# Personality Prediction Through CV Analysis using Machine Learning

1st N *Mohmmed Ashraf*
*Dept. of Computing Technologies*
*School of Computing, College of Engineering and Technology,*
*SRM Institute of Science and Technology, Kattankulathur,*
*Chennai, India*

mm0207@srmist.edu.in

*2nd Dr.Pongiannan R K, Professor*
*Dept. of Computing Technologies*
*School of Computing, College of Engineering and Technology,*
*SRM Institute of Science and Technology, Kattankulathur,*
*Chennai, India*

pongiank@srmist.edu.in

**Abstract— Personality of a person plays a crucial role in the organizational progress and also in the self-development process in an individual's life. One of the typical ways to predict the person's personality is either by a standard review or by scrutinizing the Curriculum Vitae of the candidate. The Conventional method of recruiting the candidates involves manual short listing of job seekers resumes according to the requirement of the company. In this work, a system that automates the task of segregating candidates based on eligibility criteria and personality evaluation in a recruitment process is proposed. To meet this requirement an online application is developed for the registration of candidate's details and analysis of candidate's personalities through an online Multiple-Choice Question (MCQ) test containing personality quizzes. Then the system analyzes professional eligibility by comparing the uploaded Curriculum Vitae trained datasets. This system employs a machine learning algorithm namely "Logistic Regression" which helps to choose fair decisions to recruit a suitable candidate. Thus, the final results of the personality quizzes will be sent to both candidates and the admin.**

**Keywords—Automated e-recruitment process, Online test, Big five model of Personality traits, Machine Learning, Personality prediction &Classification.**

## I. Introduction

The Job Characteristic Model (JCM) has significantly enriched Human Resource Management by offering a framework for designing jobs that meet the needs of both employers and employees. With the rise of advanced information technologies and widespread access to digital and internet technologies, HR management practices have been enhanced globally. This has led to the development of an E-HR system that integrates the principles of the JCM. The goal of this new system is to create a model that is effective in the digital era, aligning job design with modern technological capabilities and organizational objectives. This integration aims to optimize job satisfaction and performance by leveraging digital tools to implement the core dimensions of the JCM.

This study analyses and classifies the personalities of a given set of people using machine learning algorithms and prediction classifiers. Most of the e-recruitment system merely scans the Curriculum Vitae (CV) that have been submitted and shortlists the candidates via various processes that analyse the candidate's technical and communication skills only, wherein the proposed system administers personality tests, estimating the candidate's personality and short-listing him based on his talents and decision-making capacity.

It also forecasts new user personality based on personality data saved via categorization of prior user data. The test dataset was predicted using a Logistic Regression model to predict output class labels for dependent categorical data and a Random Forest classifier to classify the data sets into specific categories, predicting the candidate's personality and shortlisting him based on his abilities and capacity to make decisions.

## II. Literature survey

To overcome the drawbacks of Natural Language Processing, a system was developed in [1] in which based on the candidate's online test scores, selects the right candidate for desired job profile and to supply professional workforce for the firm, and to select the appropriate candidate for specified job profile (NLP). The purpose of [2] is to look into the causes for e-recruitment at SAT Telecom in India, as well as its success. Savings of roughly 44% on recruiting expenditures have been observed, according to the publication. Recruitment consultants and search agents are no longer as important. The average time it takes to fill a vacancy has decreased from 70 days to 37 days and also the cost of hiring is decreased.

The system in [3] implemented an automatic candidate ranking. Based on observable criteria, the candidate's information was gathered from his or her LinkedIn profile. The recruiter used the Analytical Hierarchy Process (AHP) to determine the candidate's rank based on individual selection criteria, while the recruiter regulated their weight (admin).

In [4] Human resources are seen as a critical source of growth for businesses; nevertheless, for the optimum use of human resources, it is necessary to match employees' knowledge, skills, and abilities to their allocated positions in the organization. The relationship between person-organization fit, job happiness, and job performance is explored in this study. According to the findings, there is a link between person-organization fit, job happiness, and job

performance. While there is a favourable association between job happiness and job performance. Person-to-organization fit is a significant predictor of performance. Employee job performance is a critical aspect in determining an organization's performance. Logic dictates that the better a person fits into his or her employment, the less adjusting he or she will have to do.

Using the Big Five Model, the study in [5] examines social media data to predict major personality traits, i.e. attributes or characteristics unique to an individual. Business intelligence, marketing, sociology, and psychology are just a few of the disciplines where the forecasts can be used. For analytic purposes, a parallelism between an individual's personality features and his or her linguistic information is investigated. The General Factor of Personality (GFP) was discussed in [6] that define the interrelations among the Big Five personality factors (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) per to test for the existence of a GFP. As this paper concludes, the meta-analysis shows that there was a GFP at the highest hierarchical level and that the GFP had a substantial component as it relates to supervisor-rated job performance. However, it is also recognized that the existence of a GFP does not imply that other personality variables down in the hierarchy have lost their significance.

The personality prediction system in [7] was constructed by dataset consisting of photographs, quality characteristics, personality measurement, and intellect testing are used to construct an end-to-end network for personality prediction that can consistently predict self-reported personality traits from an image. The experimental results reveal the candidate's or person's personality attributes, such as imagination, creativity, leadership, mutual respect capability, and many more. A personality trait is made up of traits that are revealed in a specific pattern of behaviour in a variety of settings. The paper also suggests that in inter-personal connections, determining an individual's personality attribute and intelligence from his or her face is important, yet it is unreliable. Personality factors such as social interaction, mutual respect, inventiveness, and a variety of other attributes cannot be determined just through personal contacts, according to the findings.

Emotional Stability and more energized forms of Extraversion and Conscientiousness were projected to be a major function of personality characteristics and sub-factors that match its emotional and motivational aspects in [8]. Predictions were confirmed in correlation and regression analyses across three studies. The personality analysis from

[9] can help to diagnose youth's mental health. Apart from technical knowledge, mental health also paves a crucial part in organization's development. In [10], a study on the big 5 personality traits were discussed and a study on howto decide the right people for right job based on these five traits was done. Importance of conscientiousness trait in educational and professional field which the aptitude skills fail to measure was done in [11]. How personality is inter-related to 3 types of learners - deep, surface and strategy and the indications of their true knowledge about the required skill set and their adaptability towards learning new things was detailed in [12]. This work was inspired from this E-Gen system - (Automatic Job Offer Processing system for Human Resources) for inventing e- recruitment system for CV analysis based on the study done in [13]. Using self and observer personality assessments, this article [14] presents experimental results for recognizing all Big Five personality traits in both conversation and text. Although statistical models outperform the baseline, ranking models outperform them overall. The study in [15] looked into the possibility of predicting anti-social personality traits based on Twitter usage. This was done by comparing 2,927 Twitter users' Dark Triad and Big Five personality traits to their profile attributes and language use. These factors have some statistically significant connections, according to the analysis.

## Methodology

The main objectives of this work project are (i) Not only should attention be paid to qualifications and experience,but also to other key aspects such as human personality, which are essential for a specific professional role. (ii) To assist the HR department in selecting the best candidate for a specific job description, resulting in an expert workforce for the company. The motivation behind this work is to simplify the immense task on the HR department to select right employee from a broad pool of job seekers for a certain job profile, resulting in an expert workforce for the firm.

“Personality as a Prognosticator of Future Job Success”, Many Scientific studies prove that personality is the authentic prognosticator of job performance and it also helps in examining the behavioural temperament of the candidates in the work environment that allows recruiters or employers to understand whether they will excel in their career and befit the culture of the organization. A personality test beats the conventional methods of collecting the information and providing the recruiters with target insights that substantially improve the selection process of the job seeker. Personality can be measured by setting the questionnaires based on the scientifically proven Big Five theoretical model popularly known as OCEAN model, which provide intuitive information about how the candidates’ personality will impact their workplace behaviour allowing recruiters to understand how candidates connect to the fellow persons, their approach to solving the problems and finally how they cope with their emotions.

“Finding the most suitable candidates increases employee quality and retention” - It is pivotal to sustain the talented employee and to reduce the rate of employee leaving a business. With personality assessment, you can screen candidates more efficiently for ability and personality. It also helps to examine whether a candidate is likely to stay in the role and fit in with the culture of the company. Automation of HR round in e-recruitment process which conserves the time for the employer as well as the candidates. Other than automation in the e-recruitment process, this project can be altered according to the need; onecan predict psychological problems of the patients by setting psychological questions along with personality questions in the medical field. This work helps the HR department to recruit qualified candidates without consuming much time as all the processes are automated with the help of machine learning techniques. It also helps candidates to self-evaluate their potentials and their interests hence they can make clear decisions about the career they want. Choosing the right candidates by the organization is as much essential as choosing the right job from the candidate’s side. The quality of the candidate decides the growth of the Organization.

The proposed work will prove to be a more effective way to shortlist the candidates by conducting a personality test and predict whether they will be suitable for the job using Logistic Regression and Random Forest Classifier. This is legally permissible. According to a scale from 1 to 8, the system ranks the necessary experience and skills for a specific job role on a scale from 1 to 8. The HR department will be able to quickly shortlist candidates using the system's expected ranking policy.

## III. BIG 5 PERSONALITY TRAITS

The Big Five model was actually developed in the latter half of the 20th century, specifically in the 1950s, rather than in 1990. Lewis Goldberg, a researcher at the Oregon Research Institute, coined the term "The Big Five" to describe this model. It has since become widely accepted and is extensively used in both business and psychological research. The Big Five Personality Traits Model evaluates five distinct aspects of an individual's personality.

### *A. Openness*

Openness is also known as "Imagination" or "Intellect," and it assesses your level of creativity as well as your thirst for knowledge and new experiences.

### *A. Conscientiousness*

Conscientiousness examines your level of care in your personal and professional life. If you have a high conscientiousness score, you are likely to be organized and meticulous, as well as know how to establish and carryout plans. If you get a low score, you'll probably be sluggish and disorganized.

### *A. Extraversion / Introversion*

Extraversion / Introversion determines how sociable you are. For example, are you gregarious or reserved? Do you get your energy from crowds, or do you struggle to collaborate and interact with others?

### *B. Agreeableness*

Agreeableness metric assesses how well you get along with others. Are you considerate, helpful, and eager to work with others to reach an agreement? Or do you prioritize your own needs over those of others?

### *B. Neuroticism*

Neuroticism also known as "Emotional Stability" measures a person's emotional reactions to a specific situation. A person with a low level of neuroticism is thought to have good emotional control. People witha high Neurotic value, on the other hand, are much more worried and sensitive.

These 5 personality traits are commonly abbreviated as “OCEAN”.

## IV. MACHINE LEARNING ALGORITHM

**Logistic Regression:** The model operated to foretell the test dataset is “Logistic Regression” because Logistic regression is a pervasive model to predict output class labels for qualified categorical data.

**Random Forest Classification:** It works in four steps. In the first step choose random samples from a set of data. The second step comprises of creating a decision tree for each sample and use the decision tree to get an augury result. As a third step, put each augury result up for a vote. In the fourth step, to conclude the forecast, choose the augury result with the most votes.

## V. PROCESS WORKFLOW

A. Data collection

A. Perform data preparations

B. Select a model

C. Training

D. Evaluation

E. Tuning hyper parameters

F. Augury

### *A. Data Collection*

The accumulation of data is the first phase in the machine learning process. To acquire the necessary results, the data must be very clear. This phase is critical for the system to produce the desired results.As an input to the system, these data are immediately delivered in tabulation format.

### *B. Perform Data Preparation*

This stage transforms the obtained data into information that has been processed. The acquired data is processed, with missing values checked and the data sorted according to the desired criteria. For the system's input, the aggregated data is tabulated. For the system's input, the aggregated data is tabulated. The data has now been split into two sections. The first component will be used to train our model, which will take up the majority of the dataset, and the second will be used to evaluate the trained model's results. This step includes all other types of adjustment and processing, such as normalization, error correction, and so on.

### *A. Select a Model*

The next step in the process is to choose a model from among the many that researchers and data scientists have developed throughout time. The easiest way to choose a model is to collect as much data as possible, which can be nearly endless depending on the problem's complexity.

### *B. Training*

The training phase is a time-consuming process since the data obtained is used as input, which is then used to train the system. The more data there is, the more accurate the system becomes. If the system is properly educated, it will be more efficient in selecting the appropriate task flow, and the system's outcome will be better. The trained system is tested with the rest of the data. The system is checked for mistakes, malfunctions, and other issues, and it corrects itself so that it becomes more efficient.

### *A. Evaluation*

Following the training phase, the evaluation phase begins, during which the system is tested and evaluated to see how it performs in real-world applications, how it learns in the future, how it corrects errors, and how it improves its efficiency. The test data accounts for 30% of the whole data set, allowing for proper training and mistake correction. The

a system that has been reviewed is ready for real-time applications and can be implemented.

*F. Tuning Hyper Parameters*

The settings are tweaked in this phase of the process to expand the range of inputs and add new ones. During the testing phase, the input parameters can be changed after the training. The system should adapt to the change so that the system's parameters are tuned in such a way that variations in input parameters after the training period have no effect on the system's output. This step is extremely important for the system's accuracy. Because data is always continuous and dynamic in real life, the system should be designed insuch a way that it adjusts flawlessly to the present circumstances and does not provide any incorrect outputs.

*A. Augury*

Machine learning is used to forecast the correct outcome and prophesy the prophecy. Finally, the system process is verified to ensure that it follows the prescribed procedure. The result is compared to the predicted result. How far has the system learned, how accurate it is, and how much has the result varied from the intended result is studied in this phase. The block diagram of personality prediction system is shown in Fig1.

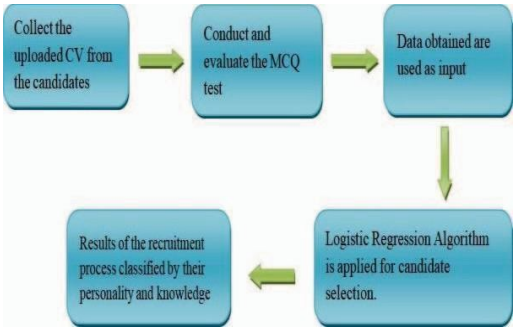
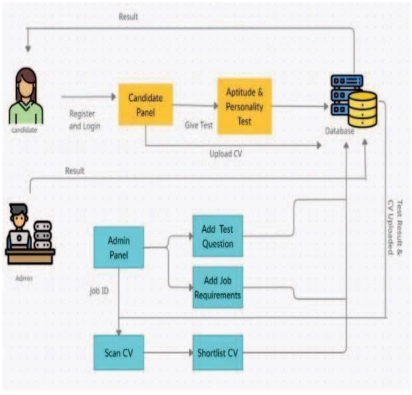



Fig. 1. Block diagram of personality prediction system

In the front end on the candidate's part is that, first; the candidate must register on our website as they are new to our website. Then login with the just created account. The link for taking a personality assessment is selected. Take the assessment. Wait for the result and check it in the email. The front end on the employer's part is that, theassessment response is collected in the excel sheet as a .csv file. The software predicts the result. Then the result ispublished to the candidates through their corresponding mail.




Fig, 2. Front and backend process outline

In the backend the system is trained with numerous data like trained and test data sets. The data of the candidate is collected using a form and it is given as input to the trained system. Then the machine learning algorithms is applied by using Logistic Regression to predict and Random Forest classifier to classify the data sets into specific categories. Thus, various outputs of given data sets are classified, which helps the employer to make fair decisions regarding recruiting the right people for the required job profile. The process flow of front and back end is shown in Fig 2.

This system can be used in companies which have an assortment of employee with varying personalities for various jobs. This system will allay the proceedings of the human resource department to label the right candidate for a right job profile which in turn provides a consummate team for the organization. System Admin or any employee head can easily screen a candidate based on their online personality test marks and can nominate an appropriate job for the candidate.

## VI. RESULTS AND DISCUSSION

As important as talent, the personality of an employee is also very important for a stable organization. With personality assessment, you can classify the candidate's personality based on his capability and adaptation to the company's requirements. Our project helps the company to select the right candidate in an easier and efficient way. We use a Machine Learning algorithm to assess the candidate personality, by conducting online aptitude and personality tests. Here, a form containing multiple choice questions will be given to the candidate and each question is given with four options to choose from, based on the answers they choose their personality is assessed and predicted. Since it is multiple choice questions, every option is given with a weightage ranging from 1 to 8 as we said in our proposed system. After completing the test each answer selected by the candidate is aggregated and brought under five class label values i.e, extraversion, lively, serious, responsible, and dependable. Based on their predicted personality, the HR team can easily filter the candidates and pick out the desired candidate they intend to hire. If a company wanted to hire a person for a Public Relations job, then the candidate should have alively, outgoing personality and capability to tackle social situations. If a person has more responsivity traits, he/she is likely to be fit for Customer Care Representative Job, teacher, paramedic, etc.. The results are shown in Fig 3 and Fig 4, where Fig 3 represents the sample output matrix and Fig 4 represents the Output csv file sample image. Our proposed system is fully automated with more transparency and legibility, thus help in reducing the burden of HR, time consumption, and more productivity. The attributes and its classification are shown in Fig 5.

```
[[ 1 17  7 ... 7  3  2]
 [ 1 19  4 ... 4  6  6]
 [ 0 18  7 ... 4  5  5]
 ...
 [1 22  5 ... 3  6  1]
 [1 19  5 ... 5  7  5]
 [0 18  5 ... 7  6  5]]
[[ 0 20  7 ... 9  5  5]
 ...
 [1 18  7 ... 6  2  7]
 [1 23  6 ... 5  4  3]
 [0 18  5 ... 3  5  6]
```

Fig. 3. Output matrix

| Person No | Pesonality Predicted |
|---|---|
| 1 | dependable |
| 2 | serious |
| 3 | serious |
| 4 | serious |
| 5 | responsible |
| 6 | serious |
| 7 | serious |
| 8 | serious |
| 9 | serious |
| 10 | serious |
| 11 | dependable |
| 12 | responsible |
| 13 | responsible |
| 14 | serious |
| 15 | lively |
| 16 | extraverted |
| 17 | serious |
| 18 | serious |
| 19 | extraverted |
| 20 | serious |

Fig. 4. Output csv file sample image

Attribute Description: No. of attributes are 7 as listed below.

| S NO | ATTRIBUTE | TYPE | RANGE |
|---|---|---|---|
| 1 | Gender | nominal | Male/Female |
| 2 | Age | numeric | 17-28 |
| 3 | Openness | numeric | 1-8 |
| 4 | Neuroticism | numeric | 1-8 |
| 5 | Conscientiousness | numeric | 1-8 |
| 6 | Agreeableness | numeric | 1-8 |
| 7 | Extraversion | numeric | 1-8 |

**Class label description:**

No .of class label: 5

Type: Nominal

Values:

- Extraversion
- Serious
- Responsible
- Lively
- dependable

Fig. 5. Attributes

## VII. CONCLUSION

In the work presented, using a self-learning system, the recruitment process can shorten the pool of candidates by using psychometric analysis to determine an applicant's emotional quotient and other technical eligibility criteria from their online resumes. The OCEAN model analyses the candidates' linguistic and personality traits. The accuracy obtained by the proposed system is higher than the previously implemented systems. The Random Forest Classifier is used to add extra accuracy to the existing system. In addition, there is room for improvement in the algorithm, like improving the personality test and adding a facial tester while the personality testing is done to find the candidates 100% accurate personality without their answering options. EGP employs supervised algorithms that have been trained based on the prior recruitment data. The future scope is to add a game- based interface in the website to improve the analysis of the candidate's traits that plays a crucial role in the organization's development and to see how the candidates reacts to a situation like that and how they solve it. Uploading CVs in a specific format can also help with the CV standardization, which also improves the system's overall efficiency.Looking ahead, the system aims to introduce a game-based interface on the website, enhancing the analysis of candidates' traits and observing their reactions and problem-solving abilities. Standardizing CV formats can further boost efficiency.

## VIII. REFERENCES


1. Namrata Ramesh Satelkar, Arati Ashok Sawant, Madhuri Vishnu Vengurlekar, Vaibhavi Ganpat Sawant, *"Personality Prediction System through CV Analysis"*, International Journal of Scientific Development and Research (IJSDR) (April 2019).
2. Sharon Pande, "*E-Recruitment Creates Order out of Chaos at SAT Telecom: System Cuts Costs and Improves Efficiency"*, Human Resource Management International Digest (June 2011).
3. Evanthia Faliagka, Athanasios Tsakalidis, Giannis Tzimas, *"An Integrated E-recruitment System for Automated Personality Mining and Applicant Ranking"*, Internet Research, vol. 22(5) (October 2012).
4. Sdd Safia Farooqui, Asha Nagendra, *"The Impact of Person Organization Fit on Job Satisfaction and Performance of the Employees"*, Procedia Economics and Finance, vol. 11, Pages 122-129 (2014).
5. Mayuri Pundlik Kalghatgi, Manjula Ramannavar, Dr. Nandini S. Sidnal, *"A Neural Network Approach to Personality Prediction based on the Big-Five Model"*, International Journal of Innovative Research in Advanced Engineering (IJIRAE), ISSN: 2349-2163 Issue (8), vol 2 (August 2015).
6. Dimitri van der Linden, Jan te Nijenhuis, Arnold B. Bakker, *"The General Factor of Personality: A Meta- Analysis of Big Five Intercorrelations and a Criterion-Related Validity Study"*, Journal of Research in Personality vol. 44, Issue (3), Pages 315-327 (June 2010).
7. Gayatri Vaidya, Pratima Yadav, Reena Yadav, Prof. Chandana Nighut, *"Personality Prediction by Discrete Methodology"*, IOSR Journal of Engineering, International Conference on Innovative and Advanced Technologies in Engineering (March 2018).
8. Ilke Inceoglu, Peter Warr, *"Personality and Job Engagement"*, Journal of Personnel Psychology, vol. 10, No.4(Dec 2011).
9. P. Asendorpf, Susanne Wilpers, *"Personality Effects on Social Relationship"*, Journal of Personal and Social Psychology, vol. 74, No. 6 (June 1998).
10. *Murray R. Barrick, Michael K. Mount, "The Big Five PersonalityDimensions and Job Performance: A Meta- Analysis", PersonnelPsychology (March 1991).*