

Decoding Job Candidates: Forecasting Personas Using Resume/Curriculum Vitae Analysis.

¹Pranit Wankhede, ²Vipul Nikam, ³Srushti Dandale,
⁴Gayatri Darshane, ⁵Mrs. Nikita Khawase

^{1,2,3,4}UG Students, ⁵Professor
Department of Information Technology
RMD Sinhgad School of Engineering
Pune, India.

Abstract- Persona of an individual strikes out as a very crucial part in the development and growth of an organization as well as one's individual growth. Out of many, one of the stereotypical strategies of speculating an individual's personality is either by the usual general inspection or by inspecting an individual's Curriculum Vitae. The traditional method for the recruitment procedure of a candidate is non-automatic (manual) pre-selection of the individuals resume trying to seek job with respect to the prerequisite specified by the organization. With this work, the goal primarily is to design a system that carries out the operation of separating candidates based on eligibility criteria and persona estimation in a recruitment process automatically. Hence to satisfy the requirements of the work proposed above, a webpage that operates online is advanced for the enrolment of candidate's information and investigation of an individual's personality via a persona questionnaire in the form of an online multiple-choice questions test. With respect to all of this the proposed system then inspects proficient aptness by analyzing the datasets that are trained of the CV/Resumes uploaded by the applicants. The indicated work incorporates two machine learning algorithms which are "Logistic Regression" and "Random Forest Classifier" which fairly help to select a candidate for the recruitment procedure. Consequently, the outcomes of the persona questionnaires are to be sent to the candidate as well as the governor of the indicated system respectively.

Index Terms- Automatic recruitment procedure, Persona questionnaires, OCEAN, Machine Learning, Persona investigation, Random Forest Classifier.

I. INTRODUCTION

Personality can be an important factor for potential employers when they are considering your job application. Although technical skills and experience are essential, employers are often, looking for employees who fit well with the company culture and can work effectively with others. Therefore, including some information about your CV/Resume can be beneficial. However it's important to note that it's not overdone and make your personality the main focus of your CV/Resume. Instead you should highlight a few key traits that are relevant to the job that one is applying for. For example, if one is applying for a service role, you might want to emphasize your communication skills, empathy and ability to work well under pressure. You can showcase your personality in a few different ways on your CV/Resume such as through your personal statement or by highlighting one's hobbies and personal interests. By doing so one can increase their chances of standing out from other candidates and finally landing the job that one wants to have.

The indicated study inspects as well as segregates the personas from a certain group of individuals incorporating machine learning algorithms i.e. Logistic Regression and Random Forest Classifier. Mostly every automatic- recruitment system completely and broadly browses and examines the Curriculum Vitae (CV)/ Resumes that are attached by the candidates and filters the applicants through various processes necessary to inspect the applicant's knowledge (practical and theoretical) in the domains such as the technological and soft skills completely, whereas the proposed work incorporates a bunch of persona questionnaires, targeting the applicant's persona as well as selecting that particular individual established on that individual's capabilities, controlling and managing abilities. Likewise the indicated system predicts the current users and recent users persona through the personality data that is stored through grouping from the user data of the former applicants. A Logistic Regression model was used to predict output class labels for dependent grouped data in the test dataset. Additionally, a Random Forest classifier was employed to classify the datasets into specific categories in order to forecast a candidate's persona and shortlist them based on their capacity to make significant, confident choices.

II. LITERATURE SURVEY

[1] In this paper "Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process", the authors G. Sudha, Sasipriya K K, Sri Janani S, Nivethitha D, Saranya S, Karthick Thyagesh G, proposes to use machine learning algorithms to assess a CV and forecast personality traits like openness, conscientiousness, extraversion, agreeableness, and neuroticism in job applicants. In order to extract pertinent information from the resume, the system employs natural language processing techniques. After that, machine learning models are used to forecast personality traits. According to the study, the proposed system may automate the e- recruitment process and predict personality traits with a high degree of accuracy. In addition to reducing human bias in the selection process, the authors contend that their system can assist firms save time and dollars throughout the recruitment process.

[2] In this paper "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network", the authors are M.F. Mridha, Rabeya Basri, Muhammad Mostafa Monowar, Md. Abdul Hamid. The present information technology craze, particularly in Artificial Intelligence, centers on the pursuit of human things through machines (AI). By extending its branches

to many dimensions and levels, machine learning has become popular as a component of AI. One of the newest technologies, deep learning, is used in sectors that involve huge amounts of processing. In the field of NLP, several tools and technologies have been developed and dissipated for their relative benefits and drawbacks. The job profiles of candidates can be screened by reviewing their CVs and resumes for selection using the most sophisticated and intellectual procedure in accordance with HR requirements. One by one, each profile is read, and it is noted whether or not it has been chosen for the preliminary stage. This work aims to automate this screening process through a thorough analysis of NLP to revolutionize the entire recruitment process by making employer tasks easier than ever before. This work is also capable of ranking the CVs of the individuals by matching the total fields in the CVs with the required fields. Also, compared to other methods, our classification model exceeds them with an accuracy of 74% for the BDJOBS site in terms of precision, recall, and f1-score.

[3] The paper "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer" is published by authors R.G.U.S.Gajanayake, M.H.M.Hira, P.I.N.Gunathunga, E.G.Janith Supun, Anuradha Karunasenna, Pradeepa Bandara. Choosing the best applicants for interviews is a crucial step for businesses because it can impact their general productivity. The standard procedure for hiring new staff for a long time has been for recruiters to review Curriculum Vitae (CV), shortlist applicants, and then contact them for interviews. Pre-screening procedures are now used by companies to cut down on the time needed for the aforementioned process. Yet for those systems to evaluate the candidate, they require enough data. For instance, while hiring a software engineer, recruiters are considering candidates' programming skills, academic standing, and personality attributes. In this study, a pre-screening method is suggested for selecting candidates for the position of software engineer. Candidates are selected using information from initial call transcripts, GitHub profiles, LinkedIn profiles, CVs, academic transcripts, and letters of recommendation. The Big Five personality traits, insights from the CV and GitHub, the candidate's skills, background, and capabilities from the recommendation letters, as well as programming skills and knowledge from the Academic transcript and Linked Profile are all identified using this approach, which extracts textual features of various dimensions based on Natural Language Processing. The findings from the various domains are displayed, demonstrating how the chosen supervised machine learning algorithms and methodologies may be utilized to assess the most qualified individuals.

[4] The paper "Real-Time Resume Classification System Using LinkedIn Profile Descriptions" published by authors Mr. Ramraj S, Dr. V. Sivakumar, Kaushik Ramnath G. For both job seekers and recruiters, precise job and resume categorization is essential in the world of online job recruitment. In order to train the model with texts, classify them into labels, and then compare the results. This study aims to have developed an automatic text classification system that makes use of a number of techniques, including Term frequency-inverse document frequency with machine learning and convolution neural networks. The applicants shall be divided into various categories using the resume data that has been provided. This employs domain adaptation because resume data is so delicate. A huge dataset of job description excerpts is used to train a classifier, which is subsequently used to the classification of resume data. Consistent classification performance is seen despite the limited dataset.

III. METHODOLOGY

The primary goals of this effort are to (i) identify the best candidate for a specific job description, (ii) facilitate the HR department in doing so, and (iii) assist the business to create a highly skilled workforce by taking into account not only skills and expertise but also meaningful personal traits that are necessary for a particular job. The purpose of this project is to make it easier for the HR department to choose the best applicant from a big pool of candidates for a specific job opening, creating a workforce of specialists for the firm. One's personality qualities may be utilized to evaluate their likelihood of succeeding in a profession or job. To put it in other words, employers or recruiters may be able to tell if a person would do well in a certain function or work just by looking at their personality. Characteristics including work engagement, emotional stability, openness to new experiences, and agreeableness are frequently taken into account in this context. Employers may be able to make better recruiting decisions and develop a stronger and more successful staff by utilizing personality as a predictor of job performance. Several scientific research demonstrate that personality is the most reliable indicator of job success and that it is equally helpful for evaluating applicant's behavioral temper in various different situations and hence this further permits recruiters or employers to have an understanding of will the candidate do well in their employment journey ahead. Over the traditional approach of hiring, using a personality test as a tool can have a number of benefits. Here are a few ways it may benefit hiring managers. Personality assessments may shed light on a candidate's working methods, communication preferences, and other crucial characteristics that may affect how well they perform in their jobs. Conventional hiring practices are susceptible to prejudice because recruiters may unintentionally prefer applicants with similar backgrounds or experiences. On the other hand, personality tests are frequently standardized. Comparing personality tests to more conventional hiring practices, including performing several rounds of interviews, can save time and money. Early on in the hiring process, personality tests can assist recruiters in eliminating individuals who are not likely to be a good match for the position.

When a candidate is a good fit for the job and has the necessary skills and personality traits, they are more likely to perform well in the role. This can lead to increased productivity. When employees are a good fit for their job, they are more likely to be satisfied with their work. This can lead to better engagement, increased motivation, and a greater sense of fulfillment in their job. Employee retention rates are higher when there is a strong fit between their position and the corporate culture. This can lower the price of hiring new staff to replace departing ones for the company. Finding the best applicants can eventually result in greater job performance, more job satisfaction, better working relationships, lower recruiting expenses, and higher employee quality and retention. Hence the intended system will justify that Administering a personality test during the recruitment process can be a more effective approach to screening candidates and determining their suitability for a job. This method allows for the prediction of a candidate's job fit based on their personality traits, potentially resulting in a more accurate shortlisting process. The persona test that this system agrees to provide will necessarily be multiple choice question test which will have four options and the candidate appearing for the test must select one. Then the further proceedings will be up once and all the results are out.

IV. BIG 5 PERSONALITY TRAITS

The Big 5 Personality Individual differences in personality, conduct, and social interactions can be understood using traits. They are frequently used to forecast job performance, academic success, and other crucial life outcomes in a variety of sectors, including psychology, education, business, and human resources. It is also known as the Five-Factor Model (FFM). The Big 5 model was released by the end of 1950. It consists of five broad characteristics or dimensions that can be used to sum up someone's personality and they are as follows:

Openness- Openness is a personality trait that describes how much a person enjoys thinking about complex and abstract ideas. People who score high on openness are usually creative and like to try new things. Alternatively, individuals that are low on openness are mostly who are more practical, focused on things they already know, that is they are not usually eager to learn new things and prefer conventional ways of doing things. They don't usually like to take risks or try new things.

Conscientiousness- Conscientiousness is a personality trait that refers to a person's ability to control themselves and their behavior in all situations in order to achieve their goals. People who score high in conscientiousness are ideally very organized. On the other hand, people who score low on conscientiousness are often impulsive and easily distracted.

Extraversion / Introversion - Extraversion refers to how social a person is in his/her life and how soon can he/she build a conversation or get up with people. Those who are extroverts, they actively participate with others and easily make friendship. Unlikely introverts are the ones who are mostly a little shy and do not express much or talk much in social environment.

Agreeableness – Agreeableness is a trait of one's personality which judges the individual in terms of how eager he/she to prioritize helping someone else when in need over his/her own needs. People who have a high score in Agreeableness are empathetic and often find fulfillment in caring for and helping others than prioritizing their own stuff. Unlikely, individuals with low score of Agreeableness tend to have less empathy and prioritize their own needs above those of others.

Neuroticism- Neuroticism is a personality trait of an individual's approach towards discovering negative emotions. Although a lot of people face these emotions, some individuals might not be able to have control on them. High Neuroticism scores indicate more chances of reacting to conditions with an emotional approach like fear, anger, and sadness, while the ones with low Neuroticism scores are more likely to take their setbacks in positively and move ahead in life.

The above mentioned are 5 personality traits that is commonly abbreviated as "OCEAN".

V. ALGORITHMS

Logistic Regression:- A supervised classification machine learning algorithm is the logistic regression algorithm. In a particular classification problem, the target or the output variable, (y), can take only discrete values for a given set of inputs (X). Logistic regression assumes that the relationship between the predictor variables and the outcome variable is linear on the logit scale. It is widely used in various fields such as medicine, finance, marketing, and social sciences for analyzing the relationship between a binary outcome and its predictors.

Certainly! Logistic regression can be used in the context of a personality quiz to predict the likelihood of a candidate being shortlisted for a job based on their personality traits. In logistic regression, the goal is to predict a binary outcome, such as whether or not a candidate will be shortlisted for a job. The model works by using a set of predictor variables, such as the candidate's responses to a personality quiz, to estimate the probability of the binary outcome. In the case of a personality quiz for job shortlisting, the logistic regression model might use the candidate's responses to questions about their personality traits (e.g. introversion/extroversion, conscientiousness, emotional stability, etc.) as predictor variables. The model would then estimate the probability of the candidate being shortlisted for the job based on their personality traits. It's important to note that logistic regression is just one of many statistical techniques that can be used to analyze data from a personality quiz. Other techniques, such as decision trees or neural networks, may also be effective depending on the specific application and data available. Additionally, it's important to use caution when using personality quiz data for hiring decisions, as there is a risk of unconscious bias and discrimination.

Random forest classifier:- Here are 5 steps to understand how Random Forest Classifier works. 1. Sampling Data Randomly. 2. Selecting Features at Random. 3. Creating Decision Tree's. 4. Aggregating the Results. 5. The final and concluding step of Random forest classifier is Prediction of output. A random forest classifier is a machine learning algorithm that is commonly used in classification tasks, such as predicting whether a job seeker would be a good fit for a particular job based on their personality traits. The algorithm works by building multiple decision trees and combining their results to make a final prediction.

To use a random forest classifier in a CV/resume shortlist job seeker based on a personality quiz, the following steps could be taken:

1. Develop a personality quiz that asks questions about different personality traits relevant to the job.
2. Collect the quiz responses from job seekers and use them to train a random forest classifier. The classifier would use the quiz responses as input and the job seeker's suitability for the job as the output.
3. Use the trained classifier to predict the suitability of new job seekers based on their quiz responses.

Overall, using a random forest classifier can help to automate and streamline the job matching process, making it more efficient and effective.

VI. PROCEDURE

- A. Collecting candidate data- The first step is to design a registration form so that users can register on the website. This form must

be filled out online by applicants; it is not a physical document. Basic information including name, contact information, educational background, employment history, and any other pertinent data needed for the position may be collected on the registration form. It is important to note that the registration form is user-friendly and simple to complete in order to motivate applicants to register on the website and take the test.

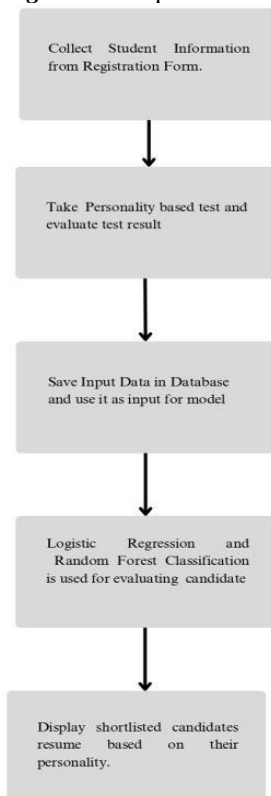
B. Taking Assessment- The next step is to conduct a personality and aptitude test to the applicants when the registration process (which includes signing in) is concluded. The execution of this test will only be done online; there will be no in-person testing. The test is meant to evaluate the applicant's personality traits, communication abilities, analytical capabilities, and other significant qualities necessary for the position. Candidates that have the qualifications and personality types needed for the positions should be found by analyzing the test results.

C. Storing data into database- The next stage is to store the data in a database after the personality and aptitude tests have been concluded. For both the machine learning model's input, this data will be used. The database is set up such that it can record the candidate's personal information, educational background, professional experience, and the outcomes of the aptitude and personality tests. The machine learning model will be trained using this data in order to increase its capability of correctly predicting if candidates are eligible.

D. Evaluation- This stage involves analyzing the data and using the machine learning algorithm to predict how well the candidate's personality will fit a certain position. In order to evaluate the candidates' eligibility, two algorithms—random forest and logistic regression—are employed. Many decision trees are used by the decision-making process known as random forest to estimate how well-suited each candidate is. A statistical method called logistic regression uses a candidate's biographical information, educational background, and employment history to predict the probability of their personality being suitable for a certain job profile.

E. Displaying results- A list of candidates who have been filtered down is then produced by the machine learning model based on their appropriateness according to the personality test that the candidate has appeared for. The candidate's eligibility is determined by the model, which also takes into account numerous aspects like personality traits, aptitude, and education. To find the best applicant for the job, the shortlisted individuals are then further assessed through interviews and other selection processes. This phase is crucial to make sure the best candidate is chosen for the position, as doing so can boost productivity and the organization's overall success.

Figure 1 : Steps involved



VII. CONCLUSIONS

Personality plays an equally significant role in maintaining a stable organisation as talent does. With the use of personality testing, you are able to categorise a candidate's personality in accordance with his aptitude and capacity to fit the needs of the firm. With the aid of our initiative, the business is better able to choose the ideal individual. Online aptitude and personality tests are used in conjunction with a machine learning algorithm to evaluate the applicant's personality. Each topic in this test has four possibilities, and the candidate will be given a form with multiple choice questions. Based on the candidate's selections, the personality test will measure and predict their personality. Although it is a multiple-choice question, each answer choice is given a weight that ranges from 1 to 4, as we said in our suggested system. Following completion of the exam, each response chosen by the candidate is combined and categorised into one of five values that are supposed to be considered, including extraversion, vivacious, deliberate, authoritative, and trustworthiness. The HR team may quickly narrow down the applicant pool based on expected personality traits and choose the ideal candidate they wish to hire. An applicant should be energetic, vivacious, and able to handle social situations if a corporation wished to hire them for a position in public relations. Someone who is more responsive is more likely to fit for jobs such as customer service representative,

teacher, paramedic, etc. The fully automated approach we've suggested is more transparent and readable, which will help it be more productive and less time-consuming for HR to manage.

REFERENCES:

1. Narmada 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. M.F. Mridha, Rabeya Basri, Muhammad Mostafa Monowar, Md. Abdul Hamid, "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network" (2021).
3. R.G.U.S.Gajanayake, M.H.M.Hira, P.I.N.Gunathunga, E.G.Janith Supun, Anuradha Karunasenna, Pradeepa Bandara "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer" (2020).
4. Mr. Ramraj S, Dr.V. Sivakumar, Kaushik Ramnath G "Real-Time Resume Classification System Using LinkedIn Profile Descriptions" 29- July-2020.