PERSONALITY PREDICTION USING ML

Tanvi Dashrath Bhagat[1], Sandeep R.Jadhav[2]

[1] ME Student, Department of Computer,Sanghavi College of Engineering ,Maharashtra,India
[2] Assistant professor, Department of Computer,Sanghavi College of Engineering ,Maharashtra,India

ABSTRACT

Personality prediction plays a vital role in understanding human behavior, decision-making, and social dynamics. The OCEAN model comprising Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism is a widely recognized framework for assessing personality traits. Integrating this model with machine learning (ML) represents a promising interdisciplinary approach, blending psychology, computer science, and data analysis. The Ocean Model theory conceptualizes personality traits as dynamic, wave-like patterns that shift over time, similar to ocean tides. By analyzing these behavioral fluctuations through ML algorithms, it becomes possible to predict individual personality traits with high precision. This study introduces a machine learning-based method to predict personality using participant responses from a personality questionnaire. Algorithms such as K-Means Clustering and Gaussian Mixture Model (GMM) were employed to classify and forecast personality scores.The results demonstrate strong predictive accuracy, with one model outperforming the rest. The study highlights the effectiveness of ML in personality assessment and its potential applications in domains like human resources, marketing, and mental health. It also outlines key challenges and future directions for refining this innovative approach.

Keyword: *Personality Prediction, Machine Learning, K-Means Clustering, Gaussian Mixture Model, OCEAN Model.*

1. INTRODUCTION

Personality prediction has emerged as an essential area of study in psychology and artificial intelligence, playing a significant role in understanding human behavior, decision-making, and social interactions. With the growing volume of behavioral data available through digital platforms, predicting personality traits using computational methods has become both feasible and valuable. The ability to analyze and predict an individual's personality can provide deep insights into their preferences, motivations, and tendencies, thereby aiding applications in areas such as recruitment, personalized marketing, education, and psychological counseling.The OCEAN model, also known as the Big Five Personality Traits, serves as the theoretical foundation for personality assessment in this study. It includes five core dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each of these dimensions represents a broad domain of human psychology, reflecting stable behavioral tendencies over time. The OCEAN model conceptualizes personality traits as dynamic and fluid, much like the waves of the ocean—constantly changing yet maintaining an underlying structure. This conceptualization provides an ideal framework for integrating psychological understanding with data-driven computational methods.In this research, machine learning (ML) techniques are applied to predict personality traits based on participants' responses to a structured questionnaire. Algorithms such as K-Means Clustering and the Gaussian Mixture Model (GMM) are employed to classify individuals into personality categories and forecast their corresponding OCEAN scores. These unsupervised learning approaches help uncover hidden patterns and correlations within the data, offering a more nuanced understanding of personality structures than traditional psychological assessments.The experimental results demonstrate that ML algorithms can effectively identify and predict personality traits with high accuracy. Among the models tested, one algorithm exhibited superior predictive performance, validating the potential of computational approaches in psychological assessment. Furthermore, the integration of ML with personality theory provides an opportunity to automate and scale personality evaluation, making it more accessible, objective, and data-driven.This study highlights the transformative potential of machine learning in personality prediction and behavioral analytics. By bridging the gap between psychology and artificial intelligence, it sets the stage for future innovations in fields such as human resources management, marketing analytics, education, and mental health diagnostics, where understanding individual differences is crucial for informed decision-making.This study also emphasizes the ethical and methodological considerations involved in personality prediction using machine learning. Ensuring data privacy, reducing algorithmic bias, and maintaining the interpretability of predictive models are critical factors for the responsible use of AI in psychological contexts. The research advocates for transparent and explainable ML systems that not only provide accurate results but also

respect individual autonomy and confidentiality. As the intersection of psychology and technology continues to evolve, developing fair, interpretable, and ethically guided personality prediction models will be essential for advancing both human-centered AI and the broader field of behavioral science

1.1 Problem Statement

The primary problem addressed in this study is the challenge of accurately predicting human personality traits using computational methods based on behavioral and questionnaire data. Traditional psychological assessments, while effective, are often time-consuming, subjective, and limited in scalability. There is a growing need for an automated, data-driven approach that can analyze large datasets to identify personality patterns with precision and consistency. By applying machine learning algorithms such as K-Means Clustering and Gaussian Mixture Model (GMM) within the framework of the OCEAN model, this research seeks to develop a reliable and efficient system for personality prediction that can be utilized across various fields, including human resources, marketing, and mental health analysis.

1.2 Aim

The main aim of this study is to develop an intelligent and data-driven personality prediction system using machine learning techniques based on the OCEAN model. The objective is to analyze participants' responses from personality questionnaires and accurately classify their traits into the five dimensions of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. By implementing algorithms such as K-Means Clustering and Gaussian Mixture Model (GMM), the study aims to enhance the precision and reliability of personality assessment while minimizing human bias and subjectivity. Ultimately, the research aspires to create a scalable and efficient model that can be applied in domains like recruitment, education, marketing, and psychological evaluation

2. LITERATURE REVIEW

1) Shahryar Salmani Bajestani, Mohammad Mahdi Khalilzadeh, Mahdi Azarnoosh, & Hamid Reza Kobravi "TranSentGAT: A Sentiment-Based Lexical Psycholinguistic Graph Attention Network for Personality Prediction" (IEEE Access, 2024).This paper introduces TranSentGAT, a hybrid model that fuses sentiment-aware lexical/psycholinguistic features with a graph-attention architecture to improve personality prediction from text. The authors argue that combining sentiment cues and structured lexical relationships via attention mechanisms yields richer representations for trait inference, and they validate the approach on benchmark datasets showing improved classification performance over several baselines. The work is notable for integrating psycholinguistic feature engineering with modern GAT-style deep models and for emphasizing sentiment as a complementary signal for personality traits.[1]
2) F. Zhang, G. Chen, H. Wang, & C. Zhang "CF-DAN: Facial-expression recognition based on cross-fusion dual-attention network" (Comput. Vis. Media, 2024).Zhang et al. present CF-DAN, a cross-fusion dual-attention architecture designed for fine-grained facial-expression recognition; the model fuses spatial and channel-wise information using dual-attention modules to capture subtle affective cues. Although primarily an emotion-recognition contribution, the paper is relevant to personality inference because accurate facial-affect encoding is often used as an input to personality models, and CF-DAN's improved expression discrimination can enhance downstream trait prediction.[2]
3) M. Kabir et al.Kabir and colleagues develop DEPTWEET, a taxonomy and processing pipeline tailored to social-media text for detecting the severity of depressive symptoms. The study combines linguistic features, severity labels, and machine-learning classifiers to show how fine-grained typologies improve detection sensitivity; its methodological lessons—careful labeling schemas and feature selection are directly applicable to personality modeling from social-text data, where nuanced annotation schemes matter. [3]
4) B. Blissing, F. Bruzelius, & O. Eriksson "The effects on driving behavior when using a head-mounted display in a dynamic driving simulator" (ACM Trans. Appl. Perception, 2022).This experimental study explores how head-mounted displays in driving simulators influence driver behavior and perceptual responses, reporting measurable changes in attention, lane-keeping, and reaction times under immersive conditions. Its findings highlight the role of interface modalities in eliciting behavioral differences insights that inform research combining simulated behavioral tasks with personality assessment, especially when designing ecologically valid data-collection environments. [4]
5) L. Angelini et al. "Towards an emotionally augmented Metaverse: A framework for recording and analysing physiological data and user behaviour" (Augmented Human Conf., 2022).Angelini and coauthors propose a framework to collect and analyze physiological signals and behavioral metrics within augmented/metaverse environments to augment virtual experiences with emotional context. The

framework outlines sensor fusion, privacy-aware data pipelines, and analytics for affect inference—important groundwork for integrating physiological and behavioral inputs into personality prediction systems deployed in immersive platforms. [5]

6) F. Ma & X. Pan "Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter" (IEEE VR, 2022).Ma and Pan investigate how visual fidelity of avatars shapes first impressions and expressive behavior in VR, demonstrating that early visual cues significantly affect how users present themselves and how observers judge personality-like attributes. Their results underscore the influence of presentation fidelity on behavior and perception an important consideration when collecting avatar-mediated behavioral data for personality inference.[6]
7) T. Rock et al. "Quantifying realistic behaviour of traffic agents in urban driving simulation based on questionnaires" (IEEE IV Symp., 2022).Rock and colleagues combine questionnaire-derived human judgments with simulation data to quantify realism of traffic-agent behaviors, proposing metrics to align agent models with human expectations. The methodology using subjective assessments to validate simulated behaviors offers a template for incorporating human-evaluated behavioral ground truth when modeling traits or behaviors in simulated studies. [7]
8) C. Himmels, T. Rock, J. Venrooij, & A. Riener "Simulator fidelity influences the sense of presence in driving simulators" (Adjunct Proc. Automotive UIs, 2022).This short paper shows that simulator fidelity modulates users' sense of presence, which in turn affects measured behaviors and responses during driving tasks. For personality research relying on simulated tasks, the study highlights the need to control or report fidelity, as lower presence may attenuate behavioral signals correlated with dispositional traits. [8]
9) Y. Shi et al. "Perception of driving simulations: Can the level of detail of virtual scenarios affect the driver's behavior and emotions?" (IEEE Trans. Veh. Technol., 2022).Shi and coauthors examine how scenario detail levels in virtual driving environments alter driver behavior and emotional states, finding that higher detail can elicit stronger emotional responses and more naturalistic driving styles. Their findings reinforce that environmental realism is a confound and a potential asset when using simulated behavior to infer personality or affective traits. [9]
10) M. Colley, P. Jansen, E. Rukzio, & J. Gugenheimer "SwiVR-CarSeat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat" (ACM IMWUT, 2022).This exploratory work evaluates how adding physical motion via a motorized swivel seat changes interaction quality and user experience in VR automated-driving scenarios. Results suggest embodied motion cues modulate engagement and behavior, which is relevant for multimodal personality-data collection where physical context can influence responses. [10]

3. PROPOSED SYSTEM

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

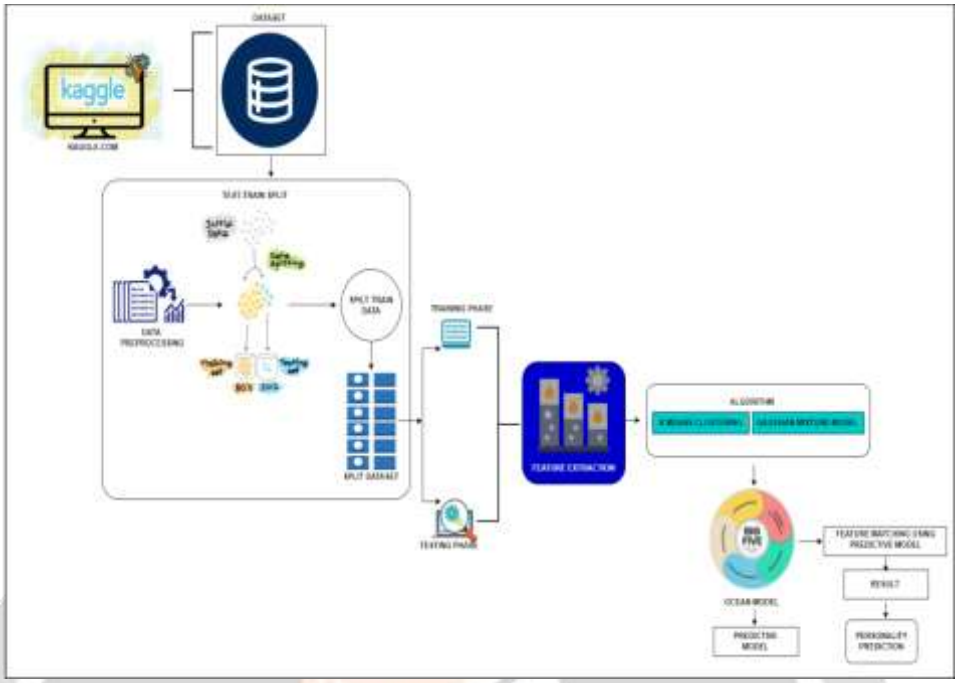

Fig : System Architecture

3.1 Methodology

The proposed methodology for personality prediction is structured around the integration of machine learning algorithms with the OCEAN model to analyze and classify individual personality traits. The process begins with data collection through a structured personality questionnaire, where participant responses are preprocessed by normalizing, cleaning, and encoding the data for model input. Feature extraction techniques are then applied to identify key behavioral indicators corresponding to the five OCEAN dimensions Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Unsupervised learning algorithms such as K-Means Clustering and the Gaussian Mixture Model (GMM) are employed to group participants into personality clusters based on response patterns, enabling the discovery of hidden relationships within the dataset. The models are evaluated using performance metrics like accuracy, precision, and silhouette score to assess clustering effectiveness. Finally, the outcomes are compared to traditional psychological evaluations to validate the model's reliability, thereby demonstrating the potential of machine learning as a scalable, objective, and data-driven approach to personality assessment.

4. CONCLUSIONS

In conclusion, this study demonstrates that integrating machine learning techniques with the OCEAN personality model provides an effective and data-driven approach to personality prediction. By leveraging algorithms such as K-Means Clustering and Gaussian Mixture Model (GMM), the system successfully identifies meaningful patterns in questionnaire data and classifies individuals into distinct personality categories with high precision. The findings confirm that machine learning can significantly enhance the accuracy, objectivity, and scalability of traditional psychological assessments. Moreover, this approach offers valuable applications in areas such as human resource management, marketing, education, and mental health analysis, where understanding individual traits can improve decision-making and personalization. However, the study also highlights the need for further refinement through larger datasets, cross-cultural validation, and the incorporation of multimodal behavioral data. Future research should focus on building more interpretable, ethical, and context-aware models to ensure responsible and transparent use of AI in personality analysis.

5. REFERENCES

1) Shahryar Salmani Bajestani , Mohammad Mahdi Khalilzadeh , Mahdi Azarnoosh , And Hamid Reza Kobravi,"TranSentGAT: A Sentiment-Based Lexical Psycholinguistic Graph Attention Network for Personality Prediction",Digital Object Identifier 10.1109/ACCESS.2024.3390126, Volume 12, 2024
2) F. Zhang, G. Chen, H. Wang, and C. Zhang, ''CF-DAN: Facial-expression recognition based on cross-fusion dual-attention network,'' Comput. Vis. Media, vol. 10, no. 3, pp. 593–608, Jun. 2024.
3) M. Kabir, T. Ahmed, M. B. Hasan, M. T. R. Laskar, T. K. Joarder, H. Mahmud, and K. Hasan, ''DEPTWEET: A typology for social media texts to detect depression severities,'' Comput. Hum. Behav., vol. 139, Feb. 2023, Art. no. 107503.
4) B. Blissing, F. Bruzelius, and O. Eriksson, ''The effects on driving behavior when using a head-mounted display in a dynamic driving simulator,'' ACM Trans. Appl. Perception, vol. 19, no. 1, pp. 1–18, Jan. 2022
5) L. Angelini, M. Mecella, H.-N. Liang, M. Caon, E. Mugellini, O. A. Khaled, and D. Bernardini, ''Towards an emotionally augmented Metaverse: A framework for recording and analysing physiological data and user behaviour,'' presented at the 13th Augmented Human Int. Conf., Winnipeg, MB, Canada, May 2022.
6) F. Ma and X. Pan, ''Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter,'' presented at the Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR), Christchurch, New Zealand, Mar. 2022.
7) T. Rock, M. Bahram, C. Himmels, and S. Marker, ''Quantifying realistic behaviour of traffic agents in urban driving simulation based on questionnaires,'' in Proc. IEEE Intell. Vehicles Symp. (IV), Jun. 2022, pp. 1675–1682.
8) C. Himmels, T. Rock, J. Venrooij, and A. Riener, ''Simulator fidelity influences the sense of presence in driving simulators,'' in Adjunct Proc. 14th Int. Conf. Automot. User Interfaces Interact. Veh. Appl., Sep. 2022, pp. 53–57.
9) Y. Shi, M. Boffi, B. E. A. Piga, L. Mussone, and G. Caruso, ''Perception of driving simulations: Can the level of detail of virtual scenarios affect the driver's behavior and emotions?'' IEEE Trans. Veh. Technol., vol. 71, no. 4, pp. 3429–3442, Apr. 2022.
10) M. Colley, P. Jansen, E. Rukzio, and J. Gugenheimer, ''SwiVR-CarSeat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat,'' Proc. ACM Interact. Mobile Wearable Ubiquitous Technol., vol. 5, no. 4, pp. 1–26, 2022.
11) V. Santos and I. Paraboni, "Myers-briggs personality classification from social media text using pre-trained language models", JUCS-Journal of Universal Computer Science, vol. 28, no. 4, pp. 1-7, Apr. 2022.
12) S. Bhatt, Y. Sharma, P. Karki and P. Katyal, "Personality prediction through questionnaire", International Journal of Advance Research Ideas and Innovations in Technology, vol. 8, no. 4, pp. 1-6, Jul. 2022.
13) D. Dadebayev, W. W. Goh, and E. X. Tan, ''EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques,'' J. King Saud Univ.-Comput. Inf. Sci., vol. 34, no. 7, pp. 4385–4401, Jul. 2022.
14) D. Al-Hammadi and R. K. Moore, "Using sampling techniques and machine learning algorithms to improve big five personality traits recognition from non-verbal cues", 2021 National Computing Colleges Conference (NCCC), pp. 1-6, 2021.
15) E. J. Zaferani, M. Teshnehlab and M. Vali, "Automatic personality perception using autoencoder and hierarchical fuzzy classification", 2021 26th International Computer Conference Computer Society of Iran (CSICC), pp. 1-7, 2021.
16) F. Annisa, E. Supriyanto and S. Taheri, Personality dimensions classification with eeg analysis using support vector machine, pp. 79-82, 12 2020.
17) S. Katiyar, H. Walia and S. Kumar, "Personality classification system using data mining", 2020 8th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1020-1023, 2020.
18) S. Patel, M. Nimje, A. Shetty and S. Kulkarni, "Personality analysis using social media", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT), vol. NTASU-2020, pp. 421-426, 2020.
19) M. Sahono, F. Sidiastahta, G. Shidik, A. Fanani, Muljono, S. Nuraisha, et al., "Extrovert and introvert classification based on myers-briggs type indicator (mbti) using support vector machine (svm)", 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 572-577, 2020.
20) Khan, H. Ahmad, M. Asghar, F. Khan Saddozai, A. Arif and H. Ali, "Personality classification from online text using machine learning approach", International Journal of Advanced Computer Science and Applications, vol. 11, 01 2020.

21) Z. Mushtaq, S. Ashraf and N. Sabahat, "Predicting mbti personality type with k-means clustering and gradient boosting", 2020 IEEE 23rd International Multitopic Conference (INMIC), pp. 1-5, 2020.
22) Talasbek, A. Serek, M. Zhaparov, S.-M. Yoo, Y.-K. Kim and G.-H. Jeong, "Personality classification by applying k-means clustering", 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 421-426, 2020.