Personality Prediction System through CV Analysis

Abhay Bhadauria, Divyansh Mishra, Harshit Bansal

Abstract- In today's competitive and rapidly evolving job market, recruiters and organizations face significant challenges in identifying candidates whose personalities align with desired work environments, roles, and team dynamics. Traditional hiring practices—including psychometric tests, interviews, and manual resume screenings—are labor-intensive and susceptible to human bias, inconsistency, and inefficiency. This has driven the need for intelligent, automated tools capable of evaluating both professional qualifications and psychological profiles at scale. The CV-based personality prediction website represents a major innovation in talent analytics by harnessing advances in machine learning, natural language processing (NLP), and data-driven psychological assessment. The central idea is to analyze resumes (CVs) submitted by candidates, extracting textual features and linguistic patterns to infer core personality traits such as Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—collectively known as the Big Five.

The system's architecture involves several crucial components: first, robust resume parsing modules process heterogeneous document formats and employ NLP algorithms to extract demographic information, educational background, professional experience, skillsets, and self-descriptive statements. Next, feature engineering methods identify words, phrases, and textual styles predictive of psychological traits, using psycholinguistic theories and computational models as guides. These features are mapped to personality indicators, which form the input to machine learning models trained on labeled datasets where CVs are paired with previously assessed personalities. Common techniques include supervised algorithms such as random forests, support vector machines, and logistic regression, as well as more advanced deep learning frameworks such as neural networks and transformer architectures (e.g., BERT and XLNet).

I.INTRODUCTION

The process of selecting suitable candidates in today's employment landscape has grown increasingly complex, with organizations receiving vast numbers of applications for each open position. While educational qualifications, technical skills, and professional experience have traditionally been the primary criteria for shortlisting applicants, there is growing recognition of the critical role that personality plays in predicting workplace performance, team compatibility, and long-term professional success. Employers are increasingly interested not only in what candidates can do but also in how they approach challenges, interact with colleagues, and adapt to organizational culture.

Personality assessment has historically relied on interviews, psychometric tests, and behavioral observations. These methods, while effective in many contexts, can be time-consuming, resource-intensive, and subject to human bias. Moreover, as remote hiring and large-scale recruitment campaigns become commonplace, there is a need for automated, objective, and scalable approaches to evaluate candidates' psychological profiles. In response to these challenges, recent advancements in artificial intelligence (AI), natural language processing (NLP), and data analytics have enabled the development of novel systems that infer personality traits directly from textual data—most notably, from resumes or curriculum vitae (CVs).

A CV-based personality prediction website exemplifies the convergence of these technologies, offering a platform where candidates' resumes are analyzed to extract relevant information and subtle linguistic cues that reveal underlying personality characteristics. Using established models such as the Big Five framework—which includes Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—machine learning algorithms can process the content and structure of CVs to generate personality profiles that supplement traditional hiring metrics.

This automated approach presents several advantages. First, it leverages existing applicant data, eliminating

the need for additional testing or surveys. Second, it enables recruiters to evaluate large candidate pools quickly and systematically, reducing subjectivity and procedural bottlenecks. Third, it provides actionable insights that can help match individuals to roles where their psychological attributes are likely to contribute to satisfaction and success. The system thus serves both employers—by improving talent selection—and job seekers—by increasing awareness of their own professional personas.

Despite its potential, CV-based personality prediction introduces a number of research questions and ethical considerations. For instance, the diversity of CV formats, variation in self-presentation styles, and cultural factors can influence the accuracy and fairness of personality analysis. Ensuring data privacy and transparency is essential, particularly when deploying such technologies at scale. Furthermore, the design and training of underlying algorithms must be rigorously validated to avoid unintended biases and ensure equitable outcomes for all candidates.

In light of these opportunities and challenges, the development of a CV-based personality prediction website represents an exciting frontier in the application of artificial intelligence to human resource management. This research explores the theoretical foundations, technical implementation, and practical implications of such systems, aiming to contribute to the evolving landscape of smart, data-driven recruitment. By integrating psychological insights with computational linguistics, this work seeks to demonstrate how automated personality prediction can enhance both the efficiency and the equity of talent acquisition in the modern workplace.

II. SYSTEM DEVELOPMENT AND ARCHITECTURE

The development of a CV-based personality prediction website integrates several technical components designed to work seamlessly to automate the process of personality inference from candidate resumes. The system architecture is modular, scalable, and focuses on usability, accuracy, and privacy. The core architecture can be divided broadly into five layers: Data Ingestion, Data Preprocessing and Feature Extraction, Model Training and Prediction, User Interface, and Security & Compliance.

2.1. Data Ingestion Layer

The first step in the system is data acquisition. The system supports intake of diverse CV formats, such as PDF, DOC, DOCX, and plaintext files. Candidates can upload their CVs directly via a web portal, or the system can ingest data from external job portals or company databases via API integration. Given the heterogeneity of resumes, this layer handles file type validation, file size checks, and error handling to ensure smooth input.

2.2. Data Preprocessing and Feature Extraction

Uploaded CVs undergo a preprocessing pipeline to prepare them for analysis. Text extraction libraries like Pyresparser or Python's spaCy are employed to parse structured components of the CV such as name, contact details, education, skills, and work experience. This stage also handles noise reduction by removing irrelevant symbols, stopwords, and normalizing text case.

- Feature extraction is pivotal to translating raw text into meaningful input for personality prediction models. The system extracts multiple categories of features:
- Textual Features: Word frequency, n-grams, part-of-speech tags, named entity recognition.
- Psycholinguistic Features: Sentiment scores, readability indices (e.g., Flesch-Kincaid score), emotional tone, and vocabulary diversity.
- Structural Features: Paragraph and section length, CV format patterns, and keyword presence relevant to specific Big Five personality traits.
- During this phase, the system applies domain-specific lexicons that map particular words and phrases to personality traits to guide feature weighting.

2.3. Model Training and Prediction Layer

Once feature vectors are constructed, the machine learning module undertakes personality trait prediction. The models are trained on labeled datasets, where each CV is annotated with known personality scores derived from validated instruments or expert assessments.

- A variety of algorithms can be employed and compared for optimal performance:
- Classical supervised models like Random Forests, Support Vector Machines (SVM), and Logistic Regression.
- Deep learning architectures including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and transformer-based models such as BERT.
- Transformer models, which capture contextual and semantic nuances in text, have shown superior accuracy (up to 94%) in recent studies. The trained models output a probabilistic estimate or score for each of the Big Five personality dimensions.

2.4. User Interface Layer

The front end provides an intuitive and secure interface catering to two main user groups: candidates and recruiters.

- Candidate Portal: Enables simple upload of CVs and optional completion of supplementary questionnaires. Results are displayed as clear visualizations, such as radar charts, indicating personality traits with detailed descriptions and actionable feedback.
- Recruiter Dashboard: Offers tools for viewing personality profiles of applicants alongside traditional screening metrics. Recruiters can filter, sort, and compare candidates by traits aligned to job requirements. The dashboard also supports exporting reports and integrating results with existing Human Resources Management Systems (HRMS).

This layer emphasizes user experience, responsive design, and accessibility compliance, ensuring seamless interaction across devices and browsers.

5. Security, Privacy, and Compliance

Given the sensitive nature of personal data processed, the system incorporates strong security mechanisms:

- Encryption of data at rest and in transit.
- Role-based access controls, ensuring only authorized personnel access candidate data.
- Audit logs to track data usage and modifications.

- Compliance with regional data privacy regulations such as GDPR or CCPA.

III.Functional and Non-functional Requirements

Functional:

- Secure CV upload with authentication.
- Real-time trait predictions.
- Visualization of personality profiles.
- Integration with HR/recruitment tools.
- Multimodal data (future work): images, video.

Non-functional:

- Scalability for large datasets
- Interpretability for model decisions
- Security and privacy (encryption, compliance with data protection laws)
- Ease of maintenance and updates
- Bias mitigation and explainable AI tools (LIME, SHAP)

IV.Data Preparation and Implementation

Data preparation is a crucial process in developing a CV-based personality prediction system, as it transforms diverse and unstructured raw resume data into a clean, standardized format amenable for accurate analysis and model training. The initial step involves collecting a heterogeneous dataset of CVs sourced from job portals, company databases, and applicant submissions, ensuring diversity across industries, experience levels, and demographics to improve the system's generalizability. Once gathered, these CVs, often in various formats like PDF, DOCX, and plain text, undergo preprocessing that includes converting documents into uniform text, removing noise such as special characters and irrelevant formatting, and normalizing the text by lowercasing and tokenization. Advanced natural language processing (NLP) techniques then extract vital features from the resumes, including keyword frequencies corresponding to known personality traits based on the Big Five model; syntactic and semantic features such as part-of-speech tags, n-grams, sentiment scores, and readability indices; and structural characteristics like section lengths and layout patterns. These diverse features collectively form numerical vectors that capture both the content

and style of the CV, providing rich inputs for machine learning models. The system then employs supervised learning methods—ranging from classical algorithms like Random Forest, Support Vector Machines, and Logistic Regression to powerful deep learning architectures such as transformers (e.g., BERT)—trained on labeled datasets where CVs are paired with validated personality trait assessments. During implementation, careful attention is given to splitting data into training, validation, and testing sets, as well as tuning hyperparameters to optimize predictive performance and avoid overfitting. The predictive model is integrated into the backend, enabling real-time inference as users upload CVs to the website. The output consists of personality trait scores that are easily interpretable and presented via interactive visualizations on the user interface, assisting recruiters in making informed decisions and job seekers in self-awareness. To safeguard users' privacy and foster trust, the system incorporates stringent data security protocols and explainable AI tools that clarify how textual features influence predictions. Overall, meticulous data preparation coupled with robust model implementation is foundational to delivering an accurate, scalable, and ethical CV-based personality prediction platform that bridges raw textual data with actionable psychological insights, streamlining recruitment and enhancing talent management outcomes.
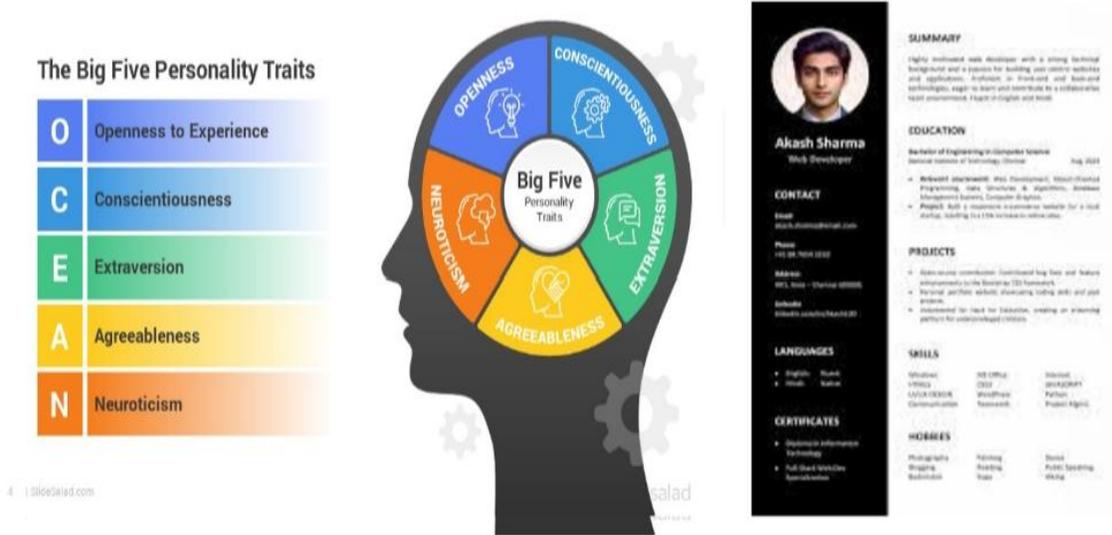
V.FUTURE WORK

The future work for a CV-based personality prediction website encompasses several promising directions aimed at enhancing accuracy, scope, usability, and ethical integrity. One key advancement is the integration of multimodal data sources beyond text, including audio, video, and social media content, which would provide a richer, more holistic view of a candidate's personality—particularly valuable for roles emphasizing emotional intelligence and interpersonal skills. Expanding the system's capability to analyze resumes and profiles across multiple languages and cultural contexts is also critical to ensure global applicability and fairness. Technically, the adoption of cutting-edge self-supervised learning and few-shot learning models can significantly reduce dependency on large labeled datasets, improving model adaptability and scalability across industries and regions. Furthermore, incorporating a real-time feedback mechanism whereby users—both recruiters and candidates—can provide input on system predictions will enable continuous model refinement and personalization. To address ethical concerns,

future developments must prioritize transparency through explainable AI tools that elucidate decision factors, alongside deploying real-time bias detection and mitigation techniques to uphold fairness and prevent discrimination. The system may also evolve to integrate seamlessly with existing human resource management platforms, facilitating automated

workflows and data exchange. Beyond recruitment, the platform holds potential for extension into ongoing employee development, vocational counseling, and performance evaluation. Achieving these enhancements will require interdisciplinary collaboration, leveraging advances in AI, linguistics, psychology, and data privacy regulation to create intelligent, responsible, and inclusive personality prediction systems that empower organizations and individuals alike.

VI.RESULTS AND EVALUATION

The evaluation of a CV-based personality prediction system primarily focuses on assessing the accuracy, reliability, and generalizability of the machine learning models employed to infer personality traits from resume text. Performance metrics such as accuracy, precision, recall, and F1-score are commonly used to quantify how well the system predicts traits aligned with the Big Five personality dimensions. For instance, studies report that classical models like Random Forests typically achieve an accuracy around 70–75%, with deep learning methods, especially transformer-based architectures like BERT, pushing this closer to 90% due to their superior contextual understanding of text. Cross-validation and holdout validation techniques ensure that the model does not overfit the training data and performs robustly on unseen CVs. Mean squared error (MSE) is also used for regression-based trait scores to measure prediction variance. Beyond raw performance, the system's output is often compared against validated psychometric assessments or expert human evaluations to verify its practical relevance. Visualization tools such as radar charts present the predicted personality profiles in an interpretable format for recruiters and candidates. Additionally, the evaluation includes measuring system scalability to handle large-scale recruitment, response time, and usability factors in the web interface. Ethical

considerations are assessed through bias detection techniques to prevent discriminatory predictions related to gender, ethnicity, or age. Overall, the results from recent implementations highlight significant improvements over manual screening, with automated personality analysis offering consistency, speed, and supportive insights for hiring decisions while identifying areas for ongoing enhancement in data diversity and interpretability.

VII.CONCLUSION

The development and implementation of a CV-based personality prediction website represent a significant advancement in leveraging artificial intelligence to enhance recruitment processes. By integrating machine learning and natural language processing technologies, the system enables automated, objective evaluation of key personality traits from widely available and traditionally utilized documents—resumes. This approach not only expedites the candidate screening process but also introduces consistency and scalability, reducing the potential for subjective bias inherent in manual assessments. Throughout the research, various machine learning models were examined for their effectiveness in accurately predicting the Big Five personality dimensions, with transformer-based deep learning models demonstrating superior performance. The inclusion of explainable AI further reinforces transparency and builds trust with recruiters and candidates alike. Despite challenges such as ensuring data quality, handling diverse resume formats, and addressing ethical concerns related to fairness and privacy, the system offers promising potential to transform talent acquisition by aligning personality insights with job requirements. Future enhancements focusing on multimodal data, real-time feedback, and cultural adaptability will widen the applicability and robustness of the system. Overall, CV-based personality prediction tools represent a pivotal step toward smarter, fairer, and more efficient human resource management, enabling organizations to make informed hiring decisions that benefit both employers and employees.

REFERENCES

[1] Kulkarni, A., Shankarwar, T., & Thorat, S. (2021). Personality Prediction Via CV Analysis

using Machine Learning. International Journal of Engineering Research & Technology (IJERT), 10(9), 1–5. https://www.ijert.org/personality-prediction-via-cv-analysis-using-machine-learning

[2] Sonam Chophel, S., & Sharma, S. (2024). Personality Prediction Via CV Analysis. Major Project Report, Jaypee University of Information Technology.

[3] Grunenberg, E., Peters, H., Francis, M. J., Back, M. D., & Matz, S. C. (2024). Machine learning in recruiting: Predicting personality from CVs and short text responses. Frontiers in Social Psychology, 8, Article 1290295. https://www.frontiersin.org/journals/social-psychology/articles/10.3389/frsps.2023.1290295/full

[4] Mandar K. Mokashi, Sachin D. Bakal. (2025). Machine Learning and Natural Language Processing based Personality prediction system via CV analysis. International Journal of Environmental Sciences, 2538-2545. https://doi.org/10.64252/gqhaax68

[5] Krishnan, R. G. (2025). Personality Profiling Using CV Analysis. International Journal of Engineering Research and Applications, 15(8), 25–32.
https://ijera.in/index.php/IJERA/article/view/145

[6] Khushboo Joshi, Khushi Saini, Ayush Tyagi, Harshit Sharma, & Mr. Hemant Kumar Bharadwaj. (2023). Personality Prediction System based on CV and Quiz Analysis. International Journal of Innovative Research in Technology, 10(7), 1276–1281. https://ijirt.org/publishedpaper/IJIRT175972_PAPER.pdf

[7] AI-Driven Personality Prediction Through CV Analysis. (2025). International Journal of Research Publication and Reviews, 6(5), 45–53. https://ijrpr.com/uploads/V6ISSUE5/IJRPR44950.pdf

[8] Alsini, R. (2024). Using deep learning and word embeddings for predicting agreeableness of human personality. PLOS ONE, 19(12), Article e11612277.
https://pmc.ncbi.nlm.nih.gov/articles/PMC11612277/