

PERSONALITY PREDICTION THROUGH CV ANALYSIS

Akshaya Deshmukh.

Computer science and Technology

Shivaji university

Kolhapur, Maharashtra, India.

Pratiksha Sangale.

Computer science and Technology

Shivaji university

Kolhapur, Maharashtra, India.

Ketki Tajane

Computer science and Technology

Shivaji university

Kolhapur, Maharashtra, India.

Shrija Saswade.

Computer science and Technology

Shivaji university

Kolhapur, Maharashtra, India

Abstract

Personality plays a crucial role in deciphering a person's calibre and work ethic. A person's personality becomes a vital resource for the organisation that he/she is employed in. Personality plays a major role in a person's life as well as in the development of an organisation. One of the ways to judge human personality is by observing the behaviour for a long time period or we can now use CV for analysing the personality. The recruiters manually shortlist/filter a candidate's CV as per their requirements. In this report, we present a system that automates the eligibility check of candidates in a recruitment process. To meet this, an online application is developed for the analysis of personality tests and candidate's CV. The system analyses professional personality based on the uploaded CV. The output of our system gives a prediction of the candidate. Further, the resulting predictions help in evaluating the qualities in the candidates by analysing the personality and expertise in different areas. The personality of any candidate makes it easier to evaluate his/her performance in the work and helpful in analysing the CV properly. Thus, the system provides a helping hand for the recruitment process so that the candidate's CV will be shortlisted and the fair decision will be made.

INTRODUCTION

Overview

In this report we have proposed a Personality Prediction Through CV Analyzer. Our project's primary goal is to make personality predictions based on a person's BIG FIVE TEST score. There are five major characteristics that define human personality known as Big 5 personality traits are classified as openness, conscientiousness, extroversion, agreeableness, and neuroticism [1]. Personality is a crucial factor that comes into play while adjudicating an individual. An apt judgement about a person's mindset is time consuming but we have put forward an approach that will predict an individual's personality and also aid in providing a recommendation concerning hiring and selection. We are proposing a Machine Learning (ML) and NLP (Natural Language Processing) based method to get an individual's personality score as well as identify their work and interpersonal skills.

As far as employment is considered, selecting the right candidate for the recruitment process from a vast pool of candidates has been a fundamental issue [2]. Many job seekers will apply for a position when the business offers specific employment requirements and information.. Therefore, job hopefuls fill out their online CV first before taking the test. In essence, the test we utilised is the BIG FIVE TEST. Based on the scores of each domain, we will get to know the personality of a person, i.e., openness, conscientiousness, extroversion, agreeableness, and neuroticism. To extract information from the CV like name, age, gender, etc., we used a simple resume parser. And we also used an important natural language processing tool, i.e., NLPT. There, after extracting information from the CV and score from the test, we predict the personality of the person.

Need of Project

Today's recruitment process is totally based on CV. CV plays a major role for show casting your skills and abilities. But nowadays it is difficult to choose the right candidate for the job profile as there are too many job

seekers applying for the same job. Personality plays a vital role in choosing the right candidate. Each personality type has different strengths and weaknesses that come with it and most jobs require a certain degree of interpersonal skills. Some careers are simply made for specific personality types such as a managing position requires an extrovert personality who knows how to communicate with ease. It needs lots of labour to go through all the CVs manually for choosing the right candidate for the job. They require some kind of filter for analysing the CVs. There are many CV analysers available in the market, but they only check for the completeness of the candidate's CV.

In this project we are trying a new approach of CV analysis that is prediction of personality. To solve the rising issue in the co-operative society we have chosen this topic "Personality prediction through CV analysis". This topic is recent and beneficial for society. So, we have chosen this topic

LITERATURE REVIEW

There are so many online recruitment web-sites that ask the candidates to upload their CV on their website. Some of those websites don't have techniques for resume screening. So company recruiters have to go through all of those CV's manually. This process is too lengthy and tiresome for them in order to recruit the capable candidates for the subsequent job profile.

It is advised that this arrangement should start with building an analyzer to sort through CV . Composing a workable analyzer will differentiate and separate the competitors that we are occupied with, it will help in getting rid of those applicants that are not suitable for the position from the very starting point. Thus, building an arrangement for ability recognition proof will be the basic requirement each framework ought to be work particular.

Firstly, resumes or CV's are uploaded into the system and candidates are shortlisted based on the administrator's request. The shortlisted candidates receive personality and ability test links, which they need to answer, and then they receive their scores. Based on

the scores and the department's requirements, candidates are shortlisted. Developed using a machine learning technique this system estimates the applicant's emotional aptitude through a psychometric analysis and predicts personality by using the OCEAN model.

It is a system wherein the candidate has to go through personality tests formulated by the organisation. The system uses the Big Five Personality dataset. A personality trait is a characteristic pattern of thinking, feeling, or behaving that tends to be consistent over time and across relevant situations. The Big Five—Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience—are a set of five broad, bipolar trait dimensions that constitute the most widely used model of personality structure.[3] A significant body of research has tested personality stability and change across the lifespan, as well as the impact of personality traits on important life outcomes, in terms of the Big Five.

We use the Big Five Personality Model for the prediction of personality. Generally this Big Five Personality Model is based on the "OCEAN" values as, "O: Openness, C: Conscientiousness, A: Agreeableness, E: Extraversion, and N: Neuroticism". The OCEAN values of a person remain quite stable throughout their lifetime. The reason behind why our system uses the Big Five Personality Model as the base methodology for conducting the psychometric analysis as it will give stable and accurate output values. The questionnaire has questions that target these 5 factors in question and takes inputs as values with reference to the model. [4]

Through these steps, we can save our time and energy dedicated to candidates who are not worthy for the position. CV analysis can be sometimes one of the most time-consuming tasks of a recruiter looking for considerable skill to perform accurately yet quickly. As always, the number of applicants grows as compared to recruiting resources which remains the same, or sometimes even reduces, thus most recruiters are opting for resume analysis software.

METHODOLOGY

This project is based on a classification problem so here we are going to use one of the classification algorithms in machine learning. The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observations into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc. Classes can be called as targets/labels or categories[13].

There are five labels: *Extraverted, Serious, Dependable, Responsible, Lively*. We need to classify the data of a particular person into these categories to predicate the personality.

The dependable values are above mentioned and the independent variables are age, gender, Openness, Conscientiousness, Extraversion, Agreeability, Neuroticism. The Problem statement falls under supervised Learning so the training of data is done through a dataset containing the above parameters.

For this we are going to use the following classification algorithms:

- **Logistic Regression (LR)**

A logistic regression model predicts a dependent data variable by analysing the relationship between one or more existing independent variables. Logistic regression uses the concept of predictive modelling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.

LR is similar to the linear regression model but LR uses *sigmoid function*.

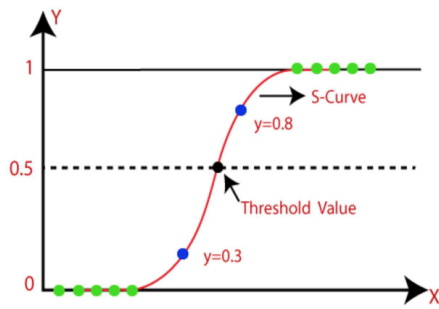


Fig.: LR graph

On the basis of the categories, Logistic Regression can be classified into three types:

- ❖ Binomial Logistic regression: There can be only two possible types of the dependent variables, such as 0 or 1.
- ❖ Multinomial Logistic regression: There can be 3 or more possible unordered types of the dependent variable, such as "banana", "orange", or "mango"
- ❖ Ordinal Logistic regression: There can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High"

As we have more than two labels of classification and they are independent variables so we can use the multinomial logistic regression for our project.

• K-Nearest Neighbour (KNN)

It is a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data. The KNN algorithm assumes that similar things exist in proximity. i.e., similar things are near to each other.

KNN algorithm can be used for Regression as well as for Classification, here we are going to use it for our classification problem

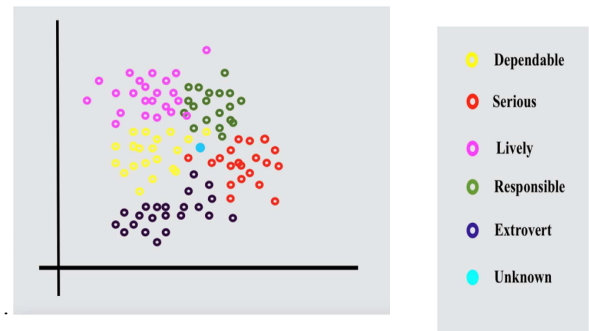


Fig.: KNN graph

First of all we have to specify and initialise K to a chosen number of neighbours. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances.

Pick the first K entries from the sorted collection and get the labels of the selected K entries. It returns the mode of the K labels.

System Architecture

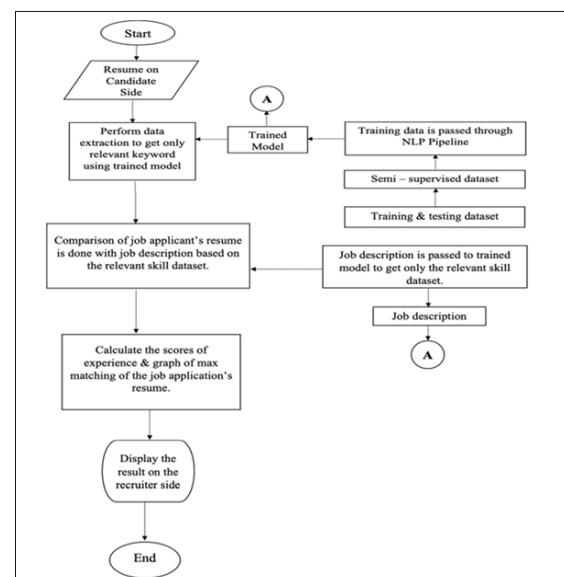


Fig.: Architecture of System

The above figure shows the basic architecture and blueprint of the project. This is a conceptual model that defines the structure, behaviour, and more views of our project.

SpaCy

spaCy is a popular open-source Python library for natural language processing (NLP) and text processing tasks. It provides efficient and scalable tools for a wide range of NLP tasks, such as tokenization, part-of-speech (POS) tagging, named entity recognition (NER), dependency parsing, text classification, and more. You also need to download the appropriate language models for the NLP tasks you want to perform using spaCy. You can install spaCy using pip, the Python package manager

Overall, spaCy is a comprehensive and powerful Python library for NLP tasks, with a large community of users and developers, extensive documentation, and numerous tutorials and examples available to help you get started.

EXPERIMENTATION

We have used python language for coding and python's tkinter library for creating a GUI. The techniques we have implemented for coding is that we have divided the code into two files based on coding and GUI parts. The "mPreditPersonality.py" file contains all the code for creating the prediction module and the "PersonalityPrediction.py" file for GUI. We have imported the "mPreditPersonality.py" file in the "PersonalityPerdiction.py" to use the functions in the "mPreditPersonality.py."

Histogram

A histogram is a graph showing *frequency* distributions. It is a graph showing the number of observations within each given interval.

We have plotted this graph for getting a better idea of the distribution of data in our data set. In Matplotlib, we use the hist() function to create histograms.

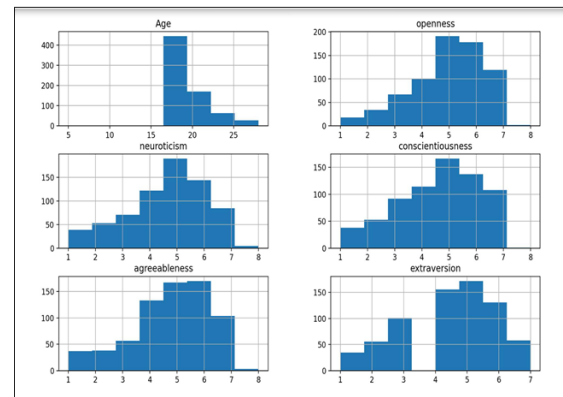


Fig.: Histogram

Correlation heatmap:

A correlation heatmap is a heat map that shows a 2D correlation matrix between two discrete dimensions, using coloured cells to represent data from usually a monochromatic scale. The values of the first dimension appear as the rows of the table while of the second dimension as a column. The colour of the cell is proportional to the number of measurements that match the dimensional value. This makes correlation heatmaps ideal for data analysis since it makes patterns easily readable and highlights the differences and variation in the same data. A correlation heatmap, like a regular heatmap, is assisted by a colour bar making data easily readable and comprehensible.

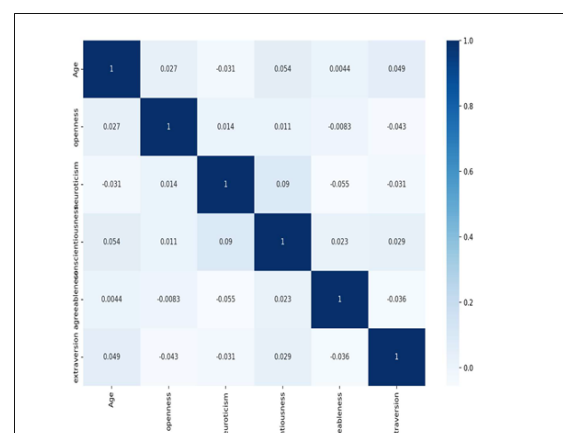


Fig.: Correlation heat-map

CONCLUSION

The traditional method is manual sorting of the resumes, where the recruiting team must go through a large number of applications. This is a time-consuming and challenging process for the recruiting team. However, they can know the details of candidates such as skills, experience and so on by going through their CVs or resumes. But they cannot identify the personality of the candidate. In today's age, the applicants and the employers can easily carry out the application and the recruiting process respectively.

Recommendations using machine learning techniques have been used for the analysis of the CV. One of the approaches has been mentioned which use a tool called "Career Mapper" for the recommendations of the CV. It checks for the completeness of the user profile [1]. But our project focuses on the personality of the candidate. Personality traits are the relatively enduring patterns of thoughts, feelings and behaviour that reflect the tendency to respond in certain ways under certain circumstances.

Here the system analyses 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.

Bibliography / References

- [1] Sudhir Bagade, Jayashree Rout, Pooja Yede, Nirmiti Patil "Personality Evaluation and CV Analysis using Machine Learning Algorithm." E-ISSN: 2347-2693 Vol.-7, Issue-5, May 2019
- [2] Seren Başaran and Obinna H. Ejimogu "A Neural Network Approach for Predicting Personality From Facebook Data" SAGE Open July-September 2021: 1–15
- [3] Mohammad Hossein Amirhosseini and Hassan Kazemian "Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator" MDPI, Received: 22 December 2019; Accepted: 12 March 2020; Published: 14 March 2020
- [4] Rutuja Narwade, Srujami Palkar, Isha Zade, Nidhi Sanghavi "Personality Prediction with CV Analysis" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue IV Apr 2022
- [5] M. H. Bornstein, M. E. Arterberry, K.L.Fingerman, & J. E. Lansford (Eds.) "Big Five Personality Traits" Soto, C. J. 30 March 2018.
- [6] Rutuja Narwade, Srujami Palkar, Isha Zade, Asst Prof. Nidhi Sanghavi "Personality Prediction with CV Analysis." International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 09 Issue: 04 , Apr 2022
- [7] Gayatri Vaidya, Pratima Yadav, Reena Yadav, Prof.Chandana Nighut "Personality Prediction By Discrete Methodology" IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 Volume 14 March-2018
- [8] Ilke Inceoglu (SHL Group, London) and Peter Warr (Institute of Work Psychology, Sheffield) "Personality and Job Engagement" journal of Personnel Psychology · January 2011
- [9] Jens B.Asendorpf and Susanne Wilpers "Personality effect on Social Relation" journal of Personnel and Social Psychology 1998, Vol. 74, No. 6,1531-1544
- [10] Avisha Anand, Mr. Sandeep Dubey "CV Analysis Using Machine Learning" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022

[11] <https://pypi.org/project/sklearn/> “Scikit-Learn (SKLearn)”

[12] <https://docs.python.org/3/library/tkinter.html>
“tkinter - Python interface to Tcl/Tk”

[13] <https://www.javatpoint.com/classification-algorithms-in-machine-learning>