AI in Recruitment: Personality Prediction through CV Analysis for Smarter Hiring

N.Bhoomika Sai[1], D. Harshita[2], K.Chandra Sekhar[3], V. Ramyasree[4], P. Srinivasa Rao[5]
[1,2,3,4,5]*MVGR College of Engineering, Vizianagaram, India*

***Abstract*—The personality of an individual plays an important role in driving organizational development and improving their self-progress. Identifying a person's personality traits and professional suitability has traditionally been achieved through interviews or by analyzing their Curriculum Vitae (CV) manually. However, these traditional recruitment methods, which involves shortlisting the candidates manually based on specific requirements of the company, are often time-intensive and prone to bias. This system is designed as an online platform that facilitates candidate registration and conducts personality assessments using Multiple-Choice Question (MCQ) tests. These personality quizzes aim to provide insights into the behavioral and psychological traits of candidates. Simultaneously, the system evaluates professional eligibility by comparing uploaded CVs against a dataset trained using various ML algorithms. The key algorithms involved are Logistic Regression, Random Forest, Support Vector Machine and Decision Tree which were chosen for its strong decision-making capabilities and effectiveness in classification tasks. This ensures that recruitment decisions are data-driven and unbiased.**

***Index Terms*—Personality Prediction, Machine Learning Models, Big Five Model (OCEAN), Term Frequency Inverse Document Frequency (TF-IDF), Natural Language Processing (NLP).**

I. INTRODUCTION

Recruitment is a crucial process for organizations to identify and familiarize the best talent suited for their needs. When a particular role is chosen, shortlisting the right candidate for the company from a pool of candidates competing for the role has become a major issue[1][2].When there is a vacancy for a particular job role in the company, they handle many number job requests and it will be a time taking process for them to go through all those CV's[3][4].The traditional recruitment process involved time-intensive methods such as manually going through resumes, conducting multiple rounds of in-person interviews, and relying on subjective evaluations of candidate's qualifications and personalities. These processes were not only labor-intensive but also prone to human biases, inefficiencies, and inaccuracies [5][6].

When a job announcement by an organization is done, it results in numerous applications collected through diverse recruitment sources. The HR department of the organization struggles to handle resume screening and candidate selection through traditional recruitment practices like interviews, discussions, and specialized testing. During the initial phase, they assess the applicants across multiple criteria considering their eligibility for the role, their ability to manage the role and their skills. To address this issue in the process, we came up with an idea reducing the complexity of process for the organizations by extracting candidate attributes through CV's and basic personality tests [7][8].

The advent of e-recruitment systems has introduced features like automated resume parsing, online aptitude and personality assessments, and machine learning-driven candidate shortlisting. These systems enable recruiters to handle a larger volume of applicants efficiently while improving the accuracy and fairness of candidate evaluations. Personality is a very important factor to be considered while selecting an individual for a particular role. Judging a person's capability and their mindset is an utmost difficult thing to do but we came up with an idea that will evaluate the personality of an individual and also provide recommendation for selecting the particular candidate [9][10]. The primary approach of our study is to develop a system ensuring a fair candidate selection. By replacing time-consuming traditional practices

with automated workflows, this project demonstrates how modern e-recruitment systems can improve the efficiency and reduce biases.

II. LITERATURE SURVEY

Some researchers are working on various machine learning models to analyze an individual's personality traits and make use of them in the recruitment process. Xinyu Shui., et al. [11][12], aimed to determine the individual's Big-Five personality traits by physiological signals. A bracelet was attached to the people which tracks their heart rate and their daily activities were observed. They used linear regression analysis to predict an individual scores. Their findings shows that individuals show a good response in multiple situations where they are dependent on people rather than in single faced situation.

Nandani Agarwal., et al. [13][14], developed an automated system that predicts personality traits from resumes using various ML algorithms like Decision Tree, Random Forest, k nearest neighbor and Naive Bayes for effectively predicting personality through CV (Curriculum Vitae) analysis. Big five personality models are used to analyze personality.

Jayashree Rout., et al. [15][16], introduced a model which checks the candidate's personality and aptitude scores. They used Five Factor Model to predict the personality and TF-IDF to identify the significant text from a document. This helped the recruiters to make fair decisions based on the test scores.

Lakshyajit Thapa., et al. [17][18], addressed a system for predicting the personality which helped the automated system. This framework collected the data of the user from various platforms like CV's, social media and their responses to the tests conducted and predicted their traits by using Big Five, Myers Briggs and HEXACO personality models. They also compared various ML algorithms to develop models which can predict and help the recruiters to select the candidate similar to their job specifications.

G.Sudha.,Sasipriya. et al. [19][20], developed an online platform for registering the candidate details and analysing their personality through an online test containing MCQ type personality quizzes. Then their CV is matched to the job requirements by using a trained dataset which is trained by Logistic Regression. This platform helped with better hiring satisfaction for the recruiters.

Atharva Kulkarni., et al. [21][22], predicted the personality based on the score obtained from labels of Big Five model. This approach parsed the entire CV and looks for texts matching to OCEAN model. In order to identify the text NLP. NLTK, SpaCy were used and for predicting KNN, SVM, Naïve Bayes and Random Forest algorithms were used which resulted in low accuracy.

Muthu Selvi M., Angeline Ranjitha Mani., et al. [23][24], designed a framework which reduces the workload on humans. This framework helped the recruiters to identify suitable candidate for a required role. The administrator can easily select a candidate based on the test results by using various clustering techniques without even an option to go through the CV.

Amal Khalifa Al Aamer., et al. [25], study found that AI technology can efficiently improve the process of hiring with many advantages considering limited costs, enhanced accuracy and reduced human work for scrutinizing the candidates.

Anum Jaffar., et al. [26], proposed system that determines the traits using nonverbal movements like facial expressions, head and body postures. Few questions with expert analysis and a dataset which recognizes the facial movements is trained using CNN.

Madhura Jayaratne, Buddhi Jayatilleke.,et al. [27], gathered data from 46,000 applicants who attempted an online interview through chat which also included personality questions based on HEXACO model. Using this HEXACO and Random Forest, they built a model that predicts the personality.

Hans Christian., et al. [28], commenced a DL model with BERT along with NLP features considering sentiment analysis. This model combined multiple online sources to increase the dataset and achieved a maximum accuracy of 86.2% on Facebook, 88.5% on Twitter data.

Amol Chincholkar.,et al. [29], automatically classified the personality based on the dataset given at backend using machine learning model,Python,NLP which helped the recruiters to shortlist the correct individual for a given position within less time for a company.

Darpan Jain.,Rajesh Kumar., et al. [30], examined the significance of various machine learning algorithms used for predicting the personality traits from candidates' CVs. This study also showed the importance of personality assessment in the recruitment process and evaluated various ML models to discover the productive approach for predicting personality to improve the accuracy.

Yash Mor and Rupali Sawanth [31] proposed a system that ranked the suitable candidates for a particular job role based on tests taken. This system used the OCEAN Model to predict the personality. Initially, CVs of specified candidates whose skills match with the job role were shortlisted using machine learning and this was followed by ranking them based on their psychometric test scores. This system achieved to provide a legal and justified CV ranking system.

Mihir Satra.,Faisal Mungi.,et al. [32], developed a system where the candidates were asked questions related to CV and personality. BERT, machine leaning and NLP techniques are used to predict personality and Logistic Regression with 85.71% accuracy is achieved.

The studies heavily rely on structured CV data, which may not fully capture a candidate's personality. The accuracy of these predictions mainly depends on the quality and diversity of training datasets, often limiting generalizability.So by overcoming all those,our methodology enhances personality prediction by incorporating a more diverse and extensive dataset, reducing biases through improved feature selection and preprocessing techniques. We utilize a refined TF-IDF-based text analysis approach combined with multiple machine learning models to ensure more accurate and interpretable results. Additionally, our system integrates a standardized assessment framework to validate predictions against real-world recruitment outcomes.The proposed methodology and the system model is explained below in detail.

III. METHODOLOGY

This article involves a systematic procedure integrating NLP and various ML techniques to extract information from CV and analyze personality traits. Multiple ML models are trained and deployed using Flask within an E-Recruitment System allowing users to upload CV's and take tests.

Here, this approach automates candidate shortlisting, making hiring more efficient, unbiased and data-driven.

The methodology develops a machine learning model and UI design for interaction. Our proposed system comprises two modules. They are:

A. Admin Module

1. Admin has an authorized login.
2. Manage aptitude and personality questions.
3. Add job details and requirements.
4. View details of the user test results,CVs and the shortlisted candidate.

B. User Module

1. Register to the system.
2. Upload CV.
3. Attend the aptitude and personality tests.
4. View the test score.

C. CV Analysis

1. Preprocess the CV text and apply the trained model which predicts the personality, match the job requirements and shortlist the candidates.

We have used different ML algorithms like Logistic Regression, Random Forest, Decision tree and Support Vector Machine for personality prediction of a candidate.

- Logistic Regression - A classification algorithm used efficiently for linearly separable data which works by estimating probabilities using the sigmoid function and applying a decision threshold.
- Random Forest - An ensemble algorithm which integrates number of decision trees to enhance the accuracy. It can handle large datasets and missing data efficiently which reduces the overfitting as in decision trees and results in averaging predictions.

- Decision Tree - A tree-based model which makes decisions by splitting the data based on feature importance. It is easy and simple to understand.
- Support Vector Machine (SVM) - A classification algorithm which maximizes the margin and finds the best fit hyperplane to separate classes and also uses kernel tricks for non-linearly separable data.This is effective for high-dimensional spaces.

The algorithms are evaluated using some metrics which determines the effectiveness of them.

Evaluation Metrics

1. Accuracy: It refers the quality of the model being correct.
2. Precision: It concentrates on the quality of model's positive predictions showing how many of the predicted positive instances were actually positive.
3. Recall: It evaluates the model ability to identify all instances correctly.
4. F1 Score: It is the harmonic mean of precision and recall.

The clear view of our idea is written in an algorithmic form and the steps to be followed.

3.1 Algorithm: JobFit AI- Intelligent Candidate Evaluation Algorithm

START
INITIALIZE system and database connection
LOAD ML model
DISPLAY login options
GET user input
IF user is Admin, THEN
Grant access to dashboard
WHILE logged in DO
Manage assessments
Modify job details and CVs
Shortlist CVs
View details
END WHILE
LOG out
ELSE IF user is User, THEN
WHILE session active DO
Attend tests
Upload CV
GET test results
END WHILE
LOG out
Process CV using NLP
Analyze responses using ML model
PREDICT personality traits
Match job requirements
STORE Results
DISPLAY Shortlisted Candidates
LOG interactions
END

The algorithm outlines how an AI-driven recruitment system starts by initializing the system and database connection, followed by loading the ML models. Users can log in as either Admin or Candidate where admin gains access to a dashboard where they can manage questions, modify job details, oversee assessments, and shortlist candidates. Candidates can upload their CVs, take aptitude/personality tests and receive test results.After logout, the system processes the CV using NLP techniques, analyzes test responses with ML models, predicts personality traits, and matches candidates with job requirements. Finally, it stores results, displays shortlisted candidates, and logs interactions for future reference. The system model for the above algorithm is as follows.

3.2 System Model

This system model represents an AI-driven recruitment framework designed to assess candidates based on personality traits and job suitability. The process involves multiple stages, where a candidate submits their CV and completes an assessment test. The system processes the responses using machine learning (ML) models to extract key personality attributes and map them with job requirements. The analyzed data is then reviewed by recruiters or HR personnel, who use the insights to make informed hiring decisions.It is constructed on the process depicted in Figure 1.

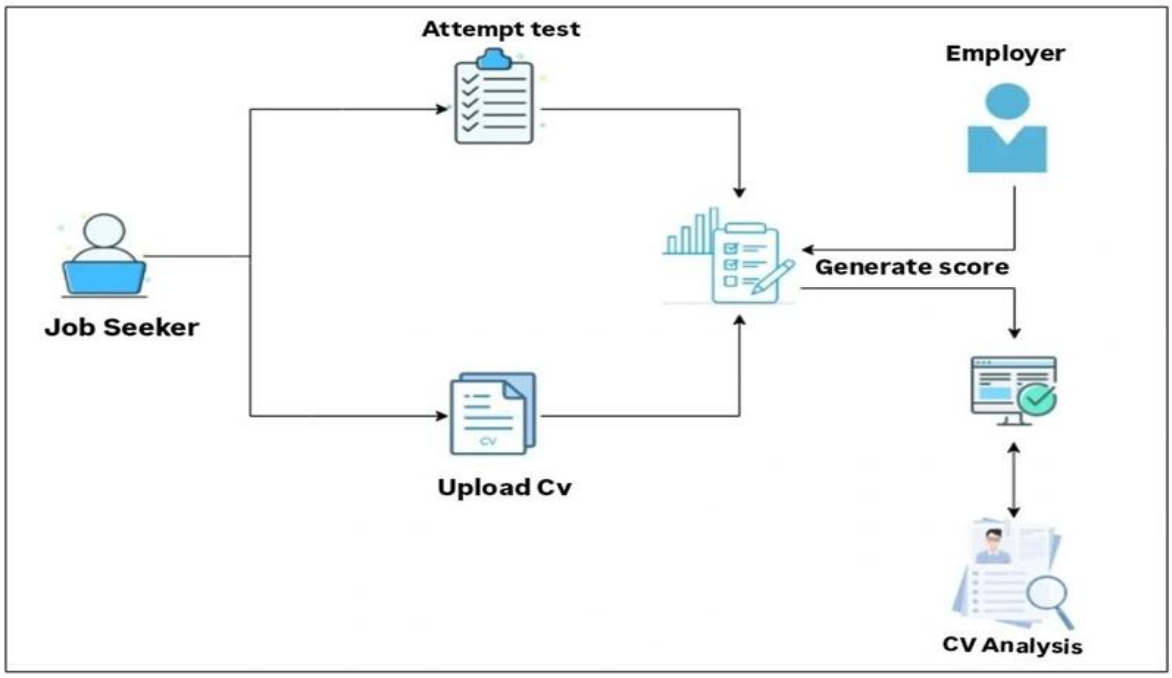


Figure1: System Model

The above figure illustrates the workflow of an AI-driven recruitment system, where candidates upload their CVs and take assessments. The system processes the CV data using ML models and NLP techniques to predict the respective personality traits and match candidates to suitable job roles.

To implement all these steps,there is a need for an environment where all this work can be done and accessed for use.

IV. ENVIRONMENTAL SETUP

This Personality Prediction System was designed incorporating both software and hardware components. It is compatible with Windows, Linux, and macOS, utilizing Python 3.12.0 for backend development with Flask for API handling, while HTML/CSS/JavaScript manage the frontend. The system employs MongoDB for database management and Tornado server for server-side logic. Essential libraries include Scikit-learn, Numpy, Pandas for machine learning, NLTK for natural language processing, and TF-IDF for CV data extraction. The system demands a processor (Intel Core i7 or AMD Ryzen 7), 8GB+ RAM, SSD storage (500GB+). This can be accessed by any devices like computer,phone,laptop that have a stable connection.This is created to make the task of recruiters easy.

V. RESULTS & DISCUSSION

We experimented with several ML algorithms like Logistic Regression,Support Vector Machine, Random Forest and Decision Tree. After training the model and uploading of CV, we observed that Logistic Regression and Random Forests have the same accuracy. We handpicked the best two models and the class-wise performance metrics for Logistic Regression and Support Vector Machine are displayed in Table 1 and Table 2.

Table 1: Logistic Regression Metrics Accuracy - 96.36%

Class	Precision	Recall	F1-score
0	0.99	0.99	0.96
1	0.99	0.95	0.97
2	0.90	0.99	0.94
4	0.99	1.00	0.99
5	0.97	0.96	0.96
Macro Average	0.97	0.96	0.96
Weighted Average	0.97	0.96	0.96

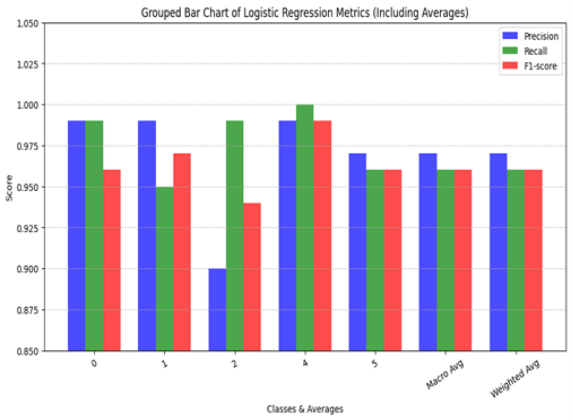


Figure 2: Bar Graph of Logistic Regression Metrics

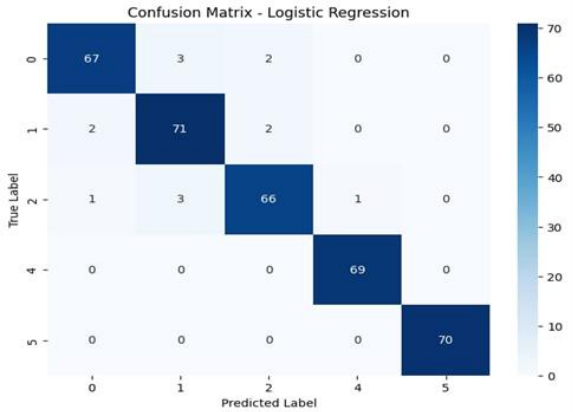


Figure 3: Logistic Confusion Matrix

Figure 2 represents the grouped bar chart of the Logistic Regression model illustrating the precision,recall and F1-score where it compares classification performance across different personality classes. The chart highlights the model's strong performance showing consistently high scores across all metrics.Figure 3 represents the confusion matrix of Logistic Regression visualizing the performance of model and no of correctly and incorrectly classified personality classes.

Table 2: Support Vector Machine Metrics

Accuracy - 96.08%

Class	Precision	Recall	F1-score
0	1.00	0.93	0.96
1	0.95	0.95	0.95
2	0.96	0.93	0.94
4	0.99	1.00	0.99
5	0.92	1.00	0.96
Macro Average	0.96	0.96	0.96
	0.96	0.96	0.96

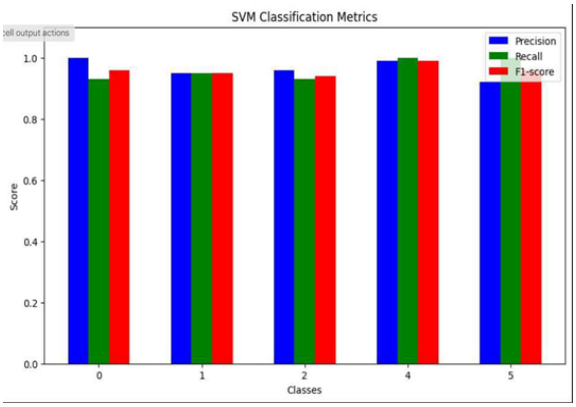
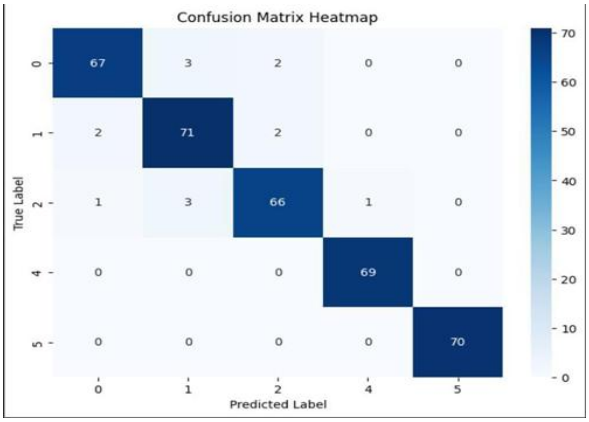


Figure 3: Bar Graph of SVMFigure



4: Confusion Matrix of SVM

Figure 3 represents the grouped bar chart of SVM model illustrating precision, recall and F1-score for various personality labels predicted using the SVM model. Figure 4 depicts the confusion matrix visualizing the performance of SVM model for personality predicting. The elements which are present diagonally represent correctly classified instances while remaining elements indicates misclassifications. The high values along the diagonal signifies strong classification accuracy.

By observing all the resultant bar charts and table values,the comparison between both the algorithms can be tabulated as shown in Table 3 and can be visualized in Figure 5 and Figure 6.

Table 3 Performance Metrics

Model	Accuracy	Precision	F1-score	Recall
Logistic Regression	0.9636	0.97	0.96	0.96
Support Vector Machine	0.9608	0.96	0.96	0.96

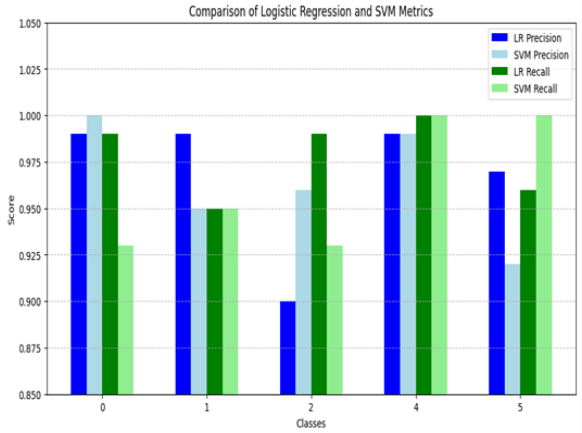


Figure 5: Bar Graph of Logistic and SVM models

6: Line Chart of Both Models

Figure 5 shows the bar graph obtained after comparing both Logistic Regression and SVM models with precision and recall scores highlighting the best model. Figure 6 represents a line chart which presents the overall comparison of accuracies of Logistic Regression and SVM models considering 20 samples showing a variation in accuracy between the two classifiers indicating Logistic has a highest accuracy when compared.

Through this paper, it is observed that both Logistic Regression and Support Vector Machine (SVM) models performed well for predicting the personality traits from CV text data. Logistic Regression attained highest accuracy of 96.36% while SVM closely followed. The high values for precision, recall and F1-scores suggest that models effectively capture personality traits from CV. Considerably, Class 4(Extraverted) achieved a perfect recall of 1.00 in both models indicating strong recognition.

VI. CONCLUSION AND FUTURE SCOPE

This paper successfully demonstrated the working of an E-Recruitment System in helping the recruiters to shortlist the candidates required.It has concluded that ML models can effectively predict the personality from CV text. These models can also be useful in automated recruitment, candidate profiling and personality assessment, reducing bias and improving hiring efficiency. It also requires shortspan of hiring process reducing the manual work.

Future work can also be done by enhanced improvements, expanded scope and collaborations for better solutions in the modern recruitment world.We can add additional personality traits than those used in this paper.Multi-model approach and the expanded dataset can be included in the future.The advancement in this system can ensure privacy and ethical considerations and can also be deployed into real-world to automate personality analysis for hiring candidates.

REFERENCES

[1] Mrs.Vijayanirmala B., Spoorthi D., Varsha B H., Madhialagi M., Meena G.," PERSONALITY PREDICTION SYSTEM"., IJARIIE-ISSN(O)-2395-4396., Vol-10 Issue-2 2024.

[2] T.V. Madhusudhana Rao, Suresh Kurumalla, Bethapudi Prakash, "Matrix Factorization Based Recommendation System using Hybrid Optimization Technique, EAI Endorsed Transactions on Energy Web,Volume:5,issue:35,2021.

[3] Nuthalapati Leena., Normala Sandhya., Reddem Sai Archana Reddy., Shaik Kasim Saheb.,"PERSONALITY PREDICTION THROUGH CV ANALYSIS".,International Research Journal of Modernization in Engineering Technology and Science., Volume:04., Issue:07., July-2022.

[4] T.V. Madhusudhana Rao,P.S. Latha Kalyampudi, "Iridology based Vital Organs Malfunctioning identification using Machine learning Techniques", International Journal of Advanced Science and Technology, Volume: 29, No. 5, PP: 5544 – 5554,2020.

[5] Riya Dagli.,Muskaan Shaikh.,Swapnil Bhadre.,Prof.Suchetadevi

Gaikwad.,"Personality Prediction and CV Analysis System".,International Research Journal of Engineering and Technology (IRJET)., Vol:09, Issue:03, March 2022.

[6] S.Vidya sagar Appaji, P. V. Lakshmi, "Maximizing Joint Probability in Visual Question Answering Models", International Journal of Advanced Science and Technology Vol. 29, No. 3, pp. 3914 – 3923,2020.

[7] Rishi Jain, Krishna Jaiswal, Pratham Jain, Tanay Jain, Jainivas Anandhan, Jainil Vaidya.,"THE SCIENCE OF HIRING : Predicting Personality through CV Parsing & Machine Learning".,INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT 07(11):1-11.,November 2023.

[8] Vidya sagar Appaji setti, P. V. Lakshmi, "A Novel Scheme for Red Eye Removal with Image Matching", Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 13-Special Issue, 2018.

[9] Rutuja Narwade, Srujami Palkar, Isha Zade Nidhi Sanghavi.,"Personality Prediction with CV Analysis".,International Research Journal of Engineering and Technology (IRJET)., Volume: 09, Issue: 04, Apr 2022.

[10] Madhusudhana Rao, T.V. Srinivas, Y, "A Secure Framework for Cloud Using Map Reduce",Journal of Advanced Research In Dynamical and Control Systems (IJARDCS), Volume:9, Sp-14, Pp:1850-1861, ISSN:1943-023x, Dec, 2017.

[11] Xinyu Shui.,Yiling Chen.,Xin Hu., "Personality in Daily Life: Multi-Situational Physiological Signals Reflect Big-Five Personality Traits" , IEEE Journal of Biomedical and Health Informatics , Volume: 27, Issue: 6 ,2023.

[12] Lolla Kiran Kumar, S Sreenivasa Rao," A Framework for Early Recognition of Alzheimer's Using Machine Learning Approaches", Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore,PP: 1-13,28/10/2022.

[13] Nandini Agarwal.,Justina.,M.J.Akshaya.,Nihar Shetty.,"Machine Learning driven Personality Prediction System using the concept of CV Analysis" ., International Conference on Electrical Electronics and Computing Technologies (ICEECT) ,Conference Paper.,2024.

[14] Bharathi Uppalapati, S Srinivasa Rao," Application of ANN Combined with Machine Learning for Early Recognition of Parkinson's Disease", Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore,PP: 39-49, 28/10/2022.

[15] Lakshyajit Thapa; Aniket Pandey; Deepanshu Gupta; Ayush Deep; Rakesh Garg.,A Framework for Personality Prediction for E-Recruitment Using Machine Learning Algorithms., 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence).,21 March 2024.

[16] Palli, P., Mishra, S.," Inferring Compound Similarity: A Clustering Approach in Drug Discovery" 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing, IC-CGU 2024.

[17] Jayashree Rout.,Sudhir Bagade.,Pooja Yede.,Nirmiti., "Personality Evaluation and CV Analysis using Machine Learning Algorithm"., International Journal of Computer Science and Engineering., Volume. -7., Issue-5., May 2019.

[18] Krishna Prasad, M.H.M., Thammi Reddy, K, "A Efficient Data Integration Framework in Hadoop Using MapReduce" Published in Computational Intelligence Techniques for Comparative Genomics, Springer Briefs in Applied Sciences and Technology, ISSN:2191-530X, PP 129-137, October 2014.

[19] G.Sudha.,Sasipriya., Sri Janani S.,Nivethitha D.,Saranya S., Karthick Thyagesh G., "Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process"., International Conference on Computing and Communications Technologies (ICCCT)., Conference Paper 2021.

[20] Nagesh Vadaparhi, Srinivas Yarramalle, "A Novel clustering approach using Hadoop Distributed Environment", Springer, (Applied Science and Technology), ISSN:2191-530X, Volume:9, pp:113-119, October 2014.

[21] Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat.,"Personality Prediction Via

CV Analysis using Machine Learning"., International Journal of Engineering Research &Technology (IJERT)., Volume 10, Issue 09.,September 2021.

[22] Balajee Maram, Guru Kesava Dasu Gopisetty, "A Framework for Data Security using Cryptography and Image Steganography",International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September,2019.

[23] Muthu selvi M., Angeline Ranjitha Mani.,"Personality Prediction System Through CV Analysis"., International Journal on Cybernetics & Informatics., Vol 10(02) ,2020.

[24] Suribabu Naick B,Prakash Bethapudi "Malware Detection in Android Mobile Devices by Applying Swarm Intelligence Optimization and Machine Learning for API Calls", International Journal of Intelligent Systems and Applications (IJISAE), Volume:10, issue:3s, 27 Dec2022.

[25] Amal Khalifa Al Aamer, Allam Hamdan & Zaher Abusaq.,"The Impact of Artificial Intelligence on the Human Resource Industry and the Process of Recruitment and Selection International Conference on Business and Technology"., Conference paper., March 2023., pp 622–630.

[26] Anum Jaffar; Sara Ali; Khawaja Fahad Iqbal; Yasar Ayaz; Ali R. Ansari; Muhammad A. B. Fayyaz.,"A Comprehensive Multimodal Humanoid System for Personality Assessment Based on the Big Five Model".,IEEE Access.,Journal Article.,Volume: 12,2024.

[27] Madhura Jayaratne; Buddhi Jayatilleke.,"Predicting Personality Using Answers to Open-Ended Interview Questions".,IEEE Access.,Journal Article.,Volume: 8,2020.

[28] Hans Christian, Derwin Suhartono, Andry Chowanda & Kamal Z. Zamli.,"Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging".,Journal of Big Data., Article number: 68 (2021).

[29] Amol Chincholkar, Dipti Bhosale, Shivanjali Adsul, Anjali Bodkhe,Rutuja Kadam.,"A Comprehensive Survey on Personality Prediction Using Machine Learning Techniques".,International Journal of Advanced Research in Computer and Communication Engineering.,Vol. 12, Issue 11, November 2023.

[30] Nongmeikapam Thoiba Singh., Abhay Chanana., Darpan Jain.,Rajesh Kumar.,"Personality prediction through CV analysis using machine learning techniques".,Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)., January 2023.

[31] Yash Mor.,Rupali Sawanth.,"Personality Prediction Using Psychometric and CV Analysis".,Global Conference on Information Technologies and Communications (GCITC).,Conference Paper.,2023.

[32] Mihir Satra.,Faisal Mungi.,Jinit Punamiya.,Kavita Mahesh Kelka.,"Personality Prediction System to Improve Employee Recruitment".,Computational Intelligence for Modern Business Systems (pp.295-308).,November 2023.