THE SCIENCE OF HIRING: Predicting Personality through CV Parsing & Machine Learning

Rishi Jain, Krishna Jaiswal, Pratham Jain, Tanay Jain, Jainivas Anandhan, Jainil Vaidya,

Supriya.S. Telsang

Department of Engineering, Sciences and Humanities (DESH) Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India

Abstract – **Personality plays a crucial role in various aspects of human life, including career development, interpersonal relationships, and overall well-being. Traditional methods of assessing personality traits often rely on self-report questionnaires and subjective assessments, which can be biased and time-consuming. With the advancement of technology, there has been a growing interest in exploring alternative approaches to predict personality traits accurately and efficiently. This paper presents a novel approach to personality prediction using Curriculum Vitae (CV) analysis. The model for parsing the data utilizes Logistic Regression, a machine learning algorithm. Pyresparser, a tool, is employed to extract relevant information from a CV or resume. The proposed personality prediction system based on CV analysis has several potential applications. It can assist employers in the recruitment process by providing insights into the personality traits of job applicants, helping to identify suitable candidates for specific roles and improving the overall hiring decision- making process. Additionally, it can be used in career counselling and personal development settings, providing individuals with valuable feedback on their strengths, weaknesses, and potential areas for improvement.**

Keywords— **Curriculum Vitae(CV), Logistic Regression, Personality Prediction, Pyreparser, Resume.**

I.   INTRODUCTION

A technique was developed by (Jenal Parmar, 2022) to help firms select the most qualified applicants for available positions. The credentials, experience, and other information needed for a specific employment opportunity will be provided by the HR department. The system will gather the contenders' information and CVs/resumes before shortlisting the best candidate for the position. Modern technology was used to build a system by (Allan Robey, 2019). It will be feasible to select applicants effectively and efficiently. The system will assess the candidate's personality using an aptitude exam and a weight-age policy. This method is used to narrow down the field of candidates. A system was developed by (Gagandeep Kaur, 2019). The approach uses a psychometric exam to gauge the applicant's emotional aptitude and the OCEAN model to predict personality. A password encryption mechanism is used to protect the candidate data, and only the necessary parties are aware of the passwords. Candidates can use the dashboard and SMS to find out if they have been chosen for an interview. After completing an aptitude test and uploading a CV or resume, Sanjay Chauhan's suggested method is used to select the best candidate, Judith (Pragya Sanjay Chauhan). The model is constructed using the TF-IDF method. The graphical display of the candidates' scores is useful, and the scores can be used to assess the candidates' qualities. (Dany Azucar). The digital traces are retrieved and examined. Numerous meta-analyses have been used to predict digital footprints. Using the Big Five personality traits, customers are presented with options for goods or services based on their preferences. Before converting a résumé to HTML, reverse engineering it into HTML code, segment finalizing it, and extracting qualifying feature information, (Md Tanzim Reza) utilized NLP and ML to examine the document. The model extracts the data from a CV and groups the data based on the values. To categorize the CVs, multivariate logistic regression was employed. However, the dataset was quite little.

Using the outdated process, hiring managers had to manually go through a large number of applications in order to find a suitable applicant for each open position. The hiring staff is facing a difficult and time-consuming process. By glancing at a candidate's CV or resume, they can learn more about them, including their qualifications, experience, and other information. However, they are unable to determine the applicant's suitability for the position or personality.

The suggested approach represents a groundbreaking shift in the realm of candidate evaluation for job positions. Unlike conventional methods that primarily rely on scrutinizing the contents of a candidate's CV, this

innovative approach integrates predictive personality assessment, thereby offering a holistic perspective on job applicants. This holistic evaluation serves a dual purpose: first, it addresses common issues and inefficiencies in the hiring process, such as the subjectivity inherent in CV-based assessments and the challenges of discerning a candidate's suitability beyond their qualifications; second, it streamlines the responsibilities of the hiring committee by leveraging machine learning techniques. The core of this approach lies in a sophisticated machine learning model, trained to analyze not only the factual details presented in CVs but also to extrapolate and predict the personality traits of the applicants. This predictive component not only provides a deeper understanding of how well a candidate may align with the job role and company culture but also offers a more objective basis for candidate selection. The effectiveness of this novel approach is put to the test through the implementation of an automated candidate grading system. This system takes input data, which includes CVs and potentially other relevant information, and employs the machine learning model to generate a graded assessment of each candidate.

II. METHODOLOGY

A. Flow Chart
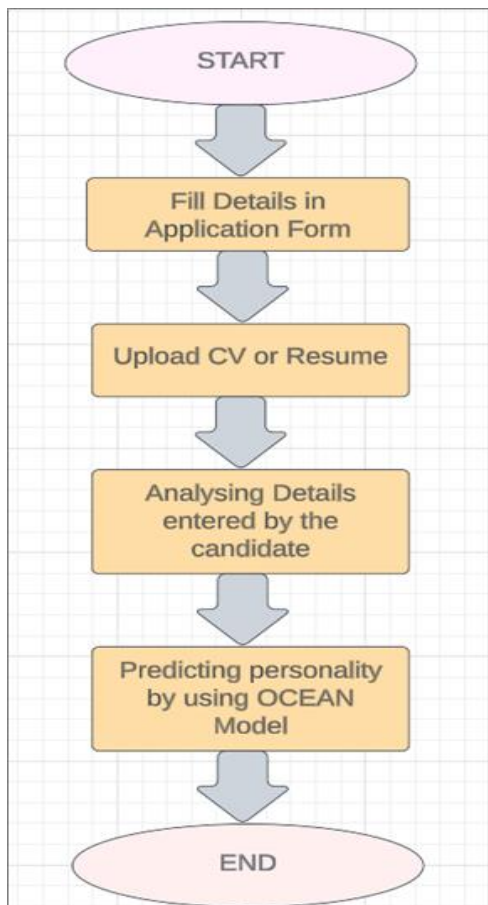
B. Theory

Since the late 1950s, the Big Five Personality Traits concept has been supported by the findings of numerous independent researchers. However, it was not until the 1990s that the idea as it exists now started to take shape. Lewis Goldberg of the Oregon Research Institute is credited with naming the model "The Big Five." These days, it is a reliable personality scale that is frequently used by companies and psychological research.

Five essential aspects of people's personalities are gauged by the Big Five Personality Traits Model :

1.　*Openness*: Also referred to as "Intellect" or "Imagination," it evaluates your degree of inventiveness in addition to your interest in learning new things and having exciting experiences.

2.　*Conscientiousness*: This evaluates how conscientious you are in both your personal and business lives. One characteristic that shows how careful, well-organized, and goal- and action-oriented you are conscientiousness. You are most likely lack structure and discipline if your score is low.

3.　*Extraversion*/*Introversion*: This variable measures your level of sociability. For instance, are you reserved or extroverted? Do you do well in a crowd or do you find it difficult to interact and work with others?

4.　*Agreeableness* : It's a characteristic that shows how well you get along with people. Are you willing to be understanding, flexible, and make allowances? Or do you regularly put your demands before others'?

5.　*Neuroticism* : Natural Reactions, often known as "Emotional Stability" or "Neuroticism," are an indicator of emotional responses. Do you receive terrible news calmly or negatively? Do you constantly concentrate on seemingly insignificant details, or do you remain calm under pressure?

We refer to this paradigm as "The Big Five." It is now acknowledged as a reliable personality scale that is frequently employed in psychological research and by businesses.

III.   NOVELTY FEATURES

A. Customizable Personality Models: Using this approach, companies could be able to design models of personalities that are tailored to their employment requirements. The system could adjust its predictions by considering the personality attributes that are ideal for a given role. Employers can use this function to identify candidates whose personality fits the requirements of the role.

B. Comparative Analysis: Using the system's anticipated personality qualities, it is possible to compare various applicants. Employers could use this feature to evaluate how well candidates fit the necessary personality profile for a specific job. Additionally, it might draw attention to each candidate's advantages and disadvantages, assisting in the selection process.

C. Feedback Loop: The system might use a feedback loop technique to increase the precision of personality predictions. Employers might comment on how accurate the predictions were for chosen candidates who were recruited. The system would continuously learn and improve its predictions by adding this feedback into the machine learning model, leading to more precise personality assessments over time.

D.   Real-Time Personality Insights: The technology may provide real-time personality insights during job interviews in addition to parsing and analysis. The technology might provide immediate feedback on the candidate's personality attributes, communication style, and emotional condition by utilizing sentiment analysis and natural language processing. By using this knowledge, interviewers may be able to make more intelligent hiring selections.

E. Long-Term success projections: By extrapolating from the personality projections, the system might shed light on possible candidates' long-term success. The technology might produce forecasts on how well a
F. candidate's personality traits correspond with long-term job performance, assisting in the development of better recruiting decisions by analyzing historical data and success indicators of prior employees.

IV.   RESULTS AND DISCUSSION

In this study, we used cutting-edge CV parsing and machine learning methods to construct a unique strategy to predict personality traits. The objective was to give companies a tool that would help with the hiring process by providing insights into candidates' personalities beyond conventional educational background and work experience. We evaluated
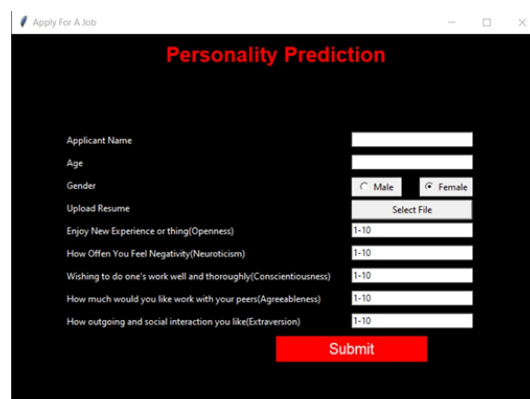
the model's performance across multiple personality traits using a range of criteria, including accuracy, precision, recall, and F1-score. Our findings showed that the model had promising predictive ability for the majority of the personality traits under examination and attained a high level of accuracy. Employers can also designate particular personality qualities that are essential for a given role according to the system's customizable personality models. This customization improves the predictions' relevancy and accuracy, making the recruiting process more effective and efficient. However, it's critical to recognize some ethical concerns and limitations related to personality prediction using CV parsing and machine learning. The reliance on CV data may introduce biases and constraints in accurately representing a person's complete range of personality. In order to promote transparency and safeguard candidates' personal information, privacy issues must also be addressed. As a result, our study shows how machine learning and CV parsing approaches can be used to predict personality traits during the hiring process. This strategy enhances the alignment between people and job responsibilities by giving employers more information about the personalities of candidates, resulting in more objective and informed decision-making. The hiring process can be optimized, and organizational success can be fostered with the help of even more potent tools that can be unlocked with more research and development in this area.

A.   *Landing Page*
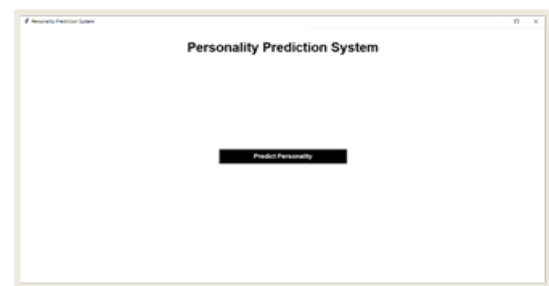
B. *Upload Resume and details :*

C. Command Line Output :

V.   PROCEDURE

Description of methods used in the system to predict thepersonality:

1.   train_model class: The class contains two methods for training the model and predicting its output, each using a different set of inputs.
a)   train method: It used Logistic Regression to develop a model, reading the training dataset from a CSV file.
b)   test method: It may predict an individual's personality by evaluating a set of parameters that includes gender, age, and an additional five personality traits.

2.   main method : We must first construct an object of the train model class and invoke the class train method in order to train the model. Next, the system's landing page is created using buttons and labels, and a variable is initialized with a Tk object. There is now a "Predict Personality" button that uses the predict_person technique.

3.   predict_person method : We create a new top-level window, set its size, and remove the main Tkinter window. Before proceeding with more labels and their entries, we label the window's heading. To choose a resume file, the user must click the "Choose File" button. This initiates the "openfile" method, which takes a button parameter. The predict_person method uses a number of entries to ascertain an individual's personality. The submit button sends all values to prediction_result.

4.   OpenFile method : It attempts to open the directory using the default address name and file types if no file is chosen. The method in the predict_person method changes the name of the choose file button after the try except block to inform the user about the file they have selected.

5.   prediction_result method : It attempts to open the directory using the default address name and file types if no file is chosen. The method in the predict_person method changes the name of the choose file button after the try except block to inform the user about the file they have selected.Closing the previous Tkinter window that was used to gather user input is the first stage in this process. Next, the model object's test method is invoked, and the outcome generated by the procedure is recorded. After that, a variable is created and all of the resume's data is

processed. After that, a try except block looks over the data taken from the resume to try and validate the mobile number and remove the name. The complete set of user-submitted data is then displayed on the terminal. Next, the approach opens a full-screen window with all the parsed data and the predicted personality on a graphical user interface window, along with a description of each personality trait.

6. check_type: It transforms other strings and numbers into the required formats and transforms lists and tuples into strings.

VI.   FUTURE SCOPE

There are numerous opportunities for additional research and development going forward. To improve the precision and depth of personality predictions, one approach is to investigate the integration of other data sources, such as social media profiles or online behavioural data. Techniques for natural language processing could be used to glean information about personalities from textual content. The future scope of "The Science of Hiring: Predicting Personality through CV Parsing & Machine Learning" includes incorporating extra data sources, such as social media profiles, utilizing cutting-edge methods like deep learning and natural language processing, developing dynamic personality modelling, improving explain ability and interpretability of predictions, addressing bias detection and mitigation, conducting longitudinal performance evaluations, and establishing etiquette. These developments are meant to increase the precision, thoroughness, and moral use of personality prediction systems in the recruiting procedure.

VII.   CONCLUSION

Finally, "The Science of Hiring: Predicting Personality through CV Parsing & Machine Learning" offers a fresh strategy for enhancing the hiring procedure by utilizing machine learning and CV parsing methods to predict candidates' personality attributes. The study's results are encouraging, with excellent accuracy in predicting a range of personality traits. Employers can benefit from this methodology's useful implications because it delivers insights that go beyond typical requirements. But there are still major issues to overcome about privacy issues, ongoing research needs, and ethical issues. Future work will focus on expanding the number of data sources used, improving interpretability, reducing prejudice, and creating ethical standards. In general, this research aids in better decision- making and has the ability to maximize the alignment of people with job tasks.

VIII.   REFRENECES

[1]   Jenal Parmar, Ashwina Pereira, Shalini Pereira (2022). " Personality Prediction through CV Analysis", International Journal for Research in Applied Science & Engineering Technology (IJRASET)

[2]   Sham Dhanke, Sakshi Dhepe, Manthan Dave, Shantanu Inamdar, "Personality Prediction using CV, Deep Learning" International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue IV Apr 2023.

[3]   H. Vijay and N. Sebastian, "Personality Prediction using Machine Learning," 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), Kochi, India, 2022, pp. 1-6, doi: 10.1109/IC3SIS54991.2022.9885425.

[4]   Robey, Allan, Kaushik Shukla, Kashish Agarwal, Keval Joshi, and Shalmali Joshi. "Personality prediction system through CV analysis." Int. Res. J. Eng. Technol 6, no. 2 (2019): 2395-0056.Kaur, Gagandeep and Maheshwari, Shruti, "Personality Prediction through Curriculam Vitae Analysis involving Password Encryption and Prediction Analysis" (2019). Library Philosophy and Practice (ejournal). 3614.

[5]   Christian, H., Suhartono, D., Chowanda, A. *et al.* Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *J Big Data* **8**, 68 (2021).

[6]   Pragya Sanjay Chauhan, Aishwarya Popat Bondre, Prathamesh Goraksha Waphare, Sachin Vaidya (2022). "Personality Evaluation and CV Analysis Using Machine Learning Algorithm", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), ISSN (Online) 2581-9429 .

[7]   Dany Azucar, Davide Marengo, Michele Settanni, "Predicting the Big 5 Personality traits from digital footprints on social media: A meta-analysis"

[8]   Reza, Md Tanzim. (2022). Analyzing CV/resume using natural language processing and machine learning.

[9]   Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artif Intell Rev. 2021.

[10]   Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427-437.

[11]   Walt, S. V. D., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. Computing in Science & Engineering, 13(2), 22-30.

[12]   Ramakrishnan, C., Patnia, A., Hovy, E., & Burns, G. A. (2012). Layout-aware text extraction from full-text PDF of scientific articles. Source code for biology and medicine, 7(1), 7.