AUTOMATED PERSONALITY PREDICTIVE MODEL FOR E-RECRUITMENT USING LOGISTIC REGRESSION TECHNIQUE

Oluyinka I Omotosho[1*], Motunrayo H. Yusuf [2]

*[*1]Department of Cyber Security Science, Ladoke Akintola University of Technology, Ogbomosho - Nigeria*
[2]Department of Computer Science, Ladoke Akintola University of Technology, Ogbomosho - Nigeria

Corresponding Author:
oiomotosho@lautech.edu.ng

Abstract

Human personality plays a vital role in individual's life as well as in the development of an organization. Common ways to evaluating human personality is by using standard questionnaires or by analyzing the Curriculum Vitae (CV). Traditionally, recruiters manually shortlist/filters a candidate's CV as per their requirements. In this work, a system that automates the eligibility check and aptitude evaluation of candidates in a recruitment process is developed. To meet this, need an automated system module is developed for the analysis of aptitude or personality test based on candidate's CV. The work presented in this paper determines the personality trait of applicants through CV analysis using Python upon which the Personality prediction Model is built. The result helps in evaluating the qualities in the candidates by analyzing personality trait and character of such candidate. The system provides serves as a better option for the recruitment process so that candidate's data can extracted from CV and shortlisted for the best decision via fair judgment.

Keywords: *Personal Traits, Personality Prediction Model, Behavioural Patterns, Logistic Regession and PYTHON.*

1. INTRODUCTION

In today's competitive environment Human Resource Management has become an integral part of the organizations because the practice, ethics and policies are directly linked with the overall client personality's performance and attainment of the goals. As far as employment is concerned, selecting the right candidate for the recruitment process from a vast pool of candidates has been a fundamental issue. Conducting personality and various technical eligibility evaluation tests, interviews, and group discussions have been traditional techniques. These traditional practices are very much time-consuming and may result in unfair choices of candidate. Also, Aptitude test followed by the interview is traditional practices for the recruitment process.

Personality is defined as the aspect with the set of perception, feeling and behavioural patterns that develop from botanical and external factors. Generally, there is no proper approval for definition of personality, mainly they focus on provocation and conceptual interactions. Even personality can be defined as traits that predict a person's behaviors [4] and it is the most important factor which reflects an individual, which keeps on varying and tackling them is a tedious task. Nevertheless, today there is a growing interest need in the personality traits of a candidate by the organization to better examine and understand the candidate's response to similar circumstances. Characteristics revealed in a certain pattern of behavior in a different bunch of situations form a personality trait.

Finding an individual personality trait and intelligence from his or her face plays a crucial role in inter-personal relationships but it is non-reliable in determining the personality trait. Mainly personality prediction depends on person's nature and results has shown that personality traits such as positive social interaction, capacity of mutual respect, creativity, and many other traits cannot be just estimated by personal interactions.

In this work, Automated personality prediction System that determines the personality traits of a recruitment candidate is developed through the analysis of submitted Curriculum Vitae (CV). Although, personality prediction depends on person's nature, the automated personality prediction is carried by comparing user's personality against the standard Five Personality Trait Model Criteria, which include, conscientiousness, extraversion, agreeableness, neuroticism and openness to experience [2]. The professional eligibility of a candidate is also checked, based on the entries in online submission of Curriculum Vitae (CV) submitted by the applicants. This system uses Natural Language Processing (NLP) Technique in PYTHON. It uses logistic regression for training the model and pyres parser module for parsing the information from resume which is built using nltk and spacy module in PYTHON.

II. SURVEY OF LITERATURE

In today's competitive environment Human Resource Management has become an integral part of the organizations because the practice, ethics and policies are directly linked with the overall client personality's performance and attainment of the goals. Attracting qualified individuals to apply for employment vacancies is a goal that organizations strive for. Periodic worker shortages and low unemployment rates have made attracting qualified employers a concern for large organizations. Although attracting qualified client personalities is an important first step, it is far from being the last stage of the recruitment process. Once employers are attracted to the organization, their interest must be held throughout the selection process, and finally, their attraction be maintained so they will accept a job application [3].

Personality is defined as the aspect with the set of perception, feeling and behavioural patterns that develop from botanical and external factors. Generally, there is no proper approval for definition of personality, mainly they focus on provocation and conceptual interactions. Even personality can be defined as traits that predict a person's behaviours. Personality identification was the old approach to identify the user's personality but now with the help of data mining techniques accuracy of this prediction has improved a way lot than old techniques.[4].

In many studies, personality structure is modeled based on factors or traits and individual personality is explained by trait value. The most commonly used is the Big Five Personality Trait model, which includes conscientiousness, extraversion, agreeableness, neuroticism and openness to experience [2].

A Realistic Job Preview (RJP) is a means by which employers, or recent hires, are exposed to the requirements of their new client personalities as opposed to the interview process, which is often used as a means of selling the position RJPs do not merely emphasize the positive aspects, they are intended to give the candidate a more balanced view of the client personalities. This includes exposure to characteristics of the client personalities that might be considered objectionable [5].

Curriculum vitaes involve the presentation of both favorable, unfavorable client personalities and related information to client personalities [6]. RJPs differ in the format, the timing, and the amount of negative information presented. Some RJPs are presented when the employer makes initial contact with the organization, others after the offer has been extended, and still others after employment begins. The amount of negative information presented can vary from medium to high [7].

Logistic regression is a supervised learning algorithm used to predict a dependent categorical target variable. It is a valuable research method because of its versatile application to different study contexts. For instance, one may wish to

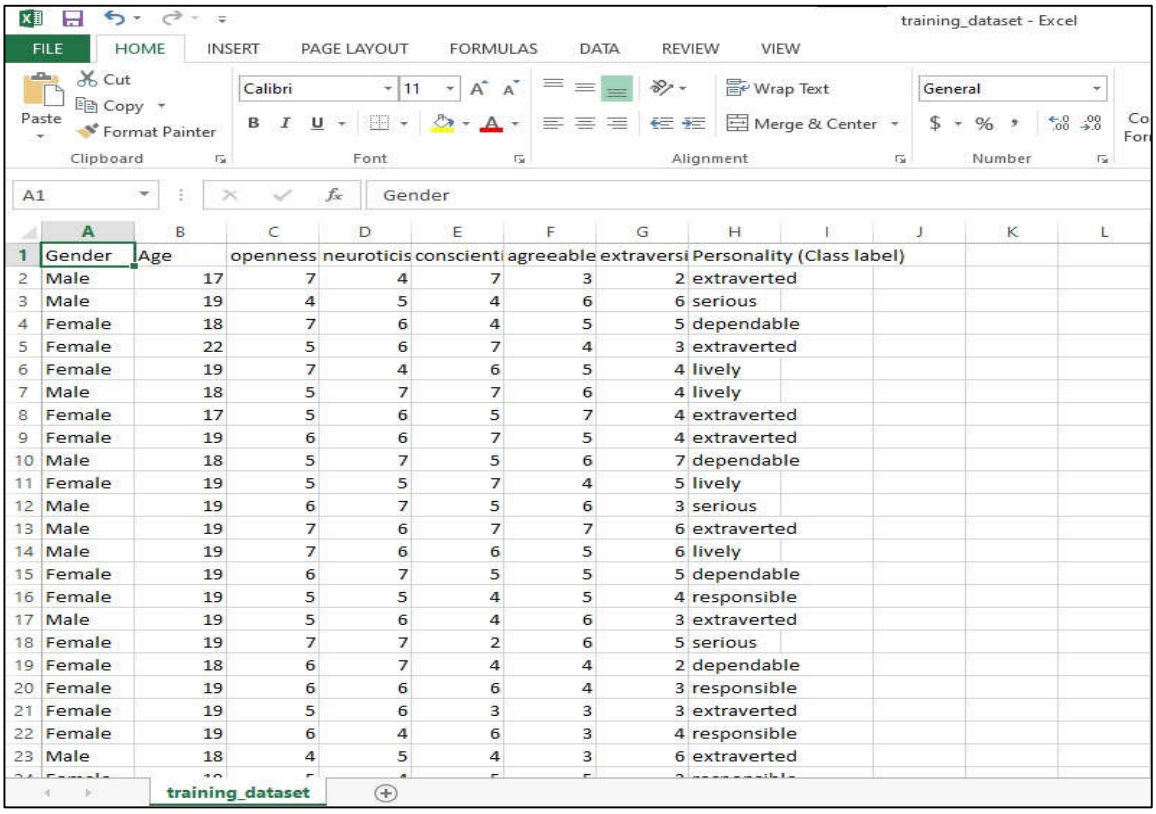
examine associations between an outcome and several independent variables (also commonly referred to as covariates, predictors, and explanatory variables), or one might want to determine how well an outcome is predicted from a set of independent variables [9][10][11].

III METHOD

In this section, the Developed Personality Prediction System through CV Analysis (in a recruiting system)" which can be used to determine and evaluate the personality trait of a potential employee is discussed. The methodologies adopted for the development of the system includes; Extraction and Integration of CV Dataset; System Model Development and System User Interface Design. Based on these the System is able to shortlisting the candidate based on the criteria of the big five personality traits and determine the personality of the candidate based on the result of the personality model.

3.1 Extraction and Integration of CV Dataset

Extracted CV dataset that was used in training this model is obtained from https://www.kaggle.com/datasets [8]. It uses seven different values for training the model, which are; gender, age, and the five personality characteristics (openness, conscientiousness, extraversion, agreeableness, neuroticism) [8]. The raw data acquired in the stage of data acquisition is then further integrated by creating a spreadsheet in Microsoft Excel 2013, as shown in Figure 1, for easy queries.



	A	B	C	D	E	F	G	H
1	Gender	Age	openness	neuroticis	conscienti	agreeable	extraversi	Personality (Class label)
2	Male	17	7	4	7	3	2	extraverted
3	Male	19	4	5	4	6	6	serious
4	Female	18	7	6	4	5	5	dependable
5	Female	22	5	6	7	4	3	extraverted
6	Female	19	7	4	6	5	4	lively
7	Male	18	5	7	7	6	4	lively
8	Female	17	5	6	5	7	4	extraverted
9	Female	19	6	6	7	5	4	extraverted
10	Male	18	5	7	5	6	7	dependable
11	Female	19	5	5	7	4	5	lively
12	Male	19	6	7	5	6	3	serious
13	Male	19	7	6	7	7	6	extraverted
14	Male	19	7	6	6	5	6	lively
15	Female	19	6	7	5	5	5	dependable
16	Female	19	5	5	4	5	4	responsible
17	Male	19	5	6	4	6	3	extraverted
18	Female	19	7	7	2	6	5	serious
19	Female	18	6	7	4	4	2	dependable
20	Female	19	6	6	6	4	3	responsible
21	Female	19	5	6	3	3	3	extraverted
22	Female	19	6	4	6	3	4	responsible
23	Male	18	4	5	4	3	6	extraverted

Figure 1: Integrated CV Data **Set (**https://www.kaggle.com/dataset)

3.2 System Model Development

The Developed System Model uses Natural Language Processing (NLP). The Model serve as the Backbone of the entire system. It involves Model and Testing;

3.2.1 Model Training was achieved by reading the dataset for training the model from a CSV file and build a model using Logistic Regression. It uses 75% of the values associated with the seven attributes for training the model, which are; gender, age, and the five personality characteristics (openness, conscientiousness, extraversion, agreeableness, neuroticism) collected.

3.2.2 Model Testing: 25% of the data was used to test the efficacy of the system in predicting the personality of a person. The array of values contained in gender, age and other five personality characteristics was passed to the trained model. This was carried out the designed User Interface.

3.3 System User Interfaces

Python GUI library named Tkinter was used for the design of the user interface where the admin can interact with the system by prompting the admin for the uploading of the Curriculum Vitae Integrated data and entry of necessary data into the system. The System User Interface includes three modules namely; the Home Page, Admin/HR inputs Pages and Output/Result Page.

3.3.1 Home page, Admin/HR Panel Figure 2 depicts the Home Page where admin prompts the personality prediction button to start the prediction of the personality of an applicant.

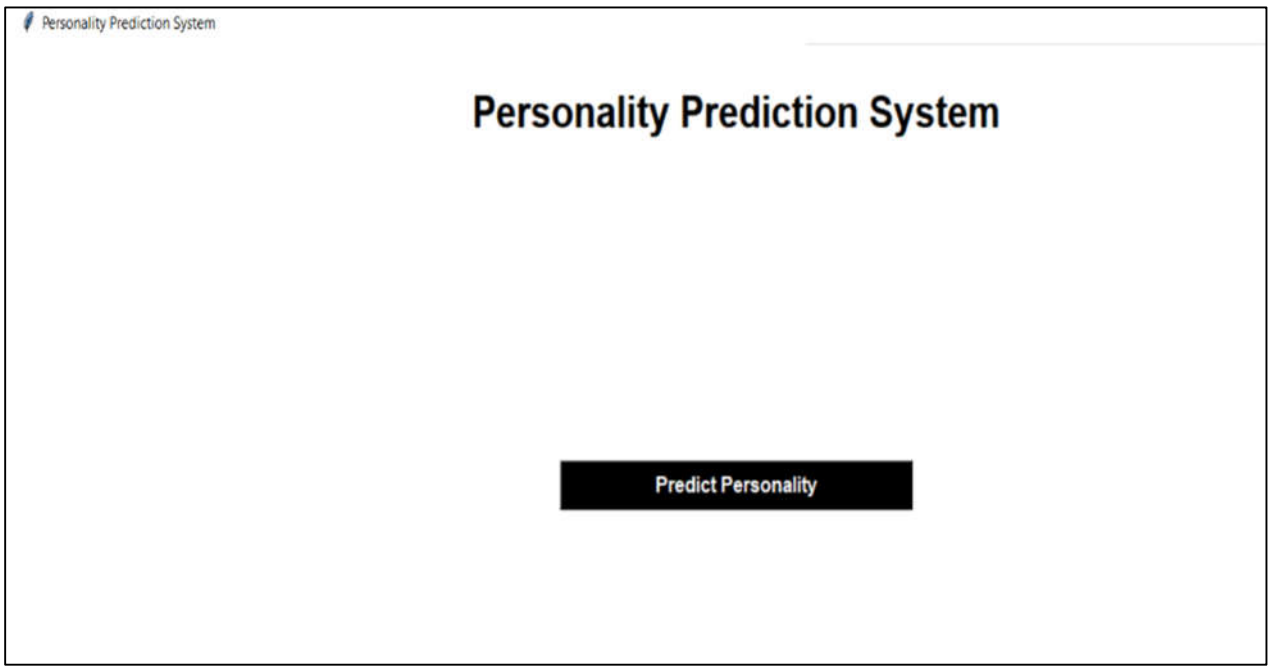


Figure 2: Home page of the system

3.3.2 Admin/HR Input Page

After the Admin has clicked the personality prediction button shown in Figure 3, Data are entered. The admin provides answers to the technical questions and use the information to predict the personality of the applicant based on the trained model.

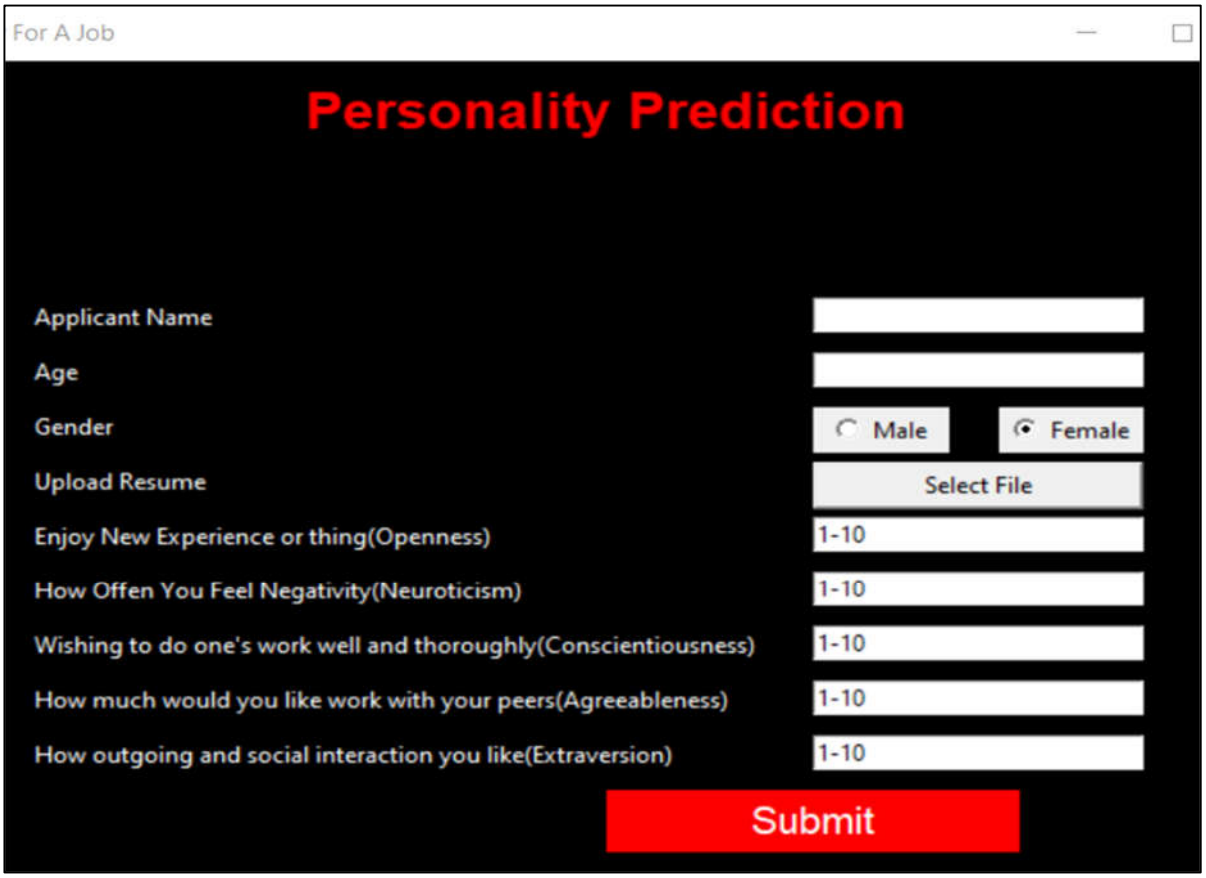


Figure 3: Admin/HR inputs Page

3.3.3 Result/Output Page: shows all the parsed information and predicted personality on the GUI window along with the definition of each personality characteristic. The Result is printed on a full screen window as shown in Figure 4.

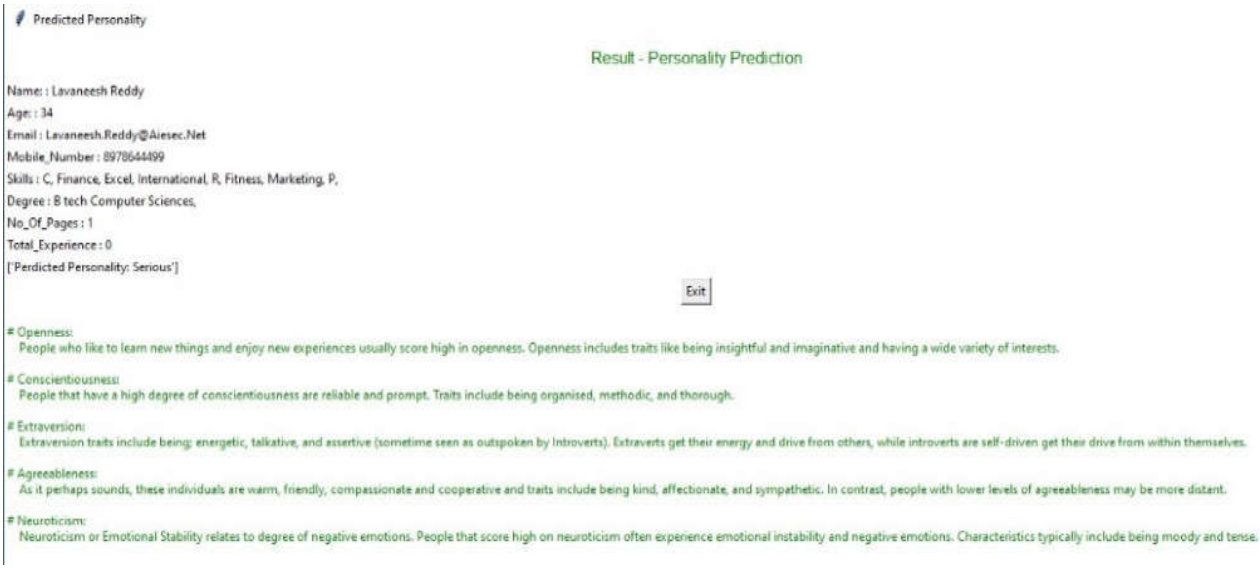


Figure 4: Resulting page after parsing the applicant's information

3.3.4 System Flowchart: The system Flow chart showing the implementation steps of the developed model is depicted in Figure 5.

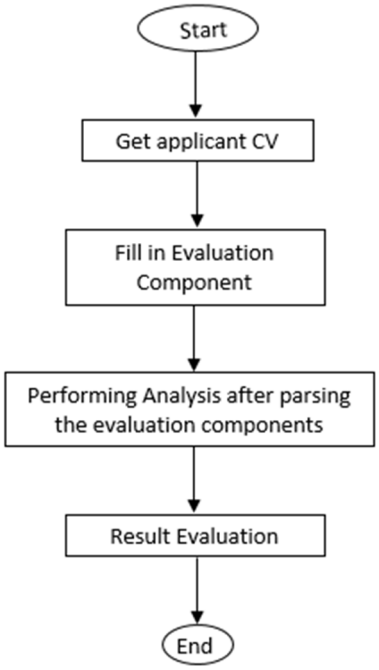


Figure 5: Flowchart of Implementation Steps

CONCLUSION

The system built in this work predicts personality of intending employees. The system can be integrated into the recruiting companies 'existing system website. Thus, the genuineness of employee guaranteed. It's also able to predict the personality of an applicant and display the category by using Natural Language Processing (NLP) technique. The User Interface of the project is done using Python GUI library named Tkinter and Logistic Regression for training the model and pyresparser module for parsing the information from resume which is built using nltk and spacy module in Python. With this system the Admin/HR can fill appropriate field and upload the CV of the candidates into the system then the system can analyze and predict the personality of the applicant. As a benefit, reduces stress, saves time and costs for the Admin.

REFERENCES

[1]. Kaulisch, M., & Böhmer, S. (2010). Inequality in academic careers in Germany: Indications from postdoctoral careers. In Understanding Inequalities in, through and by Higher Education (pp. 105-121). Brill.
[2]. De Raad, B. (2000). The Big Five Personality Factors: The psycholexical approach to personality. Hogrefe & Huber Publishers.
[3]. Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The review of financial studies, 21(2), 785-818.

[4]. Manasi Ombhase, Student, PCE, Prajakta Gogate, Student, PCE, Tejas Patil, Student, PCE, Karan Nair, Student, PCE and Prof. Gayatri Hegde, Faculty, PCE, "Automated Personality Classification Using Data Mining Techniques"
[5]. Donnelley, R. R., & Fitzmaurice, M. (2006). Designing Modules For Learning Pp. 99-110 O'neil, G., Moore, S, Mc. Mullin, B. Emerging Issues In The Practice Of University Learning And Teaching.
[6]. Rynes, S. L., Gerhart, B., & Minette, K. A. (2004). The importance of pay in employee motivation: Discrepancies between what people say and what they do. Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, 43(4), 381-394.
[7]. Wanous, M., Procter, B., & Murshid, K. (2009). Assessment for learning and skills development: The case of large classes. European Journal of Engineering Education, 34(1), 77-85.
[8]. https://www.kaggle.com/datasets
[9]. Stoltzfus, J. C. (2011). Logistic regression: a brief primer. Academic emergency medicine, 18(10), 1099-1104.
[10]. Darlington RB. Regression and Linear Models. Columbus, OH: McGraw-Hill Publishing Company, 1990.
[11]. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5th ed. Boston, MA: Pearson Education, Inc., 2007.