Personality Prediction with CV Analysis

Rutuja Narwade [1], Srujami Palkar [2], Isha Zade [3], Asst Prof. Nidhi Sanghavi [4]

[1], [2], [3] Student, Department of Computer Engineering, Atharva College of Engineering, Mumbai

[4] Assistant Professor, Department of Computer Engineering, Atharva College of Engineering, Mumbai

---***---

Abstract - *When it comes to the study of humans, adjudicating one's personality is important as it acts as a window to deciphering the individual's mindset. The personality is a vital part of an individual when he/she works for a complex organization. There are several ways to determine an individual's personality but the most sought-after and direct method is through a simple quiz. The questions in the quiz are framed in a way that they take values with reference to the big five personality model and aid the developer in framing a personality report of the individual in question. When we take a look at the current process of hiring and selection that various organizations make use of, the employers often pick out CVs in a manual way which is monotonous, time-consuming, and consumes a lot of human resources. Our approach is rendering a system that motorizes the eligibility check and aptitude evaluation of an applicant in the selection process. To target the drawbacks of the traditional recruitment system, a web application that analyzes both the personality and an individual's CV has been curated.*

Key Words: *Personality Prediction, Logistic Regression, CV Analysis, Machine Learning, Natural Language Processing, Big Five Personality Model (OCEAN), Psychometric Analysis, Recruitment Process.*

1. INTRODUCTION

The traditional recruitment process is a pretty time-consuming and tedious process. When it comes to hiring an individual, it is very important to make a sound and fair decision so that it is in the best interest of the applicant as well as the employer. The traditional recruitment process consists of aptitude tests, technical tests, group discussions, and HR interviews. With the curated system, the process has been transformed into a compact one with the same precision or even more as for that matter. The traditional process of hiring and selection may not always make a fair decision as an end result. With the help of the proposed system, the probability of making fair decisions increases by leaps and bounds. The proposed system makes use of Logistic Regression, Natural Language Toolkit, Flask, and Firebase for smooth and efficient deployment. The main methodology behind the personality questionnaire is the Big Five Personality Model which makes use of the OCEAN values namely, "Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism for predicting an individual's personality.

1.1 MOTIVATION OF THE PROJECT

Looking at the traditional way to go about the hiring and selection of the candidate(s), there are certain crucial observations to take note of. To state some of them: The process is **time-consuming**, not energy efficient (requires a team of professionals to filter through the **CVs** and make **a** sound **judgment**), there is a high probability of making an error in judgment of the candidate and also a high probability in an unfair selection of the candidate. Tackling these flaws of the traditional system is the main goal of this project.

1.2 BASIC CONCEPT

Our system gives an overall result of the candidate(s) on a psychological and professional level. To provide a smooth assessment process, our model makes use of the logistic regression algorithm and NLTK (Natural Language Toolkit). To ensure the safety of the user's data, we are using firebase as our backend service to provide authentication and safe storage.

2. BACKGROUND STUDY

The basic concept of this project came from a background study of [2] Automated Personality Classification using Data Mining Techniques paper written by Manasi Ombhas, Prajakta Gogate, Tejas Patil & Karan Nair. They have used the concept of data mining to classify the different personalities of individuals. Our objective is to predict an individual's personality using logistic regression rather than just categorizing the personality. In addition, we also integrated the predicted personality with the technical skills of the individual using CV analysis with the help of Natural Language Processing.

3. LITERATURE SURVEY

3.1 LOGISTIC REGRESSION

Logistic Regression is a type of statistical analysis often used for predictive analytics and modeling and extends its applications in machine learning. The use of Logistic Regression is mostly in statistical software to understand the relationship between the dependent variable and one or more independent variables by estimating probabilities using a logistic regression equation. The proposed framework has different modules for different functions. The role of logistic regression in the proposed framework is to

predict the personality of the individual using 7 employable values such as gender, age, openness, agreeableness, neuroticism, extraversion, and conscientiousness.
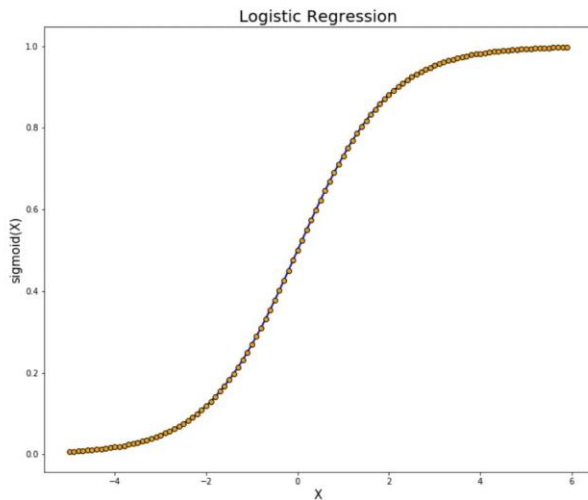

Fig -1: Logistic Regression

3.2 THE BIG FIVE PERSONALITY MODEL

The Big Five Personality Model is based on the OCEAN values namely, "Openness, Conscientiousness, Agreeableness, Extraversion, and Neuroticism". The OCEAN values of a person remain quite stable throughout their lifetime. The previous statement is the reason why our system uses the Big Five Personality Model as the base methodology for conducting the psychometric analysis as it will give output values that are stable and accurate. The questionnaire has questions that target these 5 factors in question and takes inputs as values with reference to the model.

Openness- It refers to how open an individual is to imbibing new ideologies and thought processes and how receptive he/she is to accepting said ideologies. A high value showcases the person to be highly receptive to new ideologies whereas a lower level suggests otherwise.

Conscientiousness- It refers to the organizational skills of an individual if he/she is meticulous and has good time management skills. A higher value suggests that the individual is super organized and as opposed to this a lower level indicates a sleazy/lazy attitude.

Extraversion- A person who spreads around good vibes and is very friendly with people around them has a high extraversion value.

Agreeableness- A person who genuinely cares about the people around them and is kind to everyone.

Neuroticism- Higher levels of neuroticism show anxiety and stress. Lower levels depict emotional stability.

3.3 NATURAL LANGUAGE TOOLKIT

Natural Language Toolkit (NLTK) is a Python Package used to perform Natural Language Processing (NLP). It was created as a tool for implementing NLP with ease in python-based projects. The growth of unstructured data via social media, online reviews, and voice-based human-computer interaction are some reasons why NLP has become a crucial part of modern technology. NLTK is a useful toolkit for many NLP applications such as morphological processing, syntax analysis, semantic analysis, pragmatic analysis, automatic text summarization, etc. NLTK is composed of sub-packages and modules. A typical processing pipeline in NLTK will call modules in sequence. Python data structures are passed from one module to another when using this library.

3.4 FIREBASE

When the app is connected to Firebase, the connection is not through normal HTTP. You're connecting through a WebSocket. Web Sockets are much faster than HTTP. There is no need to make individual WebSocket calls because one socket connection is enough. All the data syncs automatically through that single WebSocket as fast as the client's network can carry it. Firebase sends you the new data as soon as it's updated. When your client saves a change to the data, all connected clients receive the updated data in a negligible amount of time. Firebase Storage provides a simple and direct way to save binary files — which are most often images, but the data could be of any type — to Google Cloud Storage directly from the client. Firebase Storage has its system of security rules to protect your Google Cloud bucket from the masses while granting detailed write privileges to your authenticated clients. Firebase auth has a built-in email/password authentication system.

4. PROPOSED SYSTEM

We list some of the impediments based on surveys -

1. We came across the traditional methods which included administrating personality and professional eligibility judgment, conducting interviews, and holding group discussions. Therefore, they are tedious and can lead to disparate candidate selections. However, the most meaningful aspect that represents an individual is personality, which changes over time, and dealing with them is a tedious process. With the help of machine learning algorithms, we use to create the models that will be tested in the proposed automated candidate grading system.

2. Over the past few years, the number of manual interviews and resumes in human resources has expanded therefore we have planned a mechanism for identifying personalities and making recommendations.
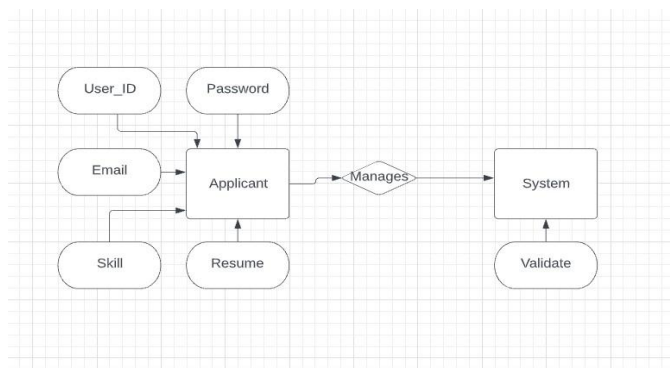
4.1 PROPOSED SOLUTION

Fig -2: E-R Diagram of the system

In our paper, we bring forward personality prediction and CV analysis using a machine learning algorithm. This method of working offers expert manpower for the company, assisting the HR department in selecting the foremost applicant for the specific job profile. Knowledge is highly valued in our culture and acquiring knowledge involves cognitive processes, communication, perception, and logical thinking. It plays an important role in a student's life in school. The incentive behind this system is to develop a web application that will assist organizations in drawing new members.

Aptitude questions test the ability of the candidate to perform tasks and react to situations at work. A psychometric test measures cognitive ability, personality or work behavior to indicate the potential of a job candidate to excel in a specific position or career. Therefore, the candidate has to fill out a form and enter the details and undertake two tests i.e., an Aptitude Test and a Personality test. In a personality test, the candidate has to rate themselves on a score of 1-10 for Openness, Neuroticism, Conscientiousness, Agreeableness, and Extraversion based on which the system predicts their personality. This is a far different approach that we have undertaken than employment websites. This makes our search process easy and this test would help find which personality category the applicant falls under and would become another parameter in finding the required category candidate.
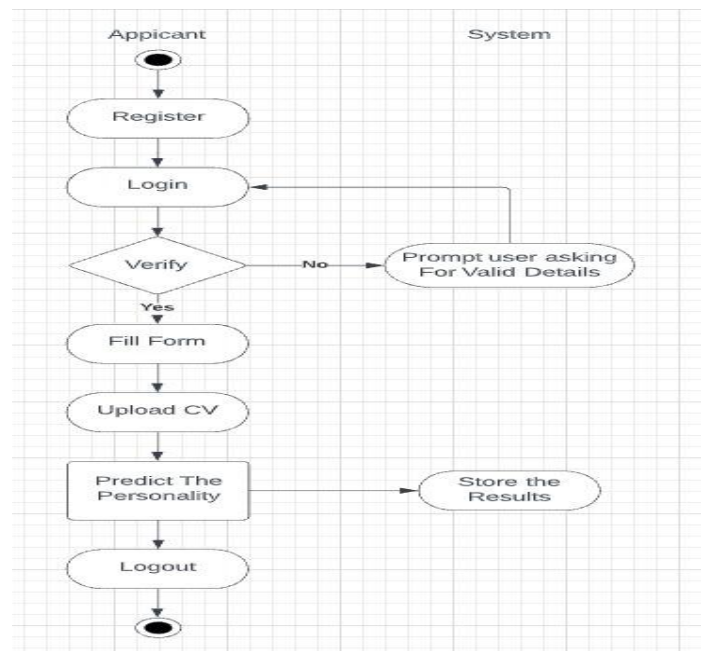

Fig -3: Activity Diagram of the system

4.2 IMPLEMENTATION

1. **Train model class:** In this model class we have two methods of working which involve giving instruction to the model and anticipating the result by providing different values.

A. **train method**: The train method looks through a CSV file with the dataset for training the algorithm and set up a model using Logistic Regression.

B. **test method**: The test method is to anticipate the personality by passing an array of values containing gender, age, and the other five personality characteristics.

2. **Main Method:** The main method starts by creating an object of the train model class and then instructs the model by referring to the class's train method. The next step is to initialize a variable with a Tk object and outline the system's landing page with labels and a press. A button is created called Predict Personality, which refers to the predict person process.

3. **Predict Person Method**: The predict person method

closes the root Tkinter window and set up a new top-level window with the relevant size and attributes. After this window's heading is labeled, followed by different labels and their entries. For choosing a resume file, the candidate must press the Choose File button, which then calls the Open file process, which requires a button argument. For predicting the personality numerous entries are used in the predict person system. By pressing the submit button, all of the values are passed to the prediction result.

Questions	Option A	Option B	Option C	Option D	Correct Option
Synonyms of Defer	Indifferent	Defy	Different	Postpone	Postpone
Antonyms of Benign	Malevolent	Soft	Friendly	Unwise	Malevolent
18,28,40,54,70,?	79	93	88	97	88
24,28,?,52,84	32	34	36	40	36
Antonyms of Exodus	Influx	Home-coming	Return	Restoration	Influx

Table 1. Sample Questions for Aptitude Test

Questions	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	Personality
I feel little concern for others	6	4	7	5	4	Extraverted
I am very prepared	4	6	4	4	7	Serious
I get stressed out very easily	5	6	4	7	4	Lively
I like multitasking	7	4	5	4	5	Dependable
I never get demotivated easily	5	7	6	6	3	Responsible

Table 2. Sample Questions for Personality Test

4. **Open File Method**: In the open file method the directory is opened using the default address, name, and file types, but ceases to function if no file is selected. After the try-except block, the method restores the name of the choose file button in the predict person method with the base name of the file so that the candidate is aware of the selected file.
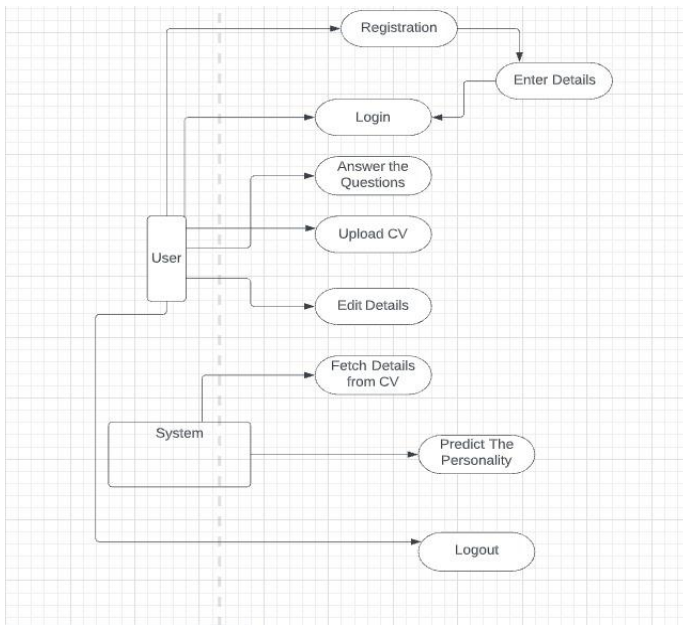
5. **Prediction Result Method**: The prediction result method closes the previous Tkinter window that was used to collect data from the candidate. After that, it invokes the model object's test method and reserves the result returned by the method. Following that, it inspects all of the information from the resume and stores it in a variable, followed by a try-except block that set out to delete the name and substantiate the mobile number from the information retrieved from the resume. The data consigned by the candidate is then printed on the console. Following that, the method exhibits a full-screen window with all of the inspected information and anticipated personality on the GUI window, as well as the description of each personality trait.

6. **Check Type Method**: The check type method modifies strings and numbers into the desired format, as well as lists and tuples in strings.

4.3 SYSTEM DESIGN

Th andidate will register her/himselfwith all the details and will also fill their own CV details into the system. Sample aptitude test questions along with their options are shown in (Table 1). After the test given by the candidates, the scores are stored in databases. The next test is of personality test. There is acommon myth which says that IQ tests measure intelligence. What an IQ test actually measures is notactual intelligence, but a person's capacity for intelligence. In this test various situations will be encountered by the candidate ranging from stronglyagree to disagree, which is provided as a drop-down list. The factors range like openness to experience, conscientiousness. (Table 2) shows the sample questions for personality test. Each question has the fix set of choices varying from strongly agree to disagre

The proposed system design is as followed :

The candidate will register himself/herself with all the details and will send their CV into the system. The candidate can also revise the details. After entering the details the candidate gets logged in and has to undertake two tests. Sample aptitude questions along with their options are shown in (Table 1). After completing the test, the scores are stored in databases. The next test is a personality test. Sample personality questions are shown in (Table 2). In this test, the candidate will come across various situations and have to rate themselves ranging from scores 1 to 10, which is provided as a drop-down list. The factors range from openness to experience, and conscientiousness. (Table 2) shows their personality type and OCEAN score. For parsing CVs, we have used pyresparser- a simple resume parser used for extracting important features such as name, email id description, and skills from CVs. Pyresparser supports file types such as PDF and Docx files. Once the parsing of the

data is done, it is stored in a CSV file. There are many Natural Language Processing (NLP) libraries like Natural Language Toolkit (NLTK), and spaCY which could help us in parsing the resume data. The system then calls the required details and predicts the type of personality of the candidate as shown in (Table2). After undertaking both the tests the candidates then log out of the system.

The candidate will register her/himself with all the details and will also fill their own CV details into the system. Sample aptitude test questions along with their options are shown in (Table 1). After the test given by the candidates, the scores are stored in databases. The next test is of personality test. There is a common myth which says that IQ tests measure intelligence. What an IQ test actually measures is not actual intelligence, but a person's capacity for intelligence. In this test various situations will be encountered by the candidate ranging from strongly agree to disagree, which is provided as a drop-down list. The factors range like openness to experience, conscientiousness. (Table 2) shows the sample questions for personality test. Each question has the fix set of choices varying from strongly agree to disagree

The candidate will register her/himself with all the details and will also fill their own CV details into the system. Sample aptitude test questions along with their options are shown in (Table 1). After the test given by the candidates, the scores are stored in databases. The next test is of personality test. There is a common myth which says that IQ tests measure intelligence. What an IQ test actually measures is not actual intelligence, but a person's capacity for intelligence. In this test various situations will be encountered by the candidate ranging from strongly agree to disagree, which is provided as a drop-down list. The factors range like openness to experience, conscientiousness. (Table 2) shows the sample questions for personality test. Each question has the fix set of choices varying from strongly agree to disagree

5 FUTURE SCOPE

In the future, we can improve the dataset so that the model will give more appropriate results. We can also integrate an SMS service into our system so that the users can get updates on the latest job openings. Recruiters can also use our system as a job recruitment platform. Our system can also be useful for college placements. To improve the accuracy of our model, we can add more aptitude tests on-site, to increase the efficiency of our system. Our system can also be of use in matrimonial sites in the future.

6 CONCLUSION

In today's age, the applicants and the employers can easily carry out the application and the recruiting process respectively. The system that we have put forward is an online application-based approach which is an unbiased

system that selects a candidate accurately. As our system also analyzes the personality of the candidates, recruiters can hire individuals on the basis of their overall temperament as well as according to the requirements of the hiring committee. In contrast with the traditional practices, an online recruitment system eases the workload as well as, less time would be consumed. We have also implemented a personality prediction system using NLP and ML-based functionalities. Furthermore, we can enhance the existing system by integrating AI to provide upgraded features. We can also implement more machine learning algorithms to add functionalities such as recommending a job to the applicant based on the skillset the individual has. This system has been created keeping in mind the best interests of the candidates as well as the hiring committee.

ACKNOWLEDGEMENT

REFERENCES

[1] Aditi Chandrashekar, 2Santhini Nadar, 3Shreya Ganesh, 4Mrinal Khadse "Online Recruitment System and Personality Prediction" International Journal for Research in Engineering Application and Management (IJREAM), April 2021.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Manasi Ombhas, Prajakta Gogate, Tejas Patil, Karan Nair "Automated Personality Classification Using Data Mining Techniques", April 2017

[3] Rabih Haddad, Eunika Mercier-Laurent. "Curriculum Vitae Evaluation using Machine Learning Approach". Artificial Intelligence for Knowledge Management IFIP AICT 614, 2021. hal-03496596

[4] Aleksandar Kartelj; Vladimir Filipović; Veljko Milutinović." Novel Approaches to Automated Personality Classification: Ideas and their Potential""

[5] Vivian LAI, Kyong Jin SHIM, Richard J. OENTARYO, Philips K. PRASETYO, Casey VU, Ee-peng LIM, and David LO:" Career Mapper: An Automated Resume Evaluation Tool"

[6] Nikolaos D. Almalis Dept. of Informatics, University of Piraeus, Piraeus, Greece; George A. Tsihrintzis; Nikolaos Karagiannis; Aggeliki D. Strati:" FoDRA — " A new content-based job recommendation algorithm for job seeking and recruiting"

[7] Dhokley, Waheeda and Jehangir, Randeria and Sarwar, Shaikh and Rashid, Shaikh:" A Novel Approach to Predict Personality of a Person"

[8] A Nathan C. Popkins: The Five-Factor Model: Emergence of a Taxonomic Model for Personality Psychology"

[9] A Janet E. Squires, Leslie Hayduk, Alison M. Hutchinson, Lisa A Cranley, Mark Gierl, Greta G. Cummings, Peter G. Norton, and Carole A. Estabrooks, "A Protocol for Advanced Psychometric Assessment of Surveys", Nursing Research and Practice Volume 2013 (2013), Article ID156782,20 December 2012

[10] A Automated CV Processing along with Psychometric Analysis in Job Recruiting Process Firoz Ahmed, Mehrin Anannya 2, Tanvir Rahman 3, Risala Tasin Khan4Institute of Information Technology, May 2015.

[11]] L D. van der, J te Nijenhuis, et al. The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. Journal of Research in Personality 2010; 44: 315-327.

[12] Rutuja Narwade, Srujami Palkar, Isha Zade, Nidhi Sanghavi. *"Personality Prediction with CV Analysis"*, Volume 10, Issue IV, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 970-974, ISSN : 2321-9653