A Survey Paper on Personality Prediction using Machine Learning

Kunal Sahu, Kartikey Joshi, Milesh Soni and Manthan Kumar

Dayanand Sagar College Of Engineering,India

Dayanand Sagar College Of Engineering, India

Dayanand Sagar College Of Engineering, India

Dr Reshma J

Associate Professor,Dept of ISEDSCE

Abstract— **In the competitive job market, selecting the right candidate for a given position is crucial for organizational success.Traditional human-based hiring processes can be time-consuming, subjective and prone to biases. Machine learning (ML) has emerged as a promising tool to streamline and enhance the recruitment process, particularly in the realm of personality prediction. This paper explores the application of ML for personality prediction in job applicant screening, aiming to automate initial candidate assessment and reduce the burden on HR personnel. By discussing various feature extraction techniques and ML algorithms for personality prediction, emphasizing the importance of data quality and ethical considerations. The potential benefits of ML-based personality prediction in job screening include improved efficiency, reduced biases, and enhanced decision-making. However, challenges remain in ensuring fairness, transparency, and accountability in the implementation of these algorithms. Further research is warranted to refine ML models, address ethical concerns, and establish clear guidelines for responsible use in the hiring process.**

Keywords—Personality Prediction

I. INTRODUCTION

In the dynamic and competitive landscape of today's workforce, organizations face the critical challenge of selecting the most suitable candidates for their open positions. Traditional hiring practices, often reliant on human-based assessments, can be time-consuming, subjective, and potentially biased, leading to suboptimal selection decisions. To address these limitations, machine learning (ML) has emerged as a powerful tool to enhance the recruitment process, particularly in the realm of personality prediction.

Personality assessment plays a pivotal role in candidate evaluation, as it provides valuable insights into an individual's work-related behaviors, attitudes, and motivations. Conventional methods of personality assessment, such as structured interviews and self-report questionnaires, can be cumbersome and may not always yield accurate or comprehensive results. ML, on the other hand, offers a promising approach to personality prediction, leveraging sophisticated algorithms to extract meaningful patterns from vast amounts of data.

The application of ML for personality prediction in job applicant screening holds several potential benefits. By automating initial candidate assessment, ML can significantly streamline the hiring process, reducing the time and effort required for HR personnel to manually evaluate each application. Moreover, ML algorithms can mitigate biases inherent in human-based assessments, leading to more objective and unbiased selection decisions. Furthermore, ML-based personality prediction can enhance decision-making by providing a deeper understanding of candidates' personality traits, enabling organizations to identify individuals who best align with the company culture, team dynamics, and specific job requirements.

Despite the potential benefits, the implementation of ML-based personality prediction in job screening raises several challenges and ethical considerations. Ensuring fairness, transparency, and accountability in the development and deployment of these algorithms is paramount. Organizations must carefully consider the data sources used for training ML models, the interpretability of model outputs, and the potential for unintended consequences. Additionally, clear guidelines and frameworks are needed to regulate the use of ML for personality prediction in the hiring context, protecting individual privacy and preventing discriminatory practices.

Moving forward, further research is warranted to refine ML models for personality prediction, address ethical

concerns, and establish clear guidelines for responsible use in the hiring process. By harnessing the power of ML in conjunction with human expertise, organizations can make more informed and unbiased hiring decisions, ultimately leading to a more diverse, productive, and engaged workforce.

2. LITERATURE SURVEY

[1] Personality prediction, an enduring area of research, takes a transformative turn by leveraging social media data, eliminating the need for user-filled questionnaires and offering a promising avenue for researchers. This shift not only reduces time constraints but also enhances credibility by capturing users' natural behavior. The significance of personality differentiation in psychology, human resources, and real-world applications underscores the importance of this research domain. As social media usage continues to surge, generating vast amounts of textual data and images daily, the exploration of methodologies such as Linear Discriminate Analysis, Multinomial Naive Bayes, and AdaBoost on Twitter standard dataset emerges as a focal point. The literature emphasizes the potential applications of personality prediction, including personalized user experiences and improved human resources practices, while also outlining future directions, such as interdisciplinary collaboration and the integration of multiple modalities, to advance the field..

[2]In the dynamic landscape of Twitter, a platform with an exponentially growing user base and terabytes of daily-generated data, this paper addresses the underutilization of this vast resource by tackling two significant tasks. Firstly, it aims to construct a sentiment analysis model employing BERT (Bidirectional Encoder Representations from Transformers) to discern and predict user sentiments from tweets. Secondly, the paper endeavors to develop a personality prediction model utilizing various machine learning classifiers within the framework of the Myers-Briggs Personality Type Indicator (MBTI), a widely recognized psychological instrument. By leveraging MBTI, the model seeks to forecast individual traits and qualities based on users' Twitter posts and interactions. Successfully predicting personality traits on Twitter holds potential applications in diverse fields such as market research, recruitment, psychological assessments, and consulting. The integration of BERT, MBTI, and machine learning classifiers underscores the multidimensional approach employed to extract valuable insights from the abundant yet underexplored Twitter data. Index Terms: BERT, MBTI, Machine Learning, Personality Prediction, Sentiment Analysis.

[3] This literature survey delves into the burgeoning field of text-based personality prediction, focusing on the integration of multiple social media data sources and employing pre-trained language models in conjunction with model averaging techniques. With the exponential growth of user-generated content across various social platforms,

researchers are increasingly exploring methods to extract personality insights from textual data. The survey investigates the utilization of pre-trained language models, such as BERT and GPT, to capture nuanced linguistic patterns and contextual information. Additionally, it explores the effectiveness of model averaging, a strategy that amalgamates predictions from multiple models, to enhance the robustness and generalizability of personality prediction systems. By examining existing studies in this domain, the survey aims to provide a comprehensive overview of methodologies, challenges, and advancements, shedding light on the state-of-the-art techniques and paving the way for future research directions in the dynamic intersection of text-based personality prediction and diverse social media data sources.

[5] This literature survey explores the transformative impact of Data Science and AI, particularly in the realm of machine learning applications, with a specific focus on classifying individuals based on their personality traits. Recognizing the uniqueness of each person and the availability of high-dimensional data, the survey emphasizes the application of machine learning, specifically the Big Five Personality Model, in predicting and categorizing personality traits. The study highlights the significance of such personality-based classifications in optimizing marketing campaigns, enhancing product/service recognition, and increasing overall effectiveness. Moreover, the survey addresses the relevance of this technology in the context of competitive exams, where personality evaluation plays a crucial role. By analyzing existing literature on Big Five Personality Model, feature analysis, and the broader landscape of predicting personality traits, the survey aims to provide insights into the advancements, challenges, and implications of utilizing machine learning for personality prediction. The integration of feature analysis and the potential for individuals to enhance their personality based on test results further underscores the practical applications of this approach. Keywords: Big Five Personality Model, Feature Analysis, Predicting Personality, Personality Traits.

[6]This literature survey delves into the realm of Automated Text-Based Personality Assessment (ATBPA) methods, recognizing their capacity to analyze extensive text data for nuanced personality cues. However, existing approaches grapple with challenges related to interpretability, explainability, and the validation standards established by traditional questionnaire instruments. Addressing these limitations, this survey focuses on the proposed Statement-to-Item Matching Personality Assessment (SIMPA) framework, which integrates questionnaire-based and text-based methods. SIMPA employs natural language processing to identify self-referencing personality descriptions in a target's text and establishes trait-constrained semantic similarity between these statements and questionnaire items. Drawing from the realistic accuracy model (RAM), the

framework incorporates a feedback loop mechanism to enhance judgment accuracy. A proof-of-concept implementation on Reddit demonstrates the efficacy of SIMPA for unsupervised estimation of Big 5 scores and, indirectly, as features for supervised ATBPA models, showcasing state-of-the-art results in personality prediction tasks on social media platforms. This survey contributes to the evolving landscape of personality assessment, bridging the gap between traditional and automated methods, and highlighting the potential of SIMPA in enhancing the interpretability and accuracy of text-based personality assessments.

3. WORKING PRINCIPLE

SYSTEM ARCHITECTURE
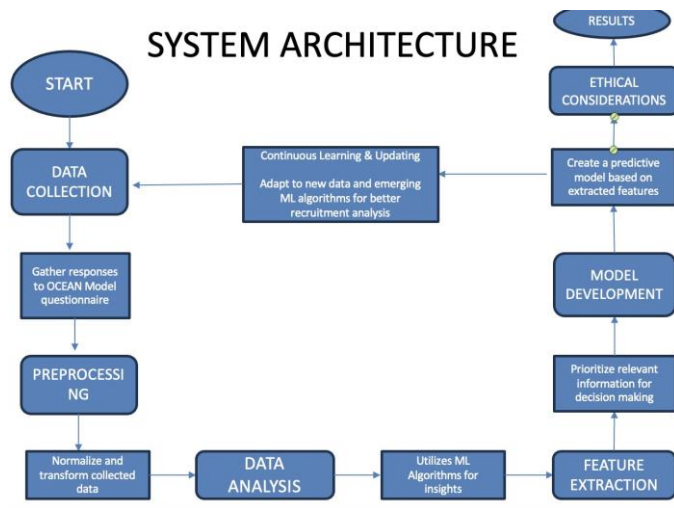
1.**Data Collection:** Gather a dataset of responses to the OCEAN model questionnaire, which assesses personality based on the Big Five personality traits.

2.**Preprocessing:** Clean and preprocess the text data, including tokenization, stemming, and removing stop words to prepare it for analysis.

3.**Feature Extraction:** Extract relevant features from the preprocessed text data, such as word frequency, sentiment analysis, and linguistic patterns associated with each personality trait.

4.**Model Training:** Utilize machine learning algorithms, such as logistic regression, neural networks, or ensemble methods, to train a model on the preprocessed data, aiming to predict personality traits based on the OCEAN model.

5. **Validation and Evaluation:** Validate the trained model using techniques like cross-validation and assess its performance using metrics like accuracy, precision, recall, and F1 score.

6. **Integration with Recruitment Process:** Integrate the trained model into the recruitment process to analyze candidate responses and predict their personality traits based on the OCEAN model questionnaire.

7. **Ethical Considerations:** Implement measures to address potential biases and ensure fair and ethical use of personality prediction in the recruitment process.

8. **Continuous Improvement:** Continuously update and improve the system by incorporating additional data, refining algorithms, and enhancing accuracy for better career guidance and recruitment analysis.

4.CONCLUSION

The survey results suggest that while HR professionals acknowledge the potential benefits of ML for improved efficiency and reduced bias, significant concerns remain about fairness, transparency, and responsible use. These findings highlight the need for further research to develop robust and ethical ML models for personality prediction, alongside clear guidelines and regulations for their implementation in the hiring process.

REFERENCES

[1] Aditi V. Kunte; Suja Panicker ,Using textual data for Personality Prediction:A Machine Learning Approach,(2022),IEEE.

[2] Myers-briggs Personality Prediction and Sentiment Analysis of Twitter using Machine Learning Classifiers and BERT,(2021),ResearchGate

[3] Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging,(2021),ResearchGate

[4] DeveshAgarwal*1, Mr. M. Karthikeyan,PERSONALITY PREDICTION USING MACHINE LEARNING,(2022),irjmets.

[5] SIMPA: Statement-to-Item Matching Personality Assessment from text,(2022),ScienceDirect.

[6] Personality Prediction Based on All Characters of User Social Media Information,(2014),Springer Link.