International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

Personality Prediction System through CV Analysis

Harshada Samel, Deepti Sawant, Raghavi Sharma, Prof P.Y Itankar

B.E Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India

B.E Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India

B.E Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India

Assistant Professor, Department Of Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India

---

Abstract

An eligibility check, an aptitude exam, and a psychometric study of potential candidates are all part of the recruitment process. The paper proposes an application in which the system predicts the applicant's personality. By scanning through the applicant's CV/Resume, the programme estimates the applicant's personality attributes. The personality predictor is modelled using machine learning approaches such as Logistic Regression and Natural Language Processing (NLP). The applicant can be chosen or rejected based on the outcome.

Keywords: *Personality, CV, Resume, Big Five Personality Traits, NLP*

1.INTRODUCTION

After completing education, a person's life moves on to the next phase, which is finding a job. The most significant thing to represent for an applicant when looking for a job is their Curriculum Vitae (CV) or Resume. In this era of technology, Job seeking has grown smarter and easier. Along with the ease of availability of jobs, the candidate must also meet the job's requirements.

Personality is the most important thing that reflects an individual, and it changes over time. Tackling them is a time-consuming process for which we have devised a method to classify the personality and provide recommendations.

The proposed system parses the submitted CVs and shortlists the candidates based on the candidate's personality. This technique will also make it easier to shortlist candidate CVs from a large number of candidates, allowing for a consistent and fair CV ranking criteria that can be legally validated. This approach will not only consider qualifications and experience, but also other crucial factors that are essential for a specific job.

2. PROBLEM DEFINITION

A Personality Prediction System using CV Analysis, this system will predict the personality of the candidates on the basis of the resume and the values of five personality values. CV analysis can be sometimes one of the most time-consuming tasks of a recruiter looking for considerable skill to perform accurately yet quickly. Hence, this system is designed to reduce the workload and for making fair choice of candidates.

This framework can be utilized as a part of numerous business divisions that may require master competitors. This framework will decrease the workload of the human asset office. Also, it will assist the human asset division with selecting the right contender for a specific employment profile which thus gives a master workforce to the association.

3. AIMS AND OBJECTIVES

The aims and objectives of the project are as stated below:

- This system can be used in many business parts and areas that may require expert candidates.
- This system will reduce the workload of the workers in the general hiring, training, and firing department.
- This system will extract the key features from the resumes and display it to the user.
- It will predict the personality of the candidate based on the Big Five Personality Traits.

- Natural Language Processing (NLP) is used to extract information from the CVS.
- This system will focus not only on qualification and inexperience but also focuses on other important aspects, which are needed/demanded for a particular job position.

4. LITERATURE REVIEW

The big companies often need to deal with hundreds of CVs/Resumes each and every day. It has become very problematic and time consuming to handle such a big number of CVs/Resumes one by one. As a result, many companies started to provide specific formats or forms where the job seekers need to fill up with required information and then the CV/Resume will be analyzed by machine with simple pattern recognition and keywords searching. Due to which, potential candidates may lose the opportunity because of ambiguous keyword matching.

4.1.First Generation Hiring System

In this system, the Hiring Team would publish their vacancies and invite applicants. Methods of publishing were newspaper, television and mouth. The interested candidates would then apply by sending their resumes. These resumes were then received and sorted by the hiring team and shortlisted candidates were called for further rounds of interviews.

4.2.Second Generation Hiring System

In this system, as the industries have grown, their hiring needs have rapidly grown. To serve this hiring needs certain consultancy units have come into existence. They proposed a system in which the candidate would upload and submit their information in a specific format to the agency. These agencies would then conduct keyword searches on the candidates. These agencies were middle level organizations between the candidate and company. While the Second Generation Hiring System reduced the workload for the employers, it increased the amount of work for the applicants significantly as they need to maintain different formats for each job they apply for. Also, these systems were not flexible as the candidate had to upload their resume in a particular format, and these formats changed from system to system. further rounds of interviews.

4.3.Third Generation Hiring System

In 2014 an Integrated E-Recruitment System for Automated Personality Mining and Applicant Ranking was proposed by Faliagka et al. An automated candidate ranking was implemented by this system. It was based on objective criteria that the candidate's details would be extracted from the candidate's LinkedIn profile. The candidate's personality traits were automatically extracted from their social presence using linguistic analysis. The candidate's rank was derived from individual selection criteria using Analytical Hierarchy Process (AHP). The drawback of the Third Generation Hiring System is that while acquiring the information of the candidate from his/her social media profiles from social media websites such as LinkedIn, GitHub etc. an employer may conduct the screening of inconsistent data related to the candidate. Also, another consequence of depending on social media in recruitment is the tendency for biases to arise. Unknowingly, an employer can be biased toward people with good social media presence, as compared to people who are not as active online.

5. PROJECT DESIGN

Personality prediction refers to the process of scientifically determining, evaluating, and interpreting an individual's personality based on their CV. An automated personality prediction system using CV is described in this study, which classifies an individual's personality traits as well as occupation automatically with the help of a Tkinter framework, without the need for human intervention in which sets of characters to classify.

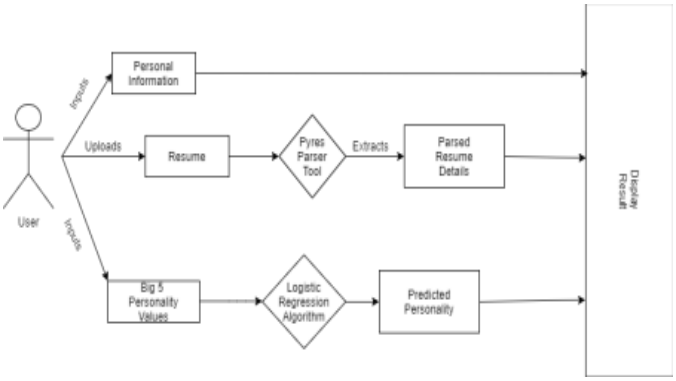


Figure 1. A basic use case diagram

The user inputs their personal details, Big 5 personality values and uploads the resume. The Big Five Personality Traits Model measures five key dimensions of people's personalities:

1. **Openness**: It sometimes is called "Intellect" or "Imagination". This measures your level of creativity, and your desire for knowledge and new experiences.
2. **Conscientiousness**: This looks at the level of care that you take in your life and work. If you score highly in conscientiousness, you'll likely be organized and thorough, and know how to make plans and follow them through. If you score low, you'll likely be lax and disorganized.
3. **Extraversion/Introversion**: This dimension measures your level of sociability. Are you outgoing or quiet, for instance? Do you draw energy from a crowd, or do you find it difficult to work and communicate with other people?
4. **Agreeableness**: This dimension measures how well you get on with other people. Are you considerate, helpful and willing to compromise? Or do you tend to put your needs before other's?
5. **Natural Reactions**: It is sometimes called as "Emotional Stability" or "Neuroticism," this measure emotional reactions. Do you react negatively or calmly to bad news? Do you worry obsessively about small details, or are you relaxed in stressful situations?

Then for resume parsing, we have use Pyresparser. Pyresparser is a python library. This python library is nothing but a simple resume parser used for extracting information from resumes. For the prediction of personality of the candidates, we have used Logistic Regression algorithm. Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. The result is then displayed to the user by extracting information from the user's CV and predicting their personality.

6. PROPOSED METHODOLOGY

Our project aims to develop an application which will analyze the personality of a candidate based on their CVs. Different modules or components created are domain establishment, data collection, parsing, ranking and database components. Parsing is the heart of our system which is created using python, nltk and libraries. This component does the morphological analysis, syntactic analysis, semantic analysis and generates the parsed data of the candidate according tohis/her skills.

The system architecture consists of a Personality prediction module which extracts information from the CVs using NLP. The system built in this project predicts personality of peoples by using their gender, age, score of openness, conscientiousness, extraversion, agreeableness, neuroticism and experience.

Following is the dataset(Source: Kaggle) that we have used for the personality prediction of the candidate.

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
0	Male	17	7	4	7	3	2	extraverted
1	Male	19	4	5	4	6	6	serious
2	Female	18	7	6	4	5	5	dependable
3	Female	22	5	6	7	4	3	extraverted
4	Female	19	7	4	6	5	4	lively
5	Male	18	5	7	7	6	4	lively
6	Female	17	5	6	5	7	4	extraverted
7	Female	19	6	6	7	5	4	extraverted
8	Male	18	5	7	5	6	7	dependable
9	Female	19	5	5	7	4	5	lively
10	Male	19	6	7	5	6	3	serious

Figure 2. Dataset based on the Big Five Personality Traits

The system consists of two methods which train the model and predict the result by giving the various values. Train method reads the dataset from a csv file and uses Logistic Regression Multinominal Classification algorithm to build the model. Test method predicts the personality of a person by passing an array of values that contains gender, age and other 5 personality characteristics.

The user interface will be simple and easy to understand even by the common man because it is made by Tkinter python module. Also, for uploading the resume, we have created an Open Upload File System using Tkinter root window. It tries to open the directory with default address name and file types and except if file is not chosen. These files can be of any extensions (.docx/.pdf/.txt etc). The framework has become more versatile and efficient as a result of the wide range of extensions available for uploading resumes.

After this it parses all the information from the resume. For this it uses pyresparser python module which is built using nltk and spaCy module in python.
Features of pyresparser:

- Extract name
- Extract email
- Extract mobile numbers
- Extract skills

- Extract total experience
- Extract college name
- Extract degree
- Extract designation

NLP Operations

1) spaCy module

Here in the spaCy module, we are using en_core_web_sm pipeline, which is used for English language. This is English pipeline that is optimized for CPU. Its components are tagger, parser, lemmatizer etc.

2) nltk module

Filtering out useless data is one of the most popular types of pre-processing. Stop words are used in natural language processing to refer to useless words (data). A stop word is a widely used word (such as "the," "a," "an," or "in") that a search engine has been configured to ignore both when indexing and retrieving entries as the result of a search query.

7. RESULT AND DISCUSSION

In this section, we will present the analysis of results obtained. Based on the algorithm the system predicts the personality of candidates. The system extracts all the important features from the resume. The personality traits and skills are taken in consideration for overall evaluation and analysis of CV. The result gives us the performance of the candidate and as per requirement of the recruiter, the candidate can be selected.

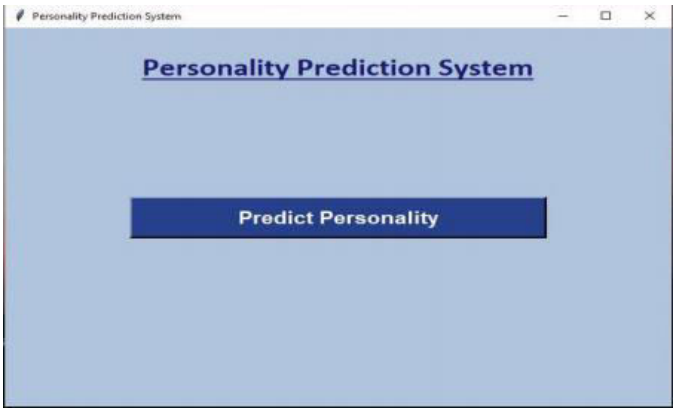


Figure 3. Home Page

On the home page, 'Predict Personality' button pops up a new window for taking various inputs from the user and submitting a prediction model which will predict the personality.

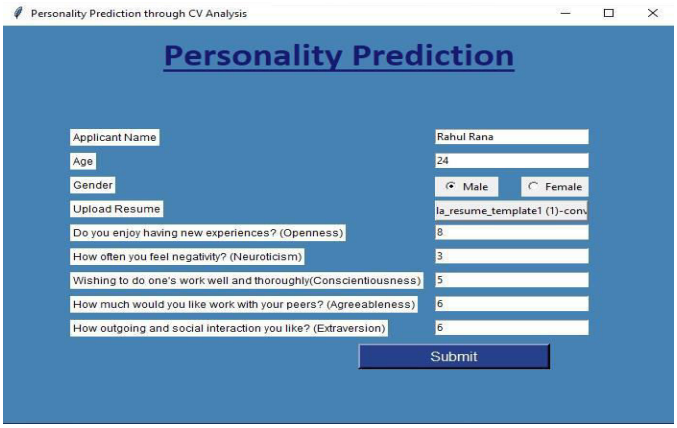
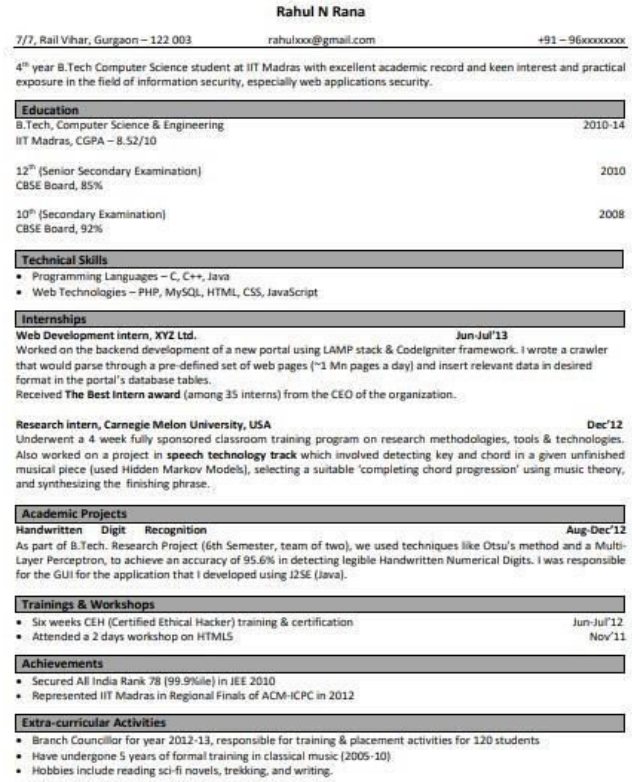


Figure 4. Personality Prediction Input Page

Once you press the "Predict Personality" button, a new pop window opens up wherein candidates have to insert their Name, Age and Gender. Then they have to answer five personality questions that are dependent on the Big Five Personality Model. Once candidate insert everything, they have to upload their resumes (.docx/.pdf/.txt etc.).



Rahul N Rana

7/7, Rail Vihar, Gurgaon – 122 003 rahulxxx@gmail.com +91 – 96xxxxxxxx

4th year B.Tech Computer Science student at IIT Madras with excellent academic record and keen interest and practical exposure in the field of information security, especially web applications security.

Education

B.Tech, Computer Science & Engineering 2010-14
IIT Madras, CGPA – 8.52/10

12th (Senior Secondary Examination) 2010
CBSE Board, 85%

10th (Secondary Examination) 2008
CBSE Board, 92%

Technical Skills

- Programming Languages – C, C++, Java
- Web Technologies – PHP, MySQL, HTML, CSS, JavaScript

Internships

Web Development intern, XYZ Ltd. Jun-Jul'13
Worked on the backend development of a new portal using LAMP stack & Codeigniter framework. I wrote a crawler that would parse through a pre-defined set of web pages (~1 Mn pages a day) and insert relevant data in desired format in the portal's database tables.
Received **The Best Intern award** (among 35 interns) from the CEO of the organization.

Research intern, Carnegie Melon University, USA Dec'12
Underwent a 4 week fully sponsored classroom training program on research methodologies, tools & technologies. Also worked on a project in **speech technology track** which involved detecting key and chord in a given unfinished musical piece (used Hidden Markov Models), selecting a suitable 'completing chord progression' using music theory, and synthesizing the finishing phrase.

Academic Projects

Handwritten Digit Recognition Aug-Dec'12
As part of B.Tech. Research Project (6th Semester, team of two), we used techniques like Otsu's method and a Multi-Layer Perceptron, to achieve an accuracy of 95.6% in detecting legible Handwritten Numerical Digits. I was responsible for the GUI for the application that I developed using J2SE (Java).

Trainings & Workshops

- Six weeks CEH (Certified Ethical Hacker) training & certification Jun-Jul'12
- Attended a 2 days workshop on HTML5 Nov'11

Achievements

- Secured All India Rank 78 (99.9%ile) in JEE 2010
- Represented IIT Madras in Regional Finals of ACM-ICPC in 2012

Extra-curricular Activities

- Branch Councillor for year 2012-13, responsible for training & placement activities for 120 students
- Have undergone 5 years of formal training in classical music (2005-10)
- Hobbies include reading sci-fi novels, trekking, and writing.

Figure 5. Sample CV

Once you submit the previous form, it will take you to the result window. This window consists of features that are extracted from the resume i.e., Email Id, Skills, Degree, No_Of_Pages, Total Experience etc. On the basis of the answers that the candidate gives in the previous window, the personality of the candidate is

predicted. On the result page, all the manipulated information and predicted result will be displayed.

Figure 6. Result Page

8. CONCLUSION

HR managers in big companies are facing difficulties in hiring the right candidate for a particular job profile due to the high volume of applications. Also, the hiring process takes lots of time and workforce. This model will predict the personality of the candidate based on the details provided in the resume/CV.

The project is easy and interactive to use and extracts all important features of the resume within seconds and easily predicts the personality of an applicant. This project would assist the human resource department in short listing the right candidate for a specific job profile.

9. FUTURE SCOPE

The project can be extended for commercial uses. It can be made more interactive where we can easily handle bulk data and represent it. We can improve the training model for various additional features that help us to predict more accurate results. We can add questionnaires' which ask some multiple-choice questions and auto calculate the various values instead of directly asking the five characteristic values.

Further, we can alter the current framework to perform a conclusion examination of web-based life information. Some more arrangement calculations of AI can be incorporated to give much better functionalities. Further, the effectiveness and execution of the application can be tried and broken down. It can be extended to various other commercial sectors where intake and elimination are in bulk like for Government Jobs.

ACKNOWLEDGMENTS

The keys to success are motivation and guidance. We would like to express our gratitude to all of our sources of inspiration.

We would like to grab this opportunity to thank Dr. S. D. Sawarkar, principal for encouragement and support he has given for our project.

We express our deep gratitude to Dr. A. P. Pande, head of the department who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide Prof. P.Y. Itankar for his/her encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. his/her concrete directions and critical views have greatly helped us in successful completion of this work.

We extend our sincere appreciation to all professors for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge them individually.

We are also thankful to all those who helped us directly or indirectly in completion of this work.

REFERENCES

[1] Barrick M. R. Mount. M. K, “Big five personality dimensions for any person to determine the personality of any reason” in 1991.

[2] Data, Vishnu M Menon Computer Rahul Nath H A, “A Novel Approach to Evaluate and Rank Candidates Recruitment Process by Estimating Emotional Intelligence through Social Media”, May 2016.

[3] Namrata Ramesh Satelkar, Arati Ashok Sawant, Madhuri Vishnu Vengurlekar, Vaibhavi Ganpat Sawant, “Personality Prediction System Through CV Analysis”.

[4] Qihang Zhang, Xiaodong Xu; “The Application of the Job Characteristics Model in E-HR System”.

[5] Juneja Afzal Ayub Zubeda, Momin Adnan Ayyas Shaheen, Gunduka Rakesh Narsayya Godavari, Sayed Zainul Abideen Mohd Sadiq Naseem, “Resume Ranking using NLP and Machine Learning”.

[6] Ashish K. Tripathi,”Personality Prediction with Social Behavior by Analyzing Social Media Data Survey”.

[7] Mr. Shaikh Rais F., Chandgude A.S.,” Personality Prediction using HR Assessment Tool” in Resincap Journal of Science and Engineering, Volume 4, Issue 7, July 2020

[8] Roshal Moraes, Larissa Lancelot Pinto, Mrunal Pilankar, Pradnya Rane, “Personality Assessment Using Social Media for Hiring Candidates” in IEEE 2020 3rd International Conference.