

A
PROJECT REPORT
ON
**UNRAVELING PERSONALITIES WITH ML-BASED CV, VIDEO, AND TONE
ANALYSIS**

SUBMITTED TO THE
SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE OF
BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY
BY

Dipti Bhosale (B190538511)

Shivanjali Adsul (B190538501)

Anjali Bodkhe (B190538515)

Rutuja Kadam (B190538552)

UNDER THE GUIDANCE OF
Prof. Shubhangi Ingale



Department of Information Technology
Zeal Education Society's
ZEAL COLLEGE OF ENGINEERING AND RESEARCH, NARHE-411041
2023-24



CERTIFICATE

This is to certify that the project report entitled
“UNRAVELING PERSONALITIES WITH ML-BASED CV, VIDEO, AND TONE ANALYSIS”

SUBMITTED BY

Dipti Bhosale (B190538511)

Shivanjali Adsul (B190538501)

Anjali Bodkhe (B190538515)

Rutuja Kadam (B190538552)

Is a bonafide work carried out by them under the supervision of Prof. Shubhangi Ingale and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Information Technology). This Project report has not been earlier submitted to any other Institute or University for the award of any degree or diploma.

Prof. Shubhangi Ingale
Internal Guide
Department of IT

Prof. Balaji Chaugule
Head of Department
Department of IT

External Examiner
Date:

Dr. A. M. Kate
Principal

Zeal College of Engineering and Research
Date:

Place:

ACKNOWLEDGEMENT

Our first experience of project has been successfully, thanks to the support staff of many friends and colleagues with gratitude. We wish to acknowledge all of them; however, we wish to make special mention of the following

First of all we are thankful of our project guide Prof. Shubhangi Ingale under whose guideline we were able to complete our project. We will forever remain grateful for his valuable help and encouragement.

We would like to thank our project coordinator Dr. Praveen T. Blessington and Head of Department Prof. Balaji Chaugule and all lab maintenance's staff for providing us assistance in various hardware and software problem encountered during course of our project.

We would like to thank our friends for listening to the ideas, asking questions, and providing suggestions for improving ideas and also for their help. We will fail in our duty if we won't acknowledge a great sense of gratitude to the Principal Dr. Ajit. M. Kate and the entire staff members in Information Technology department for their co-operation.

Dipti Bhosale (B190538511)

Shivanjali Adsul (B190538501)

Anjali Bodkhe (B190538515)

Rutuja Kadam (B190538552)

ABSTRACT

Personality plays a crucial role in understanding human behavior and decision-making processes. Predicting an individual's personality traits is a challenging but highly valuable task with numerous applications, including personalized content recommendation, targeted advertising, and psychological assessment. This research explores the development of a personality prediction model using machine learning techniques to analyze and predict an individual's personality based on their digital footprint, including social media activity, text data, and other relevant features. We employ a combination of natural language processing (NLP) and traditional machine learning algorithms to process and analyze the textual data extracted from various digital platforms. To achieve this goal, we collect and preprocess a comprehensive dataset of textual and behavioral data from social media platforms, user generated content, and psychological assessments. Feature engineering techniques are used to extract relevant linguistic and behavioral features from the data. The machine learning model is trained on this enriched dataset, utilizing various supervised learning algorithms such as Support Vector Machines, Random Forest, and Neural Networks.

Keywords: Keywords-multinomial logistic regression, python, web technology, personality, NLP, FER.

CONTENT

1. INTRODUCTION	1
1.1 Overview	
1.2 Aim	
1.3 Objectives	
1.4 Organization of project report	
2. BACKGROUND AND LITERATURE REVIEW.....	3
3. REQUIREMENT AND ANALYSIS.....	4
3.1 Problem Statement	
3.2 Functional Requirements	
3.2.1 User Interface	
3.2.2 Hardware requirements	
3.2.3 Software requirements	
3.3 Non-Functional Requirements	
3.3.1 Performance Requirements	
3.3.2 Safety Requirement	
3.3.3 Software Quality Attributes	
3.4 System Requirements	
3.4.1 Database Requirements	
3.5 Project Requirement Specification	
3.5.1 Software Requirements (Platform Choice)	
3.5.2 Hardware Requirements	
3.5.3 Assumptions and Dependencies	
3.6 Analysis Model: SDLC Model to be applied	
4. SYSTEM ARCHITECTURE.....	11
4.1 System Architecture	
4.2 Data Flow Diagram	
5. DESIGN.....	13
5.1 Use case diagram	
5.2 Activity diagram	
5.3 Class diagram	
5.4 Sequence diagram	
5.5 Website Prototype	
6. IMPLEMENTATION.....	19
6.1 System Overview	
6.2 Algorithm	
6.3 Technologies	
6.3.1 Transcribe API	
6.3.2 Video Analysis	
6.3.3 FER	
6.3.4 Tone Analysis	
7. RESULTS AND EVALUATION.....	25
8. CONCLUSION.....	26
9. REFERENCES	27

CHAPTER 1

INTRODUCTION

1.1 Overview

Personality prediction using Machine Learning (ML) is an intriguing and evolving field within the broader domain of psychology and data science. It involves leveraging computational techniques and data analysis to predict and assess an individual's personality traits, characteristics, and behaviors based on various data sources. These predictions are typically made by analyzing patterns and correlations in data, allowing us to gain insights into a person's psychological makeup.

1.2 Aim

The aim of personality prediction using Machine Learning (ML) is to develop a model that can accurately assess and predict an individual's personality traits based on available data.

1.3 Objectives

To gain insights into the psychological aspects of personality, including the Big Five personality traits and their impact on an individual's thoughts, emotions, and behaviors. To provide personalized experiences and recommendations in areas such as entertainment, based on an individual's personality traits and preferences. To identify individuals at risk of mental health issues or personality disorders, allowing for early intervention and tailored treatment approaches.

To assist counselors and coaches in offering personalized guidance and support to their clients based on their unique personality traits and challenges. To enhance security by using personality prediction as an additional layer of identity verification and authentication in various applications.

1.4 Organization of project report

- Chapter 1 - INTRODUCTION: In this chapter as shown above is the introduction of Recipe Recommendation System for Indian cuisines as well as the study of scope and software development module s while doing the project work..
- Chapter 2 - BACKGROUND: In this chapter study of literature survey and survey also.
- Chapter 3- SPECIFICATION: - In this chapter introducing problem definition, Requirement Specification and software, hardware requirement
- Chapter 4 5:- IMPLEMENTATION:- In this chapter introducing system architecture flow, ,algorithms as well as all working of each module related to the project.
- Chapter 6:- RESULTS EVALUATION :- In this chapter we explain the extent till which our project has achieved its goals.
- Chapter 7:- CONCLUSIONS AND FUTURE WORK:- In this chapter conclusion of system and future work as well as limitation of project.
- REFERENCES

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

” Prediction of Higher Education Students Academic Grades based on Personality Traits”

The quality of a country's education system has a significant impact on its growth. The education industry has seen significant transformations around the world. It is now recognized as industry, with its own set of challenges, the most significant of which is a drop in students' success rates and their abandonment of courses. Due to a desire to understand the underlying elements that determine academic achievement, predicting students' grades has become a major field of research in education. Early intervention of a student's failure may enable management to provide prompt counseling and coaching to improve success rates and student retention. In this research, we look at how personality features can be used to reinforce learning. Using temporal difference learning analytics and the big five factor model, we will examine the impacts of conscientiousness, neuroticism, extraversion, and other personality qualities on learners. The goal of this paper is to give a quick overview of how scholarly analytics can be utilized in educational institutions, what tools are available, and how institutions can anticipate student performance and achievement using academic analytics.

“Personality Prediction and Analysis using Clustering”

In this paper, K-means clustering is used to analyze human personality. Questionnaires are created by using International Personality Item Pool's public domain collection (IPIP) to collect dataset by creating a Google form. There exists some previous research work based on machine learning methods to predict human personality types. Only a handful of them, however, have good results. Data collection, data optimization, and parameter modification are the three components used in this study. As a consequence, the information is divided into five groups (OCEAN), which we refer to as personality types. In this paper, people with similar personalities are grouped.

“Personality-Assisted Multi-Task Learning for Generic and Personalized Image Aesthetics Assessment”

Traditional image aesthetics assessment (IAA) approaches mainly predict the average aesthetic score of an image. However, people tend to have different tastes on image aesthetics, which is mainly determined by their subjective preferences. As an important subjective trait, personality is believed to be a key factor in modeling individual's subjective preference. In this paper, we present a personality-assisted multi-task deep learning framework for both generic and personalized image aesthetics assessment. The proposed framework comprises two stages. In the first stage, a multi-task learning network with shared weights is proposed to predict the aesthetics distribution of an image and Big-Five (BF) personality traits of people who like the image. The generic aesthetics score of the image can be generated based on the predicted aesthetics distribution. In order to capture the common representation of generic image aesthetics and people's personality traits, a Siamese network is trained using aesthetics data and personality data jointly.

“A Study on Personality Prediction Classification Using Data Mining Algorithms”

Personality is an integral part of one's resume and life; using machine learning we can easily predict and classify one personality using intuitive input methods like questionnaire. In this research endeavor, we have looked to achieve this very motive. After reviewing past work, we concluded to use logistic regression and classifiers to classify people into different personality categories and accurately predict it. With this research, we hope to answer some real world problems by creating a research prototype in the form of a python (Flask) based webapp. The personality classification is solely dependent on the user's inputs, which are compared to the database's recorded values, which is accomplished by training the model on a training dataset, and then eventually categorizing the user's personality into the appropriate group. This is totally an automated system and we have tried to achieve our goal.[4]

” Applicant Personality Prediction System Using Machine Learning”

The proposed system, Applicant Personality Prediction System Using Machine Learning(APPS) will be predicting the personality of the candidate based on the candidate's resume details and also by using some personality tests. Personality prediction system selects the right and suitable candidate required for desired job profile by grading the resume on certain criteria like experience, skills etc. This system can also be used in other fields where they need to hire candidates by filtering processes in a very less time. Our application uses Spacy module to analyze, summarize and compare to get similarity score between resume and job description. We will be considering some other aspects like conscientiousness, openness, agreeableness etc. through which the personality of user can be predicted. For these characteristics we will give a situation so that user will be able to rate himself on a scale of 1 to 10. We will give this personality test scores to machine learning model which is developed using multinomial logistic regression which will predict the personality of job seeker.[5]

“Comparative Analysis for Personality Prediction by Digital Footprints in Social Media”

The use of social media and leaving a digital footprint has recently increased all around the world. It is being used as a platform for people to communicate their sentiments, emotions, and expectations with their data. The data available in social media are publicly viewable and accessible. Any social media network user's personality is predicted based on their posts and status in order to deliver a better accuracy. In this perspective, the proposed research article proposes novel machine learning methods for predicting the personality of humans based on their social media digital footprints. The proposed model may be reviewed for any job applicant during the times of COVID'19 through online enrolment for any organization. Previously, the personality prediction methods are failed due to the differing perspectives of recruiters on job applicants. Also, this estimation is modernized and the prediction time is also reduced due to the implementation of the proposed hybrid approach on machine learning prediction.

“Implementation of Various Artificial Intelligence Approach for Prediction and Recommendation of Personality Disorder Patient”

Personality disorder is known as chronic psychological disorders. It is a serious problem of the individual character and behavior that affect work, family and social life of a person. In many cases, the possibility of people do not realize that they have a personality disorder as a way of thinking and behaving seems natural to the patient, and the patient may blame others. To diagnose a personality disorder early, it is necessary to develop a system that is able to predict personality disorders and recommendation type of therapy patients should undergo. Artificial Intelligence is a technique that is very popular today. Various kinds of approaches and algorithms can be applied in the world of health, such as expert systems, neural networks and fuzzy logic. The aim of this study is to assess the artificial intelligence approach in predicting and recommending the type of therapy for patients with personality disorders.

“Personality and sport performance: The role of perfectionism, Big Five traits, and anticipated performance in predicting the results of distance running competitions”

Although multiple studies have analyzed the impact of perfectionism on sport-related motivation and emotions, studies analyzing its role in actual sport performance are scarce. In the present paper, in two independent studies, we analyze the effects of two features of perfectionism—perfectionistic strivings and perfectionistic concerns—on performance in a 10-kilometer street run (Study 1; n = 332) and a half-marathon (Study 2; n = 133). The participants completed a set of questionnaires online before the competition. Among perfectionistic individuals, the relationships were markedly stronger. Additionally, Study 2 showed that the effects remained significant after controlling for Big Five personality traits, so they cannot be attributed to perfectionists' elevated conscientiousness. The present paper provides pioneering evidence for a significant positive effect of perfectionistic strivings on sport performance in distance running.

CHAPTER 3

REQUIREMENT AND ANALYSIS

3.1 PROBLEM STATEMENT

Problem Statement: To develop a machine learning model to predict an individual's personality traits based on a set of input features, which may include text, demographic information, and behavioral data. The goal is to accurately classify individuals into predefined personality categories or provide continuous scores for specific personality traits.

3.2 FUNCTIONAL SPECIFICATION

3.2.1 User Interface

Application for Personality Prediction.

3.2.2 HARDWARE REQUIREMENTS

- RAM: 8GB
- As we are using Machine Learning Algorithm and Various High Level Libraries Laptop.
- The RAM minimum required is 8 GB.
- Hard Disk: 40 GB
- Processor: Intel i5 Processor
- Pycharm IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required
- IDE: Visual Studio

- Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that make typing feasible and fast.
- Coding Language: Python Version 3.5
- Highly specified Programming Language for Machine Learning because of the availability of High-Performance Libraries.
- Operating System: Windows 10

3.2.3 SOFTWARE REQUIREMENTS

- Operating System: Windows 10
- IDE: Visual Studio
- Programming Language: Python

3.3 NON-FUNCTIONAL REQUIREMENT

3.3.1 Performance Requirements

- The performance of the functions and every module must be well. The overall performance of the software will enable the users to work decently. The performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement.
- The application is designed in modules where errors can be detected and steadily.
- This makes it easier to install and update new functionality if required.

3.3.2 Safety Requirement

- The application is designed in modules where errors can be detected and fixed easily.
- This makes it easier to install and update new functionality if required

3.3.3 Software Quality Attributes

Our software has many quality attributes that are given below: -

- Adaptability: This software is adaptable by all users.
- Availability: This software is freely available to all users. The availability of the software is easy for everyone.
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
- Reliability: The performance of the software is better which will increase their liability of the Software.
- User Friendliness: Since the software is a GUI application; the output generated is much user friendly in its behavior.
- Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled
- Security: Users are authenticated using many security phases so reliable security is provided.
- Testability: The software will be tested considering all the aspects.

3.4 SYSTEM REQUIREMENTS

3.4.1 Database Requirements

- MYSQL 8.0

3.5 PROJECT REQUIREMENT SPECIFICATION

3.5.1 Software Requirements (Platform Choice)

- Operating system: Windows 7 or more.
- Coding Language: python
- IDE: Visual Studio

3.5.2 Hardware Requirement

- System: Intel I3 Processor and above.
- Hard Disk: 20 GB.

- Ram: 8GB.

3.5.3 Assumptions and Dependencies

- User must require the Python.
- User has to install the Spyder on his pc.
- User has to login to the system.

3.6 ANALYSIS MODEL: SDLC MODEL TO BE APPLIED

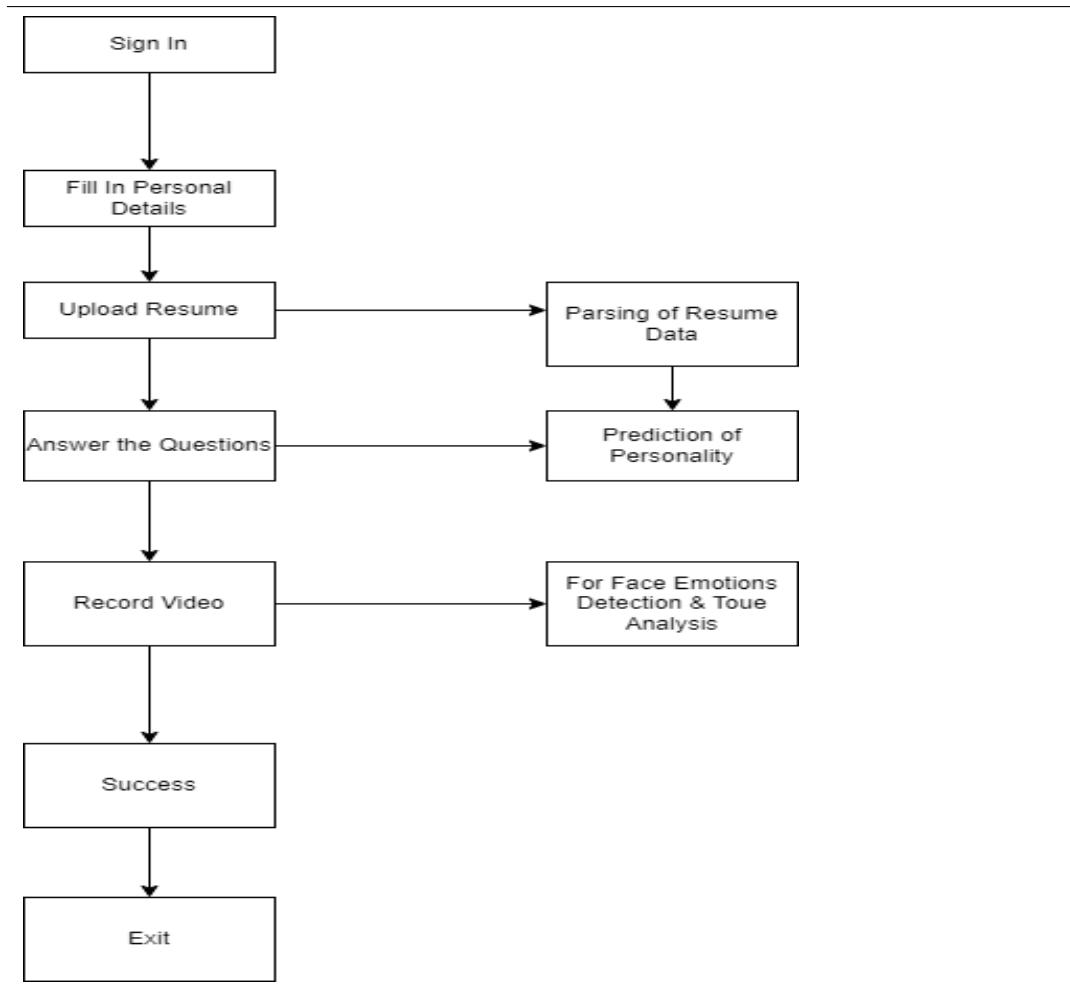
The software development cycle is a combination of different phases such as designing, implementing and deploying the project. These different phases of the software development model are described in this section. The SDLC model for the project development can be understood using the following figure. The chosen SDLC model is the waterfall model which is easy to follow and fits best for the implementation of this project.

1. **Requirements Analysis:** At this stage, the business requirements, definitions of use cases are studied and respective documentations are generated. **Design:** In this stage, the designs of the data models will be defined and different data preparation and analysis will be carried out.
2. **Implementation:** The actual development of the model will be carried out in this stage. Based on the data model designs and requirements from previous stages, appropriate algorithms, mathematical models and design patterns will be used to develop the agent's back-end and front-end components.
3. **Testing:** The developed model based on the previous stages will be tested in this stage. Various validation tests will be carried out over the trained model.
4. **Deployment:** After the model is validated for its accuracy scores it's ready to be deployed or used in simulated scenarios.
5. **Maintenance:** During the use of the developed solution various inputs/scenarios have been countered by the model which might affect the model's overall accuracy or with passing time the model might not fit the new business requirements. Thus, the model must be maintained often to keep its desired state of operation.

CHAPTER 4

SYSTEM ARCHITECTURE

4.1 System Architecture



4.1 System Architecture

4.2 Data Flow Diagram

In the Data Flow Diagram, we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system.

In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected likewise in DFD 2 we present operation of user as well as admin.

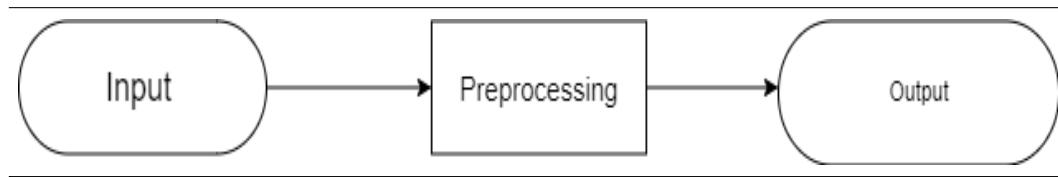


Figure 4.2: Data Flow (0) diagram

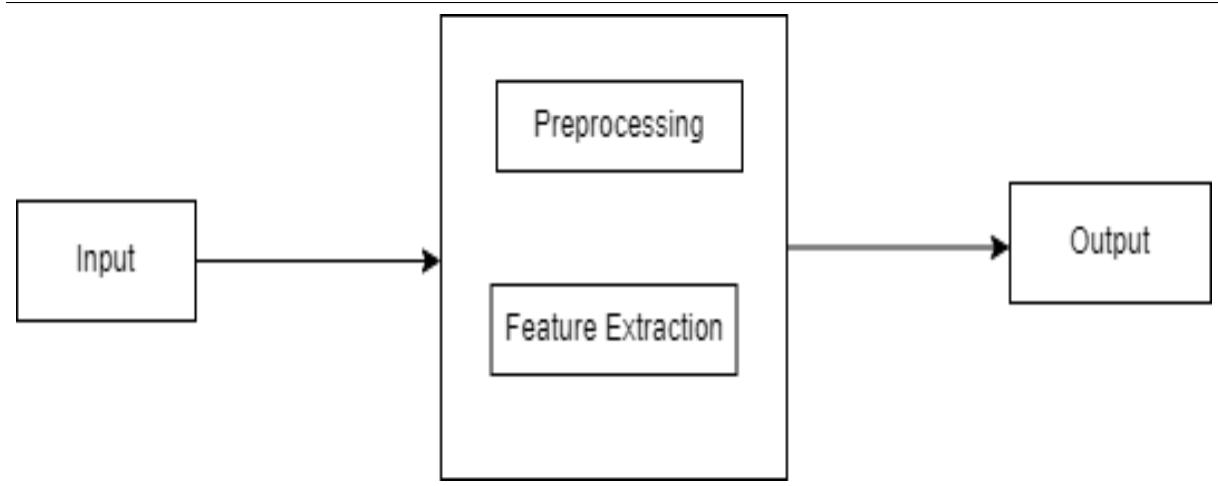


Figure 4.3: Data Flow (1) diagram

CHAPTER 5

DESIGN

5.1 Use Case Diagram

A use case diagram is a graphical representation of a user's interaction with the system and depicts the specifications of a use case. A use case diagram can show the different types of users of a system and the various ways in which they interact with the system. Use case diagrams are used to gather the requirements of a system including internal and external influences.

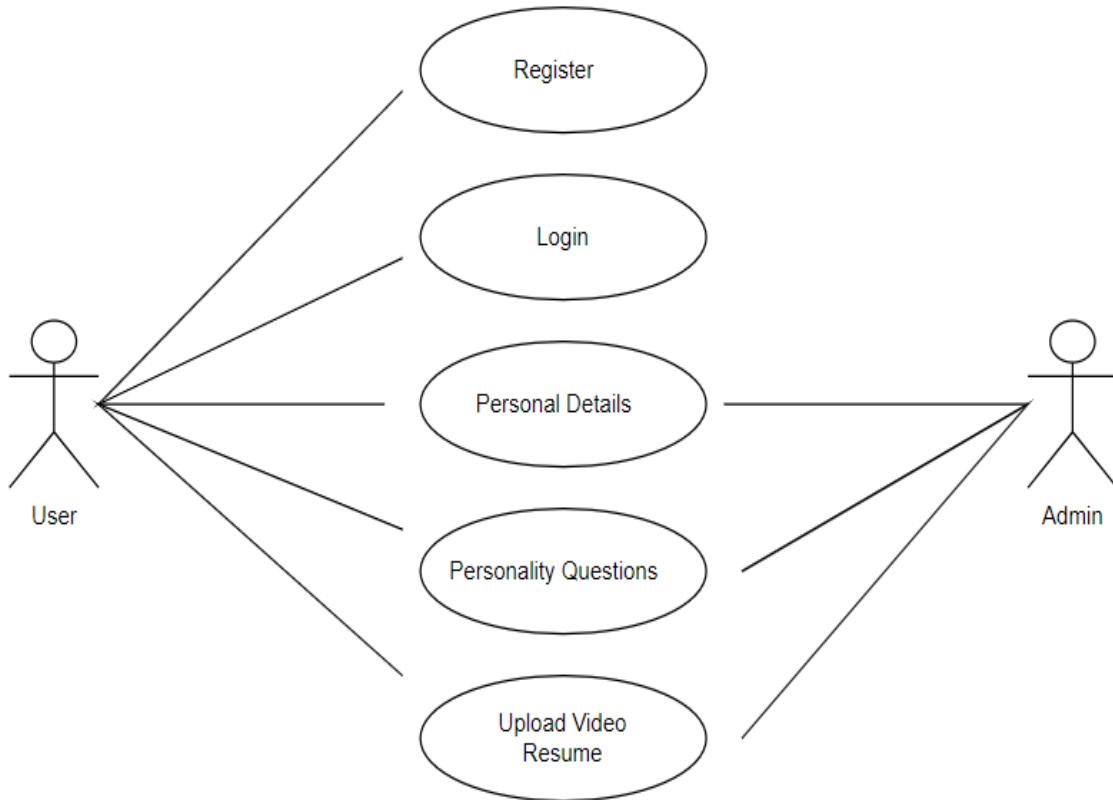


Figure 5.1: Use case diagram

5.2 Activity Diagram

Activity diagrams are graphical representations of workflows of step-wise activities and actions with support for choice, iteration, and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows). Activity diagrams show the overall flow of control.

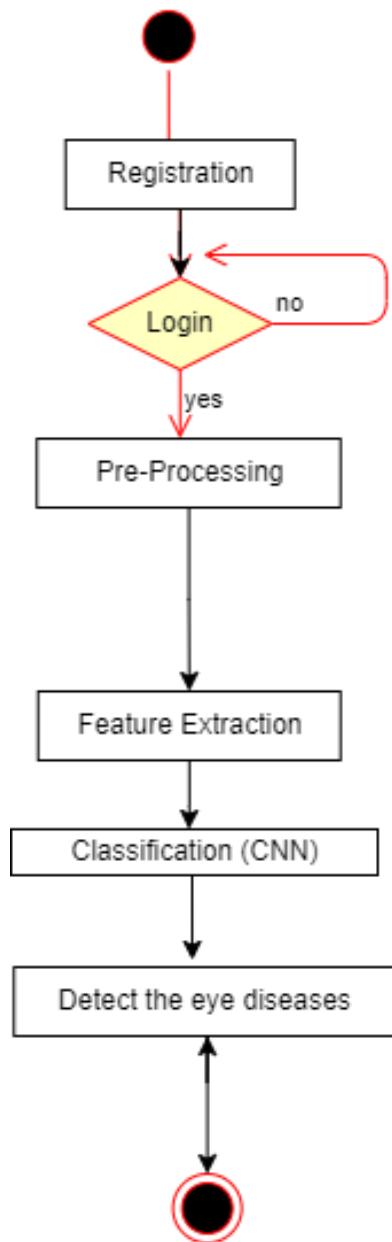


Figure 5.2: Activity diagram

5.3 Class Diagram

The class diagram is a static diagram. It represents the static view of an application.

Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application.

The class diagram describes the attributes and operations of a class and also the constraints imposed on the system.

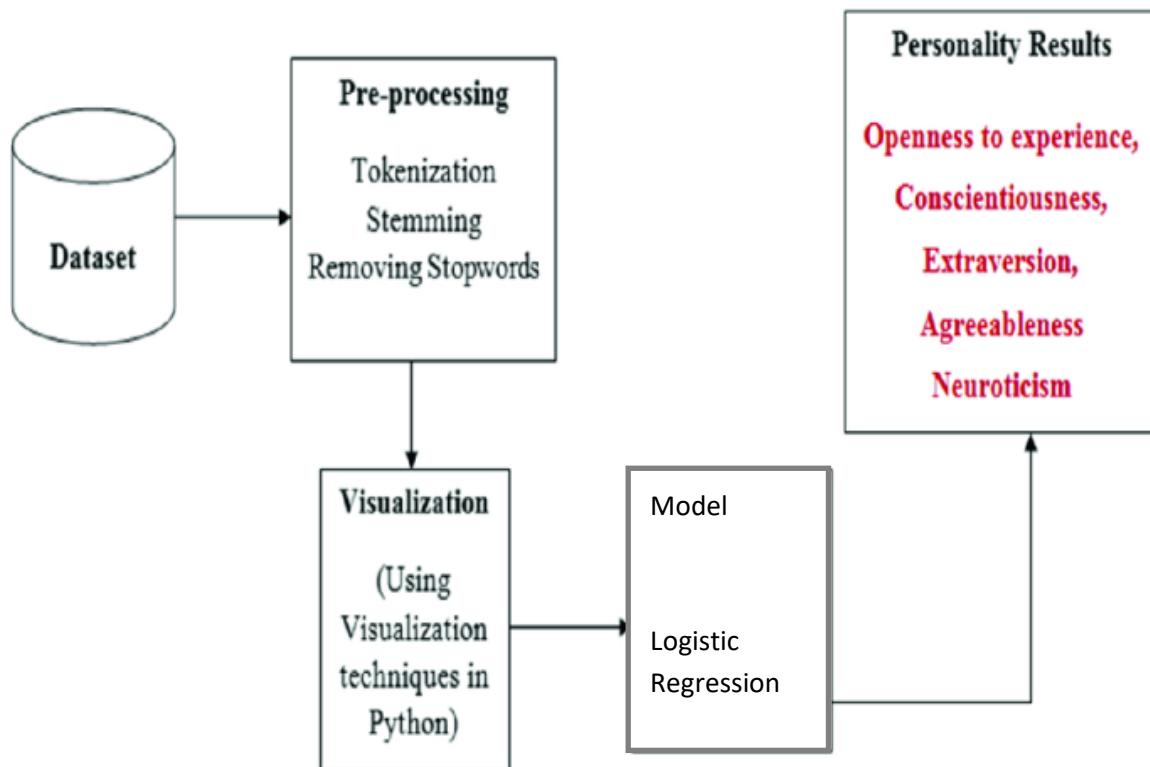


Figure 5.3: Class diagram

5.4 Sequence Diagram

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

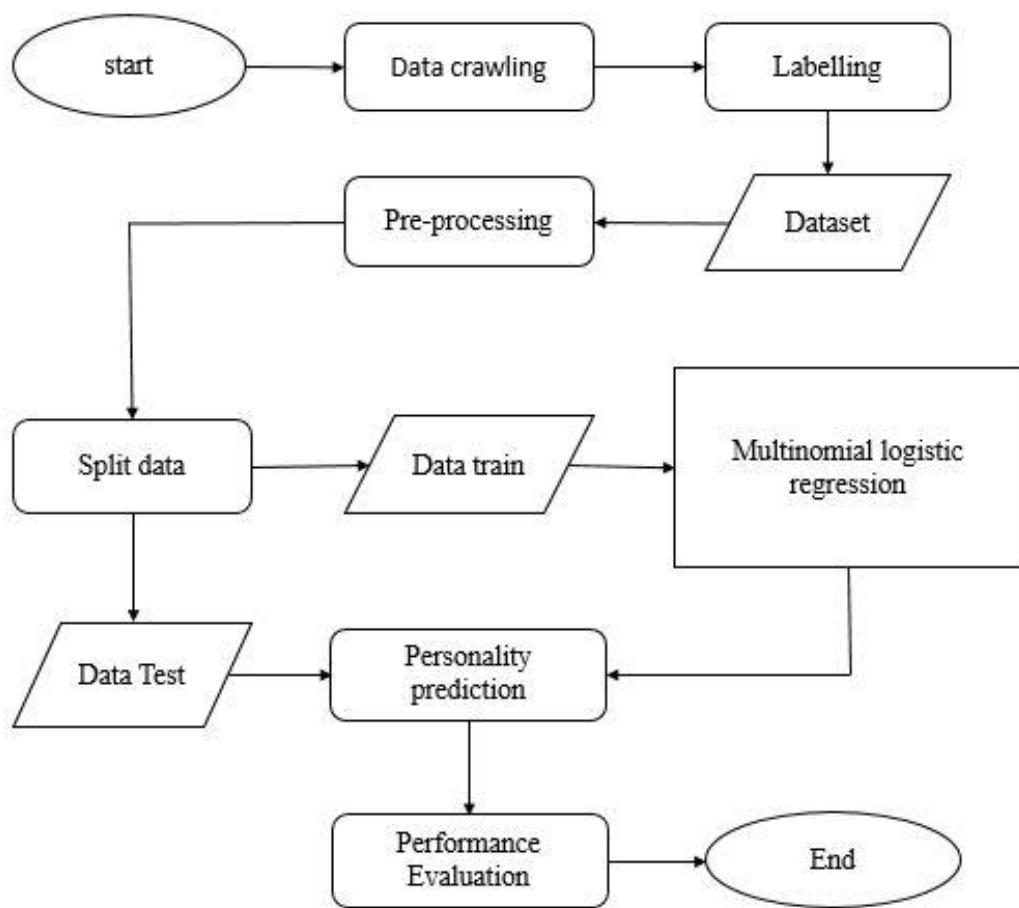


Figure 5.4: Sequence diagram

5.5 Website Prototype

WEBSITE PROTOTYPE – SLIDING LANDING PAGE

<p>Welcome Message</p> <p>For taking interview please login with company's credentials</p>	<p>Lets Start your Interview</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/></p> <p>Name Email Password</p>
--------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------

Figure 5.5.1: Sliding landing page

WEBSITE PROTOTYPE – PERSONALITY PREDICTION PAGE

NAME	GENDER	EMAIL
CONTACT DETAILS		UPLOAD RESUME
Openness	→	
Neuroticism	→	
Conscientiousness	→	
Agreeableness	→	
Extraversion	→	

GIVE RATING FROM 1-10, COMBINED WITH CV ANALYSIS PERSONALITY WILL BE PREDICTED USING MACHINE LEARNING

Figure 5.5.2: Personality prediction page

WEBSITE PROTOTYPE – RECORDING PAGE

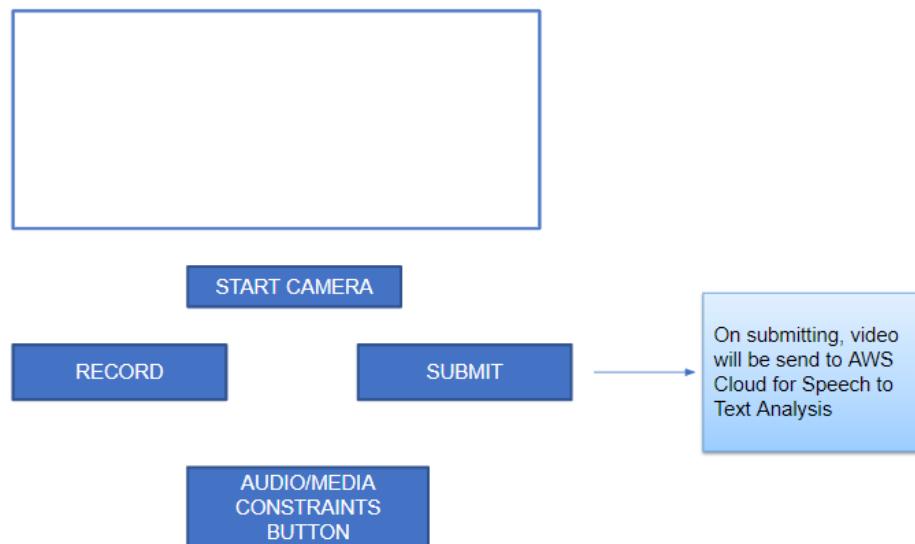


Figure 5.5.3: Recording page

WEBSITE PROTOTYPE – ADMIN DASHBOARD

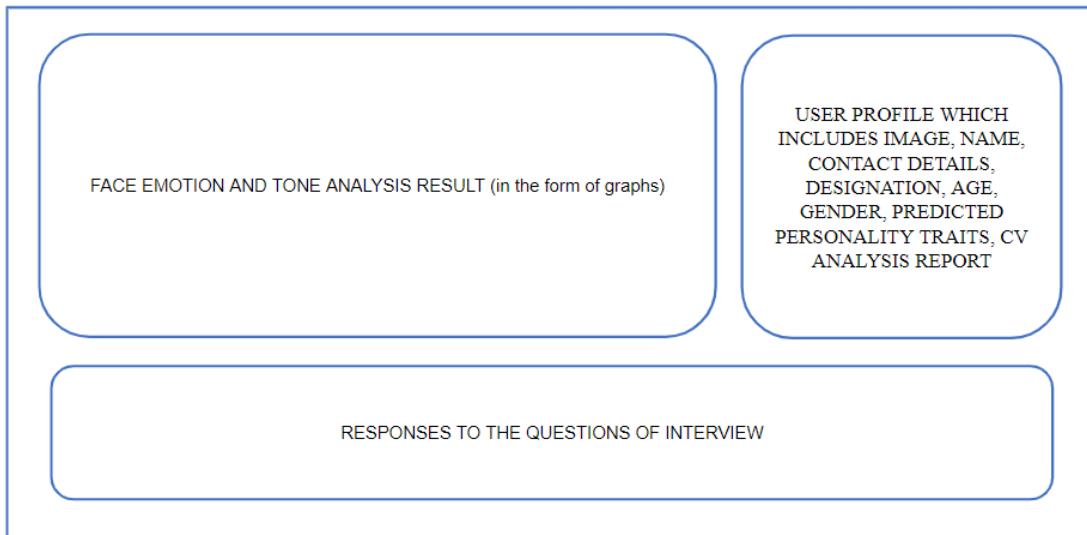


Figure 5.5.4: Admin dashboard

CHAPTER 6

IMPLEMENTATION

6.1 SYSTEM OVERVIEW

Personality prediction using machine learning is a system designed to predict an individual's personality traits based on various input features. It can find applications in diverse fields, including psychology, marketing, and human resources. Gather data from various sources, including text data, demographic information, and behavioral data.

Clean and preprocess the collected data, including handling missing values, data normalization, and feature extraction. The system aims to predict personality traits accurately, providing valuable insights for various applications while adhering to ethical and privacy considerations. It requires continuous attention to ensure that it remains effective and unbiased.

6.2 ALGORITHM

“Multinomial Logistic Regression:”

Multinomial Logistic Regression is a statistical method used for modeling the relationship between a dependent variable with more than two categories and one or more independent variables. It is an extension of binary logistic regression, which is used when the dependent variable has only two categories. In multinomial logistic regression, the dependent variable is categorical and represents more than two classes or categories.

Independent variables are features or predictors used to make predictions about the dependent variable. Like binary logistic regression, multinomial logistic regression uses the logic function to model the relationship between independent variables and the dependent variable.

“Natural Language Processing (NLP) :”

Natural Language Processing (NLP) is a subfield of artificial intelligence and linguistics that focuses on the interaction between computers and human language. NLP algorithms are designed to enable computers to understand, interpret, and generate human language. Tokenization is the process of breaking down a text into individual words or tokens. It is a fundamental step in NLP and is often the first preprocessing step.

NLP is a rapidly evolving field with a wide range of applications, including Chabot's, virtual assistants, automated translation, sentiment analysis, and much more. Researchers and developers continue to push the boundaries of NLP with the development of more advanced models and techniques.

6.3 TECHNOLOGIES

6.3.1 Transcribe API

A Transcribe API is a type of application programming interface (API) that allows developers to integrate transcription functionality into their applications or services. Transcription involves converting spoken language into written text, typically through the use of speech recognition technology. Or stream audio.

Amazon Transcribe: A fully managed automatic speech recognition (ASR) service provided by Amazon Web Services (AWS) that makes it easy for developers to add speech-to-text capability to their applications. It supports multiple languages and can transcribe audio files stored in Amazon S3 or stream audio.

6.3.2 Video Analysis

We can easily convert audio to text with Amazon Transcribe, an AWS AI service. Video analysis can recognize and classify human activities within a video.

Transcription:

Use the Amazon Transcribe API to convert spoken audio from videos into text. In this phase, you will receive the textual data needed for analysis.

Steps:

1. Install and set up the AWS CLI (Command Line Interface) on your local computer.
2. Set up the AWS CLI by running aws configure and inputting the output format, AWS Region, Secret Access Key, and AWS Access Key ID. In the AWS Management Console, under IAM
3. (Identity and Access Management) > Users > Your User > Security credentials, you may locate your Access Key ID and Secret Access Key.
4. Create an S3 bucket: To store the audio file you wish to transcribe, Transcribe needs an S3 bucket. Either the AWS Management Console or the AWS CLI can be used to create an S3 bucket.
5. Upload your audio file to the S3 bucket: To upload your audio file to the S3 bucket, use the AWS Management Console or the AWS CLI.
6. Transcribing the audio file: To begin the transcription process, use the AWS CLI.
7. Verify the status of the transcription job: The AWS Management Console or the AWS CLI can be used to verify the transcription job's status.
8. Obtain the transcription outcome: You can access the transcription result from the S3 bucket once the transcription process is finished.

6.3.3 FER

Facial Expression Recognition: FER refers to the automated process of identifying human facial expressions using computer vision and machine learning techniques. Facial Expression Recognition (FER) uses facial expression analysis to infer personality traits in personality prediction. It is based on the ability to recognize emotions, including microexpressions, which can reveal facets of a person's personality. For more precise forecasts, FER is frequently combined with other modalities like speech analysis. But given the intricacy of personality and the moral issues surrounding bias and privacy, care must be taken.

Six Facial Expressions are: -

- Happy
- Sad
- Fear
- Angry
- Surprise
- Disgust

Face Detection: The first step is to detect and locate human faces within an image or video frame. This involves identifying key facial landmarks such as the eyes, nose, and mouth.

Feature Extraction: Once the faces are detected, various features are extracted from the facial region. These features may include the positions of facial landmarks, the shape of the mouth, the curvature of the eyebrows, etc.

Feature Representation: The extracted features are then transformed into a format that can be inputted into a machine-learning model.

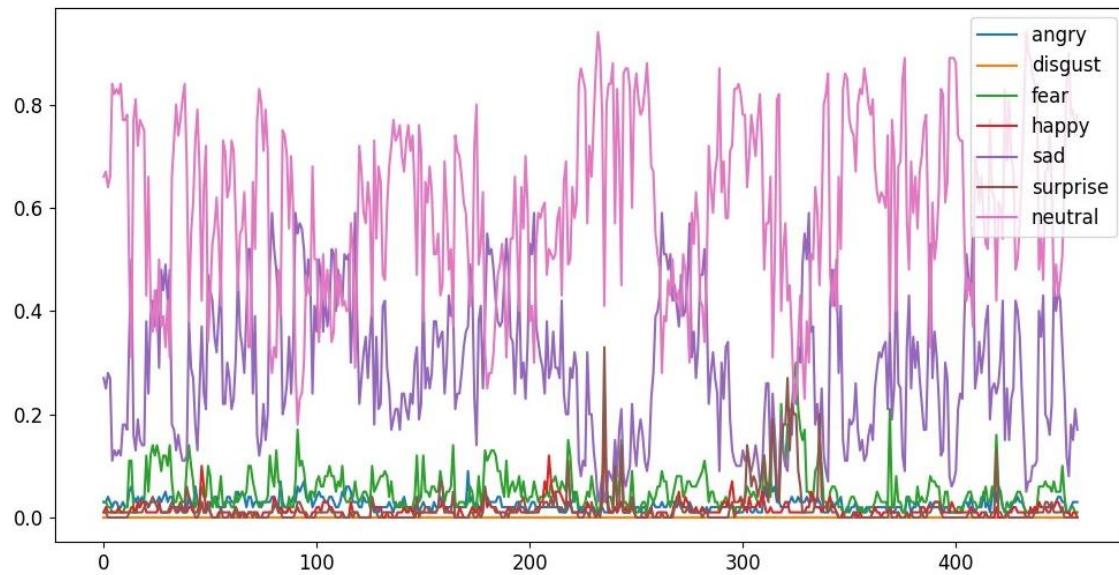


Figure 6.3.3: FER

Fig. 6.3.3 shows the output generated by Facial Expression Recognition (FER) on a

video. The FER system analyzes each frame of the video to detect and classify facial expressions in real time. Each frame is processed to identify faces, which are then analyzed to determine the corresponding facial expressions, such as ‘Angry’, ‘Disgust’, ‘Fear’, ‘happy’, ‘sad’, ‘Surprise’, or ‘neutral’.

6.3.4 Tone Analysis

Tone analysis, also known as sentiment analysis or emotion detection, is a process of computationally identifying and understanding the emotional tone or sentiment expressed in a piece of text, such as an article, social media post, email, or customer review. The goal of tone analysis is to determine the underlying sentiment of the text, whether it's positive, negative, or neutral, and sometimes to identify specific emotions like happiness, sadness, anger, or excitement.

Here's how tone analysis typically works:

Text Preprocessing: The text is first preprocessed to remove any irrelevant information, such as punctuation, stop words (commonly used words like "the," "and," "is"), and special characters. The text may also be tokenized, which involves breaking it down into individual words or phrases.

Feature Extraction: Various linguistic features are extracted from the text to capture different aspects of sentiment. These features might include word frequency, word order, grammatical structure, and syntactic patterns. Additionally, advanced techniques may involve analyzing the context in which words appear, such as identifying negations or intensifiers.

Sentiment Classification: Machine learning algorithms, such as classification models or neural networks, are trained using labeled datasets to classify text into different sentiment categories. In supervised learning, the model learns from examples where the sentiment of the text is already known. During training, the algorithm identifies patterns in the features extracted from the text and learns to associate them with specific sentiments.

Sentiment Analysis: Once the model is trained, it can be used to analyze the sentiment of new, unseen text. The text is fed into the trained model, which then predicts the sentiment or emotional tone expressed in the text. The output might include a sentiment score (e.g., positive, negative, neutral) or specific emotional categories (e.g., joy, sadness, anger).

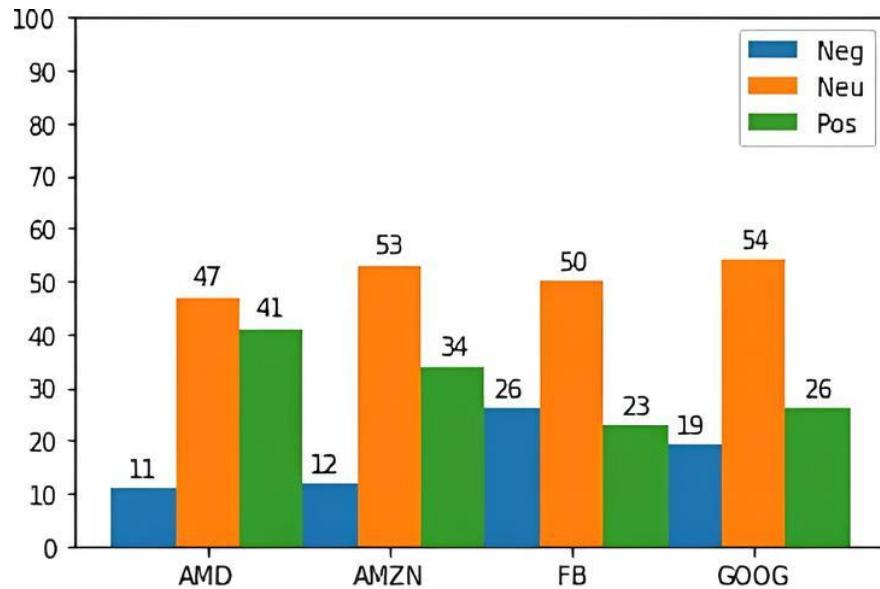


Figure 6.3.4: Tone analysis

Fig. 6.3.4 shows the output generated by sentiment analysis using the SentimentIntensityAnalyzer from the NLTK library. The system analyzes the text in real time, assigning sentiment scores for positive, negative, neutral, and compound sentiments to each sentence or document. The output includes visualizations of the text, with highlighted sentences or phrases indicating the sentiment detected. The compound score, ranging from -1 (most negative) to +1 (most positive), is also displayed for each sentence or document, providing an overall sentiment assessment.

CHAPTER 7

RESULTS AND EVALUATION

Results and Evaluation



Figure 7.1: Results and evaluation

Fig. 7.1 shows the result of Personality Prediction, Video and tone analysis. As we can see in the figure there are two charts, line chart and bar chart, depicting the video and tone analysis. At the extreme right we can see the details of the user which are filled in the form by the user, their personality predicted as a result of the five questions the user answered and the key skills parsed from the resume they've uploaded. And at the bottom we can see the questions and the answers given by the user for that specific question, which is obtained using the AWS transcribe service.

CHAPTER 8

CONCLUSION

The Big Five personality model served as the foundation for this investigation, and our predictions were made using a multinomial logistic regression model. Some intriguing patterns were seen in our model's performance for each of the Big Five personality traits: neuroticism, agreeableness, extraversion, conscientiousness, and openness. The model's total accuracy is 0.81. These findings have implications for applications such as tailored suggestions and mental health, in addition to being a contribution to the field of personality prediction.

Personality prediction through machine learning is a powerful tool with broad-reaching applications that can enhance various aspects of human life. However, its responsible and ethical use is of paramount importance to ensure that it benefits individuals and society while respecting privacy and data security. Ongoing advancements and interdisciplinary collaboration will continue to shape the field, making it an exciting and evolving area of study and application.

CHAPTER 9

REFERENCES

- [1].Pavitha N., Somesh Kamnapure, Ayush Gundawar, Ishan Gujarathi, Devang Manjramkar, Dhananjay Deore Vishwakarma Institute of Technology, Pune, Maharashtra, India.
- [2].2023 International Conference on Disruptive Technologies (ICDT) Smart-Hire Personality Prediction Using ML.
- [3].November 2023/Meiling Li; Hezi Liu Language Style Matters: Personality Prediction from Textual Styles Learning[IEEE] 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 BigFive 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." Athanasios Tsakalidis, Giannis Tzimas, and Emantha Faliagka, "An integrated e-recruitment system for automated personality mining and applicant ranking", 2012.
- [9].L. D. van der, J. te Nijenhuis, et al., "The General Factor of Personality" 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.

- [10]. Aleksandar Kartelj, Vladimir Filipovic, Veljko Milutinovic, “Novel approaches to automated personality classifications ideas & their potential”, 35th International conference MIPRO, IEEE, 2012.
- [11]. 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.
- [12]. J. Golbeck, C. Robles, K.Turner, “Predictiong personality with social media”, CHI’11 Extended Abstract on Human Factors in Computing Systems, pp.253-262, 2011.
- [13]. Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C Graesser, “Using Data Mining Techniques to detect the personality of players in an Educational Game”, Springer International Publishing, 2014.
- [14]. Randall Wad, Taghi M. Khoshgoftaar, Amri Napolitano, Chris Sumner, “Using Twitter Content to Predict Psychopathy”, 11th International conference on Machine learning & Applications, IEEE, 2012.
- [15]. Educational Game Yago Saez, Carlos Navarro, Asuncion Mochon, Pedro Isasi, 2014. “A system for personality & Happiness detection, International Journal of Interactive Multimedia & Artificial Intelligence Bayu Yudha Pratama, Riyanarto Sarno, “Personality classification based on twitter text”, International Conference on Data & Software Testing, 2015.
- [16]. Educational Game Yago Saez, Carlos Navarro, Asuncion Mochon, Pedro Isasi, 2014. “A system for personality & Happiness detection, International Journal of Interactive Multimedia & Artificial Intelligence.
- [17]. Yilun Wang, “Understand Personality through social media”, International of computer Science standford university.
- [18]. 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.
- [19]. Cantandir, I. Fernandez-Tobiaz, A. Bellogin, “Relating Personality types with user preferences in multiple entertainment domains”, EMPIRE 1st Workshop on Emotions and Personality in Personalised Services, 2013.
- [20]. C.D. Manning, P. Raghavan, H. Schutez, “Introduction to Information Retrieval”, Cambridge University Press, ISBN: 978-0-521-86571-5, 2008.



A Comprehensive Survey on Personality Prediction Using Machine Learning Techniques

Prof. Amol Chincholkar¹, Dipti Bhosale², Shivanjali Adsul³, Anjali Bodkhe⁴, Rutuja Kadam⁵

Department of Information Technology, Zeal College of Engineering and Research, Narhe¹⁻⁵

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.

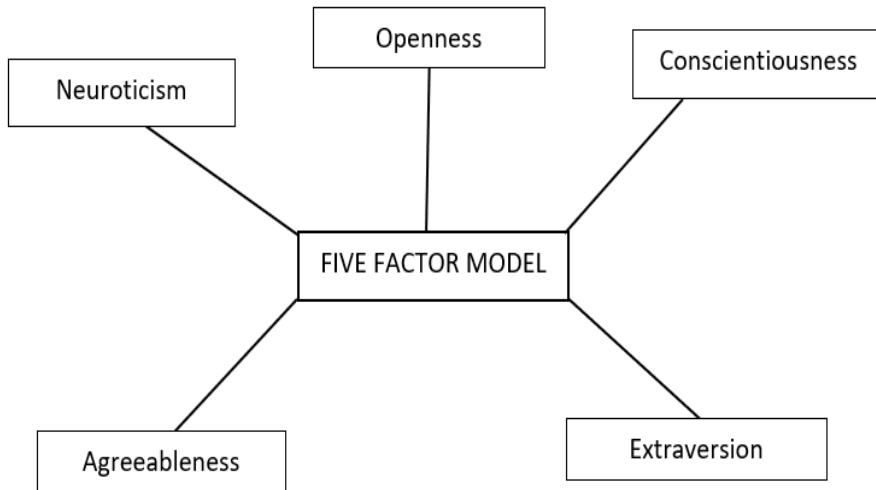


Fig. 1 Big Five Factor Model

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 more 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 Pavithra 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.



- 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.

FOR USER:

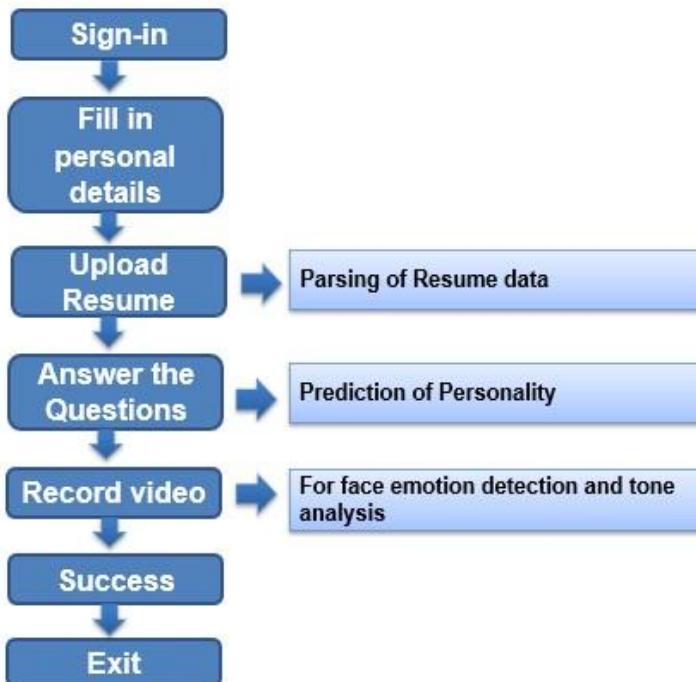


Fig. 2 For User

**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.

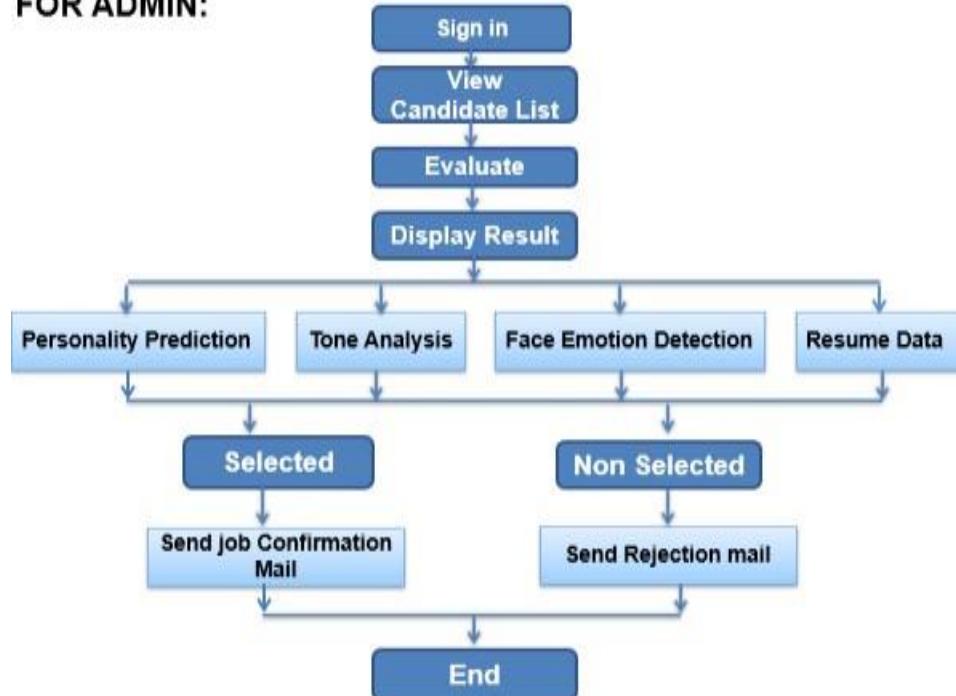
FOR ADMIN:

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.

**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."



- [9]. Athanasios Tsakalidis, Giannis Tzimas, and Emantha 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, "Using Data Mining Techniques to detect the personality of players in an Educational Game", Springer 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, Rianarto 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.



INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, NAAS Accredited Science Journal

ISO 3297:2007 Certified

Thomson Reuters ID I-8645-2017



Certificate of Publication

PROF. AMOL CHINCHOLKAR

Department of Information Technology, Zeal College of Engineering and Research, Narhe

Published a paper entitled

**A Comprehensive Survey on Personality Prediction
Using Machine Learning Techniques**

Volume 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

Certificate# IJARCCE/2023/1

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

Tejass Publishers
ORGANIZATION


Editor-in-Chief
IJARCCE



INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, NAAS Accredited Science Journal

ISO 3297:2007 Certified

Thomson Reuters ID I-8645-2017



Certificate of Publication

DIPTI BHOSALE

Department of Information Technology, Zeal College of Engineering and Research, Narhe

Published a paper entitled

**A Comprehensive Survey on Personality Prediction
Using Machine Learning Techniques**

Volume 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

Certificate# IJARCCE/2023/1

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

Tejass Publishers
ORGANIZATION


Editor-in-Chief
IJARCCE



INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, NAAS Accredited Science Journal

ISO 3297:2007 Certified

Thomson Reuters ID I-8645-2017



Certificate of Publication

SHIVANJALI ADSUL

Department of Information Technology, Zeal College of Engineering and Research, Narhe

Published a paper entitled

**A Comprehensive Survey on Personality Prediction
Using Machine Learning Techniques**

Volume 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

Certificate# IJARCCE/2023/1

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

Tejass Publishers
ORGANIZATION


Editor-in-Chief
IJARCCE



INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, NAAS Accredited Science Journal

ISO 3297:2007 Certified

Thomson Reuters ID I-8645-2017



Certificate of Publication

ANJALI BODKHE

Department of Information Technology, Zeal College of Engineering and Research, Narhe

Published a paper entitled

**A Comprehensive Survey on Personality Prediction
Using Machine Learning Techniques**

Volume 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

Certificate# IJARCCE/2023/1



INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, NAAS Accredited Science Journal

ISO 3297:2007 Certified

Thomson Reuters ID I-8645-2017



Certificate of Publication

RUTUJA KADAM

Department of Information Technology, Zeal College of Engineering and Research, Narhe

Published a paper entitled

**A Comprehensive Survey on Personality Prediction
Using Machine Learning Techniques**

Volume 12, Issue 11, November 2023

DOI: 10.17148/IJARCCE.2023.121120

Certificate# IJARCCE/2023/1

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns

**Prof. Shubhangi Ingale¹,
Dipti Bhosale², Shivanjali Adsul³, Anjali Bodkhe⁴, Rutuja Kadam⁵**
Zeal College of Engineering and Research, Narhe

Abstract- Abstract: This study classifies people's personalities using the logistic regression machine learning technique. Moreover, computers using Natural Language Processing (NLP) can comprehend and converse in human languages. Many previous studies have tried to determine a person's personality type automatically. Classifying people according to their personality traits is one of the most significant uses of machine learning algorithms. There are several advantages to categorizing people. In the world of boundless potential that exists today, knowing one's personality can be quite helpful. These forecasts can be used by anyone to choose a career or other interests. These days, many companies utilize personality tests to select candidates since they increase employee productivity because the candidates are doing what they do best.

Keywords-multinomial logistic regression, python, web technology, personality, NLP, IBM Watson (video analysis)

INTRODUCTION

The phrase "personality" refers to the distinct ideas, emotions, and inclinations of an individual. Personality tests cover a wide range of topics. Its goal is to ascertain your leadership style, temperament, strengths, and shortcomings.

Personality is a widely recognized paradigm in psychology that explains the basic components of temperament in living souls. A theory was developed based on all these features. The five-factor model, sometimes known as the Big 5 model, is the most frequently accepted and utilized personality model. For most of their studies in individual differences and personality psychology, students can make use of the word and the conceptual framework that it provides. Characteristics of the Mind: To determine a person's personality, the Openness system relies on a questionnaire. It can be modified even more to get candidates to provide more details. We propose a logistic regression-based personality classification system to tackle this problem.

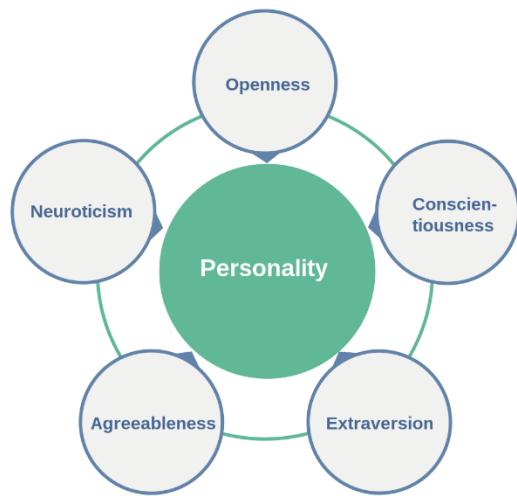
Openness: This characteristic encompasses qualities like awareness, inventiveness, insight, sensitivity, and curiosity. Individuals with high openness scores are imaginative, inquisitive, and receptive to new things.

Conscientiousness: This characteristic has to do with how thoughtful, disciplined, careful, and diligent a person is. High conscientiousness individuals typically have well-defined objectives, strong self-control, and excellent organization.

Extroversion: This trait relates to an individual's level of emotional and assertiveness. Extroverts tend to be outgoing, comfortable in social settings, and enthusiastic.

Agreeableness: This characteristic is linked to collaboration and charity in people. Many who are agreeable are also compassionate, dependable, and kind.

Neuroticism: This trait relates to an individual's ability to experience stable emotions as well as their tendency to experience sad and anxious sentiments. Those who score highly on neuroticism are more vulnerable to stress and mood fluctuations.



The system's objective:

To create a technique that will make it easier to recognize an applicant's personality traits and learn more about them without having to meet them in person. Because it will have a greater understanding of the individual, the company will be able to choose the best-qualified applicant for the open position.

SUGGESTIVE MODEL

1) Data Collection

The data set was gathered through discussions with potential employees and a variety of websites. The questions and answers were submitted to Google Forms and stored as a CSV file for easy data

retrieval and training. The picture below displays some questions on Openness, one of the Big Five personality traits. There is a predetermined range of responses for each topic, which goes from strongly agree to strongly disagree.

2) Data Analysis

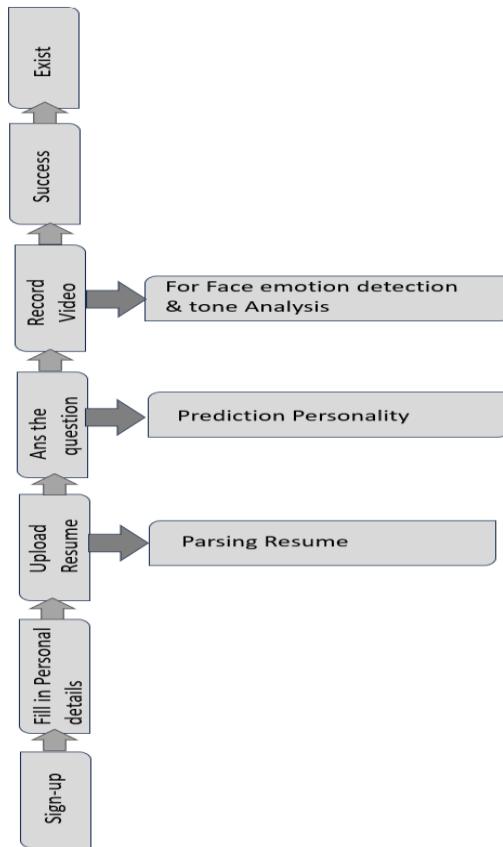
We scale the test dataset using the Standard Scaler from the Scikit Learn Library after dividing it into x- and y-tests. Our model was run on a Jupyter Notebook computer system. Among the Python libraries we have used are pandas, sklearn, numpy, re, seaborn, and matplotlib.

SYSTEM ARCHITECTURE

A) User

To rate one's personality forecast using OCEAN values and CV analysis (based on the Big Five Personality Traits model), one must provide personal information, submit a résumé, and complete a questionnaire, and internet video capture. The candidate needs to reply to a few questions that the HR department has put up on the portal.

Speech and facial emotion analysis might reveal details about a candidate's confidence level and personality.



B) Multinomial Logistic Regression

A classification method in statistics called multinomial logistic regression applies logistic regression to multiclass problems—that is, problems with more than two possible outcomes. (Reference:).

Put differently, it is a model that calculates the likelihood of specific events given a set of independent variables (which may be real-valued, binary-valued, categorical-valued, etc.).

A dependent variable's various occurrences are dispersed according to its kind.

Multinomial logit (mlogit), multinomial LR, multiclass LR, SoftMax regression, maximum entropy classifier (MaxEnt), and maximum entropy model conditional are

some of the other terms for multinomial logistic regression.

C) Execution

There are training and testing portions of the dataset. With the use of the Standard Sklearn Library, the dataset is further scaled. Testing makes up thirty percent, while training makes up seventy percent. In the dataset, which has 972 rows and 8 columns, the candidate's age, gender, and one of the five personality traits identified by the OCEAN Model are listed in each row. The relevant rows contain a list of the participant data.

D) Natural Language Processing (NLP)

The study of how computers and mortal language interact is the focus of the artificial intelligence (AI) subfield of natural language processing or NLP. NLP facilitates the understanding and use of mortal language by computers.

Steps:

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

E) Video

Transcription:

AWS Transcribe is a service provided by Amazon Web Services (AWS) that converts speech to text. It uses advanced machine learning technologies to recognize spoken language and transcribe it into written text.

This service can be used to convert audio and video files, as well as live audio streams, into text. We use the Amazon Transcribe API to convert spoken audio from videos into text. In this phase, you will receive the textual data needed for analysis.

F) FER (Facial Expression Recognition)

Facial Expression Recognition (FER) uses facial expression analysis to infer personality traits in personality prediction. It is based on the ability to recognize emotions, including microexpressions, which can reveal facets of a person's personality. For more precise forecasts, FER is frequently combined with other modalities like speech analysis. But given the intricacy of personality and the moral issues surrounding bias and privacy, care must be taken. One popular library for FER is OpenCV, which provides functions for facial detection and recognition. Another popular option is the Dlib library, which offers more advanced features for facial landmark detection and expression analysis.

G) Tone Analysis:

After transcription, the *SentimentIntensityAnalyzer* class in the *nltk.sentiment* module of the NLTK (Natural Language Toolkit) library is used for sentiment analysis, which involves determining the sentiment or tone of a piece of text. This class provides a simple interface for analyzing the sentiment of sentences or documents and can be used to extract sentiment scores indicating the positivity, neutrality, and negativity of the text.

RESULTS & DISCUSSION:

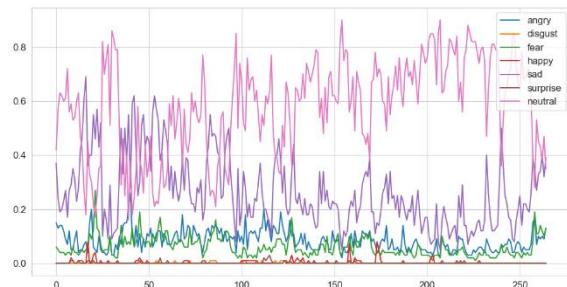


Fig: FER output

Fig.x shows the output generated by Facial Expression Recognition (FER) on a video. The FER system analyzes each frame of the video to detect and classify facial expressions in real time. Each frame is processed to identify faces, which are then analyzed to determine the corresponding facial expressions, such as 'Angry', 'Disgust', 'Fear', 'happy', 'sad', 'Surprise', or 'neutral'.

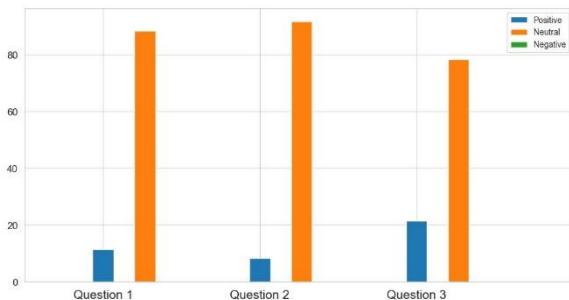


Fig: Tone Output

Fig.x shows the output generated by sentiment analysis using the *SentimentIntensityAnalyzer* from the NLTK library. The system analyzes the text in real time, assigning sentiment scores for positive, negative, neutral, and compound sentiments to each sentence or document. The output includes visualizations of the text, with highlighted sentences or phrases indicating the sentiment detected. The compound score, ranging from -1 (most negative) to +1 (most positive), is also displayed for each sentence

or document, providing an overall sentiment assessment.

CONCLUSION

The Big Five personality model served as the foundation for this investigation, and our predictions were made using a multinomial logistic regression model. Some intriguing patterns were seen in our model's performance for each of the Big Five personality traits: neuroticism, agreeableness, extraversion, conscientiousness, and openness. The model's total accuracy is 0.81. These findings have implications for applications such as tailored suggestions and mental health, in addition to being a contribution to the field of personality prediction.

REFERENCES

- [1] Prof. Waheeda Dhokley, Randeria Kaiwan Jehangir, Shaikh Nabeel Rashid, and Shaikh Almas Mohd Sarwar, "A Novel Approach to Predict Personality of a Person."
- [2] Tanuj Shankarwar, Siddharth Thorat, and Atharva Kulkarni, "Using machine learning, Personality Prediction via CV Analysis"
- [3] 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: 2349-2163 Issue 8, Volume 2, August.
- [4] Clemens Stachl, Quey Au, Ramona Schoedel, and Markus Buhner, "Predicting personality from patterns of behaviour recorded using smartphones."
- [5] Athanasios Tsakalidis, Giannis Tzimas, and Evangelia Faliagka, "An integrated e-

recruitment system for automated personality mining and applicant ranking", 2012.

[6] L. D. van der, J. te Nijenhuis, et al., "The General Factor of Personality"

[7] 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

[8] Aleksandar Kartelj, Vladimir Filipovic, Veljko Milutinovic, "Novel approaches to automated personality classifications ideas & their potential", 35th International conference MIPRO, IEEE, 2012.

[9] 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."

[10] J. Golbeck, C. Robles, K. Turner, "Predictiong personality with social media", CHI'11 Extended Abstract on Human Factors in Computing Systems, pp.253-262, 2011.

[11] Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C Graesser, "Using Data Mining Techniques to detect the personality of players in an Educational Game", Springer International Publishing, 2014.

[12] Randall Wad, Taghi M. Khoshgoftaar, Amri Napolitano, Chris Sumner, "Using Twitter Content to Predict Psychopathy", 11th International conference on Machine learning & Applications, IEEE, 2012.

[13] Educational Game Yago Saez, Carlos Navarro, Asuncion Mochon, Pedro Isasi, 2014. "A system for personality & Happiness detection, International Journal of Interactive Multimedia & Artificial Intelligence.

ISSN NO : 1434-9728



Paper Id : TSJ/1870

Technische Sicherheit

(Technical Security Journal)

Mail : editorstsjournal@gmail.com Website:https://technikwissen.eu/

Scopus And Ugc Care Group 2 Journal

Certificate of Publication

This is to certify that the paper titled

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns



Author by

Dipti Bhosale

From

Zeal College of Engineering and Research, Narhe.

Has been published in

TSJ, Volume 24, Issue 4, April 2024.

DOI:16.0006.SAH



Nach.L. v!

Wilma Nchito
Editor-in-Chief

Scopus®

ISSN NO : 1434-9728



Paper Id : TSJ/1870

Technische Sicherheit

(Technical Security Journal)

Mail : editorstsjournal@gmail.com Website:https://technikwissen.eu/

Scopus And Ugc Care Group 2 Journal

Certificate of Publication

This is to certify that the paper titled

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns



Author by

Rutuja Kadam

From

Zeal College of Engineering and Research, Narhe.

Has been published in

TSJ, Volume 24, Issue 4, April 2024.

DOI:16.0006.SAH



Nach.L. v!

Wilma Nchito
Editor-in-Chief

Scopus®

ISSN NO : 1434-9728



Paper Id : TSJ/1870

Technische Sicherheit

(Technical Security Journal)

Mail : editortsjournal@gmail.com Website:https://technikwissen.eu/

Scopus And Ugc Care Group 2 Journal

Certificate of Publication

This is to certify that the paper titled

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns



Author by

Shivanjali Adsul

From

Zeal College of Engineering and Research, Narhe.

Has been published in

TSJ, Volume 24, Issue 4, April 2024.

DOI:16.0006.SAH



Nach.L. v!

Wilma Nchito
Editor-in-Chief

Scopus®

ISSN NO : 1434-9728



Paper Id : TSJ/1870

Technische Sicherheit

(Technical Security Journal)

Mail : editortsjournal@gmail.com Website:https://technikwissen.eu/

Scopus And Ugc Care Group 2 Journal

Certificate of Publication

This is to certify that the paper titled

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns



Author by

Anjali Bodkhe

From

Zeal College of Engineering and Research, Narhe.

Has been published in

TSJ, Volume 24, Issue 4, April 2024.

DOI:16.0006.SAH



Nach.L. v!

Wilma Nchito
Editor-in-Chief

Scopus®

ISSN NO : 1434-9728



Paper Id : TSJ/1870

Technische Sicherheit

(Technical Security Journal)

Mail : editortsjournal@gmail.com Website:https://technikwissen.eu/

Scopus And Ugc Care Group 2 Journal

Certificate of Publication

This is to certify that the paper titled

AI-Based Personality Assessment: Analyzing Resumes, Videos, and Tone Patterns



Author by

Shubhangi Ingale

From

Zeal College of Engineering and Research, Narhe.

Has been published in

TSJ, Volume 24, Issue 4, April 2024.

DOI:16.0006.SAH



Nach.L. ✓

Wilma Nchito
Editor-in-Chief



Scopus®



