IJEMD-CSAI, 4 (1) (2025) https://doi.org/10.54938/ijemdcsai.2025.04.1.404

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

Personality Prediction System via Curriculum Vitae (CV) Analysis Using Natural Language Processing (NLP) and Logistic Regression

Jesse Ismaila[1*] , Sabo Victoria[1]

1.Federal University Wukari, Nigeria.

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 or 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 I 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 an automated model 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 curate. This model employs a machine learning algorithm namely "Logistic Regression" which helps to choose fair decisions to recruit a suitable candidate, and "Natural Language Processing (NLP)" uses techniques with the help of Natural Language Toolkit (NLTK) libraries to process and categorize the data. Also, the use of graphs to analyze a candidate's success makes it easier to assess his or her personality and aids in proper CV analysis. As a result, the framework lends a hand in the recruitment process, allowing the candidate's CV to be shortlisted and a reasonable decision to be reached.

Keywords: Prediction, Machine Learning, Logistic Regression, Natural Language Processing and Curriculum Vitae (CV).

INTRODUCTION

The word *personality* derives from the Latin word "*persona*" which refers to a mask worn by actors to act. However, Personality is much more than a mask now, it could determine whether a person is suitable for a particular job profile. It tells us if a human is capable enough to lead, influence,

Email Addresses: mazadujesse@gmail.com (Jesse) , victoriasabo@gmail.com (Sabo)

and communicate effectively with others. The first step of recruitment is the job application which consists of personal details, experience, and most importantly CV. Companies typically receive thousands of applications per job opening and have a dedicated team of screeners to select qualified candidates. It is very difficult to manually go through the CV of all applicants. Many candidates get filtered out in the first round itself based on suitability, improper CV not being skilled enough. Hiring the right candidate is a very difficult task as no candidate might be perfect, some might not be skilled enough or some might not have the right personality [30]

When the hiring and selection process comes into the picture, filtering through applicants and picking an individual who is apt for the job is a crucial and cumbersome process, personality is a very important factor that comes into play while adjudicating an individual. An apt judgment about a person's mindset is difficult but we have put forward an approach that will evaluate an individual's personality and aid in providing a recommendation concerning hiring and selection. In this paper, we propose a Natural Language Processing NLP and Machine Learning (ML) based method to get an individual's personality score as well as identify their work and interpersonal skills. The personality score would be identified using a personality test dataset provided by Kaggle and their skills will be determined with the help of CV analysis.

The existing software has difficulties in predicting the personality scores of an employee. 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 model. This study aims to implement a Natural Language Processing model and Logistic Regression model that can be used to predict human personality with better performance with the following objectives: To implement the Big Five Personality Trait model for personality predictions. To increase the performance of the existing models to be used with confidence when deployed to a system and to test the model.

LITERATURE REVIEW

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. Logistic Regression was used in the biological sciences in the early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical, e.g., to predict whether an email is spam (1) or not (0) (Saishruthi S., 2018). Below is a diagrammatical description of how it works.

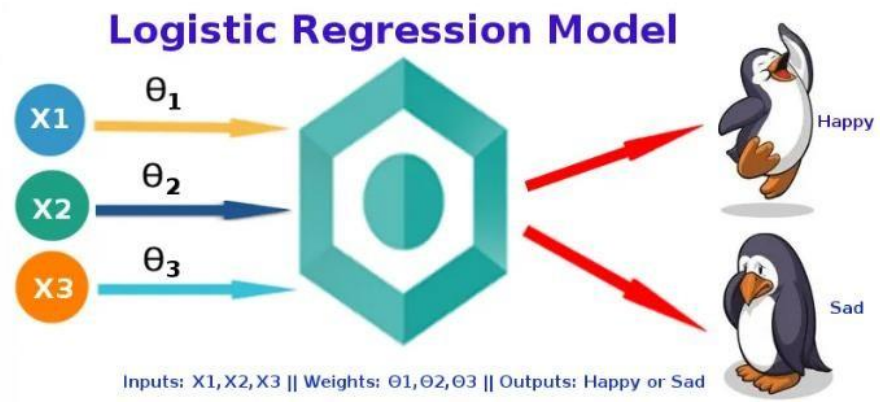

Figure 2.0 Logistic Regression model (Saishruthi S., 2018).

Types of Logistic Regression

i. Binary Logistic Regression: The categorical response has only two 2 possible outcomes. Example: Spam or Not
ii. Multinomial Logistic Regression: Three or more categories without ordering. Example: Predicting which food is preferred more (Veg, Non-Veg, Vegan) iii. Ordinal Logistic Regression: Three or more categories with ordering. Example: Movie rating from 1 to 5

Decision Boundary

To predict which class a data belongs to, a threshold can be set. Based upon this threshold, the obtained estimated probability is classified into classes.

Decision boundaries can be linear or non-linear. Polynomial order can be increased to get complex decision boundaries (Saishruthi, 2018).

Cost Function

Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a non-convex function of parameters (theta). Gradient descent will converge into the global minimum only if the function is convex.

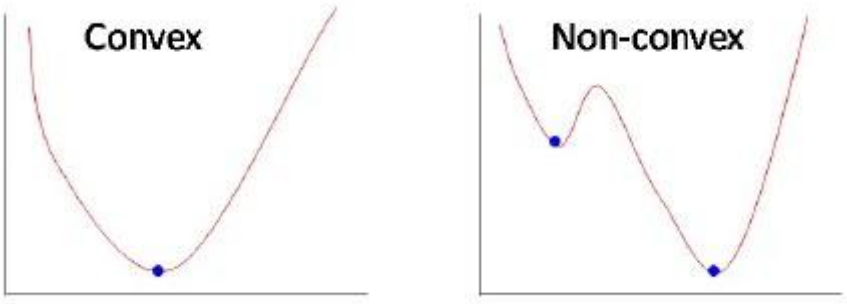

Figure 2.1 Convex and non-convex cost function (Saishruthi S., 2018).

Sigmoid Activation Function

The Sigmoid function also called the logistic Sigmoid, is defined as

$$f(x) = 1/(1+e^{(-x)})$$

It is prone to vanishing gradients around the origin, the gradient is large, but forlarge or small ***x***, the gradient is nearly 0. Hence, the Sigmoid function is not often used as an activation functionin hidden layers. However, the Sigmoid function is useful in the output layer of classification models. Itwill return a prediction of the sample belonging to each available label. (Leeuwen b.x. 2020).

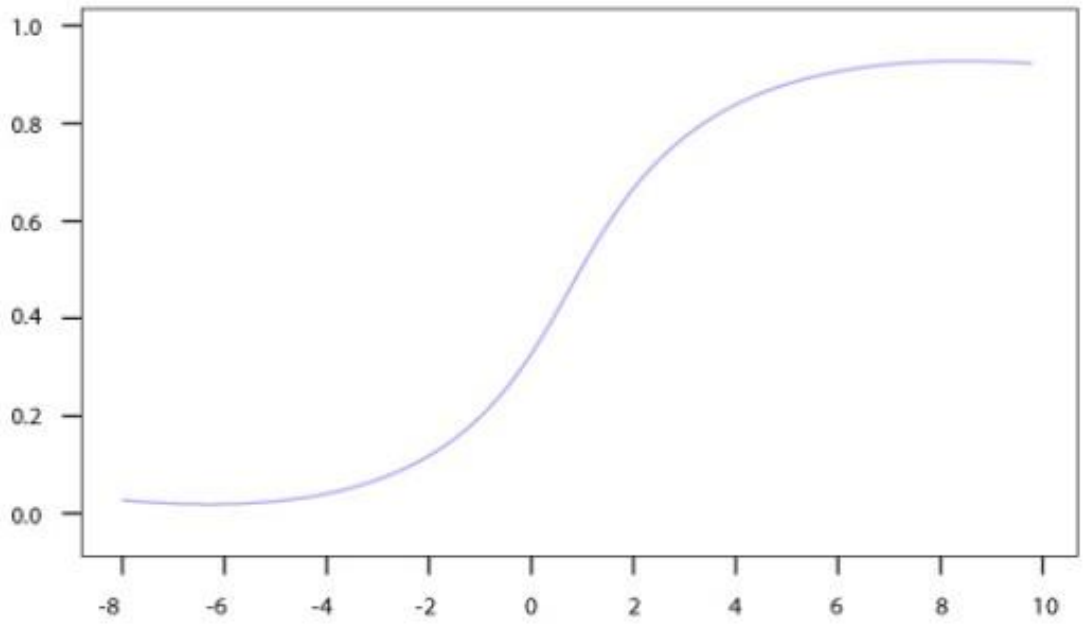

Figure 2.2. Sigmoid activation function *(*Leeuwen B.X. 2020).

Curriculum Vitae (CV)

A curriculum vitae or "CV" may be a new and unfamiliar term for you. A CV is a structured outline of your education, publications, projects, awards, and employment history. It can vary in length from one to several pages, depending upon the variety and number of your experiences. A resume, in contrast, is normally a brief, one-page overview of your job experiences. A curriculum vitae is used primarily when applying for academic, educational, scientific, or research positions. It is also applicable when applying for fellowships or grants. (Smith J, 2006).

As with a resume, you may need different versions of a CV for different types of positions. Based on your education and experience, you will need to include sections that are most relevant to your career goal. We suggest you ask a graduate advisor or other faculty member to review your vitae for specific content and format, as they are experts in their field. In addition to the basics of your name, contact information, education, and experience, a CV may include: i. Research Experience

ii. Teaching Experience
iii. Related Leadership Experience
iv. Internship or Rotation Experience
v. Presentations & Publications
vi. Awards, Grants, and Fellowships
vii. Professional Affiliations
viii. Licensure ix. Skills (lab techniques and equipment, computer programs, languages, etc.)

Things Employers look for in a Curriculum Vitae.

While each employer will have an ideal candidate in mind, the following are traits that most employers will look for and/or be cautious of when reviewing a CV:

i. Signs of achievements. ii.
Willingness to work hard. iii.
Specifics in job descriptions

iv. Patterns of stability and career direction
v. Completeness of resume
vi. Neatness/Professional presentation (including grammar and spelling)

Natural Language Processing (NLP)

Natural language processing (NLP) is the intersection of computer science, linguistics, and machine learning. The field focuses on communication between computers and humans in natural language and NLP is all about making computers understand and generate human language. Applications of NLP techniques include voice assistants like Amazon's Alexa and Apple's Siri, but also things like machine translation and text-filtering (Niklas and Mathew, 2022). Natural language processing studies interactions between humans and computers to find ways for computers to process written and spoken words similar to how humans do. The field blends computer science, linguistics, and machine learning (Niklas et al., 2022).

Why Natural Processing Language is Difficult

Human language is special for several reasons. It is specifically constructed to convey the speaker/writer's meaning. It is a complex system, although little children can learn it quickly (Niklas et al., 2022). Another remarkable thing about human language is that it is all about symbols. According to Chris Manning, a machine learning professor at Stanford, it is a discrete, symbolic, categorical signaling system. This means we can convey the same meaning in different ways (i.e., speech, gesture, signs, etc.) The encoding by the human brain is a continuous pattern of activation by which the symbols are transmitted via continuous signals of sound and vision. (Niklas et al., 2022)

Understanding human language is considered a difficult task due to its complexity. For example, there are an infinite number of different ways to arrange words in a sentence. Also, words can have several meanings and contextual information is necessary to correctly interpret sentences. Every language is unique and ambiguous. Just look at the following newspaper headline "The Pope's baby steps on gays." This sentence has two very different interpretations, which is a pretty good example of the challenges in natural language processing. (Niklas et al., 2022)

Syntactic and Semantic Analysis

The syntax is the grammatical structure of the text, whereas semantics is the meaning being conveyed. A syntactically correct sentence, however, is not always semantically correct. For example, "cows flow supremely" is grammatically valid (subject—verb—adverb) but it doesn't make any sense. (Niklas, 2022).

Syntactic Analysis

The syntactic analysis also referred to as syntax analysis or parsing is the process of analyzing natural language with the rules of formal grammar. Grammatical rules are applied to categories and groups of words, not individual words. The syntactic analysis assigns a semantic structure to text (Niklas et al., 2022).

Semantic Analysis

Semantic analysis is the process of understanding the meaning and interpretation of words, signs, and sentence structure. This lets computers partly understand natural language the way humans do. I say this partly because the semantic analysis is one of the toughest parts of natural language processing and it's not fully solved yet(Niklas et al., 2022)

Natural Language Processing Techniques for Understanding Text Parsing What is parsing? According to the dictionary, to parse is to "resolve a sentence into its parts and describe their syntactic roles." That nailed it, but it could be a little more comprehensive. Parsing refers to the formal analysis of a sentence by a computer into its constituents, which results in a parse tree showing their syntactic relation to one another in visual form, which can be used for further processing and understanding.

Stemming

Stemming is a technique that comes from morphology and information retrieval which is used in natural language processing for pre-processing and efficiency purposes. It's defined by the dictionary as to "originate in or be caused by."

Stemming is the process of reducing words to their word stem. A "stem" is part of a word that remains after the removal of all affixes. For example, the stem for the word "touched" is "touch." "Touch" is also the stem of "touching," and so on.

Text Segmentation

Text segmentation in natural language processing is the process of transforming text into meaningful units like words, sentences, different topics, the underlying intent, and more. Mostly, the text is segmented into its component words, which can be a difficult task, depending on the language. This is again due to the complexity of human language. For example, it works relatively well in English to separate words by spaces, except for words like "icebox" that belong together but are separated by a space. The problem is that people sometimes also write it as "icebox".

Named Entity Recognition

Named entity recognition (NER) concentrates on determining which items in a text (i.e., the "named entities") can be located and classified into predefined categories. These categories can range from the names of persons, organizations, and locations to monetary values and percentages.

Relationship Extraction

Relationship extraction takes the named entities of NER and tries to identify the semantic relationships between them. This could mean, for example, finding out who is married to whom, that a person works for a specific company, and so on. This problem can also be transformed into a classification problem and a machine learning model can be trained for every relationship type.

Sentiment Analysis

Using sentiment analysis, we want to determine the attitude (i.e., the sentiment) of a speaker or writer concerning a document, interaction, or event. Therefore, it is a natural language processing problem where text needs to be understood in the underlying intent. The sentiment is mostly categorized into positive, negative, and neutral categories. Through the use of sentiment analysis, for example, we may want to predict a customer's opinion and attitude about a product based on a review they wrote. Sentiment analysis is widely applied to reviews, surveys, documents, and much more.

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, and automatic text summarization; 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.

Big Five Personality Trait Model

In the latter half of 1950, the Big Five model was released. However, the current model was implemented in 1990. The prototype was given the moniker "The Big Five" by Lewis Goldberg, a research worker at the Oregon Research Institute. Currently, this prototype is thought to be accurate, and the accompanying personality gradation is routinely used by many businesses and in psychological research. The Big Five Personality Trait model, also known as the OCEAN model, is a widely used framework for assessing personality in psychology. It provides a summary of a person's overall character. (Suyash, 2022) It contains the following:

i. **Openness:** This trait encompasses characteristics such as insight, imagination, sensitivity, attentiveness, and curiosity. People who score high in openness are typically curious, creative, and open to new experiences.

ii. **Conscientiousness:** This trait relates to a person's level of care, discipline, deliberation, and diligence. People who score high in conscientiousness are typically goal-oriented and have good self-control and organizational skills.

iii. **Extroversion:** This trait relates to a person's level of emotional expression and assertiveness. Extroverted people are outgoing and comfortable interacting with others and tend to be enthusiastic and excitable.

iv. **Agreeableness:** This trait relates to a person's level of generosity and cooperativeness. People who score high in agreeableness are typically kind, trusting, and considerate.

v. **Neuroticism:** This trait relates to a person's emotional stability and tendency to experience negative emotions such as anxiety and depression. People who score high in neuroticism are more easily prone to mood swings and may be more sensitive to stress.

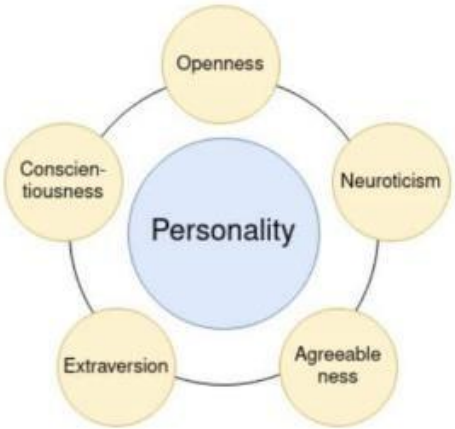

Figure 2.3. Big Five Personality Trait model [20]

Related Works

[38] proposed a system wherein the system analyses vast user temperaments and behaviors, and based on the patterns observed, it stores its user characteristics patterns in a database. This system uses the Naive Bayes algorithm, and Support Vector Machines. This system is useful to predict new user personality on a large scale based on personality data stored by the classification of previous user data.[56] presented a system that automates the eligibility screening and assessment of applicants in the process of recruiting, wherein the system goes through a candidate's CV and parses information from a CV. The proposed system uses Machine Learning Approach. The result of the system would help in shortlisting a candidate, rejecting, or asking for more information to assess the eligibility of the candidates.

[9] propose several new research directions regarding the problem of Automated Personality Classification (APC). Firstly, we investigate possible improvements to the existing solutions to the problem of APC, for which we use different combinations of the APC corpora, psychological trait measurements, and learning algorithms. Afterward, we consider extensions of the APC problem and the related tasks, such as dynamical APC and detecting personality inconsistency in a text. This entire research was performed in the context of social networks and the related data mining mechanisms. [65] presented an automated resume evaluation tool called "Career Mapper" and illustrate how Career Mapper evaluates professional resumes. Further, it demonstrated key examples of how Career Mapper makes recommendations for different sections of the user's resume by scanning through and deriving insights from a large pool of other resumes. Using this automated resume evaluation tool, users can quickly have their resumes evaluated and appropriate recommendations displayed in a user-friendly web interface.

[61] presented a protocol for advanced psychometric analysis that is based on the Standards of Educational and Psychological Testing (the Standards), considered "best practice" in instrument development and psychometrics. The authors illustrated the newly developed psychometric protocol using the Alberta Context Tool (ACT) as an exemplary survey to which it can be applied. Mehrin et al., (2015) proposed an automated job recruiting process along with psychometric analysis. The focus had been given to automating the job application and CV processing system. The authors made a social networking website for job seekers and employers are proposed to develop which will forward CVs to the desired company or organizations automatically by matching the required criteria instead of the traditional job searching and application process.

[21] proposed a novel approach to predict the personality of a person in online hiring processes, to automate candidate pre-screening. It proposed a series of procedures and strategies that will make the recruiting process easier, more reliable, and more efficient. Our key goal is to restrict the rounds of interviews and background analysis of the applicants exclusively to those capable of being recognized by the organization in terms of their attributes and the criteria of the organization. [49] examines the five-factor model, a tool used for dimensionally studying personality. Aspects of the model given attention include the specific variables in the model, other related models, and clinical applications of the model. The quality of the model is then evaluated based on five criteria: compatibility, originality, application, taxonomy, and universality. [33]presented a Neural Network Approach based on the Big Five Test to predict the personality of individuals depending on tweets published on Twitter by extracting meta-attributes from tweets. Which are used to analyze one's social behavior? The authors followed a four-step process which is Data Collection from tweets, Preprocessing, Transformation, and Classification. Although neural networks are used to predict personality there are limitations such as countering fake information, automatic analysis of tweets,

and relying on just Twitter is not enough to predict someone's personality but only user behavior and trends.

Tanuj, et al., (2022) built a system using different machine learning algorithms for predicting the personalities of the candidates using Natural Language Processing. At least, Random Forest achieves better accuracy than remaining algorithms such as KNN, Logistic Regression, Support Vector Machine, and Naïve Bayes. The system will predict the personality based on their ranking policy. It will rank the skills, experience, and other aspects of the uploaded resume. They also receive the result in the form of a graphical representation. [3] built a model using the Random Forest Algorithm, Support Vector Machine, and Weighted Majority Voting algorithm. Firstly, resumes or CVs 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. [23] developed a system using a machine-learning technique known as Logistic Regression. The system estimates the applicant's emotional aptitude through a psychometric analysis and predicts personality by using the OCEAN model. The details of the candidates are protected by using a password encryption algorithm, and the passwords are known only to the required individuals. The candidates are known only to the required individuals. The candidates can know whether they are selected for the interview via dashboard and SMS.

[54] built a proposal that evaluates the right candidates based on the edibility score obtained by attempting the aptitude test and uploading a CV or Resume. The model is built using the TF-IDF algorithm. Based on the scores, the candidate's qualities can be analyzed, and the graphical representation of the candidate's scores helps to evaluate their personalities and analyze their CV properly. [33] displayed a Neural Network Approach based on the Big Five Test to estimate a person's personality based on tweets that were published on Twitter by identifying meta-attributes from tweets, which are applied to social behavior analysis. While neural networks have been to predict personality, there are drawbacks such as detecting bogus news, automatically analyzing tweets and the inadequacy of using merely Twitter to predict users' behavior and trends rather than personality. [40] evaluated resumes of persons using Neural Language Processing and Machine Learning, then converting them to HTML, then reverse engineering them into HTML code, and finally performing segment finalization and qualification feature extraction. The model takes data from a CV and divides it into segments according to the values. Multivariate logistic regression was used to classify the CVs. However, the size of the dataset was quite little. [11] used various machine learning algorithms and analyses which one among them provides the best accuracy with a wide array of data provided. The algorithms and respective accuracies include Random Forest (71%), Logistic Regression (62%), KNN (64%), and Support Vector Machine (63%).

Summary of Literature Review and Knowledge Gap

Logistic Regression is a type of statistical analysis often used for predictive analytics and modeling and extends its applications in machine learning. It is a predictive analysis. 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. It will be used to implement the Big Five Personality Trait model. [11] used various algorithms and still got a low accuracy as some other authors. In this project Logistics Regression will be used the Big Five Personality Trait model, and NLP will be used to extract the skills from the CV supplied.

METHODOLOGY

This project proposes a method after a crucial study of previous work done in this field and insights based on the literature survey of respective research papers in context to personality prediction from Twitter and other means, which comprises of a personality prediction model based on Logistic Regression Classifier. Since personality prediction is a classification problem, our model puts forward a new approach to predicting one's personality using the Logistic Regression algorithm with a minimized error function using the Big Five Personality Trait dataset. The implementation of the model is described in a diagrammatic flowchart of various steps performed in this project.

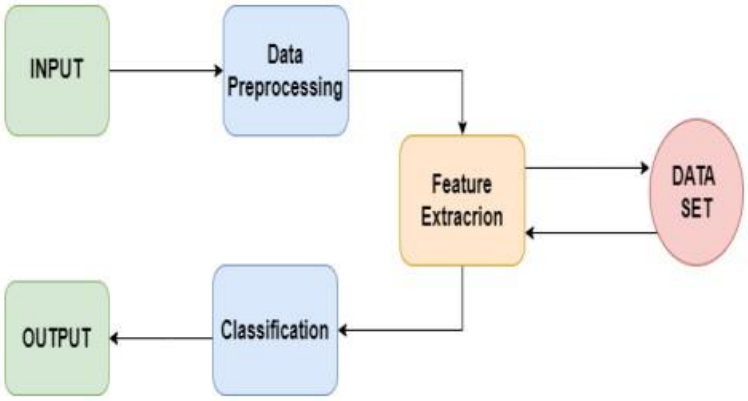

Figure 3.0 Process flow of personality classification (Mayuri 2019)

System Analysis **(Analysis of the Existing System) Table 3.0. Existing system**

Author	Published	Method	Result
Atharva K. et al	September 2021	KNN, Random Forest, Logistic Regression, SVM	64%, 71%, 62%, and 63%
Kanupriya A. et al	August 2015	Logistic Regression	72.24%

Weaknesses of the Existing System

i. [11] used ensembles of algorithms but got low accuracies. ii. The model did not perform well on the test data.

Analysis of the Proposed System

Data Collection

The accumulation of data is the first phase in the machine-learning process. To acquire the necessary results, the data must be very clear. This phase is critical for the system to produce the desired results. As input to the system, these data are immediately delivered in tabulation format.

Perform Data Preparation

This stage transforms the obtained data into information that has been processed. The acquired data is processed, with missing values checked and the data sorted according to the desired criteria. For the system's input, the aggregated data is tabulated. For the system's input, the aggregated data is tabulated. The data are in two sections. The first component will be used to train the model, which will take up the majority of the dataset, and the second will be used to evaluate the trained model's results. This step includes all other types of adjustment and processing, such as normalization, error correction, and so on.

Select a Model

The next step in the process is to choose a model from among the many that researchers and data scientists have developed throughout time. The easiest way to choose a model is to collect as much data as possible, which can be nearly endless depending on the problem's complexity.

Training

The training phase is a time-consuming process since the data obtained is used as input, which is then used to train the model. The more data there is, the more accurate the system becomes. If the model is properly educated, it will be more efficient in selecting the appropriate task flow, and the model’s outcome will be better. The trained model is tested with the test data.

Evaluation

Following the training phase, the evaluation phase begins, during which the model is tested and evaluated to see how it performs in real-world applications, how it learns in the future, how it corrects errors, and how it improves its efficiency.

Dataset

The Big Five Personality Trait dataset consists of 709 train data and 315 test data which vary between males and females of age ranging between 17 to 26 and different personality scores ranging from 1 to 9. The dataset is gotten from Kaggle.

Confusion Matrix

A confusion matrix is a performance measurement for a machine learning classification problem where the output can be two or more classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives a holistic view of how well the classification model is performing and what kinds of errors it is making. It is a table with four different combinations of predicted and actual values as shown below (Sarang N. 2021).

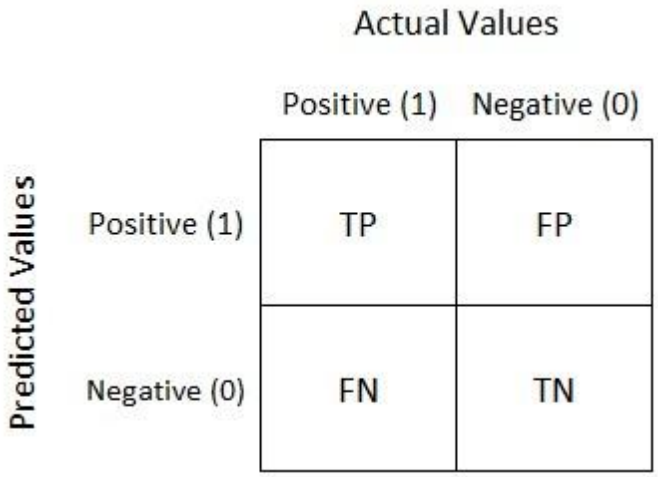

Figure 3.1 Confusion matrix table (Sarang N. 2021).

True Positive (TP): The actual value was positive, and the model predicted a positive value.

True Negative (TN): The actual value was negative, and the model predicted a negative value.

False Positive (FP): The actual value was negative, and the model predicted a positive value. Also known as Type 1 error.

False Negative (FN): The actual value was positive, and the model predicted a negative value. Also known as Type 2 error. The accuracy of a classification model is calculated using the formula below.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Recall

It tells how many of the actual positive values we were able to predict correctly with our model. It is calculated using the formula below.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision

It tells how many of the correctly predicted values turned out to be positive. It is calculated using the formula below.

$$\text{Precision} = \frac{TP}{TP + FP}$$

In a situation where there is no clear distinction between whether Precision is more important or Recall, they are combined into what is known as F1-Score. It is calculated using the formula below.

$$\text{F1-Score} = \frac{2}{\frac{1}{\quad} \quad \frac{1}{\quad}}$$

$$Recall + Precision$$

Tools and Materials

For effective implementation, this study utilized the following utilities as the requirement for the incorporation and optimization of the adopted model:

i. Visual Studio Code Insiders and Jupyter Notebook as an environment for writing codes.
ii. Python SDK (3.9.5) as a library for incorporating source code. iii. Sklearn,Pandas, matplotlib, and osas dependency supports. iv. Streamlit as a framework for the deployment of the model
v. Pyrepaser A simple resume parser used for extracting information from resumes.

RESULTS AND DISCUSSIONS

This chapter discusses the detail of experimental materials and setting used in the research process such as coding environment, dataset, python modules, parameter settings, etc. Detail of the various python modules for reading, analyzing, and visualization of data as well as a step-by-step detail of the machine learning processes implementation and results are also discussed.

Experimental Setup

The research process and experiment were set up on a Dell laptop using a 64-bit Windows 10 operating system running on a Quad-Core Intel® Core i5 CPU at 2.53GHZ processor with 8GB of RAM.

Experimental Environment

This work was carried out on the Visual Studio Code Insiders and Jupyter Notebook running python 3.9.5 with the following installed libraries.

i. **Sklearn**: scikit-learn is a key library for the Python programming language that is typically used in machine learning projects. It includes tools for mathematical, statistical, and general-purpose algorithms that form the basis for many machine learning technologies. ii. **Matplotlib**: A comprehensive library for creating static, animated, and interactive visualizations in python.
iii. **Pandas**: An open-source Python package used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.
iv. **Joblib**: is a Python library for running computationally intensive tasks in parallel. It provides a set of functions for performing operations in parallel on large data sets and for caching the results of computationally expensive functions.
v. **Streamlit:** Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers.

Machine Learning Processes Implementation

This section describes the step-by-step implementation of the research process. Data collection: the accumulation of data is the first phase in machine learning. To acquire the necessary results, the data must be very clear.

Data Description

	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

Figure 4.1 data description

Data Visualization

From the data provided, we have 328 males and 381 females as shown below.

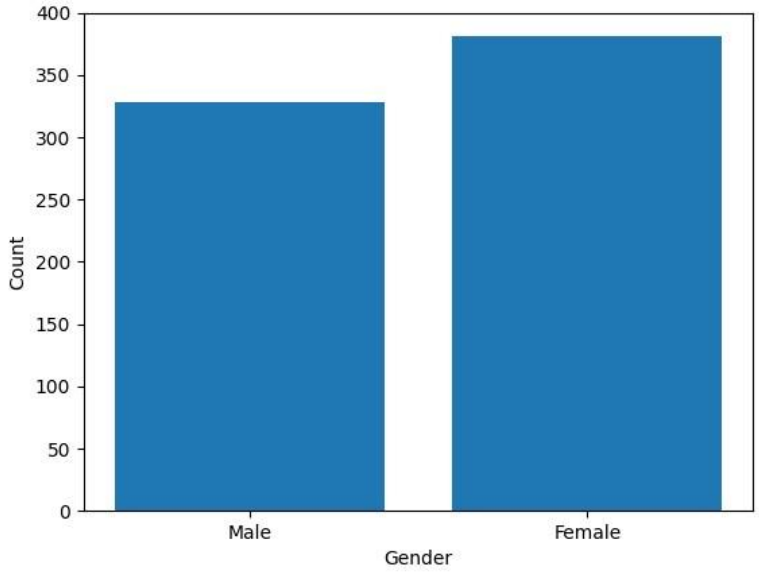

Figure 4.6 Data visualization

Data Preparation and Learning Process Splitting of data

Personalities scores, age, and gender were used as training data, and the personality values (openness, dependent, etc.) were used as validation data during the training.

```
data_frame = pd.DataFrame(data_array)

X = data_frame[[0, 1, 2, 3, 4, 5, 6]].values
Y = data_frame[7].values
```

Figure 4.5. Splitting of data

Model Architecture and Training

'multinomial' is used because the data values are dynamic, and 'newton-cg' is used as the solver to minimize the rate of error for a better result or performance of the model. The model is trained for 1000 iterations (epoch).

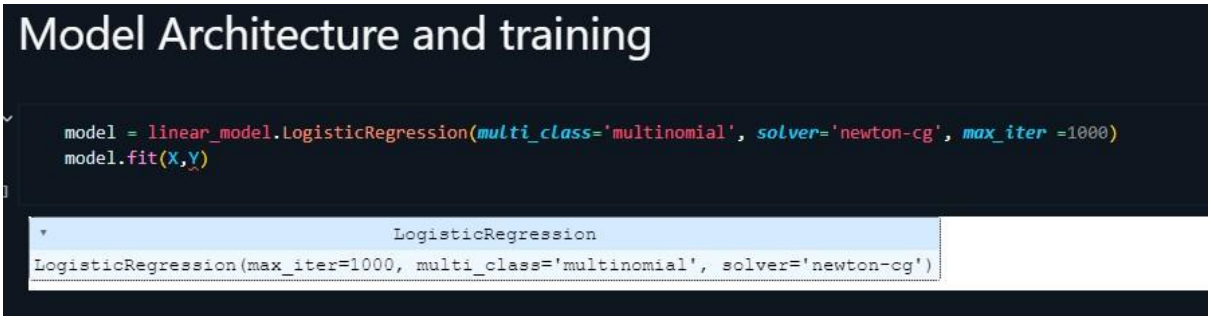

Figure 4.7 Model architecture and training

Experimental Results

A confusion matrix is a performance measurement technique for Machine learning classification. It is a kind of table that helps us to know the performance of the classification model on a set of test data for the true values known. The confusion matrix for this model is shown in fig 4.5 below.

Personality Prediction

The saved model is used on the test data provided. An accuracy of 85.71% was attained.

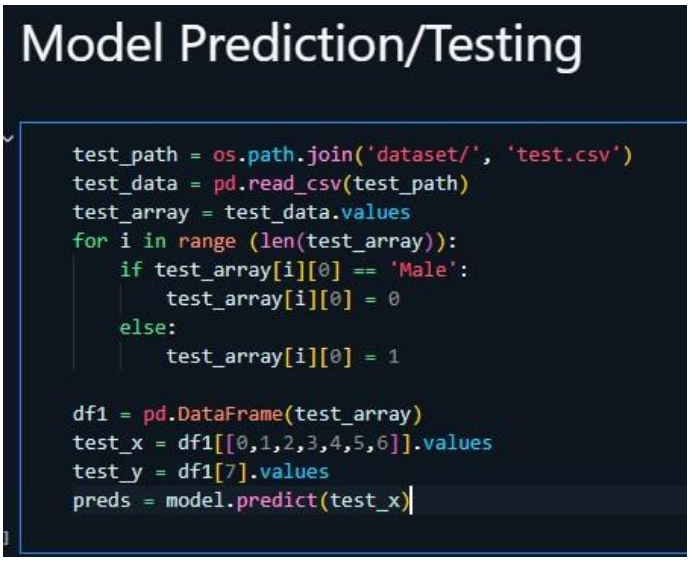

Figure 4.8 Data prediction

Figure 4.8.1 Model accuracy

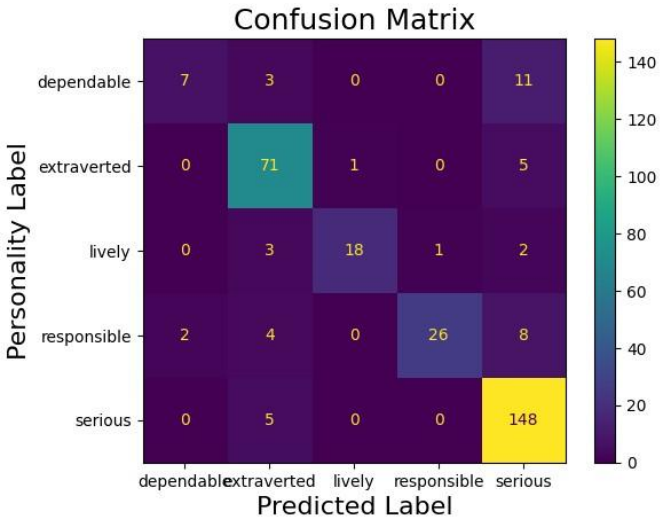

Figure 4.8.2 Confusion matrix

Performance of the Model

Performance of the model on the data

	precision	recall	f1-score	support
dependable	0.78	0.33	0.47	21
extraverted	0.83	0.92	0.87	77
lively	0.95	0.75	0.84	24
responsible	0.96	0.65	0.78	40
serious	0.85	0.97	0.91	153
accuracy			0.86	315
macro avg	0.87	0.72	0.77	315
weighted avg	0.86	0.86	0.85	315

Figure 4.8.3 Performance of the model

Deployment of the Model

The model was deployed for testing using Stream lit framework (a python-based framework provides for the machine learning process. The figure below shows the testing process.

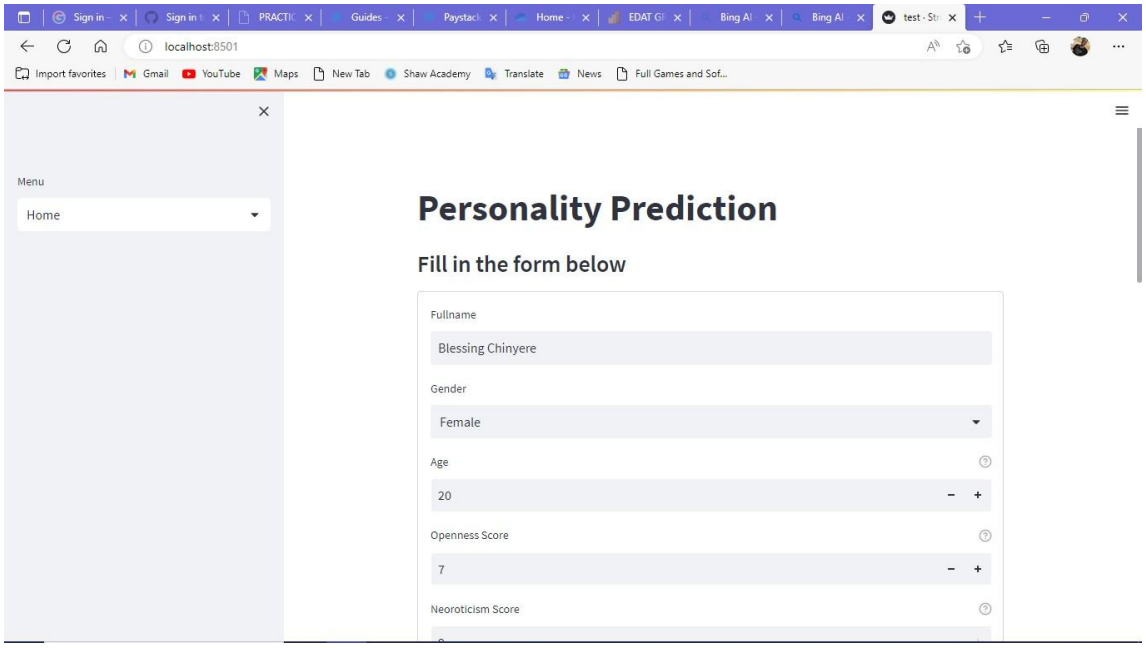

Figure 4.9. Providing details

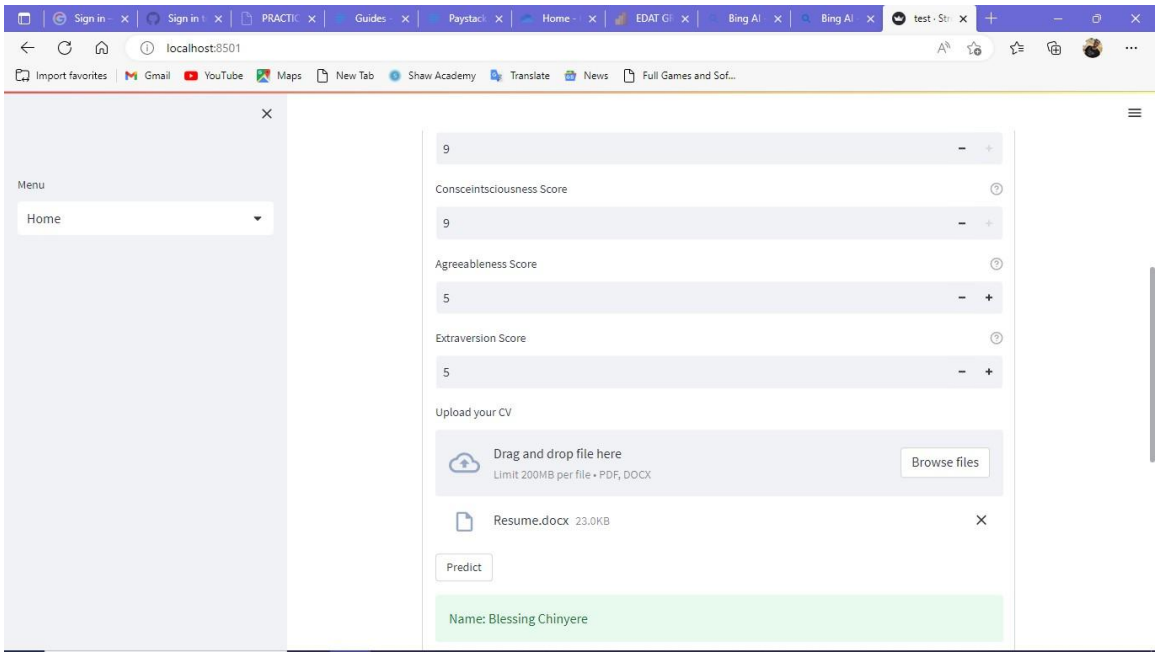

Figure 4.9.1 Providing details.

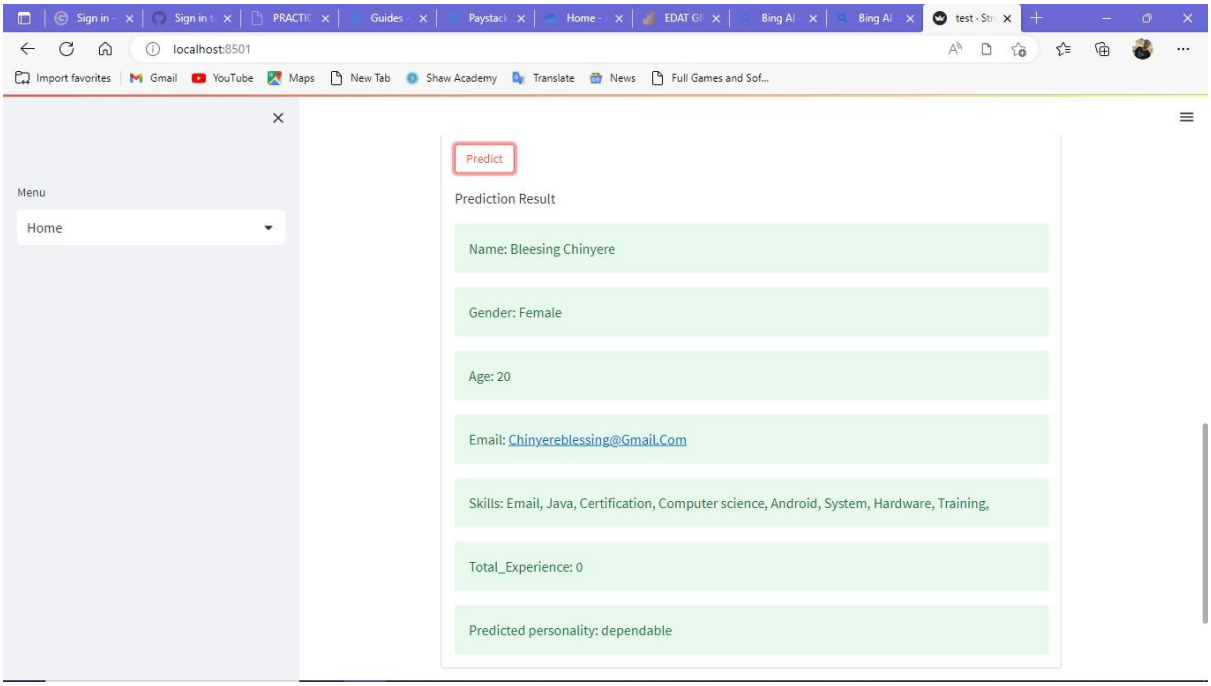

Figure 4.9.2. Prediction Result

CONCLUSION AND FUTURE WORK

We have actualized an organization situated enlistment framework that would help the human asset office in brief posting the proper candidate for a particular work profile. The framework would be utilized in numerous trade segments that will require master candidate, hence decreasing the work stack on the human asset division. In our study, we have many features. Some are similar and some different comparing to the mentioned papers. Moreover, we have developed it with full information which is more user friendly to users, admin, authentic of job authority.

References

[1] AditiChandrashekar, SanthiniNadar, Shreya Ganesh, MrinalKhadse (2021) “Online
[2] Recruitment System and Personality Prediction” International Journal for Research in Engineering Application and Management (IJREAM) April M. Young, The Technical Writer’s Handbook, Mill Valley, CA: University Science, 1989.
[3] AfrojaKhatunMonalisa, Md. Omar Kaiser Mahin (2022) “Personality Prediction System through CV Analysis” International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[4] Aggeliki et al, 2019. “As far as employment is considered, selecting the right candidate for therecruitment has been a fundamental issue, a new content-based job recommendation Algorithm for Job seeking and recruiting”. International Journal of Computer Sciences and Engineering (IJCSE), Vol. 7, 5 May 2019.
[5] Aggeliki D. Strati, FoDRA, Nikolaos D. Almalis (2019) Dept. of Informatics, University of Piraeus, Piraeus, Greece; George A. Tsihrintzis; NikolaosKaragiannis " A new content-based job recommendation algorithm for job seeking and recruiting” International Research Journal of Engineering and Technology (IRJET) pp. 3225, Vol 09, Issue 04

[6] Allan Robey, K. Shukla, K. Agarwal, K. Joshi, Professor S. Joshi (2019) “Personality prediction system through CV Analysis”, in International Research Journal of Engineering and Technology (IRJET) Vol 6, issue 02
[7] Allan Robey (2022) International Research Journal of Engineering and Technology IRJET
[8] Allan Robey, KaushikShukla, KashishAgarwal, Keval Joshi, (2022) “personality prediction system through cv analysis” International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[9] AleksandarKartelj; Vladimir Filipović; VeljkoMilutinović (2022) “Novel Approaches to Automated Personality Classification: Ideas and their Potential” Internal Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Vol 09 Issue 04
[10] Almalis, N. D., et al. FoDRA (2015) A new content-based job recommendation algorithm for job seeking and recruiting in 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA) IEEE
[11] AtharvaKulkarni, TanujShankarwar, SiddharthThorat (2022) “personality prediction Via CV Analysis using Machine Learning” International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07.
[12] Balabanović, M. and Y. Shoham, Fab (1997) content-based collaborative recommendation Communications of the ACM, 40(3): p. 66-72
[13] Barrick M.R (1991) “Proposed big five personality dimensions for any person to determine the personality of any reason and job performance” International Journal of Scientific Development and Research (IJSDR), Vol. 44, 05 May 2019, pp 1-26.
[14] BayuYudhaPratama, “Personality classification based on Twitter text using Naïve Bayes, KNN and SVM” 2022.
[15] Bergman, L., et al. (2008) International workshop on recommendation and collaboration in Proceedings of the 13th international conference on intelligent user interfaces.
[16] Data,Vishnu M Menon Computer Rahul Nath H A, (2016) “A Novel Approach to Evaluate and Rank Candidates in A Recruitment Process by Estimating Emotional Intelligence through Social Media”.
[17] DanyAzucar, Davide Marengo, Michele Settanni (2018) “Predicting the Big 5 Personality traits from digital footprints on social media: A meta-analysis”, Personality and Individual
[18] Differences, Computers in Human Behavior, pp. 150-159. International Research Journal of
[19] Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[20] Demetriou, L. Kyriakides, and C. Avraamidou (2003) “The missing link in the relations between intelligence and personality” in Journal of Research in Personality, vol. 37 issue 6, pp. 547- 581.
[21] Dhokley, Waheeda and Jehangir, Randeria and Sarwar, Shaikh and Rashid, Shaikh: (2022) “A Novel Approach to Predict Personality of a Person” International Research Journal of Engineering and Technology (IRJET) pp. 32225, Vol 09, Issue 04, e-ISSN 2395-0056 F. Evanthia, T Athanasios, et al. (2012) “An integrated e-recruitment system for automated personality mining and applicant ranking” Internet Research; pp. 551-568.
[22] Firoz Ahmed, MehrinAnannya, TanvirRahman, RisalaTasin Khan (2015). “An Automated CV Processing alongside with Psychometric Analysis in Job Recruiting Process”, Institute of Information Technology, May. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Vol 09, Issue 04, pp. 3225
[23] GangandeepKaur, ShrutiMaheshwari, (2022) “Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis” International

[24] Research Journal of Modernization in Engineering Technology and Science (IRJMETS) eISSN: 2582-5208 Vol 04 Issue 07
[25] Gayatri Vaidya, Pratima Yadav, Reena Yadav, Prof. ChandanaNighut, (2018) "Personality Prediction by Discrete Methodology", International Journal of Engineering, International Conference on Innovative and Advanced Technologies in Engineering ISSN: 2278-819 Vol. 14, pp. 10-13.
[26] Golbeck et al, (2011) "predicting personality from Twitter" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[27] Golbeck, J., Robles, C., and Turner, K. (2011) "Predicting Personality with Social Media. In Proc of the annual conference extended abstracts on Human factors in computing systems" International Journal of Research Publications and Reviews, e-ISSN: 2582-7421, Vol 3, pp.1634-1636
[28] HemalathaKallar, PetchipriyaThirumalainambi, SornamukhiSudaroli (2022) "Personality Prediction using CV Analysis" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[29] JenelParmar, (2022) "personality prediction through CV analysis (HR Helper)" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) ISSN: 2582-5208 Vol. 04, Issue 07
[30] Janet E. Squires (2019) "A protocol for Advanced Psychometric Assessment of surveys" nursing research and practice". International Journal of Computer Sciences and Engineering (IJCSE) Vol. 7, issue 5, article ID 156782.
[31] JunejaAfzalAyubZubeda et al, "Resume Ranking using NLP and Machine Learning", unpublished.
[32] J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem "Resume Ranking using NLP and Machine Learning", unpublished.
[33] Kalghatgi et al, (2015) "Neural Network approach to personality prediction based on Big Five model" in IJIRAE, vol2 issue8, pp. 56-63
[34] Kessler R. et al. (2009). Job offer management: how to improve the ranking of candidates. in International Symposium on Methodologies for Intelligent Systems. Springer.
[35] Lai, V., et al. (2016) Career Mapper: An automated resume evaluation tool, in IEEE International Conference on Big Data (Big Data)
[36] Liu, H. and L. Yu, (2005) "Toward integrating feature selection algorithms for classification and clustering" IEEE Transactions on knowledge and data engineering, 17(4): pp. 491-502 M. Kalghatgi, M Ramannavar, and Dr. N. S. Sidnal (2015) "Neural Network approach to personality prediction based on the Big-Five Model" in International Journal of Research and Engineering (IJIRAE) Vol2 issue8, pp. 56-63
[37] Madhuri Vishnu Vengurlekar, VaibhaviGanpatSawant, Namrata Ramesh Satelkar, Arati Ashok Sawant, (2019) "Personality Prediction System through CV Analysis", International Journal of Scientific Development and Research (IJSDR)
[38] ManasiOmbhas, PrajaktaGogate, TejasPatil, Karan Nair (2017) "Automated Personality Classification Using Data Mining Techniques"
[39] ManasiOmbhase et al, (2019) "Automated Personality Classification using Data Mining techniques" International Journal of Computer Science and Engineering (IJCSE) Vol. 7(5) MayuriPundlikKalghatgi, ManjulaRamannavar, Dr. Nandini S. Sidnal, (2015) "A Neural Network Approach to Personality Prediction based on the Big-Five Model", International Journal of Innovative Research in Advanced Engineering (IJIRAE), ISSN: 2349-2163 Issue (8), Vol 2

[40] Md. Tanzim Reza and Md. SakibZaman, "Analyzing CV/Resume using Natural Language Processing and Machine Learning" International Journal of Engineering Research and Technology (IJERT) vol.10, issue 09, September 2021.
[41] Menon, V. M. and H. Rahulnath (2016)"A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social media data" in International Conference on Next Generation Intelligent Systems (ICNGIS). IEEE.
[42] Mesurado, B., et al., (2014) "EXTROVERSION: NATURE, DEVELOPMENT AND IMPLICATIONS TO PSYCHOLOGICAL HEALTH AND WORK LIFE" Psychology of
[43] Extraversion, pp. 1-13.
[44] Mesurado, (2019) "Extraversion: Nature, Development and implications to Psychological Health and work life", International Journal of Computer Science and Engineering (IJCSE). Vol 7
[45] Mohammed Mehrad Sadr (2016) "The role of personality traits predicting emotion regulation strategies" International Academic Journal of Humanities, 3(4): pp. 13-24.
[46] Murray R. Barrick, Michael K. Mount, (1991) "The Big Five Personality Dimensions and Job Performance:Meta-Analysissis", Personnel Psychology.
[47] Nale R.K, PruthvirajPatankar, RajwardhanKhalate, RanjitGhorpade, SourabhBhapkar (2022) "Personality prediction system through CV analysis" International Research Journal of
[48] Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[49] Nathan Popkins C., (1998) "The five-factor model: Emergence of a taxonomic model for personality psychology" Great Ideas in Personality, Northwestern University.
[50] Ombhase, M. et al. (2017) "Automated Personality Classification using Data Mining Techniques, Pillai Institute of Information Technology"
[51] Pasquale De Meo et al., (2007)"An XML-based multi-agent system for supporting online recruitment services" IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 37(4): p. 464-480
[52] PendyalaSaiSireesha, MikkilineniRamya, MallampatiGokulchand, Motepalli Lakshmi Durga
[53] Chandrasekhar, KantapalliBhaskar, (2020) "Personality Prediction" in IJRASET, vol 8, pp. 8
[54] Pragya Sanjay Chauhan, AishwaryaPopatBondre, PrathameshGorakshaWaphare,
[55] SachinVaidya (2022) "Personality Evaluation and CV Analysis using Machine Learning Algorithm" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[56] Rabih Haddad, Eunika Mercier-Laurent. (2021) "Curriculum Vitae Evaluation using Machine Learning Approach". Artificial Intelligence for Knowledge Management IFIP AICT 614, hal-03496596
[57] RutujaNarwade, SrujamiPalkar, IshaZade, NidhiSanghavi. (2022) "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
[58] ShahzadQaiser,Ramsha Ali (2018) International Journal of Computer Applications (0975 – 8887), "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents" Volume 181, No.1.
[59] Squires Janet E. et al, (2013)"A protocol for advanced psychometric assessment of surveys Nursing research and practice"

[60] Sudir Bagade, et al (2019) "personality prediction system through CV analysis using Machine Learning Algorithm"
[61] SudirBagade, Jayashree Rout, PoojaYede (2022) "personality evaluation and CV analysis using Machine Learning Algorithm" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) e-ISSN: 2582-5208 Vol 04 Issue 07
[62] Van der Linden, D., J. teNijenhuis, and A. B. Bakker, (2010) The general factor of personality: A meta-analysis of Big Five inter-correlations and a criterion-related validity study. Journal of research in personality, 44(3): p. 315-327.
[63] Vishnu et al, "A novel approach to evaluate and rank candidates in a recruitment process". May 2016
[64] Vishnu M Menon et al, "A Novel Approach to evaluate and rank candidates in a recruitment process". International Journal of Computer Science and Engineering (IJCSE) Vol. 7, 05 May 2019
[65] Vivian LAI, Kyong Jin SHIM, Richard J. OENTARYO, Philips K. PRASETYO, Casey VU,
[66] Ee-peng LIM, and David LO "Career Mapper: An Automated Resume Evaluation Tool" International Research Journal of Engineering and Technology (IRJET) Vol 09 Issue 04
[67] VVCET-CSE (2022) "Personality Prediction System through CV Analysis" International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) eISSN: 2582-5208 Vol 04 Is