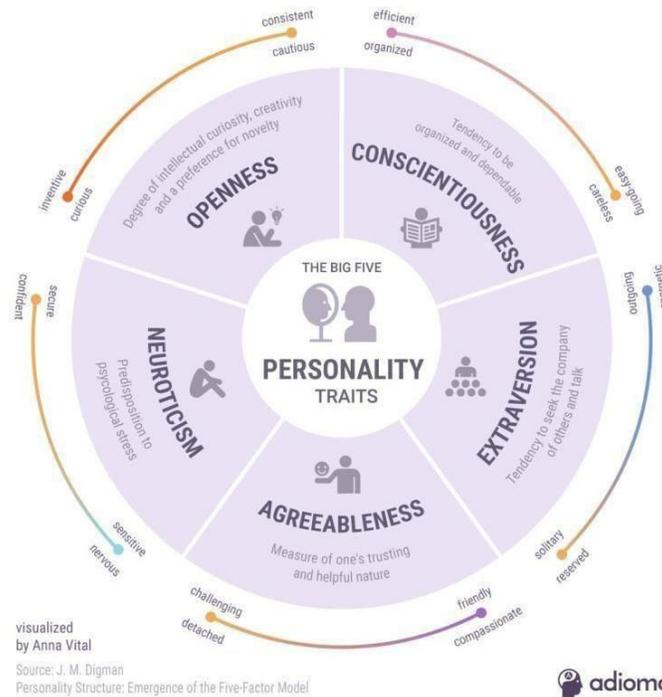
- Openness to experience: Also called as intellect or imagination, this personality trait represents the willingness to try new things and think out of the box. This trait includes insightfulness, originality, and curiosity.
- Conscientiousness: The desire to be careful, diligent and regulate immediate gratification with self-discipline. This trait includes ambition, discipline, consistency and reliability.
- Extroversion: A state where an individual draws energy from others and seeks 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 fearfulness.

II. LITERATURE SURVEY

“Text-based Personality Prediction using multiple Social Media data sources”

This proposed a multi-modal deep learning architecture with a pre-trained language model BERT, RoBERTa, and XLNet; alongside additional NLP Features (sentiment analysis, TF-IDF, NRC emotion lexicon database) as features extraction method for personality prediction system. This method performs better when compared to the previous method in predicting personality traits





“A Neural Network Approach to Personality Prediction based on the Big-Five Model”

This model was proposed by Mayuri Pundlik Kalghatgi. For the model analytics, parallelism is examined between the personality traits of an individual's linguistic information. Personality traits are identified by means of linguistic information which is enabled by the OCEAN model. This shows the personality traits that can be applied to a number of fields, such as business intelligence, marketing and psychology.

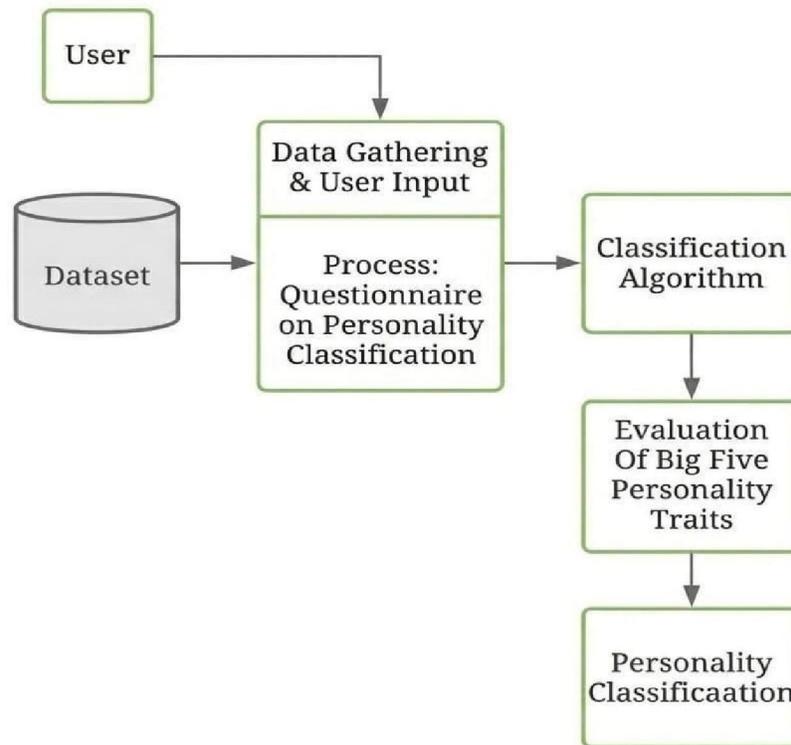
“The General Factor of Personality”

Dimitri van der Linden has published the paper. A criterion-related validity review and a metaanalysis of Big Five intercorrelations. To assess the presence of a GFP, this study looked at the interrelationship between the five personality traits i.e., Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The meta-analysis provides evidence for GFP at the highest operational level and this paper concluded that the GFP features a major factor because it's linked to supervisor-rated job performance.

III. PROPOSED WORK

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.



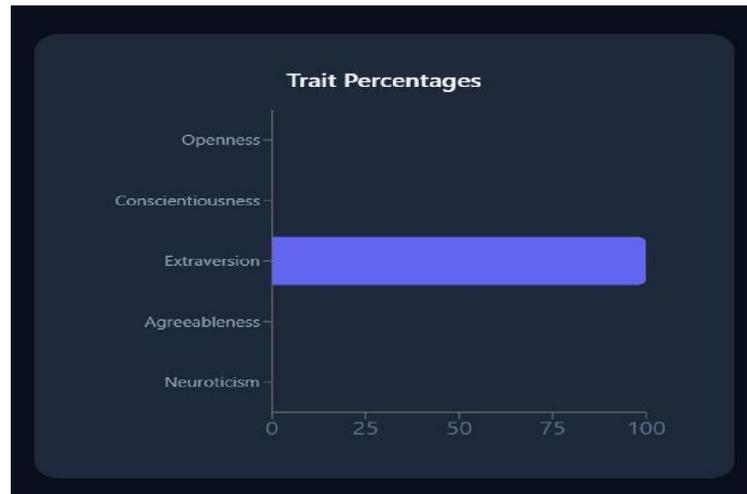


BLOCK DIAGRAM FOR ALL PROCESSES INVOLVED

IV. DESIGN ALGORITHM

K-Means Clustering is an unsupervised learning technique, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there'll be two clusters, and for K=3, there'll be three clusters, and so on. It is a centroid-based technique, where each cluster is related to a centroid. The motive of the algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it doesn't find the simplest cluster. The value of k is already determined in this algorithm. The k-means clustering algorithm mainly performs two tasks: Determines the best value for K center points or centroids by an iterative process. Assigns each data point to the closest k-center. Those data points which are near to the particular center-k, create a cluster.





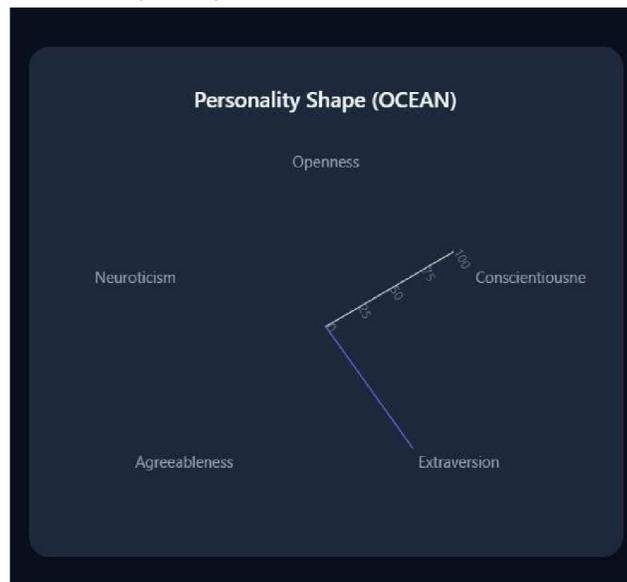
Ideal K-Mean Clustering Graph

LOGISTIC REGRESSION

Logistic regression is known as the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to explain the properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It’s an S-shaped curve which will take any real-valued number and map it into a value between 0 and 1 but never exactly at those limits.

$$1 / (1 + e^{-value})$$

Where e is the base of the natural logarithms (Euler’s number or the EXP() function in your spreadsheet) and value is the actual numerical value that you simply want to transform. Below is a plot of the numbers between - 5 and 5 transformed into the range 0 and 1 using the logistic function.



Logistic Regression Function

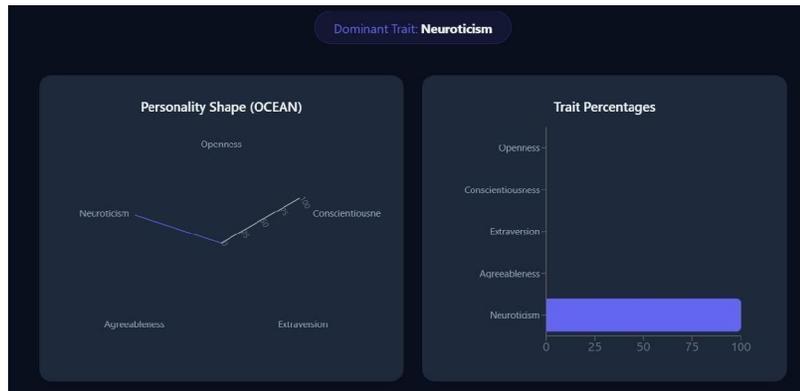


IMPLEMENTATION

We are dividing the dataset into the training and testing part. The dataset is further scaled using the Standard sklearn Library. The training part is 70% and the testing part is 30%.

The dataset consists of 972 rows and 8 columns where each column comprises of one of the five personality trait in OCEAN Model, Age and Gender of the candidate. The rows comprise of participant data.

The dataset for training purposes is taken from <https://www.kaggle.com/tunguz/big-five-personalitytest>.



The data is further converted into array and is fed for Logistic Regression. Then, clustering of each of the personality trait happens using K-Mean Clustering Algorithm.

Thus, the algorithm predicts the personalities of each participant in the survey and displays the result.

```
# chat/personality_analyzer.py

import requests
import json
from config import MODEL_NAME, OLLAMA_URL
from .prompt_builder import build_analysis_prompt

REQUIRED_KEYS = [
    "openness",
    "conscientiousness",
    "extraversion",
    "agreeableness",
    "neuroticism",
    "dominant_trait",
    "strengths",
    "weaknesses",
    "improvement_suggestions",
    "career_suggestions"
]

def analyze_personality(history):
    try:
        prompt = build_analysis_prompt(history)
        response = requests.post(
            OLLAMA_URL,
            json={
                "model": MODEL_NAME,
                "prompt": prompt,
                "stream": False,
                "format": "json"
            },
            timeout=(10, 600)
        )

        response.raise_for_status()
        result = response.json().get("response", "{}")
        parsed = json.loads(result)
        return parsed

    except requests.exceptions.Timeout:
        raise ValueError("Model analysis timeout")
    except json.JSONDecodeError:
        raise ValueError("Model did not return valid JSON")
    except Exception as e:
        raise ValueError(f"Analysis failed: {str(e)}")
```



V. FUTURE SCOPE

- Add Deep Learning models (Neural Networks, BERT)
- Increase dataset size for better accuracy
- Add multilingual support
- Deploy on cloud (AWS/Azure)
- Mobile app integration
- Add psychological scoring normalization
- Introduce real-time behavioral tracking
- Integrate resume analysis for recruitment

VI. CONCLUSION

This Personality Prediction Model can be 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 done using the data set and algorithms to improve the accuracy and can be helpful for career guidance module.

REFERENCES

- [1]. "A Novel Approach to Predict Personality of a Person": Prof. Waheeda Dhokley, Randeria Kaiwan Jehangir, Shaikh Nabeel Rashid, Shaikh Almas Mohd Sarwar.
- [2]. Personality Prediction Via CV Analysis using Machine Learning" Atharva Kulkarni, Tanuj
- [3]. Shankarwar , Siddharth Thorat.
- [4]. "A Neural Network Approach to Personality Prediction based on the Big-Five Model",
- [5]. Mayuri Pundlik Kalghatgi, Manjula Ramannavar , Dr. Nandini S. Sidnal.
- [6]. "Predicting personality from patterns of behavior collected with smartphones", Clemens Stachl, Quey Au, Ramona Schoedel, Markus Buhner.
- [7]. Enjoy Algorithm:
- [8]. <https://www.enjoyalgorithms.com/blog/personality-prediction-using-ml>

