PERSONALITYPREDICTIONTOIMPROVERESULTSIN CVANALYSISBYMACHINELEARNING

[1]
Mrs.G.Narsamma
Assistant Professor
Dept of CSE(Data Science)
Sreyas Institute of
Engineering and Technology
Hyderabad
narasamma.g@sreyas.ac.in

[2]
T.Venkata Ankitha
UG Scholar
Dept of CSE(Data Science)
Sreyas Institute of
Engineering and Technology
Hyderabad
ankithathanneru08@gmailcom

[3]
Sarabu Aishu
UG Scholar
Dept of CSE(Data Science)
Sreyas Institute of
Engineering and Technology
Hyderabad
sarabuaishu23@gmail.com

[4]
AmishaSolanki
UG Scholar
Dept of CSE(Data Science)
Sreyas Institute of Engineering and Technology
Hyderabad
amishasolunki@gmail.com

[5]
ChittimillaTanmai
UG Scholar
Dept of CSE(Data Science)
Sreyas Instituteof EngineeringandTechnology
Hyderabad
tanmai.chittimilla@gmail.com

Abstract: Personality traits play a pivotal role in determining professional performance, influencing hiring decisions, and aligning candidates with organizational culture. Traditional CV analysis primarily focuses on structured data, such as education, skills, and work experience, often neglecting deeper psychological insights. This study proposes a machine learning-based framework for personality prediction from CV data, aimed at enhancing recruitment outcomes. By leveraging Natural Language Processing (NLP) techniques, unstructured text from CV is analyzed to infer personality traits based on established psychological models such as the Big Five (OCEAN). We employ supervised learning models trained on annotated data sets that link textual features to personality profiles. Feature extraction methods, including semantic embedding and linguistic patterns, are coupled with explainable machine learning models to ensure transparency in predictions. Experimental results demonstrate improved accuracy in predicting job fit and potential performance when personality predictions are integrated into the analysis pipeline.

Keywords: *Personality Prediction, CV Analysis, Machine Learning, Natural Language Processing, Big Five Personality Model(OCEAN), Psychometric Analysis, Recruitment Process*

I.INTRODUCTION In modern recruitment processes, identifying candidates who align with organizational goals and possess the right mix of technical skills and personality traits is essential. Traditional CV analysis methods focus on explicit attributes such as education, skills, and experience but often fail to account for implicit psychological characteristics like personality, which can significantly impact job performance and cultural fit. Integrating personality prediction into CV analysis can offer a more holistic evaluation, enhancing recruitment outcomes.Machine learning (ML), particularly ensemble methods such as the Random Forest classifier, provides powerful tools to bridge this gap. XG Boost , Ada-boost known for its robustness and predictability, excels in handling highdimensional data, making it an ideal choice for analyzing complex CV data sets. By leveraging Natural Language Processing (NLP) techniques, this study employs Random Forest models to infer personality traits from CV text based on established psychological frameworks like the Big Five (OCEAN).The Random Forest algorithm is particularly suited for this task due to its ability to handle diverse features extracted from CV, including linguistic patterns, semantic embedding, and syntactic structures. Its ensemble nature ensures resilience to over fitting and provides feature importance scores, offering insights into the key textual attributes that influence personality predictions. This approach aids recruiters understanding both the model's decisions and the underlying personality traits of candidates.This paper examines the integration of Random Forest-based personality prediction into CV analysis workloads. We evaluate the algorithm’s performance in predicting personality traits and its impact on improving candidate ranking, role alignment, and decision-making accuracy. The findings

demonstrate that augmenting CV analysis with personality insights can significantly enhance recruitment efficiency and fairness

II . LITERATUREREVIEW

Machine learning has become an integral part of various domains, including personality prediction, which plays a significant role in CV analysis. Among the various algorithms available, extreme Gradient Boosting (XG Boost) has gained substantial attention due to its efficiency and accuracy. This literature review explores the use of XG Boost in personality prediction, emphasizing its applications in improving CV analysis outcomes. XG Boost is an optimized gradient-boosting algorithm designed for speed and performance. It is widely recognized for its scalability, ability to handle missing data, and robust performance in various machine learning competitions. The algorithm's use of regularization (L1 and L2), parallel processing, and tree pruning techniques ensures high accuracy and prevention of overfitting.

Personality prediction involves assessing individual traits based on data such as written text, social media activity, or physiological signals. In CV analysis, understanding a candidate's personality traits can help employers align their hiring decisions with organizational culture and role requirements. The Five-Factor Model (FFM) – comprising openness, conscientiousness, extraversion, agreeableness, and neuroticism – is commonly used in such predictions. Studies highlight the importance of feature engineering in achieving accurate personality predictions. XG Boost's ability to handle sparse and high-dimensional data makes it particularly effective for processing text-based features from CVs or social media. XG Boost consistently outperforms other algorithms, such as logistic regression, random forests, and support vector machines, in personality prediction tasks. Its ensemble nature combines the strengths of multiple weak learners, producing superior results.

XG Boost has been integrated into systems for parsing CVs, where it predicts personality traits based on textual data. This aids recruiters in shortlisting candidates who align with job profiles. By incorporating XG Boost, CV analysis systems can achieve higher accuracy in matching candidates to roles, factoring in both technical skills and personality traits. The algorithm's capacity to identify subtle patterns helps reduce biases commonly present in traditional screening methods. Several studies have benchmarked XG Boost against other machine learning algorithms in personality prediction. For example, a comparison of XG Boost, SVM, and Random Forests on text-based personality datasets demonstrated that XG Boost achieved the highest F1 scores and accuracy. Another study applied XG Boost to analyse LinkedIn profiles, finding that it outperformed deep learning models in predicting conscientiousness and extraversion.

III . METHODOLOGY

His study develops a framework for predicting personality traits from CV using the Random Forest algorithm and Natural Language Processing (NLP) techniques. The methodology is designed to analyze textual data extracted from CV, infer personality traits based on the Big Five (OCEAN) model, and evaluate the impact of these predictions on recruitment outcomes. The methodology is divided into five key stages: data collection, data prepossessing, feature extraction, model training and evaluation, and integration into CV analysis workloads.

- **1. Objective**

We aim to predict personality traits from CV using XG Boost,a machine learning algorithm, to improve the accuracy of candidate analysis in recruitment.

- **2. Data Collection**

Source: CV from online platforms or recruitment agencies, paired with personality traits from psychometric tests or surveys.
Size: For example, 10,000 CV with labeled personality traits. Privacy: Data is anonymize to protect personal information.

- **3. Data Preparation**

Cleaning: Remove irrelevant content like addresses or unnecessary formatting.
Feature Extraction:
Convert text into numbers using techniques like TF-IDF or word embedding (e.g., Word2Vec).
Identify patterns such as keywords (e.g., "teamwork") or writing style (e.g., sentence length, sentiment).

- **4 . Model Design**

Algorithm: Use XGBoost because it handles large datasets and complex patterns effectively.
Output:
Predict traits like Openness, Conscientiousness, Extra version, Agreeableness, and Neurotic ism.
Each trait can be classified as high or low, or on a scale.

- **5. Model Training**

Split the data: 80% for training the model, 20% for testing it. Adjust settings (like learning speed and tree size) to get the best results.
Measure success using metrics like accuracy and precision.

- **6. Testing and Evaluation**

Check how well the model predicts personality traits using:
Accuracy scores (how often it's correct).
Feature importance charts (showing which parts of the CV mattered most).
Compare XG Boost to other methods like Random Forests or Logistic Regression.

- **7. Practical Application**

Automation: The model can screen CV quickly and predict personality traits for better candidate matching. Capability: It works efficiently even with thousands of CV.

Algorithm

1. Define the Problem

- Objective: Use machine learning to predict personality traits from CV data to enhance job matching.

2.Data Collection

- Gather CV paired with personality assessment data (e.g., Big Five traits).
- Include job role descriptions for mapping traits to roles.

3. Data Prepossessing

- Clean CV text: remove noise, tokenize, and lemmatize.
- Encode personality traits as labels (e.g., "High", "Medium", "Low").
- Extract textual features (e.g., TF-IDF, word embedding).

4. Model Development

- Use classifiers like XG Boost
- Train and evaluate with metrics like accuracy and F1-score.

5. Personality-Job Mapping

- Map predicted traits to job roles based on known preferences (e.g., Extroversion → Sales).

6. System Integration

- Build a pipeline: CV input → Trait prediction → Job role match → CV enhancement suggestions.

7. Evaluation

- Test the system's performance using real-world recruitment success rates and user feedback.

8. Deployment

- Deploy as a web tool or API for recruiters and job seekers.

Data set collection:

TrainingDatasetCV

Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
Male	17	7	4	7	3	2	extraverted
Male	19	4	5	4	6	6	serious
Female	18	7	6	4	5	5	dependable
Female	22	5	6	7	4	3	extraverted
Female	19	7	4	6	5	4	lively
Male	18	5	7	7	6	4	lively
Female	17	5	6	5	7	4	extraverted
Female	19	6	6	7	5	4	extraverted
Male	18	5	7	5	6	7	dependable
Female	19	5	5	7	4	5	lively
Male	19	6	7	5	6	3	serious
Male	19	7	6	7	7	6	extraverted
Male	19	7	6	6	5	6	lively
Female	19	6	7	5	5	5	dependable
Female	19	5	5	4	5	4	responsible
Male	19	5	6	4	6	3	extraverted
Female	19	7	7	2	6	5	serious
Female	18	6	7	4	4	2	dependable
Female	19	6	6	6	4	3	responsible
Female	19	5	6	3	3	3	extraverted
Female	19	6	4	6	3	4	responsible
Male	18	4	5	4	3	6	extraverted
Female	19	5	4	5	5	3	responsible
Male	20	5	3	3	4	4	serious
Female	19	6	7	5	5	4	serious
Female	19	7	5	6	6	5	dependable
Female	19	7	5	7	4	6	lively
Male	18	7	5	5	4	3	lively
Female	17	5	5	4	6	4	extraverted
Female	18	6	6	5	4	4	dependable
Female	19	5	6	5	4	7	lively
Male	19	5	4	6	1	3	lively
Male	19	3	7	2	6	7	responsible
Female	18	7	6	5	5	5	extraverted
Female	19	6	5	4	5	2	responsible
Male	19	6	6	2	5	6	dependable
Female	18	7	5	5	4	5	serious
Male	19	6	6	5	5	4	serious
Female	17	6	5	5	4	4	responsible
Female	19	5	6	5	6	4	dependable
Male	19	6	5	5	6	6	responsible
Female	19	5	5	5	7	3	dependable
Male	18	6	6	4	4	3	extraverted
Male	18	7	6	6	5	7	extraverted
Female	17	5	5	5	5	3	responsible
Female	19	5	5	4	4	4	dependable
Female	18	5	4	5	6	4	responsible
Female	18	7	4	6	6	4	serious
Female	17	4	6	4	4	1	extraverted
Female	17	7	4	5	5	5	dependable
Female	19	6	4	5	5	6	lively
Female	18	5	1	3	5	2	lively
Male	18	3	6	3	3	2	extraverted

Data Processing

Data prepossessing is a critical step in preparing CV data for personality prediction using machine learning. It begins with cleaning the text by removing noise such as special characters, HTML tags, and redundant spaces, followed by standardizing the case and correcting spelling errors. Token Ionization splits the text into words or sentences, while stop words are removed to retain meaningful terms. Techniques like systematization and stemming are applied to normalize words to their base forms. Features are extracted using methods such as Bag of Words, TF- IDF, or advanced embedding like Word2Vec or BERT to capture semantic meaning. Domain-specific terms are handled through custom dictionaries or Named Entity Recognition. Missing or imbalanced data is addressed through imputation, oversampling, or class weighting. Dimensional reduction (e.g., PCA) and scaling ensure compact, normalized

data representation. Finally, data is split into training, validation, and test sets, maintaining class distributions, and ethical considerations like bias detection and optimization are incorporated to ensure robust and fair prepossessing.

IV. RESULTSANDDISCUSSION

The system leverages XGBoost to predict personality traits from CV data with 90% accuracy, aided by preprocessing and feature extraction through ResumeParser. Its user-friendly interface facilitates seamless candidate evaluation, showcasing potential for recruitment automation. Future improvements, such as larger datasets and advanced features, can enhance its accuracy and scalability.

Model	Accuracy(%)	MSE
Random Forest	88.5	0.115
XGBoost	90.25	0.0975
AdaBoost	86	0.14
Gradient Descent(LogisticRegression)	84.75	0.1525

.

CONCLUSION

The XGBoost algorithm demonstrates significant potential in personality prediction for CV analysis within machine learning. Its strengths lie in handling large datasets, managing missing values, and delivering robust gradient boosting performance. By utilizing its ensemble learning capabilities, XGBoost enhances prediction accuracy and model interpretability, offering better insights into personality traits from CV data. Overall, applying XGBoost can greatly improve the effectiveness of personality prediction models.

V.REFERENCES

[1] “Personality Prediction Through CV ” Sham Dhanke , Sakshi Dhepe , Manthan Dave , Shantanu Inamdar , 2023 .

[2] “Personality Evaluation through CV analysis using machine learning algorithm” Prof . AN Mandale , suraj Mali , Prathik Sabale , Akash Patil , 2022.

[3] “Personality Prediction Through CV analysis ” Nuthalapati Leena , Normala Sandhya , Reddem Sai Archana Reddy , Shaik kasim Saheb , 2022 .

[4] “Personality Prediction system through CV analysis” Allan Robey , Kaushik Shukla Kahish Agarwal , keval Joshi ,
Proffesor Shalmali Joshi , 2019 .