iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50718

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Personality Prediction using CV, Deep Learning

Sham Dhanke[1], Sakshi Dhepe[2], Manthan Dave[3], Shantanu Inamdar[4]
[1, 2, 3, 4]*Department of Information and Technology Vishwakarma Institute of Technology Pune, India*

Abstract: ***In this study, Deep Learning technologies are used to recognize personalities from CVs or resumes. The system for automating the evaluation of candidates' overall qualifications throughout the recruiting process is presented in this work. Based on the submitted CV or resume, which is assessed by an artificial intelligence algorithm, the system verifies the professional skills. The list comprises of terms that are pertinent to the position and best align with the applicant's qualifications. Recruiters may swiftly scan big candidate pools using the tool, which was created based on this study, to look for prospective job matches. This system's key benefit is that it automates the appropriateness evaluation procedure, freeing recruiters to concentrate on other crucial activities. Another benefit is that it gives applicant profiles to help employers and employees match up for open positions. When reviewing a CV, for instance, if a candidate hasn't completed a personality test but has been hired in the past successfully, the CV will automatically receive a score based on that evaluation.***
Keywords: ***NLP, Big five model, Classifier, Machine-learning, personality prediction***

I. INTRODUCTION

As a distinction in an individual's traits or behaviors of thinking, feeling, and acting, personality is defined. A person's personality greatly affects their life and shapes the decisions they make. In the world of employment, picking the best applicant from among a vast field of rivals is a crucial decision. The objective of this study is to anticipate a person's personality in a workplace setting for academic recruiting. Traditional research and forecasts take a lot of consumers' time. In the recruiting process, an interview is frequently followed by an aptitude test. These routine procedures take time and may result in the biased selection of job candidates. It takes time to predict personality using psychometric tests and questionnaires. The proposed system conducts an electronic aptitude test and personality assessment to predict the job seeker's personality and help select the best candidate based on his or her skills, exam performance, and decision-making abilities, in contrast to the current online recruitment system, which only reviews the submitted resume and nominates the candidates.

A. Big Five Factor Model [1]

The proposed system conducts an electronic aptitude test and personality assessment to predict the job seeker's personality and help select the best candidate based on his or her skills, exam performance, and decision-making abilities, in contrast to the current online recruitment system, which only reviews the submitted resume and nominates the candidates.
The five-factor model of personality is a five broad trait dimension consisting of the following five factors:

1) Openness to Experience- Curious by nature, intellectual person, creatively inclined and always open to new ideas.
2) Conscientiousness- A person who is organized, neat, reliable at work, with a performance-oriented approach, and trustworthy.
3) Extraversion- Person with preferences Outgoing, communicative, amiable, appreciates friendly circumstances.
4) Agreeableness- friendly, open-minded, tender, trusting, hospitable and pleasant.
5) Neuroticism- Anxious, irascible, unstable, and overemotional.

B. MBTI Personality Traits [2]

Determine a person's personality type, inclinations, and tendencies with the Myers-Briggs Personality Indicator. As a result of their work with Carl Jung's theory of character types, Isabel Myers and her mother Katherine Briggs created it. One of the most popular mental tools now in use is this one. Four categories make up the Myers-Briggs Personality Indicator (MBTI):
- Sensation/Intuition - Introversion/Extraversion
- Thinking and Feeling - Assessing and Perceiving
This test is also unreliable because it has been demonstrated that repeating the exam can change a person's personality scores.
Related work is presented in Section II, and the proposed approach as well as information on the model's training and testing are presented in Section III. We give the findings and outcomes in section IV. Section Section V - Conclusions - VI lists the references used.

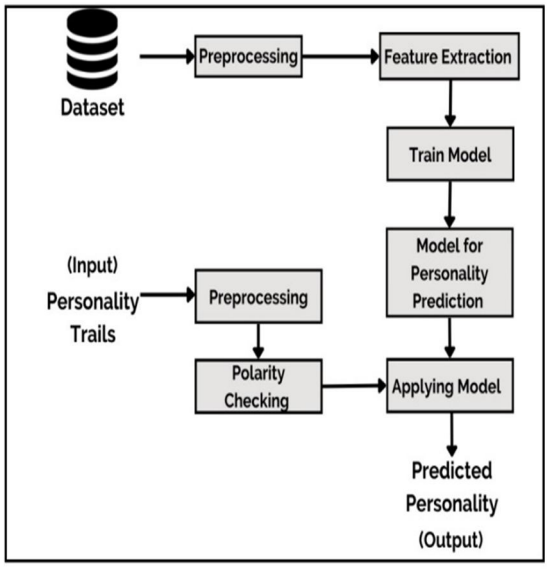
©IJRASET: All Rights are Reserved

II. RELATED WORKS

By generating meta-credits from tweets, Kalghatgi et al. [3] developed a neural network technique based on a Big Five inquiry to visualize the nature of singularities based on tweets shared through the social media platform Twitter. They are also employed in the analysis of their own social research. Data mining from social media tweets was done by the developers using a four-step procedure that included pre-processing, transformation, and bracketing. Despite the fact that neural networks are used to predict character, phoney data, programmed analysis of tweets, and reliance on Twitter are insufficient to forecast a person's character; rather, only anomalies and patterns can do so. Afzal Juneja Ayub Zubeda et al [4] discussed a business that evaluated resumes and CVs using machine learning and natural language processing (NLP). This paradigm places CV in the context provided by the organization's metrics. The authors advise organisations to take into account a candidate's GitHub and LinkedIn profiles in addition to their resume in order to gain a deeper knowledge of them and make it simpler for them to identify a suitable fit in terms of abilities, character, and especially talents. Using Natural Language Processing and machine learning, Md. Tanzim Reza et al. [5] analysed resumes of individuals by transforming the resume into an HTML page and then breaking it down into HTML code to extract parts and talents.
Afzal Juneja Ayub Zubeda et al [4] discussed a business that evaluated resumes and CVs using machine learning and natural language processing (NLP). This paradigm places CV in the context provided by the organization's metrics. The authors advise organisations to take into account a candidate's GitHub and LinkedIn profiles in addition to their resume in order to gain a deeper knowledge of them and make it simpler for them to identify a suitable fit in terms of abilities, character, and especially talents. Using Natural Language Processing and machine learning, Md. Tanzim Reza et al. [5] analysed resumes of individuals by transforming the resume into an HTML page and then breaking it down into HTML code to extract parts and talents.
According to Sudhir Bagade et al's [7] proposal, sentiment analysis in social media may be used to modify current data. Here, it is discovered that combining several categorization methods offers significantly superior characteristics; the programme can then be examined for efficiency and performance.
M. R. Barrick et al. [8] They suggested the Big Five personality dimensions in 1991 as a way to determine someone's personality. In particular, in the subfields of human choice, preparation and promotion, and execution assessment, the results discuss the benefits of applying the 5-factor model of character to gather and convey the observed results that have diverse ideas for study and practise in human psychology.

III. PROPOSED METHODOLOGY

A model for predicting personality is suggested in this study report. This model is created by first using a dataset that has been further preprocessed. A deep learning model is trained following feature extraction. Based on the input personality traits, this trained model is used to predict a person's personality. Based on this process, a workable project has been created that offers clear, user-friendly results.
Figure 1 illustrates the proposed methodology.



Fig(1) Implementation Steps

A. Dataset

We made advantage of the Kaggle and GitHub datasets as well as live data gathered from CVs and resumes. The dataset is called "mypersonality.csv". [9]

The dataset is organised into the following columns: "STATUS," "sentiNEG," "sentiPOS," "sentiNEU," "sentiCOMPOUND," "cEXT," "cNEU," "sAGR," "sCON," "sOPN," "DATE," "NETWORKSIZE," "BETWEENNESS," "DENSITY," "BROKERAGE," "NBROKERAGE," and "TRANSITIVITY." The candidate's corresponding social media post is called STATUS. It may just be some plain text or a caption. Sentiments marked with "sentiNEG", "sentiPOS", "sentiNEU", or "sentiCOMPOUND" are positive, negative, or neutral. The personality kinds are "Ext," "sNEU," "sAGR," "sCON," "sOPN," "cEXT," "cNEU," "cAGR," "cCON," and "cOPN." A person's "NETWORKSIZE", "BETWEENNESS," "NBETWEENNESS," "DENSITY," "BROKERAGE," "NBROKERAGE," and "TRANSITIVITY" are qualities.

B. Data Preprocessing

Following that, the input data set is processed to remove extraneous information, eliminate background noise, and convert characters to lowercase.

C. Feature Extraction

All of the significant characteristics that include feelings are then removed. Only the emotionally laden words are chosen.

D. Model Trainining

After using categorization methods, an emotion lexicon is also utilised. Personality is categorised using techniques like Support Vector Machine, KNN, Gaussian Naive Bayse, and LSTM (long-term short memory), a type of repeating neural network. Polarity testing is carried out based on personality attributes, and the model is trained and evaluated using the training and testing data [10]. About 66% of the input data are utilised for the training models and 33% are used for the testing models when using the Sklearn library [11].

The personality is then predicted, i.e., using the MBTI or Big Five Factor Theory. The Kaggle kernel output testaccount93/my personality -p /path/to/dest is the API that is being utilised.

We should give the personality column in our dataset a name before creating the model. We applied ML techniques including Logistic Regression, Naive Bayes, Random Forest, Support Vector Machine (SVM), and KNN to predict a participant's personality.

The difference between logistic regression and linear regression is that the former determines whether the forecast will be accurate or inaccurate. It is regarded as one of the most well-known algorithms for resolving binary classification issues (true/false, yes/no) [12].

Naive Bayes - The conditional probability is determined using the Bayes theorem. It serves as the foundation for the Naive Bayes classifier, a set of guidelines that depend on the feature's assumptions being clear. Each element within the set of rules must comply with the issues provides a comparable and objective assessment of the result.

KNN, or K-nearest Neighbour, is a technique for solving classification and regression issues. Of course, KNN and the adage "like and like" are interchangeable. The rule set presupposes that similar record factors typically occur near together [2].

SVM - To do classification and regression analysis on the presented dataset, we mainly employed Support Vector Machine (supervised), whose miles are used to handle relevant data. In the N-dimensional space (N-variety of characteristics), SVM is used to locate a hyperplane that can rapidly describe the collected factors.

1) *Random Forest*- a technique for classification and regression that is extensively utilised. To deliver the answer, more than one decision tree is used. We applied either bagging or bootstrap aggregation to train the random forest.
2) *Long-Short*- A recurrent extraordinary neural network called Term Memory (LSTM) permits learning of long-term data dependability. This is made possible by the model's recurrent module, which is the cause. It is made up of a combination of 4 layers that communicate with one another [4].

E. Personality Predictor (OCEAN)

This exam is given based on assertions that are graded from 1 to 10 with 1 being disagree, 5 being neutral, and 10 being agree.

F. Experimental Results

This exam is given based on assertions that are graded from 1 to 10 with 1 being disagree, 5 being neutral, and 10 being agree.

G. Input Personality Traits

We insert the parameters in accordance with the dataset's columns. The following terms are used: sEXT, sNEU, sAGR, sCON, sOPN, cEXT, cNEU, cAGR, cCON, cOPN, DATE, NETWORKSIZE, BETWEENNESS, NBETWEENNESS, DENSITY, BROKERAGE. Feature extraction from datasets with imputed data and data preprocessing.

Because STATUS and AUTHID are not necessary for the problem, they are not included in the data preparation. TRANSITIVITY and NBROKERAGE. Positive, negative, and neutral attitudes are primarily employed. To obtain the utmost accuracy, some features additionally undergo hot label coding.

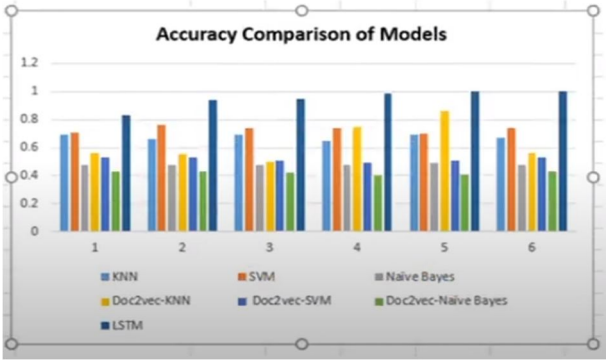
H. Polarity Checking

Whether the terms submitted are positive or negative is determined through the polarity verification process. We grade it on a scale of 1 to 10, with 1 denoting "disagree" (a bad word) and 10 denoting "agree" (a good word). The input is organised using it.

I. Personality Prediction Model

Applying the generated model to the input data ultimately results in the prediction of personality.

The model is trained using a variety of techniques, and a model is then chosen depending on how accurate the trained model is. We achieved an accuracy of 0.69 and a precision of 0.74 using KNN classifiers. We also achieved 0.74 accuracy and 0.74 precision with SVM. Gaussian Naive Bayes provided an accuracy and precision of 0.48 and 0.73, respectively. When feeding datasets into the models and evaluating the models, we were able to improve accuracy with LSTM, which resulted in a loss reduction of 0.0809. The most accurate algorithms are those using short-term memory, followed by SVM. The least reliable is achieved by Naive Bayes. Figure 2 shows the performance comparison of the models.

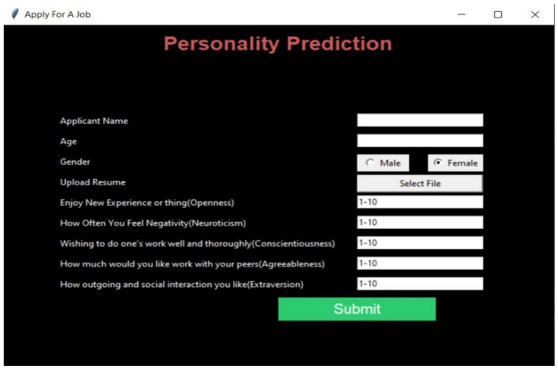


Fig(2) performance comparison of models

The model used to predict personality is the LSTM since it has the best accuracy.

J. Predicted Personality

Personality is predicted by trained models. The output is a precise number that corresponds to the personality type. Any personality type, including impolite, serious, smart, excellent, and satisfying, is acceptable. Numerous related applications can make advantage of this projected personality.



Fig(3)Application form



Fig(4) Final result

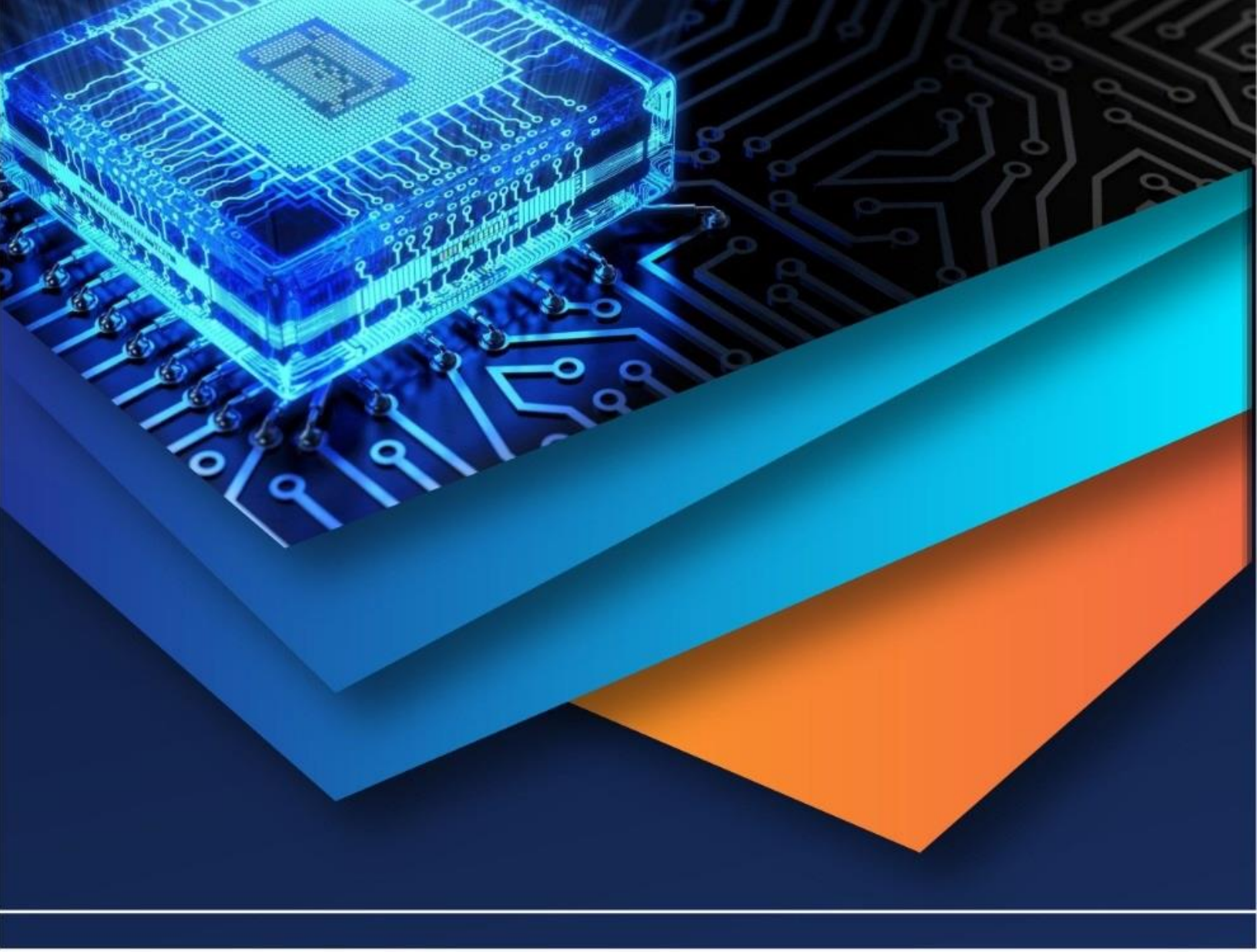
Based on the response provided in Fig 3, all the altered data and the expected outcome will be shown on the result page.

IV. CONCLUSION

We have used the standard scores provided by the HR department in accordance with the work decision measures to offer human subject evaluation in this article. The administrator can access the web-based resumes (CVs) that candidates fill out. Separate keys are supplied to each applicant for the in-person exams. The system incorporates the CVs that applicants filled out to perform CV assessment. The data was scored using an ML method. The results assist in highlighting the qualifications of the job applicants. As a result, CV is selected for consideration throughout the recruiting process, and the HR department takes a just and impartial choice. Additionally, information perception models determine how students' total performance is shown in relation to numerous factors. This test enables the administration department and recruiters to accurately assess candidates' skills. Additionally, we may modify the outdated framework to do sentiment analysis on data derived from social media. ML has many more display settings that can be combined to offer far superior functionalities. Additionally, the application's features and performance may be tried out. Furthering the growth of accuracy can benefit from deep learning calculations. The possibility to consider a multidisciplinary approach to managing expectations that takes into consideration diverse biological signals is available at this point.

REFERENCES

[1] Robey, Allan, Kaushik Shukla, Kashish Agarwal, Keval Joshi, and Shalmali Joshi. "Personality prediction system through CV analysis." Int. Res. J. Eng. Technol 6, no. 2 (2019): 2395-0056.
[2] Arora, Alakh, and N. K. Arora. "Personality Prediction System Through CV Analysis." In Ambient Communications and Computer Systems, pp. 337-343. Springer, Singapore, 2020.
[3] Kalghatgi, Mayuri Pundlik, Manjula Ramannavar, and Nandini S. Sidnal. "A neural network approach to personality prediction based on the big-five model." International Journal of Innovative Research in Advanced Engineering (IJIRAE) 2, no. 8 (2015): 56-63.
[4] Khan, Tabrez. "Resume Ranking using NLP and Machine Learning." (2016).
[5] Reza, Md, and Md Zaman. "Analyzing CV/resume using natural language processing and machine learning." PhD diss., BRAC University, 2017.
[6] Kaur, Gagandeep, and Shruti Maheshwari. "Personality Prediction through Curriculam Vitae Analysis involving Password Encryption and Prediction Analysis." (2019).
[7] Jayashree, Rout, Bagade Sudhir, Yede Pooja, and Patil Nirmiti. "Personality evaluation and CV analysis using machine learning algorithm." Int. J. Comput. Sci. Eng 7 (2019): 1852-1857.
[8] Barrick, Murray R., and Michael K. Mount. "The big five personality dimensions and job performance: a meta-analysis." Personnel psychology 44, no. 1 (1991): 1-26.
[9] https://www.kaggle.com/datasets/haisamrafid/mypersonality
[10] Kaur, Gagandeep, and Shruti Maheshwari. "Personality Prediction through Curriculam Vitae Analysis involving Password Encryption and Prediction Analysis." (2019).
[11] Sudha, G., K. K. Sasipriya, D. Nivethitha, and S. Saranya. "Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process." In 2021 4th International Conference on Computing and Communications Technologies (ICCCT), pp. 617-622. IEEE, 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)