DEPARTMENT OF COMPUTER SCIENCE

Banasthali Vidyapith

Software Requirement Specifications (SRS) and Software Design Specifications (SDS)

CV-based Personality Prediction System

Group -12

Guided By:

Ms. Pratibha Sharma

Submitted By:

Name:	**Roll No.**
Pooja Kumari	2342980
Anupriya Dadhich	2342929
Falguni Dadhich	2342944
Harshita Patni	2342947

Chapter 1 SRS

TABLE OF CONTENTS

1. Introduction

Recruiting the right candidate is a time-consuming and tedious process. It is vital to make a fair and sound decision that benefits both the applicant and the employer. The traditional recruitment process includes various stages, such as aptitude tests, technical assessments, group discussions, and HR interviews. However, this process may not always be equitable in making a final selection.

The curated system has transformed the traditional recruitment process into a more efficient and precise method. With the proposed system, the likelihood of making fair decisions is significantly heightened. The system employs Logistic Regression, Natural Language Toolkit, Flask, and Firebase to ensure smooth and efficient deployment.

1.1 Purpose

When using traditional recruitment methods, it's important to note several key issues. The process can be time-consuming and energy-draining, requiring a team of professionals to sift through CVs and make informed decisions. There is also a high likelihood of making errors in judgment and unfair candidate selection. These flaws must be addressed to improve the hiring process.

1.2 Basic Concept

Our system provides a comprehensive evaluation of candidates, encompassing both psychological and professional assessments. By harnessing the logistic regression algorithm and the Natural Language Toolkit (NLTK), we've streamlined the evaluation process. To ensure the security of user data, we've implemented Firebase as our backend service, providing authentication and secure storage.

1.4 Scope

- **Benefits:**

  - This system is applicable in various industries that require skilled employees.
  - It can reduce the workload of the HR department.
  - The system assists in selecting the right candidate for a specific job profile, leading to the employment of an expert workforce.
  - An administrator or concerned individual can easily select an appropriate candidate based on their online test scores.

.

- **Objectives of the project**

  Here are the main points to keep in mind when it comes to automating candidate shortlisting process:

  - The primary objective is to streamline the hiring process and make it more efficient.
  - The system should be designed to identify key personality traits exhibited by the applicants and gather relevant information without necessarily meeting them.
  - Consequently, this will enable employers to gain better insights about the applicants and make more informed decisions when selecting the most suitable candidate for the job.

1.3 Definitions, Acronyms, and Abbreviations.

- **CV** – Curriculum Vitae
- **NLTK** - Natural Language Toolkit
- **NLP** - Natural Language Processing
- **HR** – Human Resource
- **SVM** – Support Vector Machine
- **UI** - User Interface
- **API** – Application Programming Interface
- **kNN** – k Nearest Neighbors
- **SQL** – Structure Query Language
- **SRS** – Software Requirement Specification

1.4 Overview

In today's job market, recruiters are not just looking for candidates with specific skills; they also want people with the right personality traits. This is because personal characteristics play a significant role in determining success in both professional and personal life. However, with a high number of job seekers and a limited number of vacancies, it can be challenging to manually evaluate each candidate's CV to find the best fit for a job. To address this, our paper explores machine learning techniques to predict personality accurately through CV analysis, using Natural Language Processing (NLP) methods. Our findings indicate that the Random Forest algorithm outperforms other algorithms such as kNN, Logistic Regression, SVM and Naive Bayes in predicting personality traits.

2. The Overall Description

2.1 Product Perspective

Introducing Our Innovative Personality Prediction System through CV Analysis Our cutting-edge solution is designed to harness the power of natural language processing and machine learning techniques to uncover personality traits from curriculum vitae (CV) data. Our product is geared towards human resources professionals, recruiters, and hiring managers who are looking for deeper insights beyond traditional CV information, into a candidate's potential personality characteristics.

Providing Context with a Block Diagram For better context, a block diagram is provided showing the major components of the larger system, interconnections, and external interfaces. This diagram is not meant to be a design or architecture picture, but rather a tool to help illustrate the system's interaction with external actors. The system itself is depicted as a black box, with the internal workings reserved for the design document.

Operating within Various Constraints the following subsections describe the software's operation within a variety of constraints.

2.1.1 PRODUCT FUNCTION

1. **Enhanced Candidate Assessment:**
    - HR professionals can gain a comprehensive understanding of potential hires by supplementing traditional evaluations with insights into their personality traits.

2. **Time Efficiency**
    - Automating personality prediction through CV analysis reduces the time spent manually assessing candidate suitability, allowing recruiters to focus on more strategic aspects of the hiring process.

3. **Improved Hiring Accuracy:**
    - The system contributes to more inform hiring decisions by providing additional context beyond the standard CV content.

4. **Scalability**
    - The product is scalable to suit varying organizational sizes and recruitment volumes. It supports both small businesses and large enterprises, making it a flexible solution for diverse hiring needs.

2.1.2 HARDWARE INTERFACE

- Processor – i3
- Hard Disk – 5 GB
- Memory – 1GB RAM
- Internet Connection

2.1.3 SOFTWARE INTERFACE

- Windows 7 or higher.
- XAMPP
- Anaconda

2.1.4 COMMUNICATION INTERFACES

1. **Web-Based User Interface:**

   The primary user interface is a web-based platform accessible through standard web browsers. This interface allows users, including HR professionals, recruiters, and hiring managers, to interact with the system easily.

   Key Features:

   - Intuitive Dashboard: A user-friendly dashboard provides an overview of processed CVs, personality predictions, and system performance metrics.
   - CV Upload and Processing: Users can upload CVs directly through the interface, and the system processes them in real-time.

2. **Notification System:**

   The system incorporates a notification system to alert users about completed CV analyses, updated predictions, and any relevant system announcements.

   Key Features:

   - Email Notifications: Users can receive email notifications summarizing the results of personality predictions, providing a quick overview without the need to log in.

2.2 User Characteristics

User characteristics play a pivotal role in the effectiveness of a Personality Prediction System through CV Analysis. Candidates who are from tech background or want to pursue career in tech will grab the opportunity. Ample of Opportunity for both fresher's as well as experienced. Overall, acknowledging the uniqueness of users and their CVs is vital for creating a personalized and accurate personality prediction system.

2.3 Constraints

The quality and completeness of CV data can have an impact on the accuracy of predictions. Inconsistent or biased data can undermine the reliability of personality assessments. Furthermore, the system's effectiveness can be influenced by cultural and linguistic variations in CV content. Keep in mind that our system operates exclusively on Windows, requires an internet connection, and only accepts CVs in PDF format.

2.4 Technologies used

- **Tools Used for Model:** Jupyter notebook was the platform that we used to run our model. Pandas, numpy, re, seaborn, matplotlib, and sklearn are some of the Python libraries that we have used.
- **Database:** MySQL
- **User Interface:** PHP

3. Specific Requirements

Creating a CV-based personality prediction system involves a combination of computer vision, natural language processing, and machine learning techniques. Below is a list of specific requirements and steps you may consider for such a project:

1. **Data Collection:**
    - Gather a diverse dataset of CVs (resumes) with labelled personality traits. Ensure a representative sample across various industries and roles.

2. **Data Pre-processing:**
    - Clean and pre-process the CV data to remove irrelevant information, correct errors, and standardize the format.

3. **Feature Extraction:**
    - Extract relevant features from CVs, such as education, work experience, skills, and achievements. Use natural language processing (NLP) techniques to analyse textual content.

4. **Personality Trait Labelling:**
    - Assign personality trait labels to each CV. This can be done through manual labelling or by using external personality assessment tools.

5. **Model Selection:**
    - Choose a suitable machine learning model for personality prediction. Common choices include support vector machines, decision trees, or more advanced models like neural networks.

6. **Integration with CV Parsing:**
    - Integrate the personality prediction system with a CV parsing tool to automatically extract information from CVs.

7. **User Interface (UI):**
    - Develop a user-friendly interface for users to upload CVs and receive personality predictions. Consider the design principles to enhance user experience.

8. **Privacy and Security:**
    - Implement measures to ensure the privacy and security of the CV data. Anonymize sensitive information and adhere to data protection regulations.

9. **Error Handling:**
    - Implement robust error handling mechanisms to handle variations in CV formats, missing information, or other anomalies.

10. **Optimization:**
    - Optimize the model for efficiency, considering factors such as inference speed and resource utilization.

11. **Documentation:**
    - Provide comprehensive documentation for the project, including details on the model architecture, data pre-processing steps, and usage instructions.

12. **Testing and Validation:**
    - Conduct thorough testing and validation of the entire system to ensure its reliability and accuracy in predicting personality traits.

13. **Scalability:**
    - Design the system to be scalable, allowing it to handle a large number of CVs efficiently.

14. **Continuous Improvement:**
    - Implement a feedback loop and plan for continuous improvement. Regularly update the model based on user feedback and evolving requirements.

15. **Ethical Considerations:**
    - Consider ethical implications, such as potential biases in the model and the responsible use of personality predictions. Clearly communicate limitations and potential biases to users.

3.1 Functional Requirement:

1. **User Registration and Authentication:**
    - Users should be able to register and create accounts. Authentication mechanisms should be in place to ensure secure access to the system.

2. **CV Upload and Fill Form:**
    - Users should be able to upload their CVs in various formats (e.g., PDF, Word). The system must handle different CV structures and extract relevant information and should also fill the form for personality detection.

3. **CV Parsing:**
    - Implement a CV parsing mechanism to extract key information such as education, work experience, skills, and achievements.

4. **Personality Prediction:**
    - The system should analyse the parsed information and predict personality traits based on a predefined model. Provide users with a detailed report on the predicted personality traits.

5. **User Profile Management:**
    - Allow users to manage their profiles, update information, and view previous personality predictions.

6. **Exportable Reports:**
    - Allow admin to download and export detailed personality reports for their records or to share with others.

7. **Search and Filter Functionality:**
    - Include search and filter options to help users quickly find specific information within their CVs or personality reports.

8. **Real-time Processing:**
    - Aim for real-time or near-real-time processing of CVs to provide users with prompt personality predictions.

9. **Scalability:**
    - Design the system to scale efficiently, accommodating a growing number of users and CV uploads.

10. **Error Handling and Validation:**
    - Implement robust error handling mechanisms to address issues such as missing information in CVs or incorrect file formats.

3.2 Non-Functional Requirement:

1. **Performance:**
    - Response Time: The system should provide personality predictions within a reasonable response time (e.g., seconds) to ensure a positive user experience.
    - Scalability: The system should scale to accommodate an increasing number of users and CV uploads without a significant degradation in performance.

2. **Reliability:**
    - Availability: The system should be available for use during typical working hours with minimal downtime for maintenance.
    - Fault Tolerance: The system should be resilient to failures, and mechanisms should be in place to recover from errors gracefully.

3. **Security:**
    - Data Encryption: Personal and sensitive information within CVs should be encrypted during transmission and storage to protect user privacy.
    - Access Control: Implement access controls to ensure that only authorized users can view and interact with the system.

4. **Usability:**
    - User Interface (UI) Design: The system should have an intuitive and user-friendly interface, making it easy for users to upload CVs, and fill form.

5. **Scalability:**
    - Load Handling: The system should be capable of handling varying loads, especially during peak times when there might be a surge in CV uploads.

6. **Compatibility:**
    - Browser Compatibility: Ensure that the system works seamlessly across different web browsers to accommodate users with diverse preferences.
    - CV Format Compatibility: Support a variety of CV formats to cater to the diversity of documents submitted by users.

7. **Maintainability:**
    - Code Modularity: Design the system with modular components to facilitate easier maintenance and updates.
    - Documentation: Provide comprehensive documentation for system administrators, developers, and end-users to understand the system architecture, functionalities, and troubleshooting procedures.

8. **Performance Monitoring:**
    - Logging and Monitoring: Implement logging mechanisms to track system activities and performance metrics. Set up monitoring tools to detect and address issues proactively.

9. **Interoperability:**
    - Integration: Design the system with open standards to facilitate integration with other systems, such as HR management platforms or external APIs.

10. **Environmental Considerations:**
    - Energy Efficiency: Consider the environmental impact of the system and implement practices that promote energy efficiency in server infrastructure.

Chapter 2 SDS

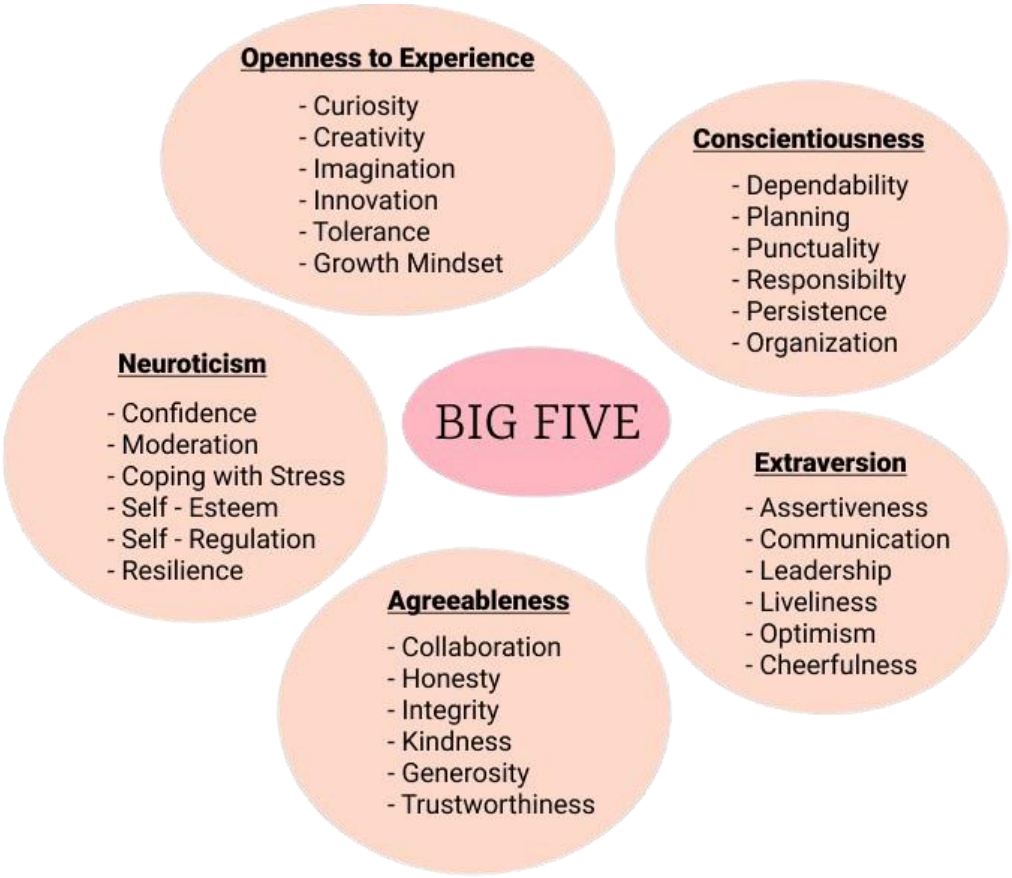
TABLE OF CONTENTS

1. Introduction

1.1 Purpose

The advent of technology in our daily lives has brought about significant changes in the recruitment process. With online platforms such as Glass door and Naukri, job-seeking has become more accessible, while businesses have adopted online management systems. Despite this, the recruitment process can be time-consuming, with personal investigations and interviews being a significant contributing factor. Recent studies suggest utilizing an e-recruitment system that analyses social media data to predict personality traits unique to each applicant. By employing this methodology, the results of the CV review can be used to screen applicants based on the organization's requirements, effectively streamlining the hiring process

The personality questionnaire is based on the Big Five Personality Model, which uses OCEAN values - Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, to predict an individual's personality. The proposed system leverages this methodology to optimize the hiring process and ensure a fair outcome.



1.2 Scope

The traditional process of hiring and selection may not always make a fair decision as an end result. With the help of the proposed system, the probability of making fair decisions increases by leaps and bounds. The proposed system makes use of Logistic Regression, Natural Language Toolkit, Flask, and Firebase for smooth and efficient deployment. The main methodology behind the personality questionnaire is the Big Five Personality Model which makes use of the OCEAN values namely, "Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism for predicting an individual's personality.

1.3 Definitions, Acronyms, and Abbreviations.

- **OCEAN** - Openness Conscientiousness Extraversion, Agreeableness, and Neuroticism
- **DFD** - Data Flow Diagram
- **ER** - Entity Relationship Diagram
- **UAT** - User Acceptance Testing
- **SDS** - Software Design Specification
- **SDD**=Software Design Document
- **PDF** – Portable Document Format

1.4 Overview

The Software Design Specification (SDS) for a CV-based personality prediction system is a detailed blueprint that outlines the architecture, components, functionalities, and interactions within the system. This document provides extensive information on the design principles, data flow, user interface elements, and underlying technologies. Moreover, it identifies the modules responsible for CV parsing, feature extraction, and personality prediction, among others.

This comprehensive SDS provides clarity on the system's architecture, functionalities, and implementation strategies, which can facilitate the systematic and coherent construction of the CV-based personality prediction system.

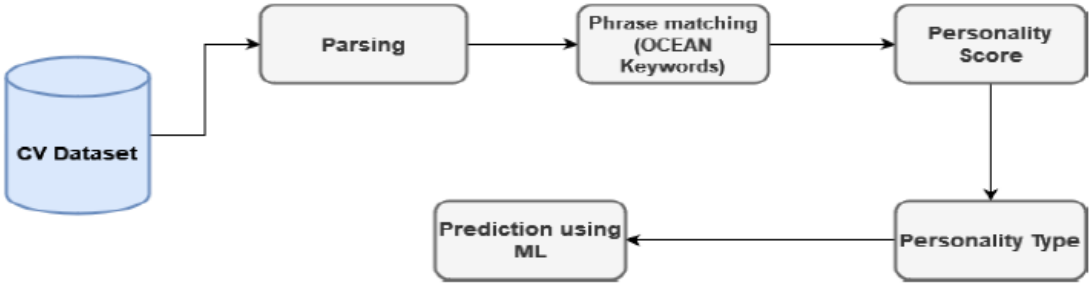
2. System Architectural Design

- We came across the traditional methods which included administrating personality and professional eligibility judgment, conducting interviews, and holding group discussions. Therefore, they are tedious and can lead to disparate candidate selections. However, the most meaningful aspect that represents an individual is personality, which changes over time, and dealing with them is a tedious process. With the help of machine learning algorithms, we use to create the models that will be tested in the proposed automated candidate grading system.

- Over the past few years, the number of manual interviews and resumes in human resources has expanded therefore we have planned a mechanism for identifying personalities and making recommendations

2.1 High-level Design Overview

Introduce the various components and systems at a high conceptual level. *A Pictorial representation of the system architecture is presented.*



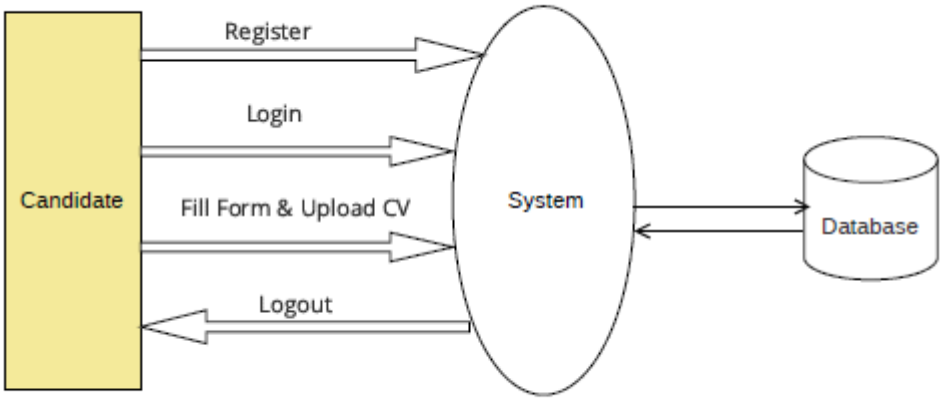
2.1 Workflow of Proposed System

2.2 Detailed Description of Components

➢ Data flow diagram (DFD):

A Data Flow Diagram (DFD) is a graphical representation that illustrates the flow of data within a system or process. It is a widely used tool in system analysis and design, providing a visual way to depict how data moves through different stages or components of a system.

❖ **Applicant**:



2.2 Applicant Data Flow Diagram

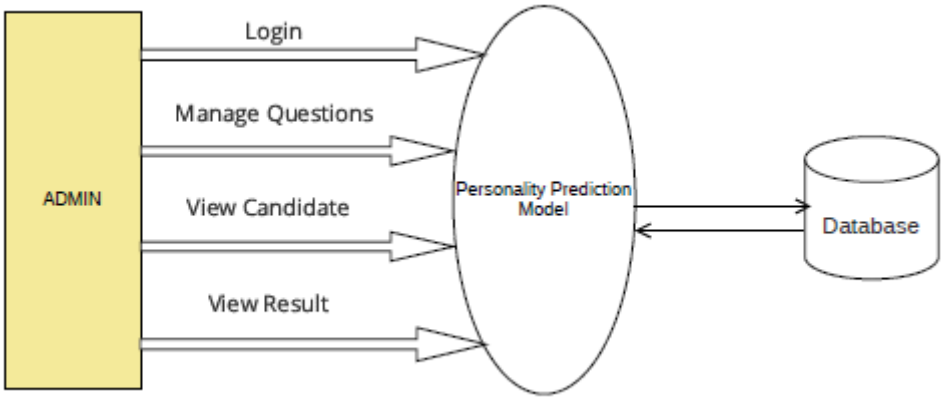
- **Registration:** To access the following sections, the candidate must fill in the registration form and generate his/her login credentials. Applicant must submit their CV during the registration process by filling out the CV form.

- **Login:** The candidate can access the sub-sections by entering the requisite credentials.

- **Test:** After successful login, an online test can be carried out on the basis of aptitude and personality. The applicant will be able to see the work details and choose the appropriate job if they satisfy the parameters set by the candidate Admin.

- **View Result:** After completing the exam, the applicant can view the results.
- **Logout:** After viewing the results, the candidate can logout from the portal.

❖ **Admin:**



2.3 Admin Data Flow Diagram

- **Login:** The admin must login to set the various system parameters and access the sub-sections of the Admin Page via his/her credentials.

- **Manage Questions:** The admin may add aptitude questions each with its own multiple-choice response on topics of his or her choice. In order to predict the candidate's personality, the admin may apply some personality-related question based on the OCEAN model to this sub-section.

- **Manage Jobs and Options:** The admin can manage the options according to the requirements of the job position and job listings available.
- **View Candidates:** The admin can view all the candidates' information.
- **View Results:** The admin can view the shortlisted candidates' evaluation results.

3.1 Database Description

A description of all data structures including internal, global, and temporary data structures.

* Temporary data structure

❖ **Login:**

The concept of "logging in" is commonly used in the context of accessing secure systems, websites, or applications.

Field	Type	Description	Constraints
Username	Varchar (20)	Name representing the user	Not null
Email	Varchar (20)	Email id used for register and login	Primary key
Password	Varchar (20)	To keep user data safe	Not null

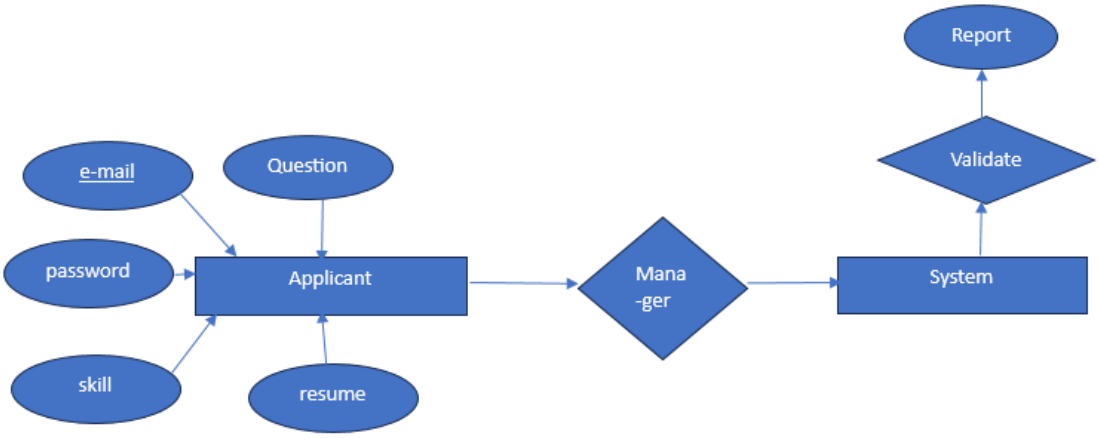
❖ **Applicant:**

An applicant, in the context of a curriculum vitae (CV) or a job application, is an individual who is seeking employment or a particular position within a company or organization. The applicant is someone interested in joining the workforce and is typically in the process of submitting application materials, such as a resume or CV, cover letter, and possibly other supporting documents, to be considered for a job opportunity.

Field	Type	Description	Constraints
Email	Varchar (20)	Email id used for register and login	Primary key
Password	Varchar (20)	To keep user data safe	Not null
Skill	Varchar (20)		Not null
Question	Varchar (20)	Email id used for register and login	Not null
Resume	LongBolb	To keep user data safe	Not null

3.2 E-R Diagram

An Entity-Relationship (ER) diagram is a visual representation of the data model that illustrates the relationships between entities within a database. It is a key tool in database design and serves to model the structure of a database system, showing how different entities relate to each other. ER diagrams are part of the entity-relationship modelling technique and are commonly used during the conceptual design phase of database development.

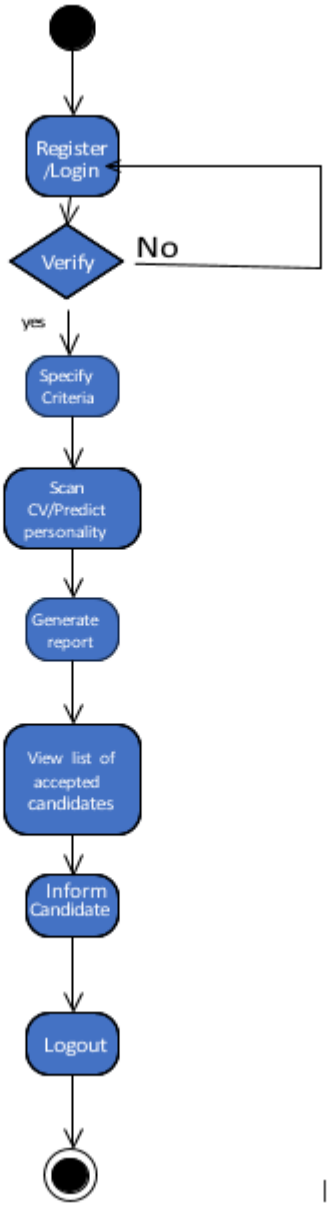


2.3 E-R Diagram of the system

3.3 Activity Diagram

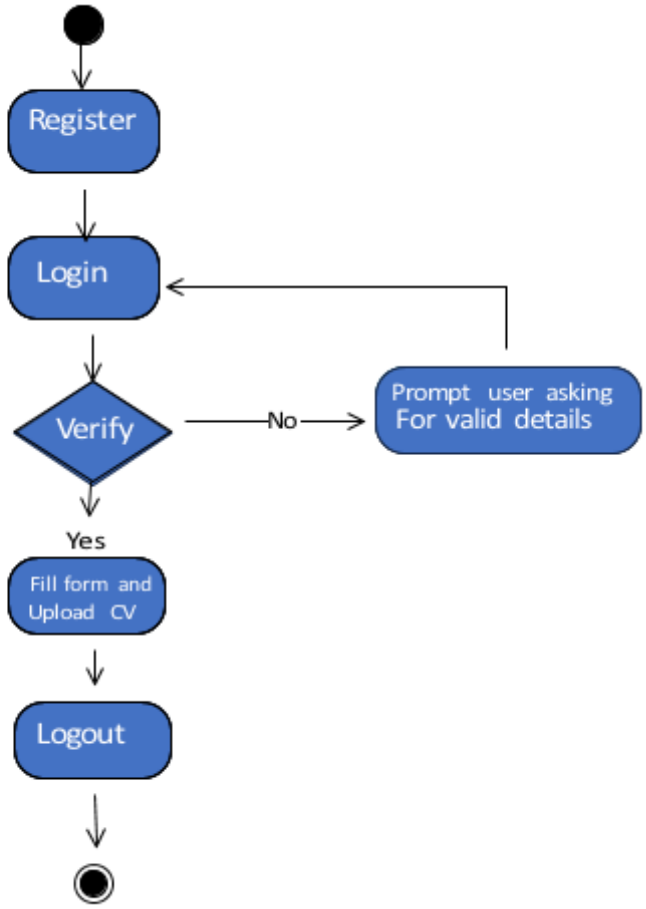
The concept of "logging in" is commonly used in the context of accessing secure systems, websites, or applications.

- **Admin:**



Admin Activity Diagram

❖ **Applicant:**



Applicant Activity Diagram

4. User Interface Design

4.1 Description of the user interface

The user interface (UI) of a CV-based personality prediction system plays a crucial role in facilitating user interaction and ensuring a positive user experience. Here's a description of the UI elements and functionalities typically found in such a system:

- **Homepage:**
  - The homepage is the primary user interface and offers an overview of the system's capabilities. It should also provide user guidance on how to begin using the system.

- **User Registration / Login:**
  - Users have the option to either register for a new account or log in using their existing credentials. This vital feature guarantees that user data and personality predictions remain secure and accessible only to authorized individuals.

- **CV Upload Interface:**
  - The system should offer users an intuitive interface for uploading their CVs that supports various file formats, such as PDFs. Users should be able to upload their CVs through drag-and-drop or by browsing their local storage.

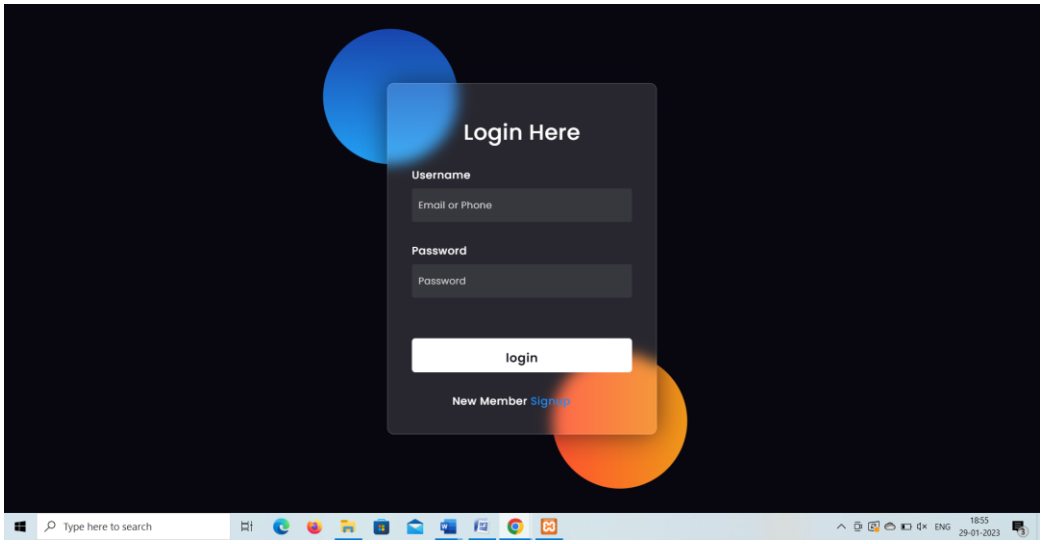
- **Personality Prediction Results:**
  - Upon completion of the personality prediction process, admins should receive detailed results displaying the predicted personality traits. These results could be presented using visually appealing formats, such as charts or graphs, to enhance comprehension.

- **Navigation Menu :**
  - The navigation menu should be well-organized, allowing users to access different sections of the system, such as CV uploads and logout.

- **Responsive Design:**
  - The user interface should be responsive and adaptive to different screen sizes and devices, including desktops, laptops, tablets, and smartphones. Responsive design helps ensure a consistent and optimal user experience across various platforms.

4.2 Screen images

Representation of the interface form from the user's point of view.



4.3 Objects and actions

In a CV-based personality prediction system, several objects and actions are involved to enable users to interact with the system effectively. Here's an overview of the key objects and corresponding actions:

Objects:

- **User Account:**
  - The system represents individual users, each with a unique account containing personal information, preferences.

- **CV (Curriculum Vitae):**
  - CVs, which contain information about a user's education, work experience, skills, achievements, and other relevant details, serve as input for the personality prediction process.

- **Personality Attribute :**
  - Personality traits, such as extroversion, agreeableness, conscientiousness, emotional stability, and openness to experience, are attributes used to describe an individual's personality.

- **Personality Prediction Model:**
  - A computational model analyses the content of CVs to predict the personality traits of individuals based on the information provided.

- **User Profile:**
  - The user profile contains information, such as personal details, contact information, preferences.

Actions:

- **CV Upload:**
  - Users initiate the personality prediction process by uploading their CVs to the system, which extracts relevant details from the content for analysis.

- **Personality Prediction:**
  - The system applies a personality prediction model to analyse the information from the CVs, taking into account factors such as language, employment history, education, and skills.

- **Result Interpretation:**
  - Upon completion of the personality prediction process, the system presents an easily understandable report to the admin. The report indicates the predicted personality traits of the user and how specific details from their CVs influenced the predictions. The admin can utilize these insights to gain a better understanding of the user's personality traits.
-

5. TYPES OF TESTS (With Implementation)

- **Unit Testing:** Test individual components or units of the system, such as parsing algorithms, feature extraction modules, or personality prediction models, to ensure they function correctly in isolation.

- **Integration Testing:** Verify that different modules or components of the system work together seamlessly. Test interactions between CV parsing, feature extraction, and personality prediction components.

- **Functional Testing:** Ensure that the system's functionalities meet the specified requirements. Test scenarios such as CV upload, personality prediction, result interpretation, and user feedback mechanisms.

- **Regression Testing:** After making changes or updates to the system, rerun tests to ensure that existing functionalities have not been negatively impacted. This helps identify unintended consequences of code modifications.

- **User Acceptance Testing (UAT):** Involve end-users in testing to validate whether the system meets their expectations and requirements. Users can assess the system's usability, accuracy of personality predictions, and overall satisfaction.

6. Conclusion

Implementing this method would aid the human resources department in selecting the most qualified candidate for a specific job vacancy, ultimately resulting in the recruitment of an expert employee for the company. The proposed scheme would streamline the selection process by rating the CVs based on various parameters such as test scores, expertise, credentials, and more. This would alleviate the workload of the human resources department and enhance their efficiency.

7. References

- "An integrated e-recruitment system forautomated personality mining and applicant ranking": Athanasios Tsakalidis, Giannis Tzimas, Evanthia Faliagka (2012).

7. Gayatri Vaidya, Pratima Yadav, Reena Yadav,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 International Conference on Innovative and Advanced Technologies in Engineering

- Mayuri Pundlik Kalghatgi, Manjula Ramannavar, Dr. Nandini S. Sidnal, "A Neural Network Approach to Personality Prediction based on the Big-Five Model", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN:

- C. Sumner, A Byers, R. Boochever, and G. J Park, ―Predicting Dark Triad Personality Traits from Twitter usage and a linguistic analysis of Tweets,‖in Proc. of 11th IEEE International Conference on Machine Learning and Applications, 2012, pp. 386-393.