PERSONALITY PREDICTION USING MACHINE LEARNING

Kanchan Sangle[1], Nitin Divate[2], Punit Sonawane[3], Mr. A. P. Patil[4]

[1,2,3]Student, Department of Information Technology, MVPS' Rajarshi Shahu Maharaj Polytechnic, Nashik, Maharashtra, India.

[4]Lecturer, Department of Information Technology, MVPS' Rajarshi Shahu Maharaj Polytechnic, Nashik, Maharashtra, India

ABSTRACT

Personality prediction is a crucial aspect of understanding human behaviour, decision-making, and social interactions. The OCEAN model, comprising Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, is a widely accepted framework for personality assessment.

The concept of using ocean models to predict personality traits based on machine learning (ML) is an emerging field in the intersection of psychology, computer science, and physics. The Ocean Model theory suggests that personality traits can be viewed as a set of interconnected waves or patterns that ebb and flow over time, much like the oceans tides. By analysing these patterns and fluctuations using ML algorithms, it may be possible to predict an individuals personality traits with a high degree of accuracy. This study proposes a machine learning approach to predict personality traits based on the OCEAN model. We collected a dataset of participants' responses to a personality questionnaire and applied various machine learning algorithms, including K- Means Clustering, Gaussian Mixture Model to predict their personality scores. Our results show that the proposed approach achieves high accuracy in predicting personality traits, with showing the highest prediction accuracy. The study demonstrates the potential of machine learning in personality prediction and provides insights into the relationships between personality traits and behavioural patterns.

The findings have implications for applications in human resources, marketing, and mental health. This abstract presents an overview of the Ocean Model theory, its potential applications in personality assessment, and the challenges and limitations associated with this approach.

Keywords: Personality prediction ,Machine Learning, K- Means Clustering, Gaussian Mixture Model, Ocean Model.

1. INTRODUCTION

Personality prediction based on the OCEAN model using machine learning and artificial intelligence is an innovative approach that combines psychological theory with advanced technology.

The OCEAN model, also known as the Big Five personality traits, represents five fundamental dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Machine learning and artificial intelligence techniques are widely used to analyze large amounts of data and detect patterns. With these techniques, personality prediction models can be developed to accurately assess an individual's personality traits based on their behavior, preferences, and other relevant data. By using machine learning algorithms, such as decision trees this models can learn from training data that includes various predictors, such as social media activity, textual data, and surveys. The trained model can then make predictions about an individual's personality traits based on their input data, even without explicitly stating their personality type.

This approach has several potential applications. For example, in psychology, it can assist therapists in tailoring personalized treatment plans for patients. In the field of human resources, it can aid in candidate selection or team building by identifying candidates or team members with specific personality traits that align with the job requirements or the existing team dynamics.

OCEAN Model:

a) Open to Experience: It involves various dimensions, like imagination, sensitivity, attentiveness, preference to variety, and curiosity.

b) Conscientiousness: This trait is used to describe the carefulness and diligence of the person. It is the quality that describes how organized and efficient a person is.

c) Extraversion: It is the trait that describes how the best candidates can interact with people that is how good are his/her social skills.

d) Agreeableness: It is a quality that analyses the individual behavior based on the generosity, sympathy, cooperativeness and ability to adjust with people.

e) Neuroticism: This trait usually describes a person to have mood swings and has extreme expressive power.

2. 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. 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

2.1 Data Preprocessing

Once the data is collected, preprocessing is performed to clean and prepare the dataset for analysis. This includes removing noise, irrelevant information, and redundant data. Text data is analyzed to extract meaningful features, and standardization of data formats and normalization of scales are applied. Missing values are handled, ensuring the data is ready for feature selection and model training.

2.2 Ethical Considerations

Throughout the process, ethical considerations are given priority, especially with respect to user privacy. The model is designed to ensure transparency and interpretability, providing insight into why certain predictions are made. Additionally, data privacy regulations are strictly followed to ensure the ethical handling of user data throughout the research.

3. MODULES

1. Data Collection and Preparation:

Dataset (Kaggle): The process begins with data acquisition from Kaggle, a popular platform for datasets. This dataset likely contains text data (e.g., social media posts, survey responses) associated with individuals.

Database: The acquired data is stored in a database for organization and management.

2. Data Preprocessing:
This crucial step involves cleaning and transforming the raw data to make it suitable for machine learning. This includes tasks like: Text Cleaning: Removing noise, special characters, HTML tags, etc. Tokenization: Breaking down text into individual words or phrases. Stop Word Removal: Eliminating common words (e.g., "the," "a," "is") that don't carry much meaning

3. Train/Test Split:
The preprocessed data is divided into two sets: Training Data (80%): Used to train the machine learning model. Testing Data (20%): Used to evaluate the model's performance on unseen data.

4. Feature Extraction and Model Training:
Feature Extraction: This module converts the text data into numerical features that the machine learning model can understand. Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings (Word2Vec, Glo Ve) might be used.

Training Phase: The extracted features from the training data are fed into a machine learning algorithm. The diagram mentions two specific algorithms:

K-Means Clustering: A partitioning algorithm that groups data points into clusters based on similarity. While typically used for unsupervised learning, it might be employed here for feature engineering or pre-clustering before

applying a predictive model.

Gaussian Mixture Model (GMM): A probabilistic model that assumes data points are generated from a mixture of Gaussian distributions.1 It can be used for clustering or density estimation and potentially for predicting personality traits based on the probability distributions.

Algorithm Selection: The choice between K-Means, GMM, or potentially other algorithms (as implied by the broader "ALGORITHM" box) depends on the specific approach and the nature of the dataset.

5. Personality Prediction and Evaluation:

Testing Phase: The trained model is applied to the held-out test data to predict personality traits. Feature Matching Using

Predictive Model: This likely refers to comparing the extracted features of the test data with the patterns learned by the model during training. OCEAN Model/Big Five: The diagram highlights the focus on the OCEAN model, which represents the five major personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The model predicts these traits based on the input text data.
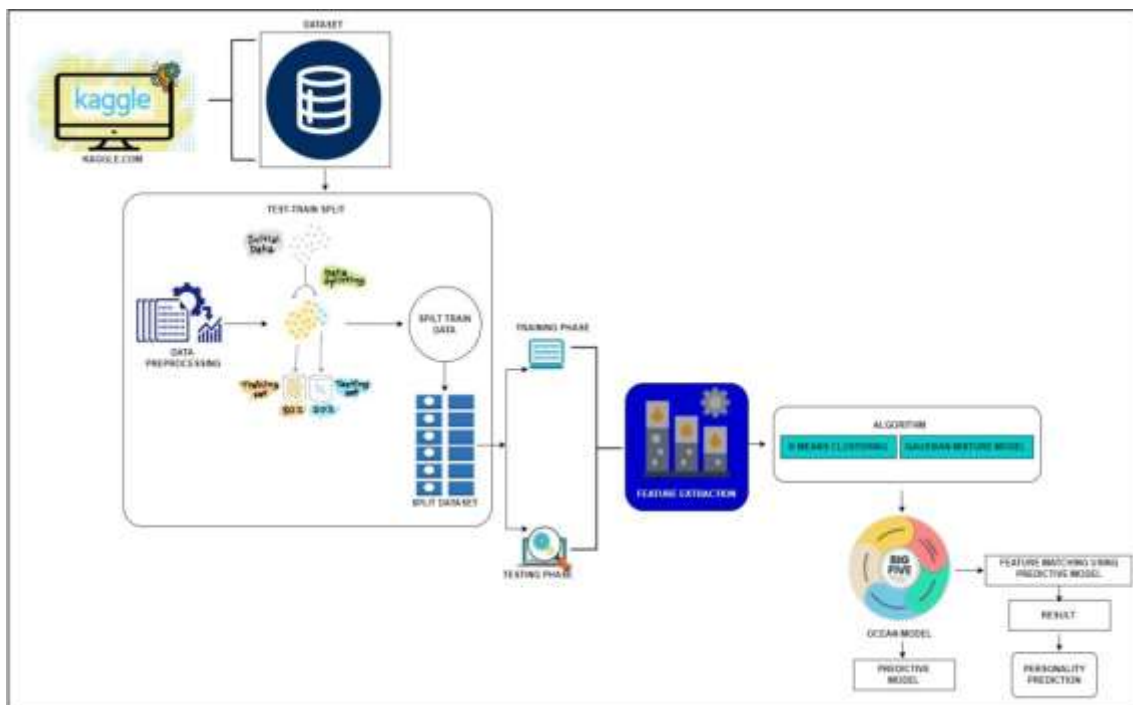
4. SYSTEM ARCHITECTURE

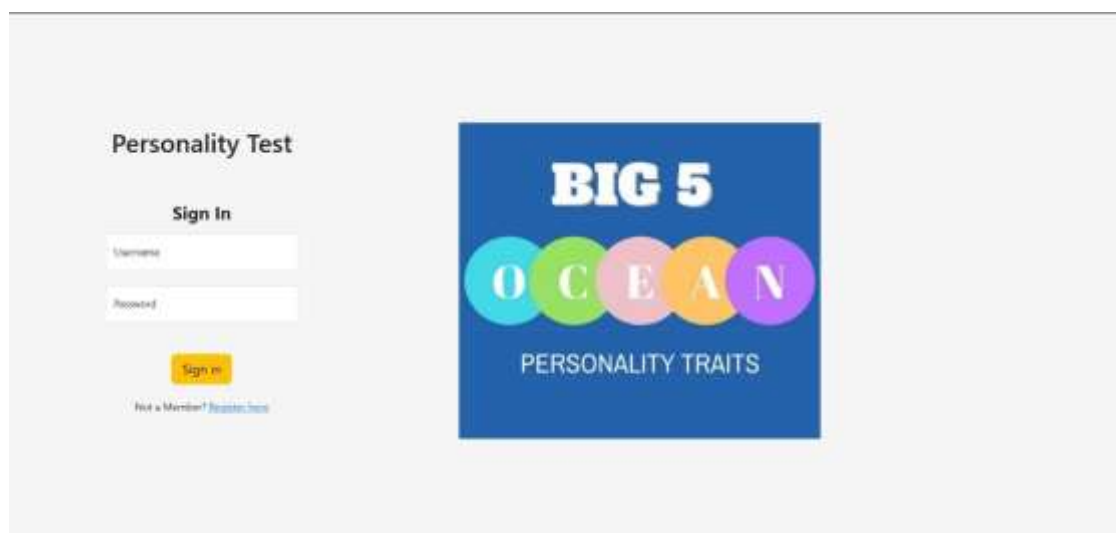**Figure 1:** System Architecture Diagram

5. RESULTS

Figure 2: Personality Prediction Sign In Page
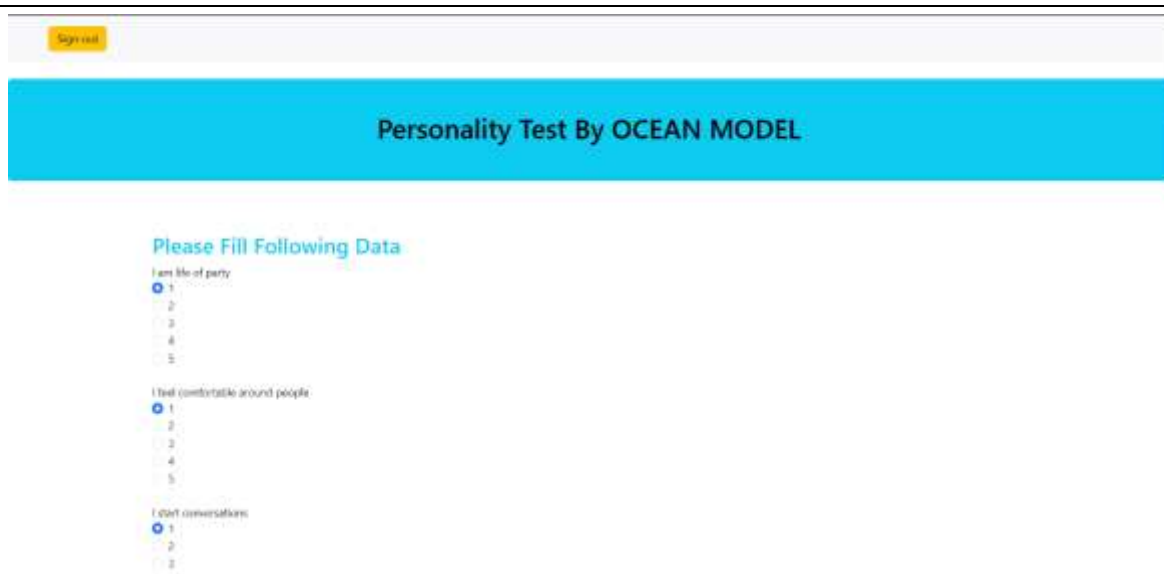
Figure 3: Personality Prediction Survey Page

Figure 3: Personality Prediction Result History
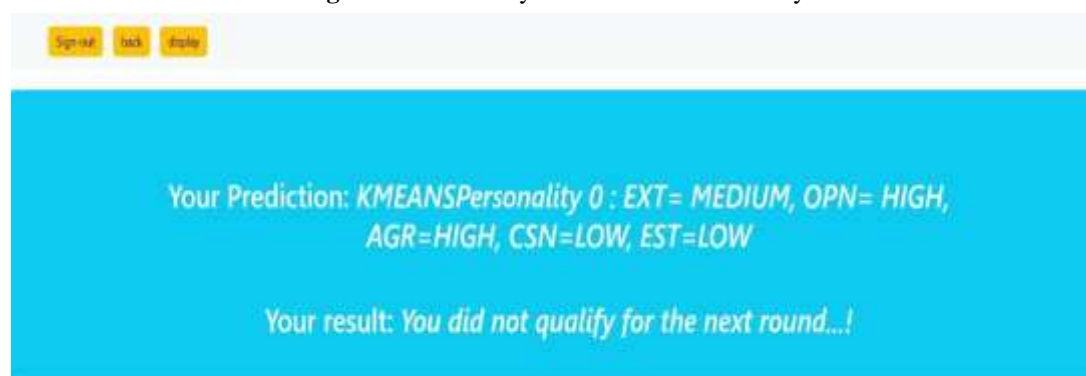

Figure 4: Personality Prediction Result

6. FUTURE SCOPE

Enhancing data collection techniques by incorporating advanced technologies such as wearable devices or sensors to gather real-time data on individuals' behaviors and activities. Integrating cognitive computing systems with machine learning algorithms to better analyze and understand complex patterns in individual personality traits. Developing advanced predictive modeling techniques that can accurately predict an individual's personality traits based on a wider range of variables, including social media activity, online behavior, and biometric data. Using AI algorithms to generate personalized recommendations and insights per individuals based on their predicted personality traits, helping them make informed decisions and improve their overall well-being. Addressing ethical concerns related to the use of AI systems in predicting personality traits, such as ensuring data privacy and security, and ensuring that predictions are not used to discriminate or harm individuals. Implementing systems that can continuously learn and adapt based on new data and feedback, improving the accuracy and reliability of personality predictions over time.

7. CONCLUSION

In conclusion, personality prediction based on the OCEAN model using machine learning and artificial intelligence has great potential. The OCEAN model, which stands for openness, conscientiousness, extraversion, agreeableness, and neuroticism, is a widely used framework for understanding and measuring personality traits. By applying machine learning and artificial intelligence algorithms to analyze large sets of data from individuals, it is possible to predict and understand their personality traits based on patterns and correlations found within the data. These algorithms can learn from past data and make accurate predictions about a person's personality traits, even without explicitly being taught. The use of machine learning and artificial intelligence in personality prediction can have several applications. It can be used in fields such as psychology, human resources, and marketing to gain insights into people's behavior, preferences, and decision-making processes. This information can be leveraged to make personalized recommendations, optimize targeted advertising campaigns, and even support mental health interventions. It is important to note that personality prediction algorithms are not without their limitations and ethical concerns. The accuracy of the predictions heavily relies on the quality and diversity of the data used for training the algorithms. Biases present in the data, such as underrepresentation of certain demographics or cultural groups, can lead to inaccurate and unfair predictions. Additionally, the use of personal data for personality prediction raises privacy concerns and requires careful handling to ensure the protection of individuals' information. Personality prediction based on the OCEAN model using machine learning and artificial intelligence is a promising field with various applications. However, it is crucial to approach it with caution, addressing the limitations and ethical considerations to ensure responsible and unbiased use of this technology.

8. 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

[3] Vol:04 /Issue:04/ April-2022, e-ISSN-2582-5208

[4] 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

[5] 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.

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

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

[8] Sudhir Bagade, Jayashree Rout, Pooja Yede, Personality Evaluation and CV Analysis using Machine Learning Algorithm VVCET - CSE, Personality Prediction System Through cv Analysis

[9] Gagandeep Kaur, Shruti Maheshwari, Personality Prediction through Curriculam Vitae Analysis involving Password Encryption and Prediction Analysis