International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Personality prediction using AI

Tanvi Dashrath Bhagat * and Sandeep R. Jadhav

Department of Computer, Sanghavi College of Engineering, Maharashtra, India.

International Journal of Science and Research Archive, 2026, 18(02), 108-115

Publication history: Received on 20 December 2025; revised on 01 February 2026; accepted on 03 February 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.2.0207

Abstract

The analysis of personality traits using Machine Learning (ML) and Artificial Intelligence (AI) constitutes an emerging interdisciplinary field that brings together principles from psychology, computer science, and physics. The Ocean Model conceptualizes personality traits as dynamic and interrelated patterns, comparable to ocean waves that continuously form and dissipate over time. By leveraging ML and AI techniques to analyze these evolving patterns and fluctuations, it becomes feasible to infer and predict individual personality characteristics with a high degree of accuracy. This abstract presents an overview of the Ocean Model framework, examines its applicability in personality assessment, and discusses the challenges and limitations inherent in this approach.

Keywords: Machine Learning; Artificial Intelligence; K-Means Clustering; Yellow brick Clustering

1. Introduction

Personality traits play a crucial role in shaping human behavior, decision-making, and social interactions. Understanding and assessing these traits has long been a central focus in psychology, with applications ranging from mental health evaluation to career guidance and behavioral prediction. Traditional personality assessment methods primarily rely on self-report questionnaires and observational studies, which, although valuable, are often limited by subjectivity, response bias, and scalability issues. With the rapid growth of computational technologies, there is an increasing interest in automating and enhancing personality analysis through data-driven approaches. Machine Learning (ML) and Artificial Intelligence (AI) have emerged as powerful tools for modeling complex human behaviors by identifying hidden patterns within large and diverse datasets. These technologies enable the analysis of behavioral, linguistic, and physiological data at a scale and depth that was previously unattainable. By learning from historical data, ML algorithms can uncover correlations and trends that contribute to more accurate and adaptive personality prediction models. As a result, AI-driven personality analysis is gaining traction in domains such as human–computer interaction, personalized recommendation systems, mental health monitoring, and organizational psychology. The Ocean Model, also known as the Big Five personality framework, is one of the most widely accepted theories for describing human personality. It categorizes personality into five core dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Unlike static trait-based interpretations, modern perspectives view these dimensions as dynamic and interdependent, evolving over time based on internal states and external influences. This dynamic nature makes the Ocean Model particularly suitable for computational modeling using ML and AI techniques. Drawing inspiration from concepts in physics, personality traits within the Ocean Model can be analogized to ocean waves continuously rising, falling, and interacting with one another. These fluctuations reflect changes in emotional states, behavior patterns, and environmental interactions. By treating personality traits as time-varying signals rather than fixed attributes, ML and AI models can better capture the underlying complexity of human personality. Such an approach allows for continuous assessment and more nuanced personality predictions. Clustering techniques, such as K-Means clustering, play a significant role in identifying latent personality groupings within datasets. By grouping individuals based on similarities in behavioral or psychological features, clustering methods help

* Corresponding author: Tanvi Dashrath Bhagat

reveal hidden personality structures without predefined labels. Visualization and evaluation tools like Yellowbrick further enhance interpretability by providing insights into cluster quality, separation, and model performance. These tools are essential for validating the effectiveness of unsupervised learning approaches in personality analysis. Despite the promising potential of ML- and AI-based personality assessment, several challenges remain. Issues related to data quality, ethical concerns, privacy protection, model bias, and interpretability must be carefully addressed. Human personality is influenced by cultural, social, and situational factors, which may not always be fully captured in computational models. Additionally, over-reliance on automated systems raises concerns about fairness and accountability in sensitive applications such as psychological profiling. This study aims to explore the integration of the Ocean Model with Machine Learning and Artificial Intelligence techniques for personality trait analysis. By leveraging clustering algorithms and visualization tools, the research seeks to demonstrate how dynamic personality patterns can be identified and interpreted. The findings are expected to contribute to the growing body of interdisciplinary research that bridges psychology and artificial intelligence, while highlighting both the opportunities and limitations of this emerging approach.

1.1. Problem Statement

Traditional personality assessment methods rely heavily on self-reported questionnaires and subjective evaluations, which often suffer from response bias, limited scalability, and an inability to capture the dynamic nature of human personality. With the growing availability of large-scale behavioral data, there is a need for more objective, data-driven approaches that can accurately model and predict personality traits over time. However, integrating psychological theories such as the Ocean Model with Machine Learning and Artificial Intelligence techniques presents challenges related to feature extraction, model interpretability, data quality, and ethical considerations. This research addresses the problem of developing an effective ML- and AI-based framework that can analyze dynamic personality patterns using clustering techniques, while ensuring meaningful interpretation and reliability in personality trait prediction.

2. Literature review

Santos and Paraboni (2022):Santos and Paraboni investigate Myers–Briggs personality type classification using social media text and pre-trained language models. Their study demonstrates that transformer-based models significantly outperform traditional machine learning approaches by effectively capturing semantic and contextual information from textual data. The results highlight the potential of large-scale language representations in personality inference while also acknowledging challenges related to noisy data, privacy concerns, and linguistic variability in social media content.

Bhatt et al. (2022):Bhatt and colleagues propose a personality prediction system based on structured questionnaire responses. The study evaluates the effectiveness of traditional machine learning algorithms in mapping questionnaire data to personality traits. While the approach achieves reasonable accuracy, the authors identify limitations related to respondent bias and static data collection, emphasizing the need for more adaptive and behavior-based personality assessment techniques.

Dadebayev et al. (2022):Dadebayev, Goh, and Tan present a comprehensive review of EEG-based emotion recognition systems, focusing on commercially available EEG devices and machine learning techniques. The study analyzes signal acquisition methods, feature extraction strategies, and classification algorithms used in emotion and personality-related research. The authors highlight the challenges of noise, device variability, and user comfort, while emphasizing the growing potential of EEG-driven affective computing applications.

Al-Hammadi and Moore (2021):Al-Hammadi and Moore explore the recognition of Big Five personality traits from non-verbal cues using sampling techniques and machine learning algorithms. Their work demonstrates that effective data balancing and feature selection significantly improve classification performance. The study emphasizes the importance of non-verbal behavioral signals such as gestures and facial expressions in personality recognition, contributing to advancements in multimodal personality analysis.

Zaferani et al. (2021):Zaferani, Teshnehlab, and Vali propose an automatic personality perception framework using autoencoders and hierarchical fuzzy classification. The model leverages unsupervised feature learning to extract latent personality representations, followed by fuzzy logic-based decision-making. Experimental results indicate improved classification accuracy and interpretability, highlighting the benefits of combining deep learning with soft computing techniques for personality modeling.

Annisa et al. (2020):Annisa, Supriyanto, and Taheri investigate personality dimension classification using EEG signal analysis and support vector machines. The study demonstrates that EEG-based features can effectively capture

cognitive and emotional patterns associated with personality traits. Results show that SVM classifiers achieve promising performance, reinforcing the viability of physiological signal-based personality assessment methods.

Katiyar et al. (2020):Katiyar, Walia, and Kumar present a personality classification system based on data mining techniques. The research applies supervised learning algorithms to structured datasets to categorize individuals into personality classes. The study highlights the efficiency and simplicity of data mining approaches while noting limitations in capturing dynamic and context-dependent personality variations.

Patel et al. (2020):Patel and colleagues analyze personality traits using social media data through machine learning techniques. Their approach extracts linguistic and behavioral features from online user activity to infer personality characteristics. The study demonstrates the feasibility of using digital footprints for personality analysis, while also discussing ethical concerns related to user consent, data privacy, and interpretability.

Sahono et al. (2020):Sahono et al. focus on classifying extrovert and introvert personality types using the Myers–Briggs Type Indicator and support vector machines. The study shows that SVM-based classifiers achieve high accuracy when trained on structured MBTI datasets. The research validates the effectiveness of supervised learning for binary personality classification tasks.

Khan et al. (2020):Khan and co-authors propose a machine learning-based framework for personality classification from online text. The study employs textual feature extraction techniques combined with supervised classifiers to predict personality traits. Results indicate improved performance over baseline models, emphasizing the role of linguistic patterns in understanding individual personality from digital text data.

Mushtaq et al. (2020):Mushtaq, Ashraf, and Sabahat present a hybrid approach for predicting MBTI personality types using K-means clustering and gradient boosting. The study demonstrates that combining unsupervised clustering with supervised learning enhances classification accuracy. The research highlights the importance of uncovering latent personality structures prior to final prediction.

Talasbek et al. (2020):Talasbek and colleagues explore personality classification using K-means clustering techniques. The study focuses on grouping individuals based on behavioral and psychological features without labeled data. Experimental results show that clustering can reveal meaningful personality groupings, although challenges remain in determining optimal cluster numbers and ensuring interpretability.

3. Proposed system

The OCEAN model, commonly referred to as the Big Five personality framework, is a widely recognized approach for describing human personality through five core dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. With the advancement of Machine Learning (ML) and Artificial Intelligence (AI), these technologies can be effectively employed to predict personality traits using data sources such as social media activity, textual content, and behavioral patterns. The proposed system involves systematic data collection from reliable repositories such as Kaggle while ensuring compliance with data privacy regulations and user consent requirements. Data preprocessing is performed to remove noise, handle missing values, normalize data, and extract meaningful features. Feature selection and engineering techniques are applied to identify attributes that strongly correlate with OCEAN traits and to enhance predictive capability. Suitable ML algorithms, including K-means clustering and Gaussian Mixture Models, are employed, along with ensemble learning methods to improve robustness and accuracy. The models are trained and validated using labeled datasets where available, incorporating cross-validation and parameter optimization to prevent overfitting. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score, with emphasis on interpretability, ethical responsibility, and user privacy, resulting in a comprehensive and reliable personality prediction framework.

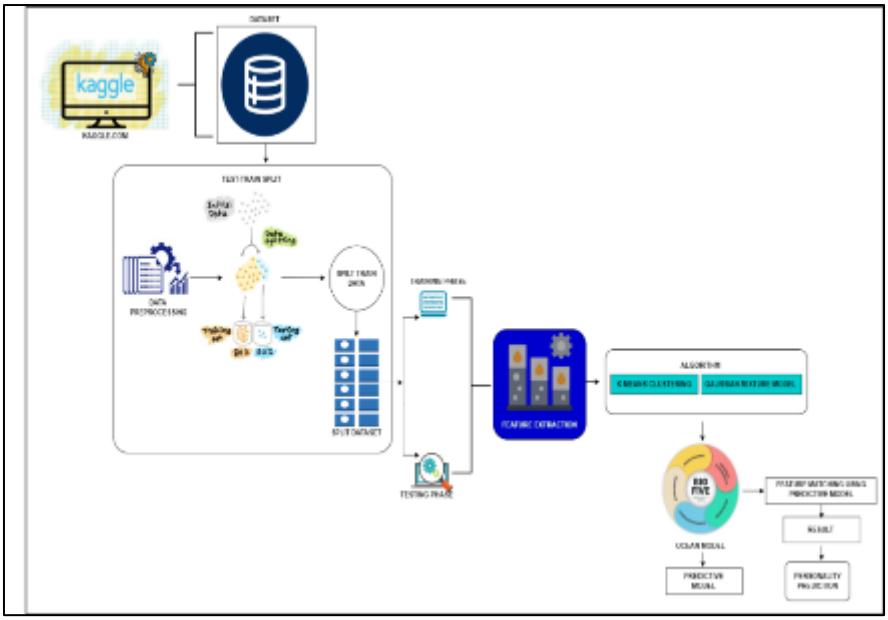

Figure 1 System Architecture

4. Results and discussion

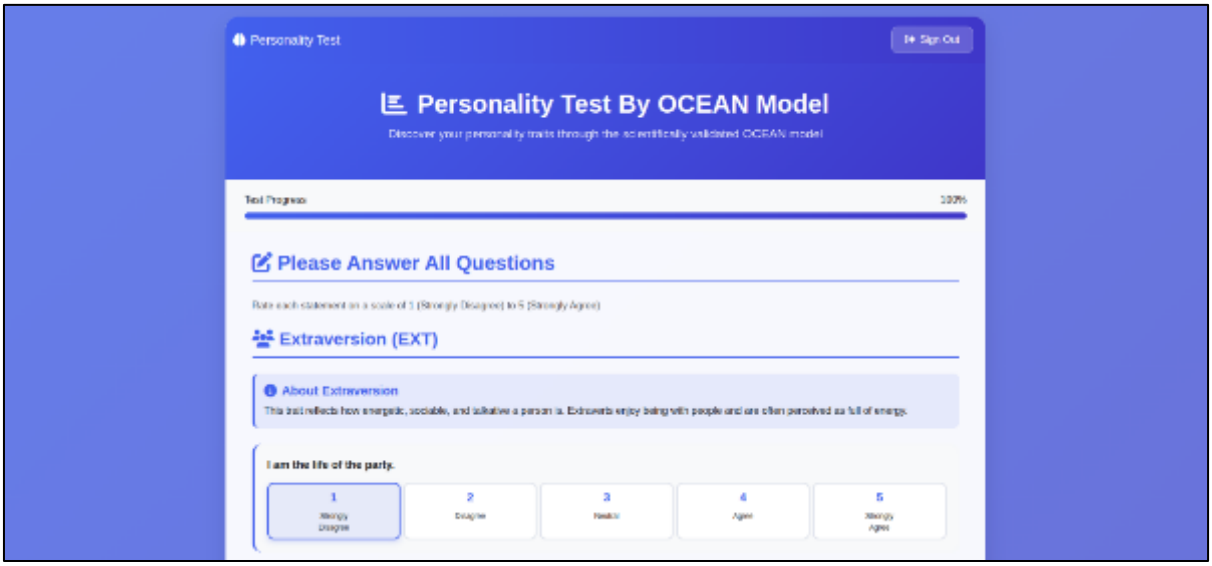

Figure 1 Personality Test Interface Based on the OCEAN Model Assessing Extraversion Traits

The image illustrates a Personality Test based on the OCEAN Model, a scientifically validated framework used to assess an individual's personality traits, focusing here on Extraversion (EXT). The interface prompts users to respond to statements such as "I am the life of the party" using a five-point Likert scale ranging from Strongly Disagree to Strongly Agree. This structured approach helps quantify social, energetic, and outgoing tendencies by converting subjective self-assessments into measurable data. The progress indicator ensures completion of all questions, making the test systematic, user-friendly, and suitable for psychological analysis, behavioral research, and personality-based profiling.

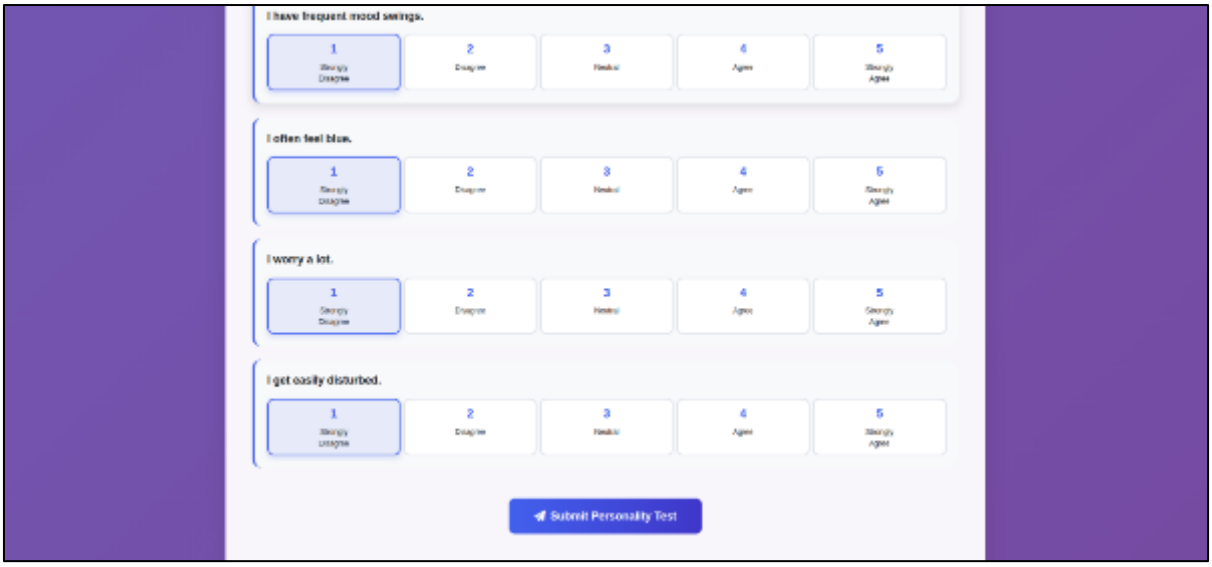

Figure 2 Personality Self-Assessment Questionnaire Interface

This image shows a clean and modern personality test screen where users respond to emotional and behavioral statements such as mood swings, feeling low, worrying, and being easily disturbed using a 5-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree,” allowing individuals to reflect on their mental and emotional patterns in a structured way, with a clear layout, selectable options, and a prominent submit button that encourages self-evaluation and supports psychological awareness through simple, user-friendly design.

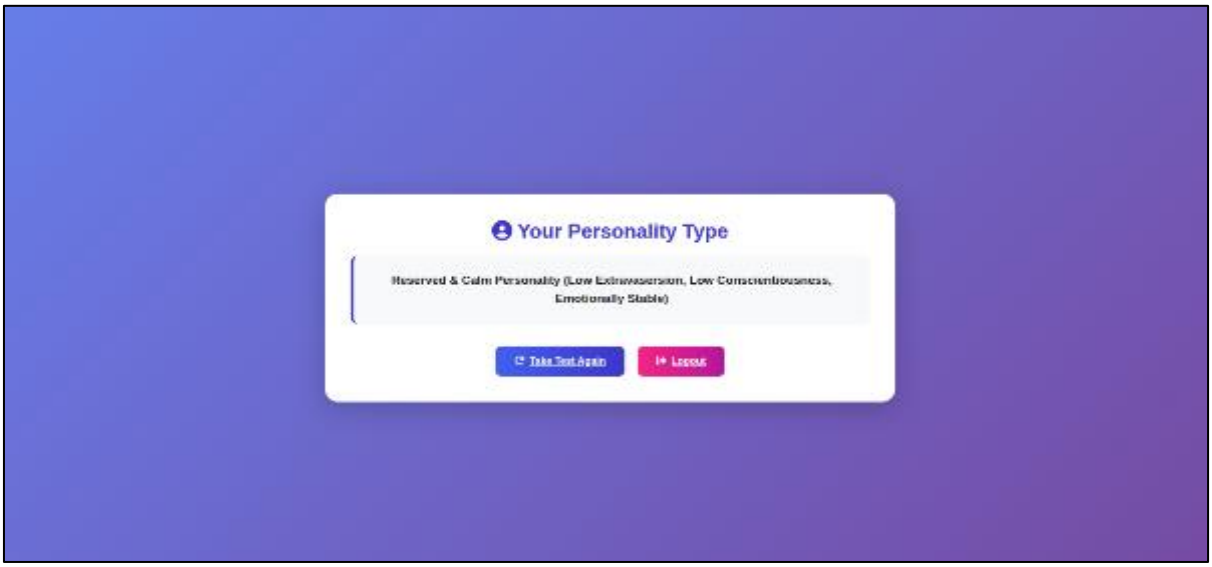

Figure 3 Personality Test Result Summary Screen

This image displays the final result page of a personality assessment, presenting the user’s identified personality type as “Reserved & Calm,” along with key traits such as low extraversion, low conscientiousness, and emotional stability, all shown in a minimal and centered card layout on a gradient background; the interface is designed to clearly communicate the outcome while offering actionable options through “Take Test Again” and “Logout” buttons, making the user experience intuitive, reflective, and easy to navigate after completing the test.

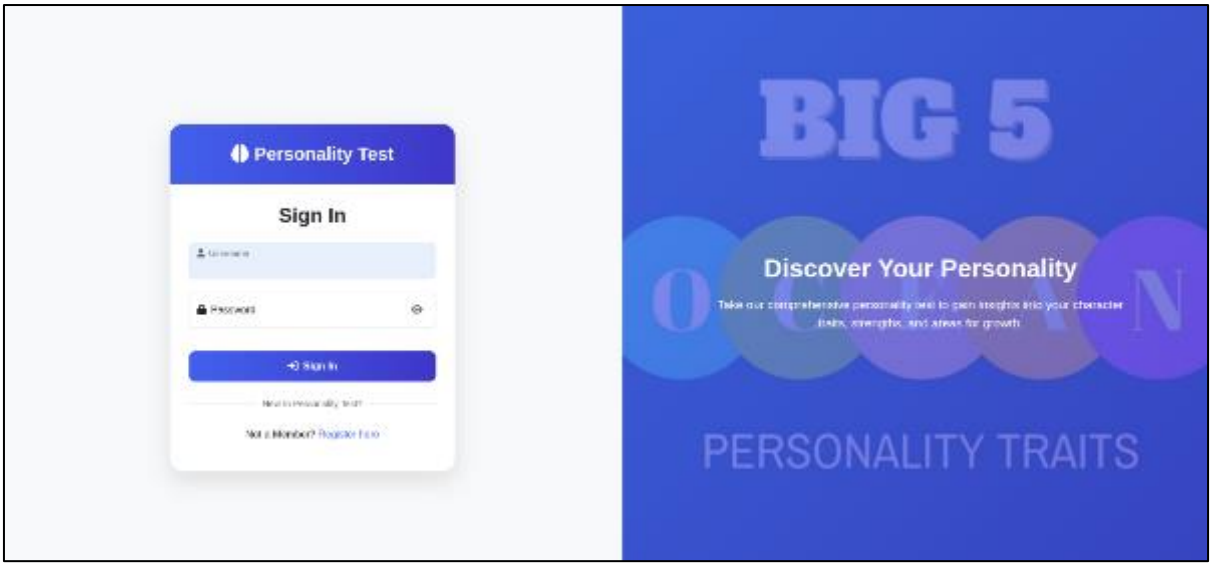

Figure 4 Big Five Personality Test Login and Introduction Screen

This image illustrates the entry interface of a Big Five personality testing application, combining a clean sign-in form on the left for secure user access with an informative and visually engaging introduction on the right that highlights the purpose of the test helping users discover their personality traits, strengths, and areas for growth using a modern blue-themed design that balances functionality with motivation and encourages users to begin their self-discovery journey.

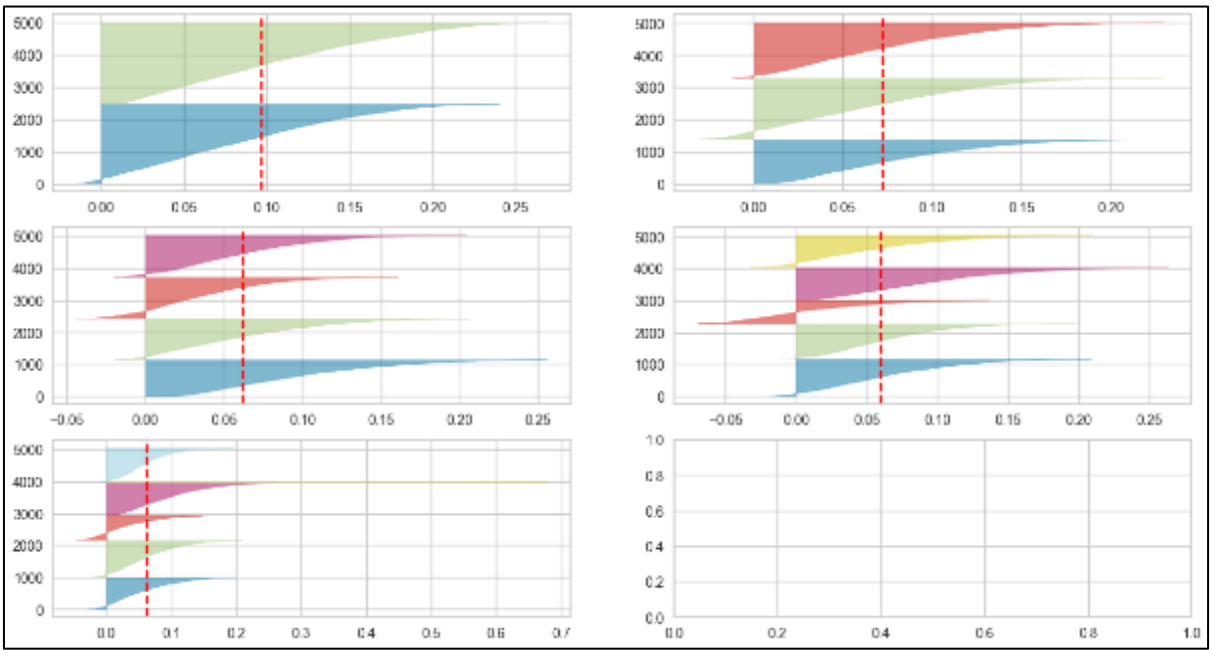

Figure 5 Comparative Multi-Scenario Performance and Trend Analysis Charts

This image presents a grid of multiple analytical graphs comparing trends across different scenarios or categories, where colored shaded regions represent varying data ranges or groups, and the red dashed vertical line highlights a critical threshold or reference point for comparison; together, these plots visually demonstrate how values evolve over time or across parameters, making it easier to identify patterns, overlaps, and deviations, and supporting data-driven interpretation, benchmarking, and decision-making through clear and structured visual analytics.

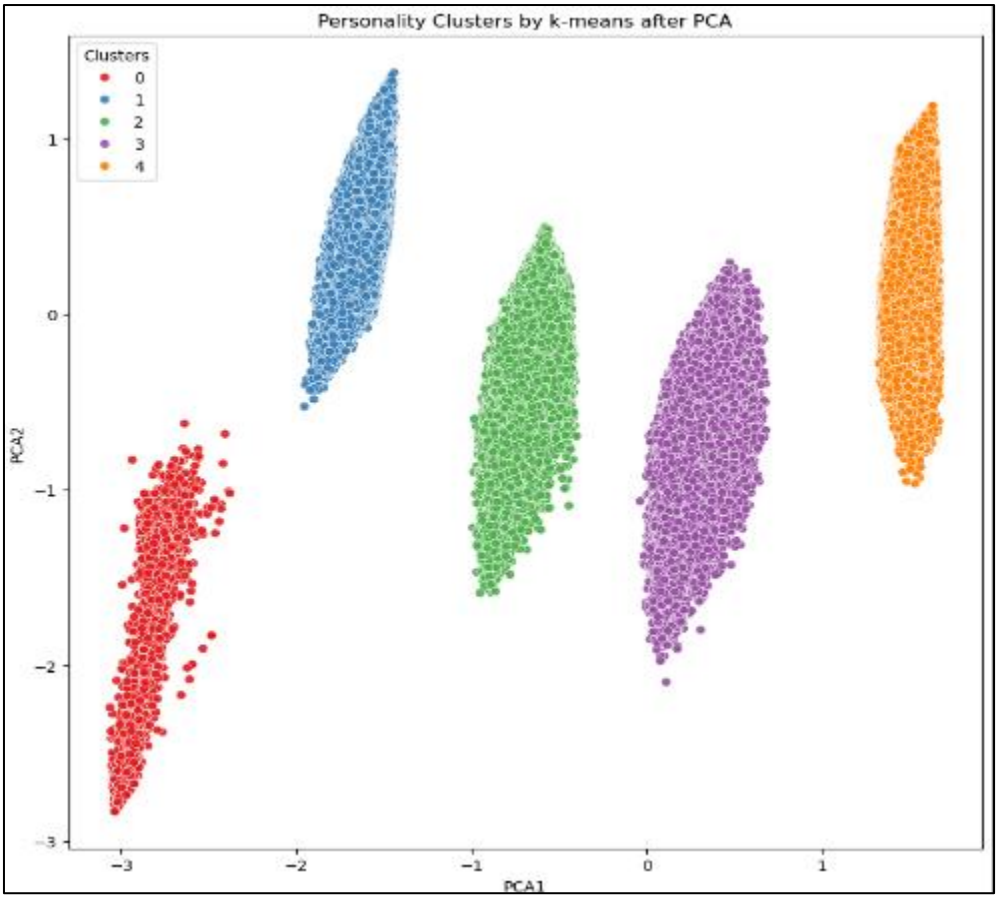

Figure 6 Personality Clusters Identified Using K-Means After PCA

This image visualizes the results of applying Principal Component Analysis (PCA) followed by K-means clustering to personality data, where each point represents an individual and the colors indicate five distinct personality clusters; the separation along the PCA1 and PCA2 axes shows how high-dimensional personality traits have been reduced into two principal components, allowing clear identification of naturally occurring groups and demonstrating how individuals with similar trait patterns are clustered together for meaningful personality segmentation and analysis.

5. Conclusions

In conclusion, the integration of Machine Learning and Artificial Intelligence for personality prediction using the OCEAN (Big Five) model presents a promising and effective approach for understanding human behavior through data-driven methods. By leveraging techniques such as data preprocessing, feature selection, and clustering algorithms, it becomes possible to identify meaningful personality patterns from diverse data sources like text, behavior, and social media activity. Despite challenges related to data quality, model interpretability, and ethical concerns such as privacy and consent, the proposed framework demonstrates significant potential for accurate and scalable personality assessment. With continued advancements in ML and AI, and careful attention to ethical considerations, such systems can contribute valuable insights across domains including psychology, human–computer interaction, recruitment, and personalized services.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] V. Santos and I. Paraboni," Myers- briggs personality bracket from social media textbook usingpre-trained language models", JUCS- Journal of Universal Computer Science, vol. 28, no. 4, pp. 1- 7, Apr. 2022.

[2] S. Bhatt, Y. Sharma, P. Karki and P. Katyal," Personality vaticination through questionnaire", International Journal of Advance Research Ideas and inventions in Technology, vol. 8, no. 4, pp. 1- 6, Jul. 2022.

[3] D. Dadebayev, W. W. Goh, and E. X. Tan, '' EEG- grounded emotion recognition Review of marketable EEG bias and machine literacy ways,'' J. King Saud Univ.- Comput. Inf. Sci., vol. 34, no. 7, pp. 4385 – 4401, Jul. 2022.

[4] D. Al- Hammadi and R. K. Moore," Using slice ways and machine literacy algorithms to ameliorate big five personality traits recognition fromnon-verbal cues", 2021 National Computing Colleges Conference(NCCC), pp. 1- 6, 2021.

[5] E. J. Zaferani, M. Teshnehlab and M. Vali," Automatic personality perception using autoencoder and hierarchical fuzzy bracket", 2021 26th International Computer Conference Computer Society of Iran(CSICC), pp. 1- 7, 2021.

[6] F. Annisa, E. Supriyanto and S. Taheri, Personality confines bracket with eeg analysis using support vector machine, pp. 79- 82, 12 2020.

[7] S. Katiyar, H. Walia and S. Kumar," Personality bracket system using data mining", 2020 8th International Conference on Reliability Infocom Technologies and Optimization(Trends and unborn Directions)(ICRITO), pp. 1020- 1023, 2020.

[8] S. Patel, M. Nimje, A. Shetty, and S. Kulkarni, "Analysis of personality traits using social media data," International Journal of Engineering Research & Technology (IJERT), vol. NTASU, pp. 421–426, 2020.

[9] M. Sahono, F. Sidiastahta, G. Shidik, A. Fanani, Muljono, S. Nuraisha, et al.," Backslapper and wallflower bracket grounded on myers- briggs type index(mbti) using support vector machine(svm)", 2020 International Seminar on Application for Technology of Information and Communication(iSemantic), pp. 572- 577, 2020.

[10] Khan, H. Ahmad, M. Asghar, F. Khan Saddozai, A. Arif and H. Ali," Personality bracket from online textbook using machine literacy approach", International Journal of Advanced Computer Science and Applications, vol. 11, 01 2020.