

## PERSONALITY PREDICTION THROUGH CV ANALYSIS USING MACHINE LEARNING TECHNIQUES

Karthi S, Deepakraja M, Dineshwaran S, Mohamed azad T

<sup>1</sup>Student, Dept. of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, IN

<sup>2</sup>Student, Dept. of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, IN

<sup>3</sup>Student, Dept. of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, IN

<sup>4</sup>Student, Dept. of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, IN

\*\*\*

**Abstract** -The proposed Personality Prediction System is a web application designed to streamline the job application process by incorporating personality prediction and skill matching. The system caters to two primary user roles: applicants and administrators (Admins). Upon logging in, applicants gain access to a personalized dashboard where they can explore company profiles, view published job listings, and apply for positions aligned with their skills and preferences. Admins, on the other hand, have the capability to publish job opportunities for their company and register the company's profile within the portal. This dual focus on skills and personality prediction aims to optimize the hiring process for both applicants and companies, fostering better job-role fit and overall satisfaction in the recruitment.

**Key Words** Personality Prediction, skills, administrators System, company profile , administrators (Admins). , hiring process, register, dual focus.

### INTRODUCTION

The Personality Prediction System is a cutting-edge web application designed to revolutionize the job application process. With distinct login capabilities for both applicants and Admins, the platform enables Admins to effortlessly publish job listings, register company profiles, and manage recruitment seamlessly. Applicants, in turn, can explore company profiles, peruse published job opportunities, and apply for roles that align with their skills and preferences.

This innovative feature aims to optimize successful job placements, making the recruitment experience more efficient and tailored for both applicants and employers. Personality is a complex construct that influences various aspects of human behavior, including career choices, job performance, and interpersonal relationships. Traditional personality assessment methods, such as self-report questionnaires and structured interviews, can be time-consuming and prone to biases. In recent years, there has been growing interest in leveraging machine learning techniques to predict personality traits from digital footprints, such as social media posts and online resumes (CVs). This paper explores the feasibility of using machine learning algorithms to predict personality traits from information contained within CVs. Your resume is more than just a list of jobs and skills. It's a window into your personality, and machine learning is the key to unlocking its secrets. In this project, we're putting on our detective hats and using machine

learning techniques to analyze the text in your CV. We'll be looking for clues like the words you use, the way you structure your sentences, and even the fonts you choose to reveal your hidden personal.

### Background of the Work

The field of personality psychology has long sought to understand and measure individual differences in behavior, thoughts, and emotions. Traditionally, personality assessment has relied on self-report questionnaires and structured interviews. However, these methods can be time-consuming, subjective, and prone to biases. With the advent of machine learning and natural language processing (NLP), researchers and practitioners have begun to explore innovative approaches to personality assessment. By analyzing large amounts of textual data, such as social media posts, emails, and resumes, machine learning algorithms can identify patterns and correlations that may reveal underlying personality traits. A CV, or resume, is a structured document that provides a snapshot of an individual's skills, experiences, and career goals. It often includes textual information that can be analyzed using NLP techniques. By examining the language used in a CV, such as word choice, sentence structure, and tone, it is possible to infer personality traits. By addressing these challenges and ethical considerations, researchers can develop robust and reliable methods for personality prediction through CV analysis, with potential applications in various fields, including human resources, education, and mental health. Machine Learning Algorithms: Classification: Employing algorithms like Support Vector Machines (SVM), Naive Bayes, and Random Forest to categorize individuals into personality types. Regression: Using algorithms like Linear Regression to predict continuous personality scores on specific traits. Deep Learning: Leveraging neural networks, especially Recurrent Neural Networks (RNNs) and Transformers, to capture complex patterns in textual data.

## Motivation and Scope of the Proposed Work

- The proposed Personality Prediction System is a sophisticated web application poised to streamline and enhance the job application and recruitment process.
- The system provides distinct login interfaces for both applicants and Admins, allowing Admins to effortlessly publish job listings, manage company profiles, and oversee the recruitment workflow.
- Applicants gain access to comprehensive company profiles and a curated list of job opportunities, enabling them to apply for roles aligned with their skills and preferences.
- The system incorporates a state-of-the-art skill-related prediction, enriching the matching process by evaluating the compatibility of applicants' skills with specific job requirements.
- This predictive feature aims to elevate the efficiency and precision of job placements, offering a tailored and optimized recruitment experience for both applicants and employers.

## 2. METHODOLOGY

The existing system relies on manual methods for job application and recruitment processes. Job seekers typically submit physical resumes or application forms to potential employers, and companies manually review and manage these documents. Paper-Based Resumes: Applicants typically submit physical resumes to companies or mail them, which can be time-consuming and less efficient. Manual Job Searches: Job seekers rely on traditional methods such as newspapers, networking, and word of mouth to find job opportunities. Time-Consuming Application Process: The application process is often manual, involving paperwork and lengthy turnaround times for both applicants and employers. Collect a diverse dataset of CVs, ensuring a wide range of industries, job roles, and educational backgrounds. Preprocess the CVs to remove noise, inconsistencies, and irrelevant information. Word frequency, n-grams, part-of-speech tags, and sentiment analysis. CV length, formatting style, and section organization. Writing style, tone, and complexity.

### System Architecture

CVs in various formats (PDF, DOCX, TXT). Extract text from CVs using tools like PyPDF2, docx2txt, or textract. Remove noise, inconsistencies, and irrelevant information. Convert text to lowercase, remove stop words, and perform stemming or lemmatization. Represent text as a bag of words.

Weight words based on their importance in the document.

Identify grammatical categories of words. Measure text readability (e.g., Flesch-Kincaid). Determine the emotional tone of the text. Total word count or number of pages. Length of different sections (e.g., experience, education). Create relevant features from extracted data. Train the chosen model(s) on the prepared dataset. Optimize model performance by adjusting hyperparameters. Deploy the trained model to a production environment (e.g., web application, API). Given a new CV, extract features and feed them to the trained model. Generate personality trait predictions (e.g., Big Five scores). Allow users to upload CVs and receive personality predictions. Provide an API for integration with other systems (e.g., HR software).

### Core Components of the Architecture

**Data Ingestion Layer:** Data Source This component is responsible for collecting CVs from various sources, such as job boards, company databases, or user uploads. **Data Preprocessing** This involves cleaning, normalizing, and formatting the CVs to ensure consistency and compatibility with the machine learning models. **Feature Extraction Layer:** Textual Features This component extracts textual features from the CVs, such as word frequency, n-grams, part-of-speech tags, and sentiment analysis. **Structural Features** This component extracts structural features, such as CV length, section lengths, and formatting style. **Stylistic Features** This component extracts stylistic features, such as readability indices and sentiment analysis. This component extracts stylistic features, such as readability indices and sentiment analysis. **Machine Learning Model Layer:** Model Selection This component selects appropriate machine learning models, such as SVM, Naive Bayes, Random Forest, or deep learning models like RNNs or Transformers, based on the problem and dataset. **Model Training:** This component trains the selected model(s) on the prepared dataset, optimizing hyperparameters to achieve the best performance. **Prediction** This component takes a new CV as input, extracts features, and feeds them to the trained model to generate personality trait predictions. **Output** This component presents the personality trait predictions in a user-friendly format, such as a report or a visualization. **User Interface Layer** **Web Interface** This component provides a user-friendly interface for uploading CVs and receiving personality predictions. This component provides an API for integration with other systems, such as HR software or recruitment platforms. This component visualizes personality profiles and compares them to job requirements.

**Fig -1- Table for parameters and key techniques**

Parameter	Description	key techniques
Data Ingestion	Collecting, cleaning, and preprocessing CV data.	Data collection, data cleaning, text extraction, data normalization
Machine Learning	Choosing appropriate machine learning algorithms.	Support Vector Machines (SVM), Naive Bayes, Random Forest, Deep Learning
Prediction and Inference	Making predictions on new CVs..	Model deployment, feature extraction, model inference
User Interface	Providing a user-friendly interface for input and output.	Web interface, API, visualization tools

## Data Ingestion

Data ingestion is the initial and crucial step in the process of personality prediction through CV analysis. It involves collecting, cleaning, and preparing the CV data for further analysis. **Data Collection** Source Identification Identify relevant sources for CV data, such as job boards, company databases, or user uploads. **Data Extraction** Extract CVs from various formats (PDF, DOCX, TXT) using tools like PyPDF2, docx2txt, or textract. **Data Cleaning** Noise Removal Remove noise, inconsistencies, and irrelevant information from the CVs.

**Formatting** Ensure consistent formatting across different CVs. **Text Normalization** Convert text to lowercase, remove stop words, and perform stemming or lemmatization. **Feature Normalization** Scale numerical features to a common range to improve model performance. **Challenges and Considerations:** **Data Quality** Ensuring the quality and consistency of the data is essential for accurate predictions. **Data Privacy** Protecting the privacy of individuals whose CVs are used for training and prediction is crucial. **Data Bias** Addressing biases in the data, such as gender, age, or cultural biases, is important to ensure fair and unbiased predictions. By carefully collecting, cleaning, and preparing the data, researchers can improve the accuracy and reliability of personality prediction models.

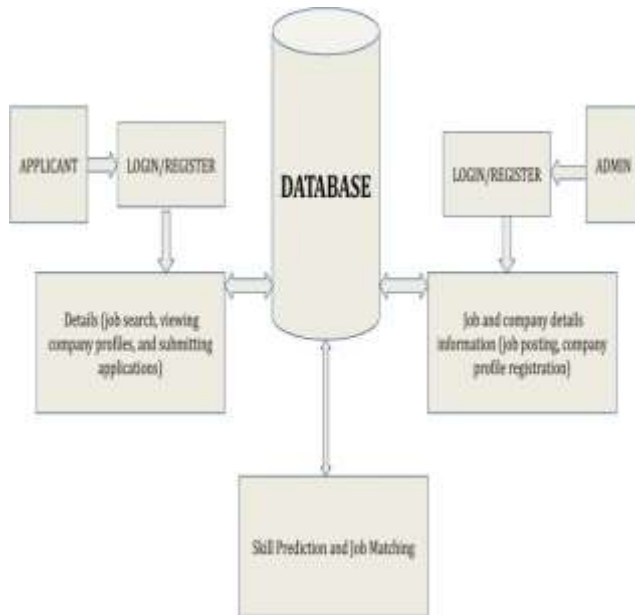
## User Interface

The user interface (UI) is the bridge between the complex machine learning model and the end-user. A well-designed UI can enhance user experience, increase user adoption, and improve the overall effectiveness of the personality prediction system. **Parallel Processing Units.** **Input Interface** File Upload Allow users to upload their CVs in various formats (PDF, DOCX, TXT). **Text Input** Provide an option for users to directly input CV text. **Output Interface** Personality Profile Display the predicted personality traits in a clear and concise manner, using visualizations like bar charts or radar charts. **Detailed Report** Provide a detailed report with explanations and insights into the predictions. **Customization Options** Trait Selection: Allow users to select specific traits of interest. **Comparison Tools** Provide tools to compare personality profiles with job requirements or ideal team members. **Design Considerations** **User-Friendliness** Keep the interface simple and intuitive. **Visual Clarity:** Use clear and concise visualizations to communicate information effectively. **Accessibility:** Ensure the interface is accessible to users with disabilities. **Security and Privacy:** Implement strong security measures to protect user data.

## Prediction and Inference

The prediction and inference phase is the core of the personality prediction system. It involves taking a new CV as input, extracting relevant features, and feeding them to the trained machine learning model to generate personality trait predictions. Extract relevant features from the new CV, using the same techniques as in the training phase. Feed the extracted features to the trained machine learning model. The model processes the input and generates predictions for the specified personality traits. Present the predicted personality traits in a clear and concise format, such as a numerical score, a categorical label, or a detailed report. **Model Accuracy** Ensuring the model's accuracy and reliability is crucial.

Fig -2-Working Flowchart



## CONCLUSION

In conclusion, the Personality Prediction System represents a significant advancement in the realm of job application and recruitment. The user-friendly platform, the system enhances efficiency for both applicants and Admins. The incorporation of a skill-related prediction algorithm not only refines the matching process but also increases the likelihood of successful job placements. The comprehensive company profiles and simplified application process contribute to a more informed and streamlined recruitment experience. The system stands as a robust solution, leveraging technology to optimize the dynamics of talent acquisition in the digital age, benefiting both job seekers and employers alike.

## REFERENCES

- [1] Ertan Taş and Mustafa Ersel Kamaşak, "Prediction Of Personality Traits From Videos By Using MachineLearning Algorithms", 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019.
- [2] W.H. Gomaa and A.A. Fahmy, "A survey of text similarity approaches", International journal of computer applications, no. 63, 2018.
- [3] M. S. Raja, M. Anurag, C. P. Reddy and Nageswara Rao. Sirisala, "Machine Learning Based Heart Disease Prediction System", 2021 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-5, 2021.
- [4] A.Nigam, A. Roy, H. Singh and H. Waila, "Job Recommendation through Progression of Job Selection", 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 212-216, 2019.
- [5] M. Jianjun, "Research on collaborative filtering recommendation algorithm based on user behavior characteristics", 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), pp. 425-428, 2020.