Personality Prediction via CV analysis & Psychometric Test

Version	Date	Description	Author	Reviewer
1.1	08 November,2021	Revised SOW	Pooja Ayre / Shariq Shaikh	Prof. Marcos Bittencourt

Table of Contents

Problem Statement 3

Objective 3

Scope 3

Design Algorithm 3

Data Pre-processing 3

Feature engineering and Statistical Analysis 4

Data Requirements and Data Sources 5

Deliverables and Timeline 5

Assumptions 5

Model Testing 5

Pre-tests 5

Post-tests 6

Change Requests 6

Problem Statement

Traditional recruitment process consumes ton of HR and candidate's time in filling forms, identifying the right fit for the job based on manually entered details. It leaves a room for favouritism in the process and bias approach impacting organizational values and vision.

Objective

Personality is an imperative attribute which reflects individual and, that keeps changing situationally. With the implementation of this machine learning application, based on 2 approaches Psychometric test and CV analysis, it will allow HR department to minimize the manual effort required to search / select candidate by maximising the operational efficiency.

Scope

Phase	Description	Release Date
Phase – 1 (Current Scope for AI Algorithm project deliverable)	a. HR team will share the application link with candidates. b. Candidates will be requested to upload CV and enter some personal information, followed by a psychometric test. c. Upon completion of this process, Candidate will be prompted to download the result of the test and share it with HR personnel.	05 December,2021
Phase – 2 (Future Scope)	a. Integration with Career website of a company b. Shall allow candidate to create login and apply for jobs c. At the time of application of each job, Phase – 1 deliverable will be integrated. If results, already exist then application will be submitted with same result or else candidate will be prompted to follow the Phase – 1 process.	05 April,2021

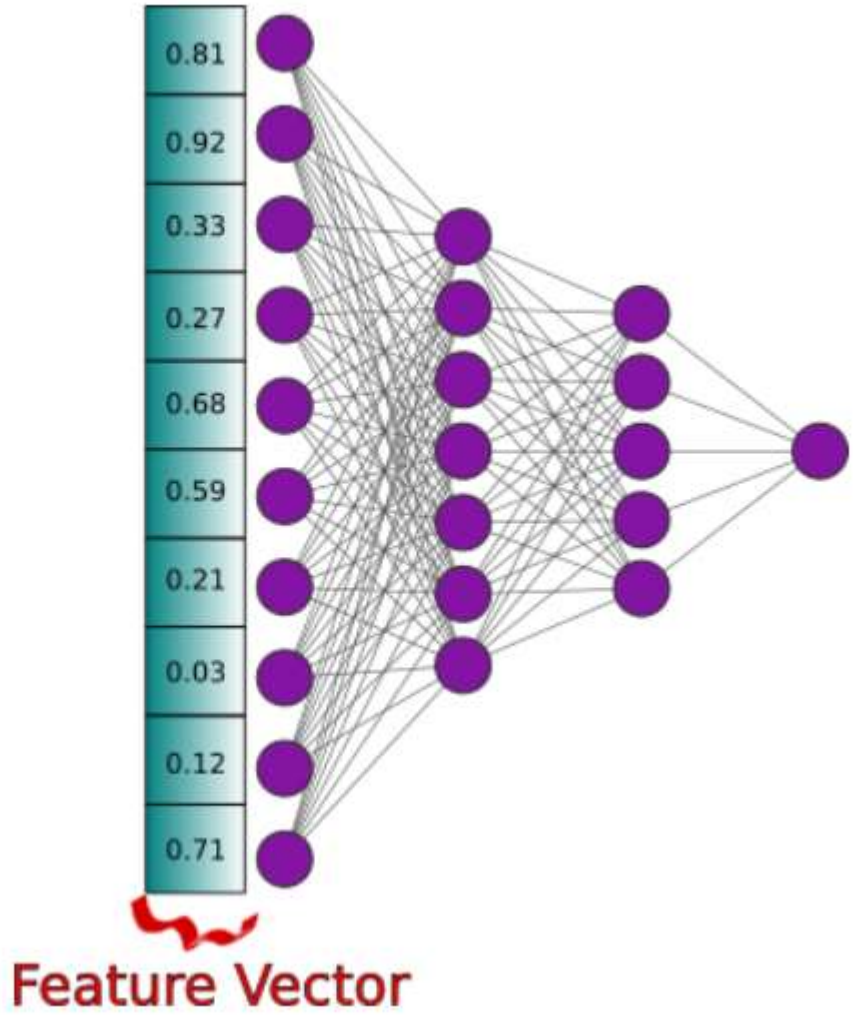
Design Algorithm

Data Pre-processing

This step allows us to clean the dataset for further processing. We are cleaning the dataset to remove special chars, hyperlinks, punctuation, URLs, or hashtags if any. Also, we are identifying STOP words to enhance the model accuracy. **Please refer to attached jupyter notebook for details**

Feature engineering and Statistical Analysis

This allows to extract features from the data post modifying existing set. For this, we will use technique called **text/word vectorization**. Vectors aids in visualizing how we can transform text into features. **Please refer to attached jupyter notebook for details.**



Data Requirements and Data Sources

- **CV** – Uploaded by Candidate
- **Psychometric question set** – Generated using online resources

Deliverables and Timeline

Milestone	Deliverable	Status
Dataset preparation and pre-processing	Data Collection	Completed
	Data Visualisation	Completed
	Data Selection	Completed
	Data cleaning, formatting	Completed
Dataset Splitting	Training set	Not Started
	Test Set	Not Started
	Validation Set	Not Started
Data Modelling	Model Training	Not Started
	Evaluation and testing	Not Started
Model Deployment		Not Started

Assumptions

- Each CV might not have a standard format hence the model will work based on keywords.
- CV should be in word/pdf format as the model will not be trained for OCR's

Model Testing

Pre-tests

- Check the shape of model output and if it aligns with the labels in dataset
- Check output ranges and if it aligns with expected output.
- Check label leakage between training and validation data sets

Post-tests

- Invariance Test
- Minimum Functionality Test

Change Requests

- Any change in the scope shall be included after an approval from Prof. Marcos Bittencourt.
- Change in scope that impacts the Phase – 1 deliverable shall be communicated to Prof. Marcos Bittencourt 7 days in advance of the deliverable end date.

Approver: Prof. Marcos Bittencourt

Approved Date: