IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

A Comprehensive Survey on Personality Prediction Using Machine Learning Techniques

Prof. Amol Chincholkar[1], Dipti Bhosale[2], Shivanjali Adsul[3], Anjali Bodkhe[4], Rutuja Kadam[5]

Department of Information Technology, Zeal College of Engineering and Research, Narhe[1-5]

Abstract: In order to categorize people's personalities, this study applies the machine learning approach known as logistic regression. Moreover, machines that use Natural Language Processing (NLP) can comprehend and interact with human language. Several earlier research projects have tried to automatically determine an individual's personality type. Sorting people according to their personality types is one of the most significant uses of machine learning algorithms. There are a lot of advantages to grouping people into categories. Knowing one's personality can be quite beneficial in the modern world with abundant opportunities. Based on these forecasts, anyone can select a job or other interests. Many firms in the modern world utilize these personality assessments to select candidates since it increases productivity because the employee is doing what he or she is best at.

Keywords: Multinomial Logistic Regression, Python, Web Technology, Personality, NLP, IBM Watson (video analysis)

I. INTRODUCTION

The term "personality" describes a person's unique thoughts, feelings, and behavior tendencies. Numerous subjects are covered in personality tests. Determining your strengths, weaknesses, temperament, and leadership style is its aim.

In psychology, the Big 5 component hypothesis is widely accepted as a paradigm for explaining the fundamental makeup of living soul temperament. Based on all these characteristics, a theory has been created.

The most widely used and acknowledged personality model is the five-factor model, popularly referred to as the Big 5 model. Students can use the word and conceptual framework it offers for a large portion of their studies in personality psychology and individual differences.

Personality Traits: Openness, the system is based on a questionnaire that ultimately predicts a person's personality. It can be adjusted further to elicit additional information from candidates. To address this issue, we suggest a personality classification system based on logistic regression.

Openness: This trait includes traits such as insight, imagination, sensitivity, attention, and curiosity. People who score high on openness are curious, creative, and open to new experiences.

Conscientiousness: This trait relates to a person's level of care, discipline, consideration, and diligence. People who score high on conscientiousness tend to have clear goals, good self-control, and good organization.

Extroversion: This characteristic has to do with how assertive and emotional a person is. Extroverts are gregarious, at ease in social situations, and frequently show excitement and enthusiasm.

Agreeableness: This trait is associated with a person's generosity and cooperation. Agreeable people are often kind, trustworthy, and caring.

Neuroticism: This characteristic has to do with a person's capacity for stable emotions and propensity to feel depressive and anxious feelings. Individuals with high neuroticism scores are more sensitive to stress and mood swings.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

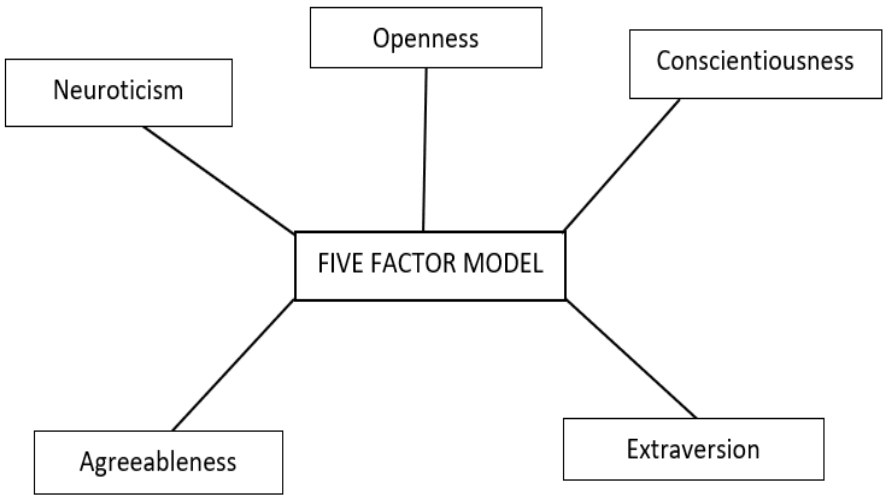

Fig. 1 Big Five Factor Model

Objective of the System:
The goal is to develop a method that will make it simpler to identify the applicant's personality features and discover m ore about them without actual meeting them. The business will be in a better position to select the most qualified candidate for the available position since it will have a deeper understanding of the individual.

II. LITERATURE SURVEY

The burgeoning field of personality prediction through machine learning (ML) has witnessed substantial growth, with researchers exploring diverse methodologies and data sources. This literature survey aims to present an overview of key contributions between June 2022 and November 2023, focusing on four significant works.

1) "Smart-Hire Personality Prediction Using ML" (May 2023) by Isha Gupta and Manasvi Jain:

This study underscores the practical implications of personality classification. It suggests that individuals, upon discovering their personality types through ML predictions, can actively engage in self-improvement efforts. The paper emphasizes the potential impact of such insights on personal and professional development.

2) "A Study on Personality Prediction & Classification Using Data Mining Algorithms" (August 2022) by Pavitha N., Somesh Kamnapure, and Ayush Gundawa:

Highlighting the importance of personality in personal and professional contexts, this work explores data mining algorithms to rapidly predict and categorize an individual's personality. The researchers advocate for integrating ML techniques, particularly through intuitive input methods like questionnaires, to enhance prediction efficiency.

3) "Language Style Matters: Personality Prediction from Textual Styles Learning" (November 2023) by Meiling Li and Hezi Liu:

This research delves into psycholinguistic literature, emphasizing the role of language styles in unveiling personality aspects. The paper contends that language styles offer insights into users' personalities, including social networks and mental health. Textual styles learning is presented as a valuable approach for personality prediction.

4) "Personality Prediction using Machine Learning" (June 2022) by Hima Vijay and Neenu Sebastian:

Acknowledging the importance of sorting individuals based on personality types, this work emphasizes the applications of ML algorithms in achieving this goal. The paper contributes to the literature by exploring the potential benefits and implications of personality prediction using ML.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

5) "Non-operative Personality Prediction Based on Knowledge Driven" (July 2022) by Huang Tao, Li Bi-Cheng, and Lin Zheng-Chao:

In contrast to traditional methods, this paper introduces a personality scoring algorithm based on vocabulary weight and word frequency. The proposed algorithm aims to address the ambiguous physical meaning associated with personality scoring in personality analysis using the dictionary method. This work provides an alternative perspective on non-operative personality prediction.

III. METHODOLOGY

1) Data Collection

A range of websites and conversations with prospective employees were used to collect the data set. For quick data retrieval and training, the questions and answers were entered into Google Forms and saved as a CSV file. Some questions pertaining to Openness, one of the Big Five personality traits, are shown in the graphic below. Every topic has a preset range of answers, ranging from strongly agree to strongly disagree.

2) Data Analysis

After splitting the test dataset into x- and y-tests, we utilize the Standard Scaler from the Scikit Learn Library to scale the data. We used a Jupyter Notebook computer system to run our model. We have utilized matplotlib, sklearn, numpy, re, seaborn, pandas, and numpy among other Python libraries.

SYSTEM ARCHITECTURE

A) User

Enter personal information, upload a resume, and complete a questionnaire asking to rate one's own personality prediction using OCEAN values and CV analysis (based on the Big Five Personality Traits model).

online video recording. The applicant must respond to a few inquiries posted on the portal by the HR department. Face emotion and speech analysis can provide information about a candidate's personality traits and degree of confidence.

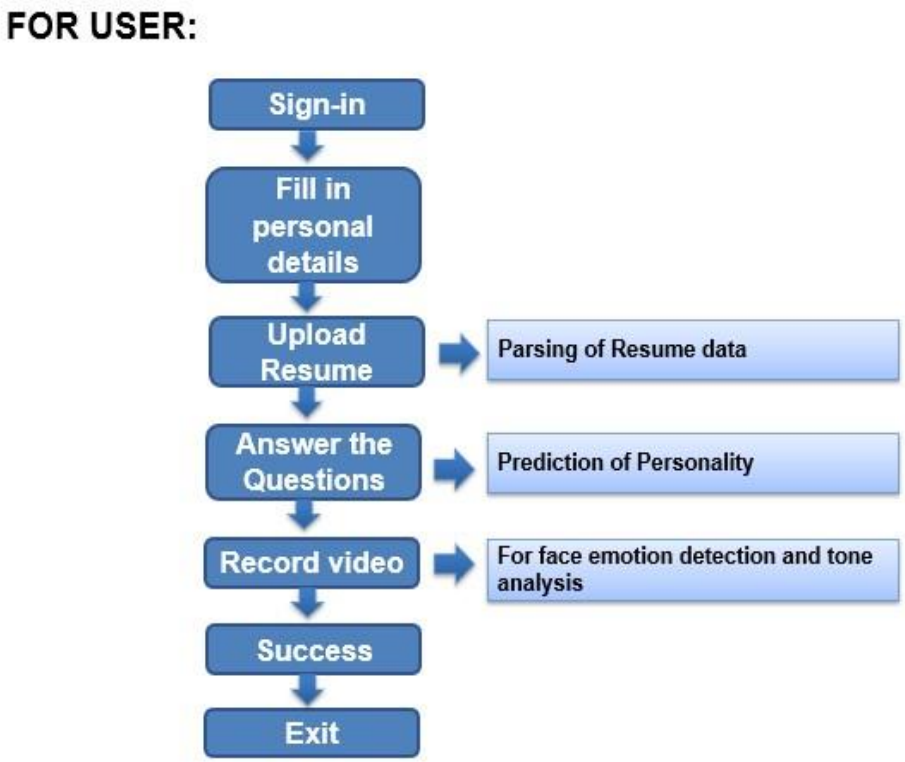

Fig. 2 For User

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

B) Admin

View the profiles of every candidate who has applied. View the summary of each candidate's profile, which includes their resume, answers to questions, technical expertise, personality traits, and the results of their tone and video analysis. Inform the candidate via phone call or one-click email about selection/rejection and the next steps in the interview process.

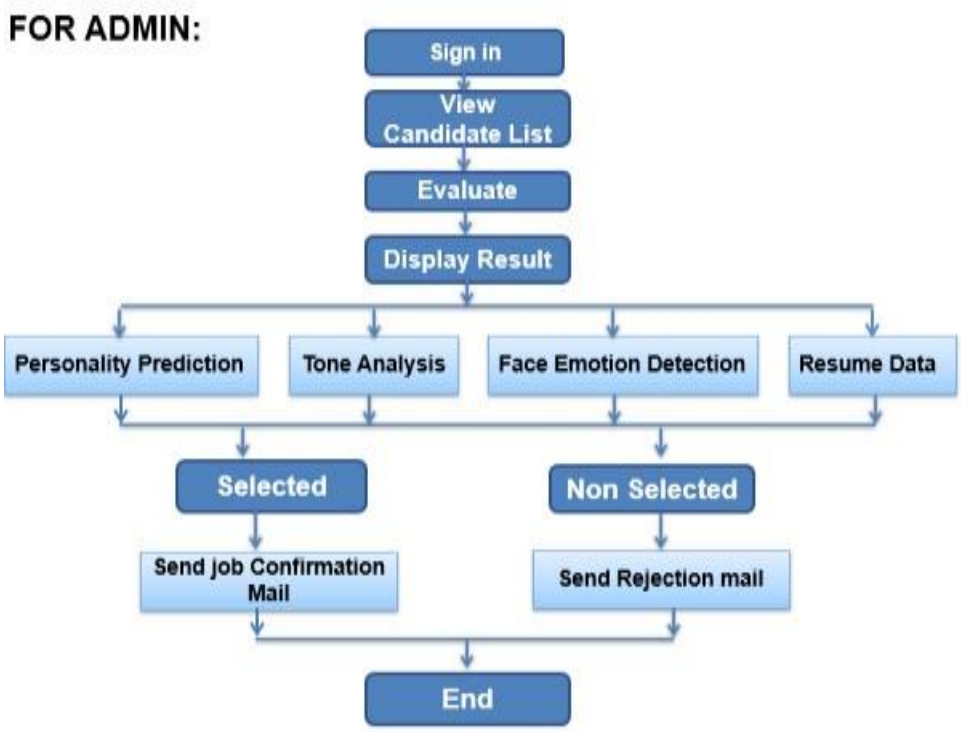

Fig. 3 For Admin

C) Multinomial Logistic Regression

Multinomial logistic regression is a classification technique in statistics that extends logistic regression to multiclass problems, that is, problems with more than two distinct possible outcomes. (Source:).

In other words, it is a model that, given a set of independent variables (which can be real-valued, binary-valued, categorical-valued, etc.), estimates the probability of certain outcomes.

The different occurrences of a dependent variable are distributed by type.

There are many other names for multinomial logistic regression, such as multinomial LR, multiclass LR, SoftMax regression, multinomial logit (mlogit), maximum entropy classifier (MaxEnt), and maximum entropy model conditional.

D) Implementation

The dataset is divided into training and testing. The dataset is further scaled with the Standard Sklearn Library's help. Thirty percent is testing and seventy percent is training.

The dataset comprises 972 rows and 8 columns, with each row containing the candidate's age, gender, and one of the five personality traits determined by the OCEAN Model. The participant data are listed in the respective rows.

E) Natural Language Processing (NLP)

The study of how computers and human language interact is the focus of the artificial intelligence (AI) subfield of natural language processing, or NLP.

NLP facilitates the understanding and use of human language by computers.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023
DOI: 10.17148/IJARCCE.2023.121120

Steps:
1. Segmenting Sentences
2. Tokenization of words
3. Climbing
4. Formulation
5. Quit analysing words
6. Parsing dependencies
7. Tagging of parts of speech (POS)

F) Video
Examining:
An AWS Artificial Intelligence (AI) service called Amazon Transcribe makes it simple for you to turn audio into text.

Transcription:
To turn spoken audio from videos into written text, use the Amazon Transcribe API. You will receive the textual data required for analysis in this step.

Sentiment Analysis:
Following transcription, you can evaluate the content's emotional tone by using a sentiment analysis model or tool. While sentiment analysis can shed light on the speaker's emotional state, it is not a substitute for personality profiling.

G) Tone Examination
IBM developed a suite of machine learning and artificial intelligence technologies called IBM Watson Examiner."

This service, which is a component of the IBM Watson Natural Language Understanding suite, is intended to evaluate the emotional content, tone, and sentiment of textual data. It can assist users in comprehending the sentimentality of written material, including emails, social media posts, and client testimonials.

IV. CONCLUSION

This personality prediction model can be used for government agencies such as the Army, navy, and Air Force, e-commerce websites, psychometric testing, competitive exams, and marriage websites. After the user attempts the survey, the system automatically classifies their personality based on the data set that was supplied at the back end. There may be more personality traits added in the future because personality analysis and prediction have grown recently. Any further adjustments that are needed to boost accuracy and enhance the career counseling module can be made using the algorithms and data gathering. By using this procedure, the human resources department would be able to select the most qualified applicant for a given position, providing the company with an experienced worker. This system would rank each CV, making it easier to select which ones.

REFERENCES

[1]. Pavitha N., Somesh Kamnapure, Ayush Gundawar, Ishan Gujarathi, Devang Manjramkar, Dhananjay Deore Vishwakarma Institute of Technology, Pune, Maharashtra, India 2023 International Conference on Disruptive Technologies (ICDT) "*Smart-Hire Personality Prediction Using ML*"
[2]. November 2023/Meiling Li; Hezi Liu "*Language Style Matters: Personality Prediction from Textual Styles Learning*" [IEEE]
[3]. June 2022/Hima Vijay; Neenu Sebastian "*Personality Prediction using Machine Learning*" [IEEE]
[4]. July 2022/Huang Tao; Li Bi-Cheng; Lin Zheng-Cha "*Non-operative Personality Prediction Based on Knowledge Driven*" [IEEE]
[5]. Prof. Waheeda Dhokley, Randeria Kaiwan Jehangir, Shaikh Nabeel Rashid, and Shaikh Almas Mohd Sarwar, "*A Novel Approach to Predict Personality of a Person.*"
[6]. Tanuj Shankarwar, Siddharth Thorat, and Atharva Kulkarni, "*Using machine learning, Personality Prediction via CV Analysis*"
[7]. Mayuri Pundlik Kalghatgi, Manjula Ramannavar, and 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: 23492163 Issue 8, Volume 2, August.
[8]. Clemens Stachl, Quey Au, Ramona Schoedel, and Markus Buhner, "*Predicting personality from patterns of behaviour recorded using smartphones.*"

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed journal & Refereed journal ⁜ Vol. 12, Issue 11, November 2023
DOI: 10.17148/IJARCCE.2023.121120

[9]. Athanasios Tsakalidis, Giannis Tzimas, and Evanthia Faliagka, "*An integrated e-recruitment system for automated personality mining and applicant ranking*", 2012.
[10]. L. D. van der, J. te Nijenhuis, et al., "*The General Factor of Personality*"
[11]. Gayatri Vaidya, Pratima Yadav, Reena Yadav, and Prof. Chandana Nighut, "*Personality Prediction by Discrete Methodology*," IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719 Volume 14, pp. 10-13
[12]. Aleksandar Kartelj, Vladimir Filipovic, Veljko Milutinovic, "*Novel approaches to automated personality classifications ideas & their potential*", 35th International conference MIPRO, IEEE, 2012.
[13]. Vishnu M. Menon, Rahulnath H. A, "*A Novel Approach to Evaluate and Rank Candidates in a Recruitment Process by Estimating Emotional Intelligence Through Social Media Data."*
[14]. J. Golbeck, C. Robles, K. Turner, "*Predicting personality with social media*", CHI'11 Extended Abstract on Human Factors in Computing Systems, pp.253-262, 2011.
[15]. Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C Graesser, "U*sing Data Mining Techniques to detect the personality of players in an Educational Game*", Spinger International Publishing, 2014.
[16]. Randall Wad, Taghi M. Khoshgoftaar, Amri Napolitano, Chris Sumner, "*Using Twitter Content to Predict Psychopathy*", 11th International conference on Machine learning & Applications, IEEE, 2012.
[17]. Educational Game Yago Saez, Carlos Navarro, Asuncion Mochon, Pedro Isasi, 2014. "*A system for personality & Happiness detection, International Journal of Interactive Multimedia & Artificial Intelligence*.
[18]. Yilun Wang, "*Understand Personality through social media*", International of computer Science standford university.
[19]. Bayu Yudha Pratama, Riyanarto Sarno, "*Personality classification based on twitter text*", International Conference on Data & Software Testing, 2015.
[20]. Manasi Ombhase, Prajakta Gogate, Tejas Aptil, Karan sNair, Prof. Gayatri Hedge, "*Automated Personality classification using Data Mining Techniques*", International Conference on Data & Software Testing, April 2017.
[21]. Cantandir, I. Fernandez-Tobiaz, A. Belllogin, "*Relating Personality types with user preferences in multiple entertainment domains*", EMPIRE 1st Workshop on Emotions and Personality in Personalised Services, 2013.
[22]. C.D. Manning, P. Raghavan, H. Schutez, "*Introduction to Information Retrieval*", Cambridge University Press, ISBN: 978-0-521-86571-5, 2008.
[23]. Jayashree Rout, Sudhir Bagade, Pooja Yede, Nirmiti Patil, 2019. "*Personality Evaluation and CV Analysis using Machine Learning Algorithm.*", International Journal of Computer Science & Engineering ISSN: 2347-2693 Issue 5, Volume 7, May.