iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70400

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Personality Prediction Using ML

Pratik Shinde[1], Tushar Malode[2], Abhishek Khemnar[3], Pruthviraj Dhale[4], Rupali R. Patil[5]
[1, 2, 3, 4]*Students, Dept. of Information Technology, SITRC, Nashik*
[5]*Assistant professor at SITRC, Nashik, Maharashtra, India*

Abstract: ***Data Science and AI are revolutionizing the planet through technical transformations. One of the most significant uses of machine learning is the classification of people based on their personality features. We can see many applications of machine learning in our daily lives. Every human on the world has distinct personality type. This is made by the availability of high dimensional data. Using the Big Five personality traits, we created a system for predicting personality. t. Daily numerous students are writing competitive exams that mainly focuses on his/her Personality. The main motive of these exams is to check the student's personality and skills. This project helps to write the personality test and check the personality of the person. From the personality classification, the person can view the type of personality and can improve the personality based upon the results***
Keywords: personality prediction, Machine Learning, K-means Clustering, Gaussian Mixture model Ocean Model

I. INTRODUCTION

The big five personality traits, often mentioned as OCEAN**. OCEAN** stand for Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. These five personality traits represent broad domains of human behaviour and account for differences in both personality and decision making. Today, the model is employed by HR practitioners to measure potential employees and marketers to understand the audiences of their products We are deploying OCEAN Model in this project for developing the algorithm

1) Openness to experience: Also called as intellect or imagination, this personality trait represents the willingness to try new things and think out of the box. This trait includes insightfulness, originality, and curiosity.
2) Conscientiousness: The desire to be careful, diligent and regulate immediate gratification with self-discipline. This trait includes ambition, discipline, consistency and reliability.
3) Extroversion: A state where an individual draw energy from others and seek social connections or interaction, as opposed to being alone. This trait includes being outgoing, energetic, and confident.
4) Agreeableness: The measure of how an individual interacts with other individuals, characterized by the degree of compassion and cooperation. This trait includes tactfulness, kindness, and loyalty.
5) Neuroticism: A tendency for negative personality traits, emotional instability, and self-destructive thinking. This trait includes pessimism, anxiety, insecurity and feature

Fig 1.1 OCEAN MODEL

II. LITERATURE SURVEY

1) Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi (2019) – Personality Prediction System through CV Analysis: A machine learning system analyses CVs using NLP and the OCEAN model to predict personality traits. Challenges include diversity in CV formats and data preprocessing.
2) Devesh Agarwal, M. Karthikeyan (2022) – Personality Prediction Using Machine Learning: Machine learning (Decision Trees, Random Forest) is used to predict personality traits from user data, with high accuracy, especially with Random Forest. Useful for recruitment.
3) Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat (2021) – Personality Prediction Via CV Analysis using Machine Learning: CV text is analysed using NLP and algorithms like Naive Bayes and SVM. TF-IDF feature extraction improves accuracy but is hindered by unstructured CVs.
4) M. Kalghatgi, M.Ramannavar, Dr. N. S. Sidnal (2015) – Neural Network Approach to Personality Prediction Based on the Big-Five Model: Neural networks outperform traditional ML methods in predicting personality traits using the Big-Five model, offering better accuracy and generalization.
5) Zubeda, M. Shaheen, G. Narsayya Godavari, S. Naseem – Resume Ranking Using NLP and Machine Learning: NLP and machine learning rank resumes based on content. Though not focused on personality prediction, the study shows the importance of structured data in recruitment.
6) Md Tanzim Reza, Md. Sakib Zaman – Analyzing CV/Resume Using NLP and Machine Learning: A system using NLP and ML to extract skills and personality traits from CVs for job compatibility. Preprocessing and feature extraction are key for accuracy.

III. PROPOSED MODEL

One of the major challenges for the project will be the collection of input datasets for the algorithm. For conducting the test, we are using K – Mean Clustering Algorithm. The dataset for testing the algorithm is collected from the participant. This is done by giving a questionnaire on personality classification. Then, the collected information is fed to the personality classification algorithm i.e., K-Mean Clustering Algorithm. Finally, the algorithm evaluates the data on the basis of the big five personality traits and displays the result. To give you an idea, here is a diagrammatical representation of the whole process we will follow for the conduction of the project.

A. Methodology

The OCEAN model, also known as the Big Five personality traits, is a widely accepted framework for understanding human personality. It comprises five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Given the rise of machine learning and artificial intelligence, these technologies can be used to predict personality traits based on data such as social media activity, text analysis, behavioral patterns, and more.

1) Data Collection: Collect data from Kaggle.com,Ensure that data collection complies with data privacy laws and obtain user consent where necessary.
2) Data Preprocessing: Remove noise, redundant information, and irrelevant data. Extract meaningful features from text data. Standardize data formats, normalize scales, and handle missing values
3) Feature Selection: Use domain knowledge and statistical methods to identify features that correlate with the OCEAN traits. Create new features that can help with personality prediction (e.g. selecting more appropriate questions).
4) Model Selection: Consider machine learning algorithms suitable for this task, such as K mean clustering, Gaussian Mixture Model algorithm. Use ensemble learning techniques to combine multiple models for improved accuracy and robustness.
5) Model Training and Validation: Use a labeled dataset (if available) to train models. Labeled datasets contain known personality traits for a set of individuals. Implement cross-validation techniques to ensure the models robustness and avoid overfitting. Optimize model parameters to improve performance.
6) Model Evaluation: Evaluate model performance using appropriate metrics such as accuracy, precision, recall, F1-score, or others relevant to the application. Ensure the model is interpretable to understand why it makes certain predictions. This architecture outlines a comprehensive approach to building a system for personality prediction based on the OCEAN model using machine learning and AI. The focus is on accuracy, ethical considerations, and respecting user privacy throughout the process.

B. Flowchart

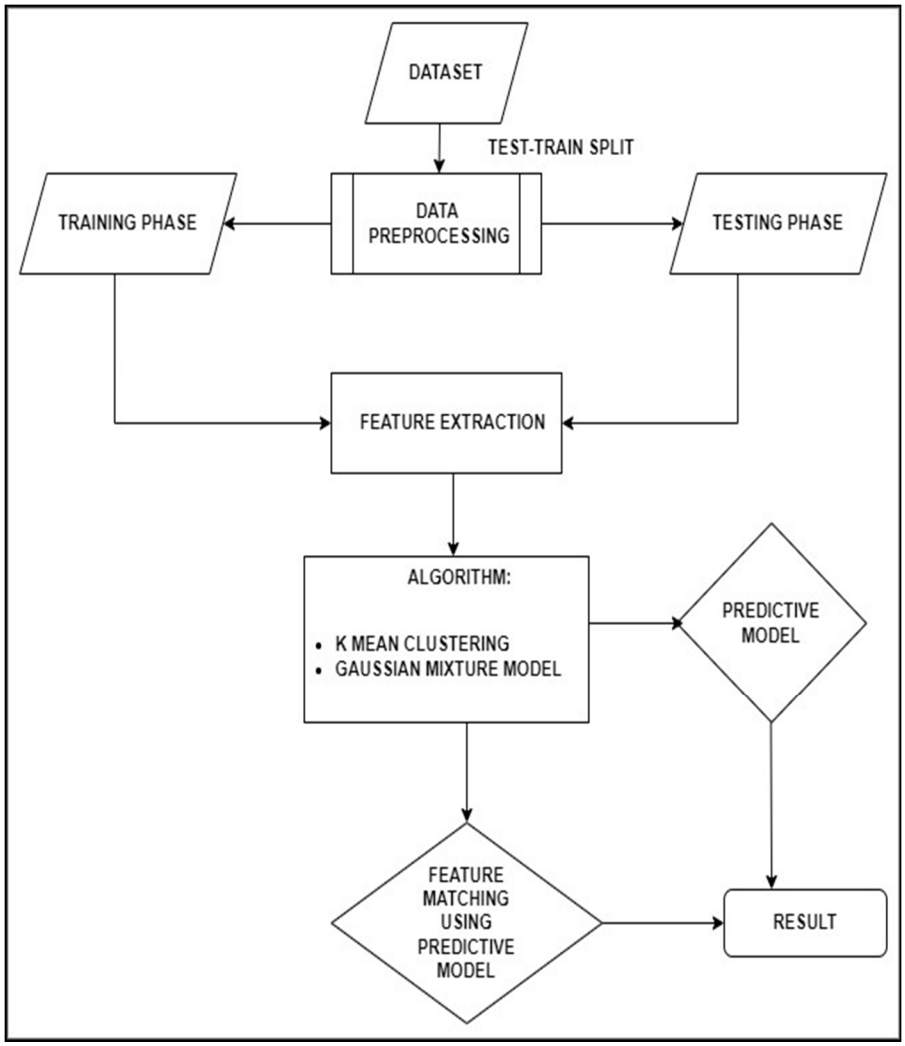


Fig 3.1 FLOW CHART

1) Dataset: The process begins with gathering a dataset that contains relevant information about individuals, such as their behaviours and preferences.
2) Test-Train Split: The dataset is divided into training and testing sets. This ensures that the model's performance can be evaluated on unseen data, preventing overfitting.
3) Data Preprocessing: Both training and testing data undergo preprocessing to handle missing values, outliers, and inconsistencies. This step also includes feature scaling and transformation for better model performance.
4) Feature Extraction: Important features are extracted from the data to simplify and optimize the input for machine learning algorithms.
5) Algorithm: The extracted features are processed using machine learning algorithms such as:
6) K-Means Clustering: Groups individuals into clusters based on similarities.
7) Gaussian Mixture Model (GMM): Provides probabilistic groupings, offering a more flexible clustering approach.
8) Predictive Model: The algorithms are used to train a predictive model, which identifies patterns and makes predictions based on the input data.
9) Feature Matching Using Predictive Model: The trained predictive model is applied to match features and generate output.
10) Result: The final output is the prediction, which represents personality traits or scores based on the OCEAN model.

IV. SYSTEM IMPLEMENTATION

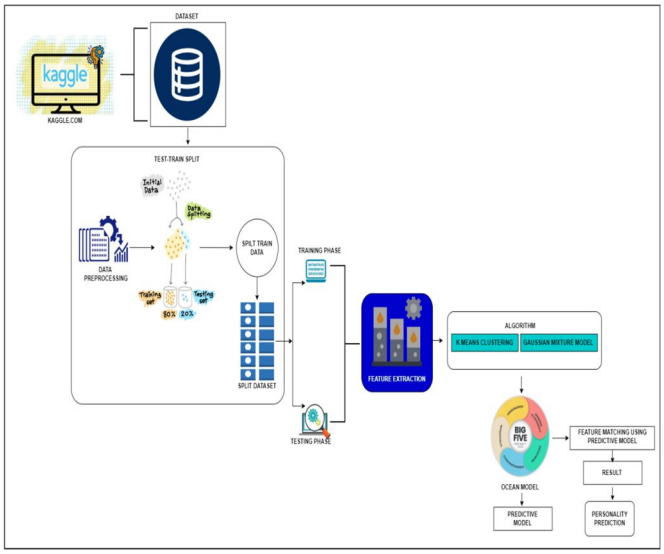


Fig 4.1 SYSTEM ARCHIITECTURE

A. *Dataset Acquisition*

The dataset is sourced from Kaggle, a platform known for providing high-quality data for machine learning projects. This ensures a diverse and representative dataset containing relevant attributes such as behaviours, preferences, and self-reported personality scores.

B. *Test/Train Split*

The dataset is divided into training and testing sets. This split is crucial for evaluating the model's performance objectively and avoiding overfitting. Typically, a common split ratio is 80% for training and 20% for testing.

C. *Preprocessing*

Data preprocessing involves cleaning the data to handle missing values and outliers. Normalization ensures that features are scaled to a similar range, which is vital for algorithms sensitive to magnitude.

Feature selection methods, such as correlation analysis or importance ranking, help in identifying the most influential features for personality prediction.

D. *Feature Extraction*

This process involves transforming raw data into meaningful input for machine learning models. Techniques like Principal Component Analysis (PCA) or clustering might be employed here to reduce dimensionality and capture the essence of the data.

E. *Algorithm Selection*

Two machine learning algorithms are employed:

K-Means Clustering: This unsupervised learning algorithm is useful for grouping individuals with similar traits.

Gaussian Mixture Model (GMM): This probabilistic model helps in identifying the distribution of traits across the dataset, potentially offering more nuanced groupings compared to K-Means.

F. *Predictive Models*

OCEAN Model: Maps personality traits to the five dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).

G. *Result Generation*

The final output is a set of predicted personality scores for an individual based on their behaviours and preferences.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

V. CLASS DIAGRAM

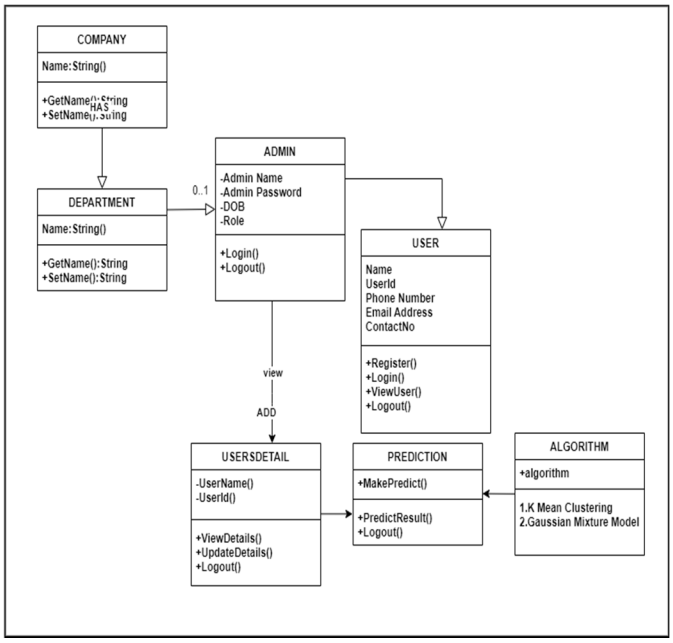


Fig 5.1CLASS DIAGRAM

A Class is a category or group of things that has similar attributes and common behaviour. A Rectangle is the icon that represents the class it is divided into three areas. The upper most area contains the name, the middle; area contains the attributes and the lowest areas show the operations. Class diagrams provide the representation that developers work from. Class diagrams help on the analysis side, too Class is nothing but a structure that contains both variables and methods. The Class Diagram shows a set of classes, interfaces, and collaborations and their relating ships. There is most common diagram in modelling the object-oriented systems and are used to give the static view of a system. It shows the dependency between the classes that can be used in our system. The interactions between the modules or classes of our projects are shown below. Each block contains Class Name, Variables and Methods.

VI. RESULTS AND DISCUSSION

In this section, we present the analysis of results obtained. Below, we discuss the results as snapshots of the output.

Fig 6.1 ADMIN PANAL

Admin handles this section and reviews the results of the given personality of the person.

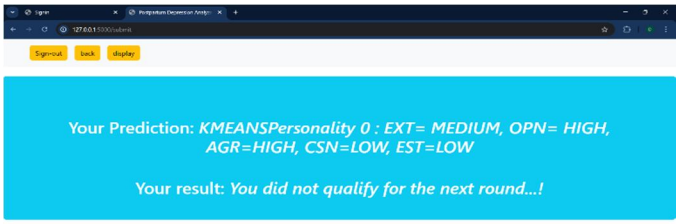


Fig 6.2 RESULT

On this page show the result of candidate

VII. FUTURE SCOPE

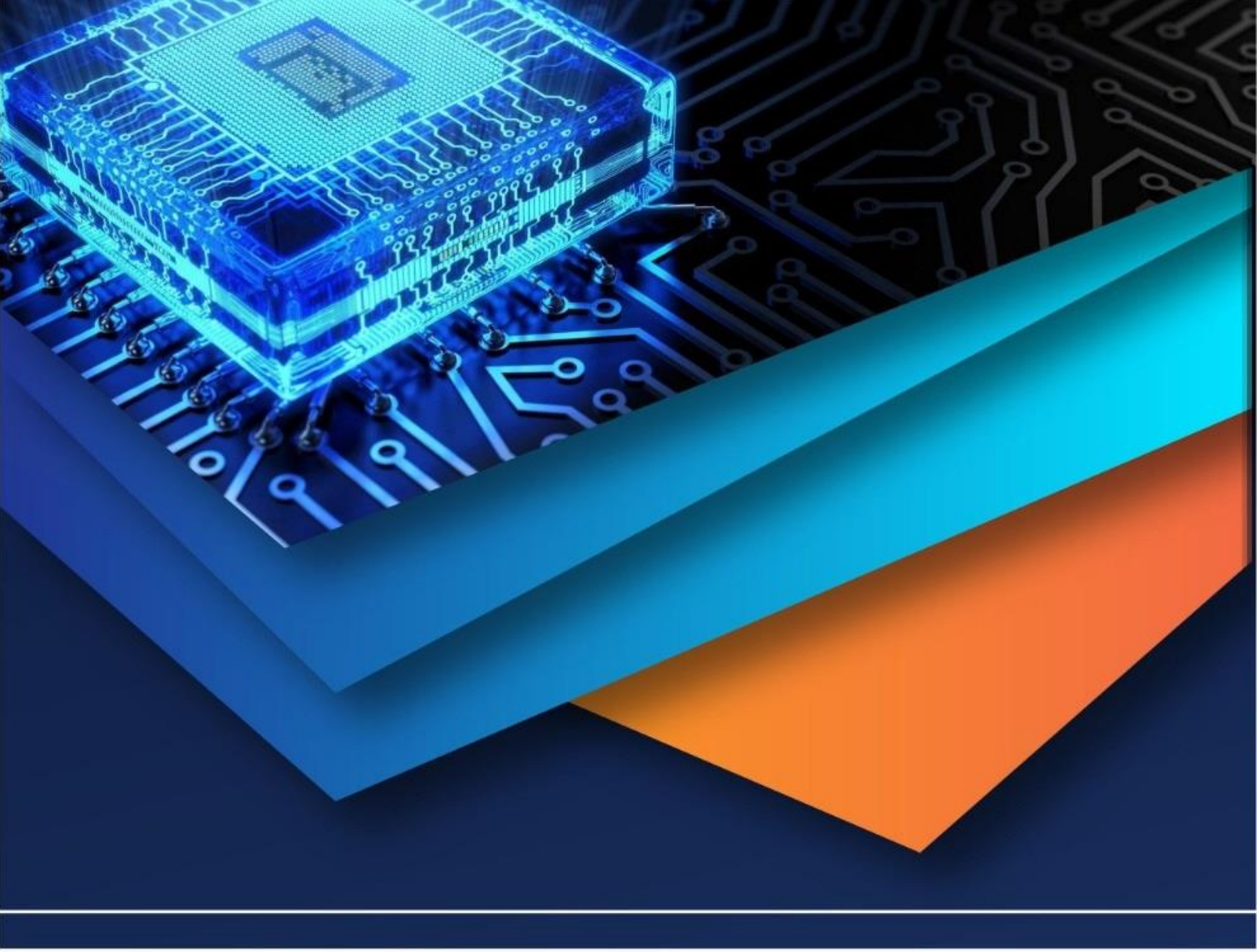
Enhancing data collection with wearable devices and sensors for real-time insights. Integrating AI and machine learning for deeper personality analysis. Developing advanced predictive models using social media and biometric data. Generating personalized recommendations based on personality traits. Ensuring ethical AI use with strong data privacy and security. Implementing adaptive systems that continuously improve with new data.

VIII. CONCLUSION

Personality prediction using machine learning and AI based on the OCEAN model holds great potential. By analysing large datasets, these algorithms can accurately predict personality traits such as openness, conscientiousness, extraversion, agreeableness, and neuroticism. This technology has applications in psychology, human resources, marketing, and mental health, allowing for personalized insights and recommendations. However, it is essential to address ethical concerns, such as biases in data and privacy issues, to ensure fair and responsible use of these predictions. Despite its promise, careful attention is required to avoid inaccuracies and protect individuals' data.

REFERENCES

[1] Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi (2019). Personality Prediction System through CV Analysis, IRJET vol 06 issue 02, e-ISSN-2395-0056, p-ISSN-2395-0072
[2] Devesh Agarwal, Mr. M. Karthikeyan (2022). PERSONALITY PREDICTION USING MACHINE LEARNING, IRJMETS Vol:04 /Issue:04/ April-2022, e-ISSN-2582-5208
[3] Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat (2021). Personality Prediction Via CV Analysis using Machine Learning, IJERT Sep 2021 Vol 10 Issue 09, ISSN – 2278-0181
[4] M. Kalghatgi, M Ramannavar, and Dr. N. S. Sidnal, “Neural Network approach to personality prediction based on the Big-Five Model” in IJIRAE, vol2 issue 8, August 2015, pp 56-63.
[5] J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem “Resume Ranking using NLP and Machine Learning”, unpublished
[6] Md Tanzim Reza, and Md. Sakib Zaman, “Analyzing CV/Resume using natural language processing and machine learning”, unpublished.
[7] Amirhosseini, M.H.; Kazemian, H. Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®. Multimodal Technol. Interact. 2020, 4, 9.
[8] Sudhir Bagade, Jayashree Rout, Pooja Yede, Personality Evaluation and CV Analysis using Machine Learning Algorithm
[9] Gagandeep Kaur, Shruti Maheshwari, Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis
[10] Afroja Khatun Monalisa, Md. Omar Kaiser Mahin, PERSONALITY PREDICTION SYSTEM THROUGH CV ANALYSIS.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY