ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 4| April 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

PREDICTING THE BIG FIVE PERSONALITIES THROUGH CV ANALYZER USING MACHINE LEARNING TECHNIQUES

Ms. Srinithi R[1], Mrs. P. Shanthi[2]

[1]221AI053, III – B.Sc Artificial Intelligence & Machine Learning, Dr. N.G.P Arts and Science College, Coimbatore.
[2]MCA., M.Phil., M.Sc(Mathematics)., NET, Assistant Professor, Department of Artificial Intelligence & Machine Learning, Dr. N.G.P Arts and Science College, Coimbatore.

ABSTRACT

The paper titled "Predicting the Big Five Personalities Through CV Analyzer Using Machine Learning Techniques" revolutionizes personality assessment by merging technology with practical utility. It enables users to upload resumes, which are analyzed using Natural Language Processing (NLP) to extract and analyze the key information. Feature Engineering refines this data for accurate personality trait predictions. Machine Learning models generate comprehensive reports detailing traits and suggesting personalized career paths. These insights are valuable for both individuals and recruiters, bridging the gap between resumes and meaningful outcomes. A statistical report summarizes analytical insights for broader applications. Overall, it demonstrates how NLP and ML drive career development and talent acquisition advancements.

INTRODUCTION

This study automates the personality analysis to aid career guidance and recruitment. It uses Natural Language Tool Kit, Logistic Regression, Machine Learning Models, and Streamlit in the Python platform. The NLTK is mainly used for extracting and analyzing texts in resumes, Logistic Regression is used for the classification of numerical values. Machine learning is proposed for the training and prediction of the models. Streamlit is a Python package that is used to build web apps. Matplotlib is used to visualize the models' data. The CVs are uploaded, transforming textual data into a structured format for analysis. The Machine learning models predict the personalities. The system generates two kinds of reports or results. This helps in aligning the career with their strengths and choices. These insights demonstrate how AI-driven analytics to enhance self–development and career plans in today's world. As an **accuracy rates 80%.**

LITERARY REVIEW

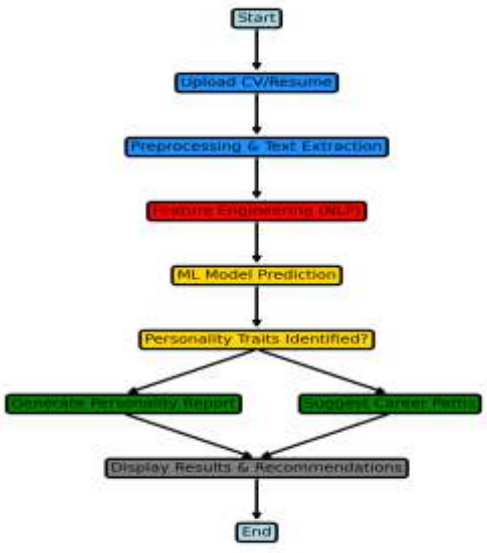
1. Machine Learning-Driven Personality Prediction System Using the Concept of CV Analysis, Nandani Agarwal, Siddharth Kumar

The study explores the use of machine learning algorithms like logistic regression, decision trees, and random forests to predict candidate personalities via CV analysis, employing the Big Five personality model. With the rising challenge of matching job seekers to roles, it highlights the importance of personality in professional success. Among the tested algorithms, logistic regression demonstrated superior accuracy and precision in predicting traits.

2. Personality Evaluation Through CV Analysis Using Machine Learning Algorithm, Suraj N Mali

The study proposed a system that automates candidate evaluation by analyzing uploaded CVs, leveraging the TF-IDF algorithm for professional eligibility and personality assessment. It generates candidate recommendations and graphical performance analyses, simplifying the recruitment process. The system ensures fair decision-making while aiding recruiters in shortlisting CVs efficiently. It leverages a database of recorded personality traits to identify and categorize personalities. It aims to help students better understand their traits and plan their actions accordingly. The innovative use of ML and NLP streamlines personality assessment for the educational process.

METHODOLOGY



1. **Start**: This is the initial phase where the process begins. Users access the system or application to kick-start the workflow.
2. **Upload CV/Resume**: In this step, users upload their resumes or CVs into the system. This document contains

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 4| April 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

important details about their qualifications, skills, experience, and more.

3. **Pre-processing and Text Extraction**: The uploaded document undergoes pre-processing, where irrelevant or unnecessary content is cleaned up, and text is extracted. This process ensures that the data is ready for further analysis, regardless of the document's format or structure.
4. **Feature Engineering (NLP)**: Using Natural Language Processing (NLP), meaningful features are extracted from the text, such as key skills, educational background, work experience, and other attributes. These features are structured to make them machine-readable and suitable for the next steps.
5. **ML Model Prediction**: A machine learning (ML) model is employed to analyze the extracted features and make predictions. For example, the model might predict personality traits, suitability for specific roles, or other characteristics based on the data.
6. **Personality Traits Identified**: Using the predictions from the ML model, the system identifies various personality traits of the individual, such as openness, conscientiousness, extroversion, agreeableness, and emotional stability. These traits are inferred based on patterns in the resume data.
7. **Personality Trait Report**: The identified traits are compiled into a report, providing a detailed overview of the person's personality. This report might include scores or qualitative descriptions for each trait and give career recommendations.
8. **Results and Recommendations**: In the final step, the system provides comprehensive results and actionable recommendations on enhancing career prospects, suggested areas for skill development, or ideal job matches.

DATASET

The dataset used for this study contains resumes and CVs collected from publicly available sources, ensuring ethical usage and compliance with privacy policies. It comprises the samples, each annotated with features such as skills, experience, education, and other relevant attributes. The data has been preprocessed to remove irrelevant elements, such as formatting artifacts and normalizing for consistency. This dataset serves as the basis for feature extraction, NLP techniques, and personality predictions using machine learning algorithms. The source of the data set is **Kaggle and GitHub**. It is being selected for its comprehensive coverage and relevance to the objectives of this study.

Traditional Assessment Methods:

Traditional personality assessment methods, such as self-reported tests like the Big Five or Myers-Briggs and manual review by HR professionals, are often subjective, time-consuming, and prone to bias. Similarly, rule-based resume screening systems rely on keyword filtering but cannot analyze linguistic patterns or predict behavioral traits, and automated visual representation of personality traits, as modern data visualization tools like Plotly or Seaborn are rarely integrated into such systems. These limitations highlight the need for more advanced and efficient approaches.

CV Analyzer Using Artificial Intelligence and Machine Learning:

This methodology harnesses AI and Machine Learning to revolutionize personality prediction and resume analysis. Logistic Regression models predict traits based on BIG FIVE scores, automating the process to reduce bias and enhance efficiency. NLP tools like pyresparser extract essential resume details such as skills, experience, and contact information. Dynamic visualizations, including bar charts(Plotly), heatmaps(seaborn), and Worldclouds, provide intuitive insights into personality distributions and resume data. Additionally, the system offers career recommendations by matching personality traits with suitable career paths, empowering job seekers to make informed decisions.

Comparing the Previous Research and Current Research's Performance Measure:

METRIC	VALUE
Precision	85%
Recall	82%
Accuracy	87%

Table 1: TF-IDF Algorithm

PERSONALITY TRAIT	MODEL ACCURACY
Openness	82%
Conscientiousness	85%
Extraversion	80%
Agreeableness	78%
Neuroticism	79%

Table 2: LOGISTIC REGRESSION

CONCLUSION

The Integration of Machine Learning and NLP in personality prediction and resume analysis showcases the potential of AI in transforming career guidance and recruitment. By automating the extraction. Analysis and visualization of data in the system reduces bias, saves time, and improves accuracy. Dynamic visualization and deleted personality reports provide actionable insights, empowering individuals to make informed career decisions. The use of python-based tools like Streamlit, NLTK, and Matplotlib further enhance the efficiency and usability of the system, This approach not only streamlines recruitment processes but also fosters self-development by aligning career paths with individual strength . Such advancements highlight the role of AI in shaping the future of career planning and human resource management.

REFERENCES

1. *G. Sudha, S. K. K, S. J. S, N. D, S. S, and K. T. G, "Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process", IEEE Xplore, Dec. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9711787 (accessed Mar. 29, 2022).*
2. *N. T. Singh, Abhay Chanana, D. Jain, and R. Kumar, "Personality prediction through CV analysis using machine learning techniques," vol. 4, pp. 1–6, Jan. 2023, doi: https://doi.org/10.1109/icaect57570.2023.10117883.*
3. *A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 11| Issue: 4| April 2025|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2025: 8.691 || ISI Value: 1.188

prediction: a survey," The International Journal of Robotics Research, vol. 39, no. 8, pp. 895–935, Jun. 2020, https://doi.org/10.1177/0278364920917446.

4. *A. Kochling and M. C. Wehner, "Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development," Business Research, vol. 13, no. 3, pp. 795–848, Nov. 2020, https://doi.org/10.1007/s40685-020-00134-w.*
5. *V. Siahaan and Rismon Hasiholan Sianipar, Customer Personality Analysis and Prediction Using Machine Learning With Python. Balige Publishing, 2023.*
6. *C. G. DeYoung, J. B. Hirsh, M. S. Shane, X.. Papademetris, N. Rajeevan, and J. R. Gray, "Testing Predictions From Personality Neuroscience: Brain Structure and the Big Five," Psychological Science, vol. 21, no. 6, pp. 820–828, Apr. 2010, doi: https://doi.org/10.1177/0956797610370159.*
7. *Dr. Y. Arora, "A Review of Machine Learning Techniques Over Big Data Case Studies," International Journal of Innovative Research in Computer Science &Technology,vol.8,no.3,May2020,doi.*