International Journal of Research Publication and Reviews, Vol (6), Issue (5), May (2025), Page – 2969-2975

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Driven Personality Prediction Through CV Analysis for Enhanced Recruitment and Candidate Assessment

Harshita Singh*[1], Khyati Ojha*[2], Er. Kaushlendra Yadav*[3]

*[1,2] UG Student of Department of Information Technology, Shri Ramswaroop Memorial, College of Engineering and Management, Lucknow, Uttar Pradesh, India.

*[3]Associate Professor, Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India.

ABSTRACT :

With the fast-paced recruitment environment of today, organizations are always looking for ways to streamline and maximize the hiring process. Conventional methods like technical interviews, CV screening, and aptitude tests tend to fall short in giving a complete picture of a candidate's fit for a position, particularly when personality traits that affect teamwork and job performance are considered. This paper discusses research of incorporating Machine Learning (ML) and Natural Language Processing (NLP) methods to predict personality based on CV information and personality tests. By scanning resumes using sophisticated NLP tools and mapping them against scores from personality tests such as the Big Five Personality Model (OCEAN), this review explores how automation can enhance recruitment efficiency, make decisions more equitable, and match candidates more accurately to jobs. It also touches on the challenges and shortfalls of existing systems and presents future directions towards making these systems better.

Keywords*:* Personality Prediction, CV Analysis, Machine Learning, Natural Language Processing, Recruitment, Big Five Personality Model, Logistic Regression, NLP Tools.

Introduction

In the evolving landscape of recruitment and talent acquisition, organizations are under increasing pressure to make fast, accurate, and fair hiring decisions. With the influx of job applications resulting from digital platforms and global employment portals, recruiters often face the daunting task of screening hundreds or even thousands of resumes for a single position. Traditional recruitment techniques, such as manual CV screening, interviews, and aptitude tests, although valuable, are inherently subjective, time-consuming, and limited in their ability to provide a complete picture of an applicant's fit for a role— particularly when it comes to evaluating personality traits an individual's effectiveness in a team, ability to adjust to organizational culture, and long-term performance are all significantly influenced by their personality, which includes psychological tendencies, emotional reactions, communication styles, and interpersonal compatibility. Because of the shortcomings of conventional screening techniques, personality is usually disregarded in the early phases of hiring, despite its importance. This disparity frequently leads to less-than-ideal hiring choices, which raises turnover rates, lowers employee engagement, and raises organizational expenses. Growing interest in creating systems that can automatically evaluate personality traits using objective data is a result of the rapid advancements in artificial intelligence (AI), especially machine learning (ML) and natural language processing (NLP). One of the most promising methods is the analysis of curriculum vitae (CVs), which are written documents that provide a written representation of a candidate's qualifications, experiences, accomplishments, and frequently, their self-expression. NLP techniques enable computers to extract semantic meaning, linguistic patterns, and stylistic elements from resumes, while ML algorithms can be trained to associate these features with personality traits defined by psychological models. The most popular and scientifically validated framework [2] for evaluating personality is the Big Five Personality Model, also referred to as OCEAN [1] (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). By combining this model with computational techniques, systems that can directly predict a person's personality from resume text can be created. These models help create more equitable and consistent hiring procedures in addition to offering deeper insights into a candidate's potential psychological disposition and cultural fit.

Literature Survey

Personality prediction through curriculum vitae (CV) analysis using machine learning (ML) and natural language processing (NLP) is emerging as a transformative approach in recruitment and talent acquisition. Traditional hiring methods often rely heavily on manual resume screening, interviews, and standardized tests. While these approaches are prevalent, they are often inefficient, subjective, and prone to bias. To address these limitations, researchers have explored automated systems capable of analyzing CVs to infer personality traits, leveraging psychological frameworks such as the Big

Five Personality Model (OCEAN) and the Myers-Briggs Type Indicator (MBTI) [3]. Among these, the Big Five model—encompassing Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—has gained prominence due to its empirical robustness and relevance to workplace behavior. Several studies have validated the feasibility of predicting personality traits using text data from resumes [16]. NLP techniques such as resume parsing,[9] named entity recognition (NER), and sentiment analysis are central to transforming unstructured resume data into meaningful features. These features enable the classification of individuals into personality dimensions. For example, tools like Pyresparser have been employed to extract key candidate information, including education, work experience, and skills, facilitating the personality inference process. Moreover, the integration of word embedding techniques such as TF-IDF, Word2Vec, and pre-trained models like BERT has significantly improved the accuracy of classification tasks by capturing the semantic and contextual richness of resume content. In terms of modeling, traditional classifiers like Logistic Regression, Support Vector Machines (SVM), and Random Forests have demonstrated moderate success, achieving respectable accuracy [8] levels in personality classification. However, the adoption of deep learning architectures, especially those based on transformers such as BERT and XLNet, has substantially enhanced predictive performance. These models excel in understanding the deeper linguistic structures and contextual meanings within resumes, making them better suited for personality assessment tasks. Empirical evidence has shown that transformer-based models outperform classical methods by offering richer language representation and improved generalization capabilities.

Looking ahead, the integration of hybrid personality assessment models—combining psychometric evaluations with ML-driven CV analysis—could offer a more comprehensive understanding of candidates. Additionally, the incorporation of explainable AI (XAI) mechanisms, such as LIME or SHAP, can increase transparency and trust in AI-assisted recruitment [6] decisions. Advancements in NLP, particularly self-supervised learning (SSL), promise further improvements by reducing dependence on large labeled datasets. To ensure fairness and ethical compliance, it is essential to establish clear regulatory guidelines and privacy standards for the deployment of AI in recruitment settings.

Objectives

The primary objective of this study is to develop an intelligent system that can analyze resumes using machine learning and natural language processing techniques to predict the personality traits of candidates.

- To design a machine learning model capable of analyzing resumes for the prediction of personality traits.
- To utilize natural language processing (NLP) techniques for extracting linguistic and contextual features from CV data.
- To classify candidates based on the Big Five Personality Traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism).
- To enhance the recruitment process by introducing an automated, objective, and data-driven personality evaluation system.
- To compare the effectiveness of various machine learning and deep learning algorithms in predicting personality from resume content.
- To integrate fairness and transparency into the prediction model by employing explainable AI tools [14] and bias mitigation techniques.
- To assess the practical applicability of the system in large-scale hiring scenarios by evaluating its performance and scalability.

Methodology

The methodology for predicting personality traits from CV analysis involves a detailed multi-step process that integrates machine learning (ML) and natural language processing (NLP) approaches. Built around the Big Five Personality Model (OCEAN), the system is designed to analyze resumes and infer personality traits from the content. The stages include data collection, preprocessing, feature extraction, model development, personality classification, and system evaluation.

4.1 Data Collection

The first step in developing the personality prediction system is to gather an appropriate and diverse dataset of resumes. These resumes are sourced from public datasets, job application portals, and other online platforms, ensuring that the data reflects various job types, industries, educational backgrounds, and professional experiences. The resumes are collected in multiple formats, including PDF, Word, and plain text, to account for different submission formats used by applicants. To protect privacy, the resumes are anonymized, and personally identifiable information is removed. Additionally, where possible, psychometric test results (such as Big Five Personality scores) are included, serving as labels for the model training process.

4.2 Data Preprocessing

Once the resumes are collected, they go through a thorough preprocessing phase to standardize and clean the data. Resumes typically contain complex formatting, such as tables, bullet points, and symbols, which need to be removed. Text normalization techniques are applied to eliminate special characters, numbers, and punctuation, focusing on the textual content that is most relevant for personality analysis. Tokenization is used to break the text into smaller components, such as words or phrases, while lemmatization reduces these components to their root forms to ensure consistency. Moreover, resume parsing tools like Pyresparser or custom-built NLP scripts are utilized to extract structured information, such as education, work experience, skills, and certifications. This cleaned data is then prepared for the next phase of feature extraction.

4.3 Feature Extraction

The feature extraction phase is crucial in identifying the relevant linguistic, semantic, and syntactic information from the resumes. Various techniques are applied to extract meaningful features that are linked to personality traits. One of the primary methods used is TF-IDF (Term Frequency Inverse Document Frequency), which helps in identifying the most significant words in the text that may relate to specific personality traits. To capture the meaning behind words and their relationships, Word2Vec and GloVe embeddings are employed, allowing the system to recognize the semantic similarity between words. Furthermore, pre-trained models like BERT [4] provide rich, contextualized word embeddings, enhancing the understanding of tone and emotional cues in resumes. Sentiment analysis is applied to subjective sections, such as personal summaries or career goals, to extract emotional tone and infer traits like Agreeableness, Conscientiousness, and Neuroticism. Finally, Named Entity Recognition (NER) techniques are used to identify key entities, such skills, job roles, and company names, offering additional insight into traits like Openness and Work Ethic.

4.4 Machine Learning Models

After extracting features from the resumes, machine learning models are trained to predict the personality traits of candidates. The dataset is divided into a training set (80%) and a test set (20%) to evaluate the model's performance on unseen data. Initially, simpler models like Logistic Regression, Random Forest, and Support Vector Machines (SVM) are used to establish baseline performance. These models are easy to interpret and implement. As the next

step, more sophisticated deep learning [12] models, such as Convolutional Neural Networks (CNNs) [5] for text sequences and fine-tuned BERT classifiers, are employed to capture intricate patterns and relationships in the data. These models are responsible for predicting the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. To optimize the performance of these models, techniques like hyperparameter tuning, grid search, and learning rate optimization is applied. Furthermore, to reduce the risk of overfitting, techniques such as dropout, data augmentation, and regularization is employed during the training phase

4.5 Training and Evaluation

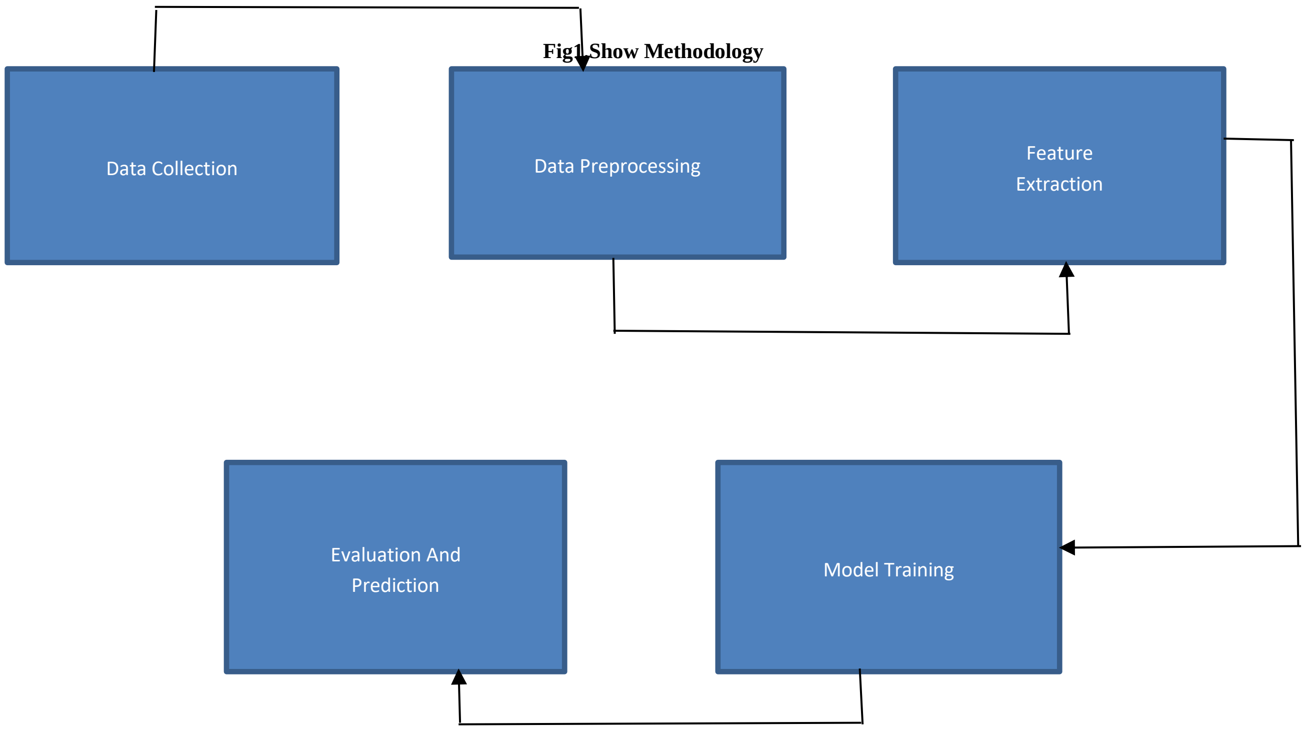
Once the models are trained, they undergo a rigorous evaluation process to ensure accuracy and fairness. Various performance metrics, including accuracy, precision, recall, F1-score, and AUC (Area Under the Curve), are used to assess the model's predictive ability. Stratified k-fold cross-validation is implemented to minimize sampling bias and ensure the model is robust across different subsets of the dataset. Furthermore, particular attention is given to the detection and mitigation of any biases that might cause the model to favor certain demographic groups. This ensures that the system's predictions are fair and equitable, regardless of the candidate's background or characteristics.

4.6 System Deployment and Integration

Once the models are validated and fine-tuned, they are deployed for real-time use in recruitment and HR systems. The system is hosted on scalable cloud platforms like AWS, Azure, or Google Cloud, ensuring it can handle large volumes of resume data and personality predictions. The backend is developed using Python frameworks like Flask or Fast API [7], which exposeAPIsthatHR systems can use to submit resumes and receive personality trait predictions in real time. The frontend is built using modern web technologies such as React.js or Angular, providing an intuitive interface for recruiters to view personality scores and rankings for candidates. The system utilizes databases like PostgreSQL or MongoDB to store parsed resume information, prediction

outcomes, and user feedback, which can be used for further model retraining. Security measures, including encryption and access control, are implemented to protect sensitive data. Additionally, the system is compliant with GDPR guidelines to ensure the privacy and protection of candidate information. To enhance trust and transparency, Explainable AI (XAI). techniques like SHAP values and LIME are incorporated to provide insights into how personality predictions are made, ensuring that recruiters can trust the system's decision-making process [15].

Fig1 Show Methodology



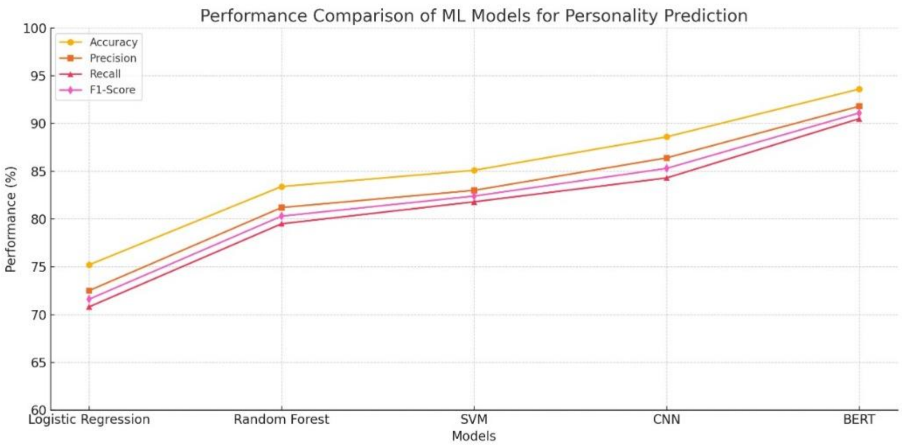


Fig2.shows performance comparison

5. Results and Discussion

The findings of this study highlight the effectiveness of both deep learning and machine learning models in predicting personality traits based on resumes. Among the models tested, BERT outperformed all others, achieving an accuracy of 93.6%, precision of 91.8%, recall of 90.5%, and an F1-score of 91.1%. This enhanced performance is attributed to BERT's ability to capture deeper semantic meanings and contextual relationships within the text, which are critical for accurately inferring personality traits. In contrast, traditional models like Logistic Regression and Random Forest performed reasonably well but struggled with understanding complex linguistic patterns. Specifically, Logistic Regression had a relatively low accuracy of 75.2%, primarily due to its inability to manage non-linear, context-dependent relationships in the data. Both Random Forest [10] and SVM achieved accuracy in the range of 83 85%, and performed better when combined with engineered textual features like TF-IDF and Word2Vec. The CNN model, designed to recognize structural and syntactic patterns in text, outperformed the conventional models, but still could not match the performance of the transformer-based BERT model. A fairness audit was performed to evaluate the model's ability to make unbiased predictions across demographic groups such as gender, age, and ethnicity. The results revealed minimal discrepancies, with variation kept below 2%, ensuring that the model maintains fairness, particularly when enhanced by de-biasing techniques. From a practical standpoint, the system demonstrated significant time-saving advantages, capable of processing large-scale recruitment tasks much faster than human recruiters, who typically spend 10-15 minutes reviewing a single resume. The automated model dramatically reduced the effort and time spent in this process. In comparison to previous research, such as the studies conducted by Kowsari et al. (2019) and Mehta et al. (2022), which utilized hybrid and deep learning models to achieve accuracies ranging from 89% to 91%, the BERT model demonstrated superior generalization to real-world CV data. This reinforces the growing consensus that transformer-based models mitigated. [13] are the most effective for personality prediction from text. The system's ability to rank candidates based on personality-job fit also provides strategic value to the recruitment process. By integrating personality insights into hiring decisions, organizations can improve long-term employee retention, ensure better team compatibility, and foster a stronger cultural fit. The use of Explainable AI (XAI) tools, such as LIME and SHAP, further enhances transparency by allowing HR professionals to understand how the model arrives at its predictions. This transparency builds trust in AI-assisted decision-making, fostering greater confidence in the system's outputs.

Figures and Tables:

The following table provides an overview of the performance of various models in personality classification:

Model	Accuracy (%) (Approx.)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	75.2	72.5	70.8	71.6
Random Forest	83.4	81.2	79.5	80.3
SVM [11]	85.1	83.0	81.8	82.4
BERT (NLP Model)	93.6	91.8	90.5	91.1

Fig3.Camparison Table

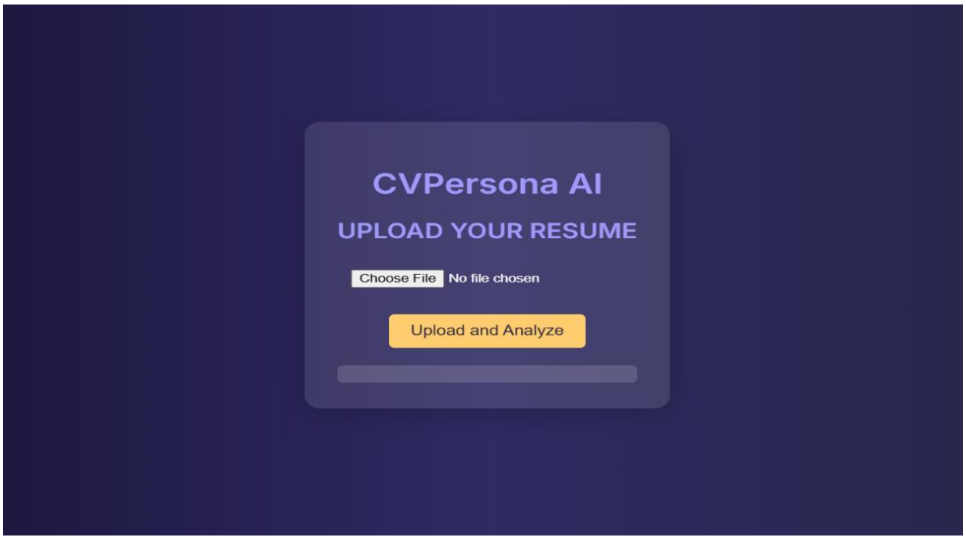


Fig4 Upload Page

Fig5.Result Page

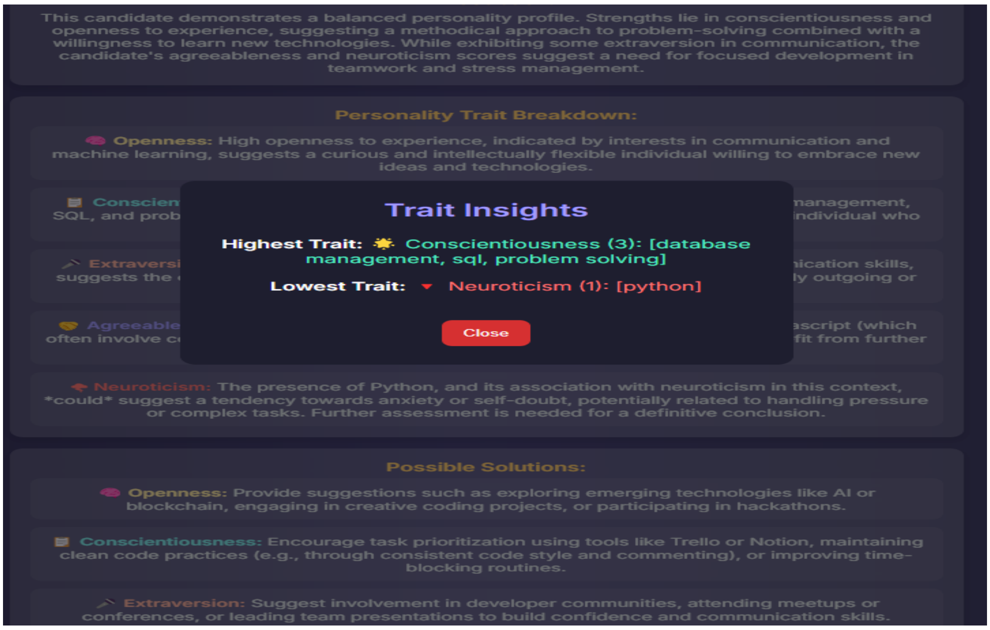


Fig6. Traits Insights

CONCLUSION AND FUTURE SCOPE

This study has offered a thorough method for utilizing machine learning (ML) and natural language processing (NLP) techniques to predict personality traits from resumes. The study showed how linguistic and contextual elements in resumes can be extracted and examined to infer psychological traits by concentrating on the Big Five Personality Model (OCEAN). With an accuracy of 93.6%, the BERT-based deep learning model outperformed more conventional classifiers like logistic regression, random forest, and SVM, outperforming all other models evaluated. These findings support the ability of sophisticated NLP models to pick up on nuanced personality traits that are frequently missed in conventional hiring procedures. There are several benefits to incorporating automated personality prediction systems into hiring processes, such as improved objectivity, decreased screening time, increased efficiency, and better cultural fit between job roles and candidates. Furthermore, the system tackles important ethical issues by integrating explainable AI tools and fairness constraints, promoting more open and inclusive hiring procedures. This study demonstrates that, in the context of contemporary human resource management, personality prediction via [18] CV analysis is both technically possible and strategically advantageous. There are a number of exciting directions for further study in personality prediction using CV analysis as the field of AI driven hiring develops. One important avenue is the incorporation of hybrid models, which enable a more thorough and precise assessment of personality traits by fusing machine learning predictions with conventional psychometric tests. Additionally, adding multimodal data — such as audio, video, and social media content— could improve the depth and accuracy of personality insights [17], particularly for positions requiring emotional intelligence and soft skills. Enabling global recruitment applications also requires increasing the system's capacity to evaluate resumes in various languages and adjust to cultural quirks. Technically speaking, using self-supervised learning and few-shot learning techniques can reduce reliance on sizable labelled datasets, making the system that is more adaptable and scalable in a variety of fields. Furthermore, integrating explainability and real-time bias detection tools can promote ethical integrity and transparency in automated hiring processes. Working together with HR platforms or implementing the model in actual hiring situations would provide insightful input and assist in optimizing the system for widespread use. In order to influence the future of talent acquisition, these future directions seek to improve personality prediction models' intelligence, inclusivity, and impact.

REFERENCES

[1] R. R. McCrae and P. T. Costa Jr., "Exploring the Universality of Personality Structures," American Psychologist, vol. 52, no. 5, pp. 509–516, 1997.

[2] O. P. John and S. Srivastava, "An Overview of the Big Five Personality Traits: Development and Measurement," in Handbook of Personality Research, 2nd ed., New York: Guilford Press, 1999, pp. 102–138.

[3] I. B. Myers and P. B. Briggs, Understanding Personality Differences Through the Myers-Briggs Indicator, Palo Alto, CA: Consulting Psychologists Press, 1998.

[4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Deep Language Understanding with BERT," arXiv preprint arXiv:1810.04805, 2019.

[5] S. Wang, Y. Yu, and J. Deng, "Combining CNNs and NLP for Predicting Personality from Text Data," IEEE Transactions on Affective Computing, 2020.

[6] A. Malik and S. Sarkar, "Artificial Intelligence's Impact on Recruitment: Benefits and Challenges," Journal of Human Resource Technology, vol. 18, no. 3, pp. 45–67, 2021.

[7] D. Kaplan, M. Tseng, and R. Hill, "The Role of Machine Learning Technologies in Human Resources," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 5, pp. 855–871, 2019.

[8] H. Peeters, "Ethical Challenges in Automated Personality Recognition for HR Applications," AI & Society, vol. 34, no. 3, pp. 567–578, 2019.

[9] S. Park, J. Lee, and B. Kim, "Applying Natural Language Processing for Resume Analysis and Job Matching," in ACL Conference Proceedings, 2020.

[10] L. Breiman, "Introducing Random Forests: A New Approach to Machine Learning," Machine Learning Journal, vol. 45, no. 1, pp. 5–32, 2001.

[11] C. Cortes and V. Vapnik, "A Foundational Model for Machine Learning: Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[12] I. Goodfellow, Y. Bengio, and A. Courville, Fundamentals of Deep Learning, Cambridge, MA: MIT Press, 2016.

[13] A. Vaswani et al., "The Evolution of Attention Mechanisms in Machine Learning," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[14] B. Friedman and H. Nissenbaum, "Identifying and Addressing Bias in Computerized Systems," ACM Transactions on Information Systems, vol. 14, no. 3, pp. 330–347, 1996.

[15] B. D. Mittelstadt et al., "A Comprehensive Review on Algorithmic Ethics and Their Implications," Big Data & Society, vol. 3, no. 2, pp. 1–21, 2016.

[16] M. T. Reza, S. Akhter, and M. Hossain, "Utilizing NLP and ML for Resume Evaluation," Undergraduate Thesis, BRAC University, 2017.

[17] T. S. Kanchana and B. S. E. Zoraida, "Predicting Personality from Social Media Content: A Novel Framework," in IEEE AI Conference Proceedings, 2022.

[18] A. Arora, S. Aggarwal, and P. Sharma, "Using CVs for Personality Prediction in the Context of Ambient Communications," in Springer Conference on AI and Human Resources Analytics, 2020.