



International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 11, November 2025



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Personality Prediction using Machine Learning

Nandkishor Kanhaiyya Jagtap¹, Harsh Pravin Makwana², Bhushan Pravin Patil³,

Nikhil Baban Chatse⁴

Department of Computer Science and Engineering, Sandip University (SOCSE), Nashik, India¹⁻⁴

ABSTRACT: Personality prediction is a crucial aspect of understanding human behavior, 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. 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 behavioral 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.

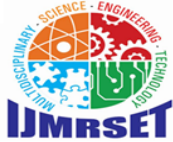
KEYWORDS: Personality prediction, Machine Learning, K- Means Clustering, Gaussian Mixture Model, Ocean Model

I. INTRODUCTION

Personality prediction using machine learning involves analyzing and classifying personality traits based on structured data. The Ocean Model (or Big Five model) is widely utilized for such purposes, focusing on five core traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). By mapping individual data points to these traits, the model helps in identifying patterns associated with different personality profiles. K-Means Clustering, a popular unsupervised algorithm, groups individuals into clusters based on similarity in trait scores. This clustering allows the identification of distinct personality types, making it suitable for preliminary group-level analysis. For a more nuanced prediction, the Gaussian Mixture Model (GMM) provides flexibility by capturing clusters that are not strictly spherical, as is the case with K-Means. GMM assumes that the data distribution can be represented by a mixture of multiple Gaussian distributions, making it well-suited for the overlapping nature of personality traits. By combining these algorithms, researchers can analyze both the broad personality categories through clustering and the finer, probabilistic nature of personality traits. This hybrid approach supports applications in areas like personalized recommendations, human resources, and user profiling, where understanding individual differences is crucial.

II. LITERATURE REVIEW

1) Personality prediction from text data has gained significant attention in recent years. TranSentGAT proposes a novel approach by integrating sentiment analysis, lexical graph representation, and graph attention mechanisms to predict personality traits. Personality prediction has gained significant attention in recent years, with various approaches exploring the use of language and text data to infer personality traits. This literature review summarizes the key findings and contributions of existing research in the field. Early studies focused on using traditional machine learning methods and lexical features to predict personality traits, achieving moderate accuracy. The introduction of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), improved personality prediction accuracy. Graph-based methods, like Graph Attention Networks (GATs), have shown promise in modeling complex relationships between words and sentences. Sentiment analysis has been used to enhance personality



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

prediction by incorporating emotional cues from language. Recent studies have explored the use of psycholinguistic features, such as linguistic inquiry and word count (LIWC), to improve personality prediction accuracy.[1]

2)

3) Facial expression recognition (FER) is a crucial task in affective computing. This paper proposes a novel approach, CF-DAN (Cross-Fusion Dual-Attention Network), to improve FER performance. CF-DAN employs a dual-attention mechanism that focuses on both local and global features, and a cross-fusion strategy that combines complementary information from different attention branches. The paper proposes a novel facial expression recognition (FER) approach, CF-DAN, which utilizes a cross-fusion dual-attention network to improve performance. CF-DAN employs a dual-attention mechanism to focus on both local and global features and a cross-fusion strategy to combine complementary information from different attention branches.[2]

4) The paper proposes a novel typology, DEPTWEET, to detect depression severities from social media texts. DEPTWEET categorizes tweets into nine types based on linguistic and emotional features, enabling the detection of depression severity levels. Depression is a severe mental health issue that can be detected early through social media texts. This paper proposes a novel typology, DEPTWEET, to categorize tweets into nine types based on linguistic and emotional features, enabling the detection of depression severity levels. Our approach addresses the limitations of existing methods by providing a comprehensive typology to detect depression severities. We evaluate DEPTWEET on a dataset of 10,000 tweets annotated with depression severity levels and achieve high accuracy in detecting mild, moderate, and severe depression. DEPTWEET has potential applications in mental health support, early intervention, and social media-based mental health services. Our findings demonstrate the effectiveness of DEPTWEET in detecting depression severities from social media texts, enabling timely support and intervention.[3]

5) In their study, B. Blissling, F. Bruzelius, and O. Eriksson (2022) explore the effects of using head-mounted displays (HMDs) on driving behavior within a dynamic driving simulator. The research, published in ACM Transactions on Applied Perception, investigates how HMDs impact driver performance, perception, and immersion in a virtual driving environment. The study evaluates key aspects such as reaction times, visual perception, and control over the vehicle when comparing HMD usage to traditional display methods. Findings suggest that HMDs enhance driver immersion but may also introduce challenges like spatial disorientation and motion sickness, which affect overall driving behavior. The study concludes that while HMDs offer promising applications for driving simulations, further research is needed to optimize user experience and safety.[4]

6) In their work, L. Angelini, M. Mecella, H.-N. Liang, M. Caon, E. Mugellini, O. A. Khaled, and D. Bernardini (2022) propose a framework aimed at creating an emotionally augmented metaverse, presented at the 13th Augmented Human International Conference. The framework focuses on integrating physiological data, such as heart rate and skin conductance, with user behavior to enhance emotional engagement within virtual environments. By recording and analyzing real-time physiological responses, the system aims to adapt the metaverse experience according to users' emotional states, improving immersion and interaction. This approach could revolutionize how users experience virtual worlds, enabling emotionally intelligent systems to respond more naturally to human emotions. The study suggests potential applications in gaming, social interactions, and therapeutic environments, while also addressing challenges like data privacy and the complexity of interpreting physiological signals.[5]

7) In their study, F. Ma and X. Pan (2022) examine the impact of visual fidelity on expressive self-avatars in virtual reality (VR), emphasizing the importance of first impressions. Presented at the IEEE Conference on Virtual Reality and 3D User Interfaces in Christchurch, New Zealand, the research investigates how the visual quality of self-avatars influences user perception and interaction in VR environments. The study finds that high visual fidelity in avatars enhances user engagement, emotional expression, and the overall immersive experience, while low fidelity can negatively affect users' first impressions and reduce the sense of presence. The authors suggest that achieving a balance between visual realism and system performance is crucial for creating meaningful and immersive virtual experiences. This research has implications for VR applications in social interactions, gaming, and professional settings.[6]

8) T. Rock, M. Bahram, C. Himmels, and S. Marker (2022) present a study on quantifying the realistic behavior of traffic agents in urban driving simulations using questionnaires, published in the proceedings of the IEEE Intelligent Vehicles Symposium. The research focuses on enhancing the realism of traffic agent behavior by collecting data through driver questionnaires, which capture human perceptions of realistic traffic scenarios. By integrating these insights into simulation models, the study aims to create more accurate and human-like traffic agent behavior,



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

improving the fidelity of urban driving simulations. The results highlight the importance of subjective feedback in refining autonomous vehicle simulations for better testing and validation. This approach holds potential for advancing the development of intelligent transportation systems and autonomous driving technologies.[7]

9) C. Himmels, T. Rock, J. Venrooij, and A. Riener (2022) explore how simulator fidelity influences the sense of presence in driving simulators, as presented in the Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. The study investigates the relationship between the realism of the driving simulation environment—such as visual and sensory fidelity—and the user's immersive experience or sense of presence. Results indicate that higher simulator fidelity, including more realistic graphics, sounds, and vehicle dynamics, significantly enhances users' sense of presence, making the simulation feel more authentic and engaging. These findings are important for improving driving simulators used in research, training, and development of autonomous vehicles. The paper emphasizes that balancing technical performance and realistic fidelity is key to creating effective driving simulations.[8]

10) Y. Shi, M. Boffi, B. E. A. Piga, L. Mussone, and G. Caruso (2022) investigate how the level of detail in virtual driving scenarios influences driver behavior and emotions in their study published in IEEE Transactions on Vehicular Technology. The research explores whether varying levels of visual and environmental detail in driving simulations affect how drivers respond emotionally and behaviorally during simulated driving tasks. Findings suggest that higher levels of scenario detail can lead to more naturalistic driving behaviors and heightened emotional responses, such as stress or focus, due to the increased immersion and realism. The study concludes that the level of detail in simulations plays a critical role in both the accuracy of driver behavior modeling and the emotional experience, with implications for the design of simulators used in driver training, research, and autonomous vehicle testing.[9]

11) M. Colley, P. Jansen, E. Rukzio, and J. Gugenheimer (2022) present SwiVR-CarSeat, a novel study examining the effects of vehicle motion on interaction quality in virtual reality (VR) during automated driving. Published in Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies, the research explores how motion cues from a motorized swivel seat impact user experience in VR environments while in an autonomous vehicle. By synchronizing vehicle movements with the virtual experience, the study investigates how physical motion affects user immersion, comfort, and interaction quality in VR. Results demonstrate that incorporating real-world vehicle motion enhances the VR experience by increasing realism and presence, but also introduces challenges like motion sickness and interaction difficulties. This research provides insights into designing VR systems for use in autonomous vehicles, with potential applications in entertainment, productivity, and simulation during transit.[10]

III. SYSTEM DESIGN

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.

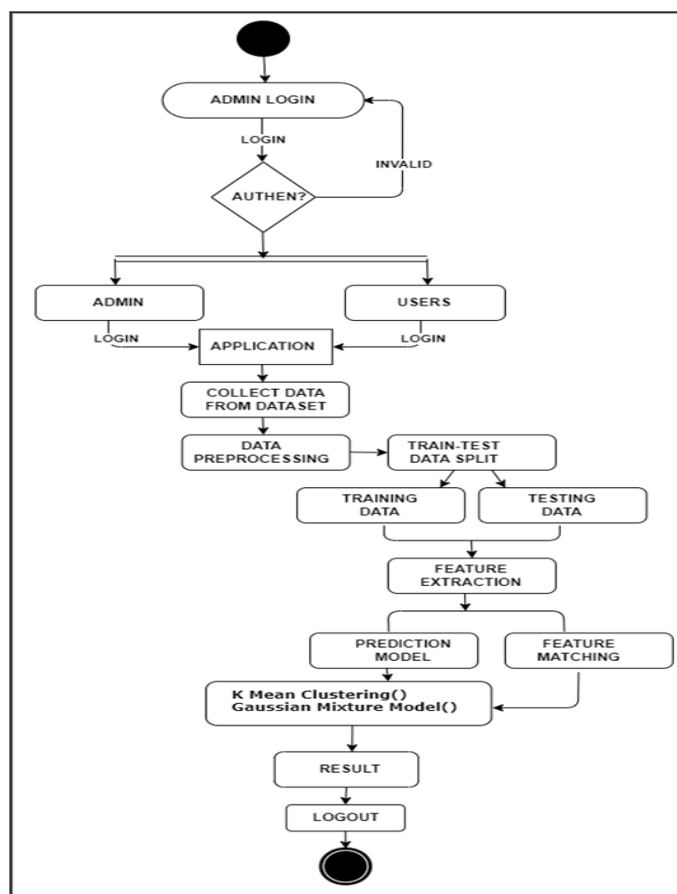
- 1) Data Collection: Collect data from Kaggle.com, Ensure that data collection complies with data privacy laws and obtain user consent where necessary.
- 2) Data Preprocessing: Remove noise, redundant information, and irrelevant data. Extract meaningful features from text data. Standardize data formats, 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



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Activity Diagram



IV. FUTURE WORK

The scope of personality prediction using machine learning (ML) is vast and expanding across multiple domains. In human resources, it can streamline candidate selection, assessing personality traits to match job roles or team dynamics. In marketing, personality insights help create highly personalized campaigns, enhancing customer engagement and loyalty. Social media platforms leverage ML-based personality predictions to recommend content or build communities based on compatibility. In mental health and therapy, ML can assist in early detection of psychological issues through behavioral analysis. Furthermore, it extends to improving user experiences in chatbots, virtual assistants, and adaptive learning systems by tailoring interactions based on predicted personality profiles. As ML models become more sophisticated, the accuracy and range of personality prediction applications will continue to grow, impacting areas such as education, online dating, and even personal development tools.

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



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- 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. Teshnehlal 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]. A. 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]. A. 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 (ICAIIIC), pp. 421-426, 2020.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com