JAYPEE UNIVERSITY OF ENGINEERING & TECHNOLOGY,
A-B ROAD, RAGHOGARH, DT. GUNA - 473226, M.P., INDIA

Aug 2022- Dec 2022

INTERACTIVE INTERVIEW BOT(IIB)

Submitted by:

ANURAG PATTER (201B056)

VIPUL KANJOLIA (201B309)

VIVEK (201B310)

Name of Supervisor- DR. MAHESH KUMAR

Project No.- 2

Submitted in partial fulfilment of the Degree of Bachelor of Technology

Department of Computer Science & Engineering

DECLARATION

We hereby declare that the work reported in 5th semester Minor project entitled "INTERACTIVE INTERVIEW BOT", in partial fulfilment for the award of the degree of B.Tech (CSE) submitted at Jaypee University of Engineering and Technology, Guna, as per the best of our knowledge and belief there is no infringement of intellectual property rights and copyright. In case of any violation, we will solely be responsible.

Anurag Patter (201b056)
Vipul Kanjolia (201b309)
Vivek (201b310)

Jaypee University of Engineering and Technology,

Raghogarh, Guna – 473226

Date: 9/12/2022

CERTIFICATE

This is to certify that the project titled "**Interactive Interview Bot**" is the bonafide work carried out by **Anurag Patter** ,**Vipul Kanjolia** and **Vivek**, a student of B Tech (CSE) of Jaypee University of Engineering and Technology, Guna (M.P) during the academic year 2020-21, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology (Computer Science and Engineering) and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar tile.

Signature of the Guide

Jaypee University of Engineering and Technology,

Raghogarh, Guna – 473226

Date: 9/12/2022

ABSTRACT

Interview process is generally associated with screening candidate suitable for available post. This process varies from company to company. However in general there two major methods remain same which are technical interview and human resource interview. Traditional process of interview is complicated and resource intensive. We can automate this process by replacing human interviewer with chatbot .Chatbot will ease the process to certain extent. Our idea is to provide an interactive chatbot which will conduct interview and generate report and base on this result candidate will be shortlisted. System is based on Natural Language Processing (NLP), Google- Text-to-speech.

ACKNOWLEDGEMENT

We would like to express our gratitude and appreciation to all those who gave us the opportunity to complete this project. Special thanks is due to our supervisor **Dr Mahesh Kumar** whose help, stimulating suggestions and encouragement helped us in all the time of development process and in writing this report. We also sincerely thanks for the time spent proofreading and correcting my many mistakes. We would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited period. Last but not the least I am grateful to all the team members of **Interactive Interview Bot**.

Thanking you

Anurag Patter (201b056)

Vipul Kanjolia (201b309)

Vivek (201b310)

LIST OF FIGURES

CONTENTS

Keywords And Glossary

Keywords:

Machine Learning (ML), Natural Language processing (NLP), Sentiment analysis, Google-text-to-speech (gtts), pandas, CSV (comma-separated values), counter, Ranking, Google Colab

Glossary:

ATS : Applicant Tracking System

DES : Data Envelopment Analysis

ML : Machine Learning

NLP : Natural Language Precessing

gtts : Google-text-to-speech

CSV : comma separated values

MySQL : SQL language

Parser : A parser is a program, usually part of a compiler.

CHAPTER - 1

INTRODUCTION

1.1 Problem Definition

1.1.1 Need of Interview Bot:

Most of the companies rely on their employee for communication part but this may result into reduced output of company. Human interviewer may be biassed or tired of interview process so it will affect the entire selection process. Using chatbot we can overcome these problems.
Chatbots are being introduced to ease the difficulties that the industries are facing today. The purpose of chatbots is to support and scale business teams in their relations with customers. It could live in any major chat applications like Facebook Messenger, Slack, Telegram, Text Messages, etc.

1.1.2 Problems and Solutions:

A bot also referred to as talkbot, is a computer program which carries out human like conversations with users via text or audio formats. It understands what the user is trying to do and send appropriate responses which will help the user's needs. Scope of Chatbot can be limited to business function or it can be as generic as possible.
Natural language processing (NLP) is a branch of AI .It can be defined as the ability of a machine to analyse, understand, and generate human speech. The main purpose of NLP and for which it is high on demand is, it give your bot a personality, it bridges the gap between human communication and computer understanding. Its major objective is to let your chatbot understand conversation in a better way.

1.2 Motivation

The current recruitment process are more tedious and time consuming which forces the candidates to fill all their skill and information manually. And HR team requires more man power to scrutinise the resumes of the candidates. So that motivated to build a solution that is more flexible and automated.

1.3 Objective and scope

1.3.1 Objective

The major objective of our system is to take the current resume ranking system to other level and makes it more flexible for both the entity.

1) Candidates, who has been hired.

2) Client company, who is hiring the candidates.

Candidates, who has been hired :

Candidates who are searching for jobs after been graduated. Out of those, major number of candidates are so much desperate that they are ready to work on any post irrelevant to their skill set and ability. The main reason behind this unemployment is like a cancer to our society, if a guy 1 is not got place after been passed out for 1yr, society include relatives starting blaming that guy. In Spite of this reason the candidate are ready to work in any condition, on any post. So they don't have to face those situation.

Where our system help such candidates to get hired by such a company or an organisation who really worth their ability and their skill sets. Where our algorithm will work in such a way that with the help of the previous result and previous ranking constraints, it will try to optimise the current result, which we called it Machine Learning.

This will make sure that the relevant candidate is been hired for that particular vacancy. You can say best possible candidate.

Client company, who is hiring the candidates :

Like I am the owner of a particular organisation, obviously my aim would be to create such a team which is the best team in the world. It is like, if there is a vacancy of a java developer in my organisation. So, I won't prefer to hire a python developer and then make him learn Java. That will be pretty useless and time consuming for both that candidate and for the organisation too.

Where our system help the organisation to make out the best possible candidates list according to their given constraints and requirement for that particular vacancy.

This kind of approach, will help our hiring sector to improve like anything and make it more efficient as the relevant person is getting a relevant job. So there would be no regrets for both the entities, client company and that hired candidate. Hence satisfaction will be achieved.

1.3.2 Scope:

As we know Indian I.T sector is second largest candidate recruiting sector of our country. It contribute about 7.5% to our Gross Domestic Product(G.D.P) Our Proposed system is initially concerned with the I.T sector of our country. It is mainly going to deal the Indian I.T industry but if you talk about the pro version of our system it can be extended to various other commercial sector where, intake and elimination are in bulk like for Governmental Jobs.

1.4 Project Overview

The project “Interactive Interview Bot” mainly focus on -

- Saving time and energy by eliminating the need to go through multiple rounds of interviews.
- Increase efficiency by conducting interviews with multiple candidates at the same time.
- Reduce the amount of time spent on each interview.
- Eliminate the possibility of human error.
- Make it easier to hire people who may have been overlooked due to lack of information or experience.

CHAPTER - 2

LITERATURE SURVEY

2.1 Case Study on talent acquisition

2.1.1 First Generation Hiring System

In this System the Hiring team would publish their vacancies and invite applicants. Methods of publishing were newspaper, television and mouth. The interested candidates would then apply by sending there resumes. These resumes were then received and sorted by the hiring team and shortlisted candidates were called for further rounds of interviews.
The whole process would take lot of time and human efforts to find right candidate suitable for their job roles.

2.1.2 Second Generation Hiring Systems

As the industries have grown, there hiring needs has rapidly grown. To serve this hiring needs certain consultancy units have came into existence. They offered a solution in which the candidate has to upload their information in a particular format and submit it to the agency. Then these agencies would search the candidates based on certain keywords. These agencies were middle level organisations between the candidate and company. These systems were not flexible as the candidate has to upload there resume in a particular formats, and these formats changed from system to system.

2.1.3 Third Generation Hiring Systems

This is our proposed system, which allow the candidates to upload their resumes in flexible format. These resumes are then analysed by our system, indexed and stored in a specific format. This makes our search process easy. The analysing system works on the algorithm that uses Natural Language Processing, sub domain of Artificial Intelligence. It reads the resumes and understands the natural language/format created by the candidate and transforms it into a specific format. This acquired knowledge is stored in the knowledge base. The system acquires more information about candidate from his social profiles like Linkedin and Github and updates the knowledge base.
Ranking Attributes are :

1) Problem Solving skills
2) Ability to work in a team
3) Strong work ethic
4) Analytical/Quantitative skills
5) Communication Skills (Written)
6) Leadership
7) Technical Skills
8) Strategic Planning skills
9) Computer skills
10) Creativity

2.2 Intelligent searching

Put simply, Artificial Intelligence or "AI" is an add-on to system, complementing to provide the online recruitment solution . As the name suggests, AI enables a combination of an applicant-tracking system and an artificial intelligence resume parsing, searching and matching engine. The result is a supercharged tool giving incredibly accurate candidate matching to jobs, and 'talent pool' searching that makes other systems look like they're from the stone-age.

2.2.1 Identifying "best" applicants in recruiting using data envelopment analysis

Selecting the most promising candidates to fill an open position can be a difficult task when there are many applicants. Each applicant achieves certain performance levels in various categories and the resulting information can be overwhelming. We demonstrate how data envelopment analysis (DEA) can be used as a fair screening and sorting tool to support the candidate selection and decision-making process. Each applicant is viewed as an entity with multiple achievements. Without any a priori preference or information on the multiple achievements, DEA identifies the non-dominated solutions, which, in our case, represent the ''best'' candidates. A DEA-aided recruiting process was developed that (1) determines the performance levels of the ''best'' candidates relative to other applicants; (2) evaluates the degree of excellence of ''best'' candidates performance; (3) forms consistent tradeoff information on multiple recruiting criteria among search committee members, and, then, (4) clusters the applicants.

2.3 Weaknesses

1. Prior systems needed lot of human efforts and time.
2. Cost of hiring is high.
3. Potential candidate may loose the opportunity because of ambiguous keyword matching.
4. Resumes needed to be in specific format.

2.4 How to overcome

1. Use of NLP to read resumes allow candidates the freedom to choose any format that's available to them.
2. Machine learning is used to rank candidates in accordance to requirements Which reduces the efforts of sorting thousands of resumes.
3. Use of NLP can be used to get mean out of ambiguous data.
4. Five benefits of A.I.
    Goes Beyond Key Words
    Fast and Accurate
    Perfect For the New World of Social Recruiting
    Customises to your Needs
    Gets Smarter

CHAPTER - 3

Requirement Analysis

3.1 Requirement specification

3.1.1 Python

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasises code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.
Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

Python libraries used in Project :

PyPDF2 : It is a python library used for performing major tasks on PDF files such as extracting the document-specific information, merging the PDF files, splitting the pages of a PDF file, adding watermarks to a file, encrypting and decrypting the PDF files, etc.

re : A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

math : The Python math module provides functions that are useful in number theory as well as in representation theory, a related field. These functions allow you to calculate a range of important values, including the following: The factorials of a number. The greatest common divisor of two numbers.

string : Like many other popular programming languages, strings in Python are arrays of bytes representing unicode characters. However, Python does not have a character data type, a single character is simply a string with a length of 1. Square brackets can be used to access elements of the string.

Counter : Counter is a subclass of dict that's specially designed for counting hashable objects in Python. It's a dictionary that stores objects as keys and counts as values. To count with Counter , you typically provide a sequence or iterable of hashable objects as an argument to the class's constructor.

Cosine Similarity : Cosine similarity is a measure of similarity between two non-zero vectors. It is calculated as the angle between these vectors (which is also the same as their inner product).

3.1.2 Jupyter notebook

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualisations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.
Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

3.1.3 Pandas

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labelled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis/manipulation tool available in any language. It is already well on its way toward this goal.

3.1.4 CSV

A CSV file (Comma Separated Values file) is a type of plain text file that uses specific structuring to arrange tabular data. Because it’s a plain text file, it can contain only actual text data—in other words, printable ASCII or Unicode characters.
The structure of a CSV file is given away by its name.

CSV files are normally created by programs that handle large amounts of data. They are a convenient way to export data from spreadsheets and databases as well as import or use it in other programs. For example, you might export the results of a data mining program to a CSV file and then import that into a spreadsheet to analyse the data, generate graphs for a presentation, or prepare a report for publication.

3.1.5 VISUAL STUDIO

Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.

3.2 Hardware Requirements

Linux: GNOME or KDE desktop GNU C Library (glibc) 2.15 or later, 2 GB RAM minimum, 4 GB RAM recommended, 1280 x 800 minimum screen resolution.

Windows: Microsoft R Windows R 8/7/Vista (32 or 64-bit) 2 GB RAM minimum, 4 GB RAM recommended, 1280 x 800 minimum screen resolution, Intel R processor with support for Intel R VT-x, Intel R EM64T (Intel R 64) Execute Disable (XD) Bit functionality.

3.3 Supportive Operating Systems

The supported Operating Systems for client include:

- Windows xp onwards
- Linux any flavour

Windows and Linux are two of the operating systems that will support comparative website.
Since Linux is an open source operating system, This system which is will use in this project is developed on the Linux platform but is made compatible with windows too.The comparative website will be tested on both Linux and windows. The supported Operating Systems for server include: The supported Operating Systems For server include Linux. Linux is used as server operating system. For web server we are using apache 2.0

CHAPTER - 4
Project Design

4.1 Design Approach

Design is the first step in the development phase for any techniques and principles for the purpose of defining a device, a process or system in sufficient detail to permit its physical realisation. Once the software requirements have been analysed and specified the software design involves three technical activities design, coding, implementation and testing that are required to build and verify the software. The design activities are of main importance in this phase, because in this activity, decisions ultimately affecting the success of the software implementation and its ease of maintenance are made. These decisions have the final bearing upon reliability and maintainability of the system. Design is the only way to accurately translate the customer requirements into finished software or a system. Design is the place where quality is fostered in development. Software design is a process through which requirements are translated into a representation of software. Software design is conducted in two steps. Preliminary design is concerned with the transformation of requirements into data.

4.2 Software Architectural Designs

Our system follows the three tier architecture . First tier consist of GUI, Processing block and the Database.

GUI:

The GUI(Graphical User Interface) in our project deals with the interface for the user where the user will login and submit his resume in any formate(pdf,doc,docx,ect.) and social profiles links. The GUI provides a platform for the user to communicate with the database. It acts as a connector as well as communicator which connects the database and helps in transfer of data between the GUI and the database.

Processing block:

Processing block is the block where the actual processing of our project is

done. This block connects the gui to the database i.e. it acts as a connector as well as communicator which connects the database and helps in transfer of data between the gui and the database. Its main function is to take input from resumes and social profile of the candidate and parse it to store the information and store it in the structured format(json), and database. After storing this information this system will give output using web application.

Database:

Database tier is the tier used for the storage of data. This tier contains all the data that is need for the processing of the whole project. The data in this tier is related to the student information gathered form his/her resumes and social profiles.

Software Architecture Design

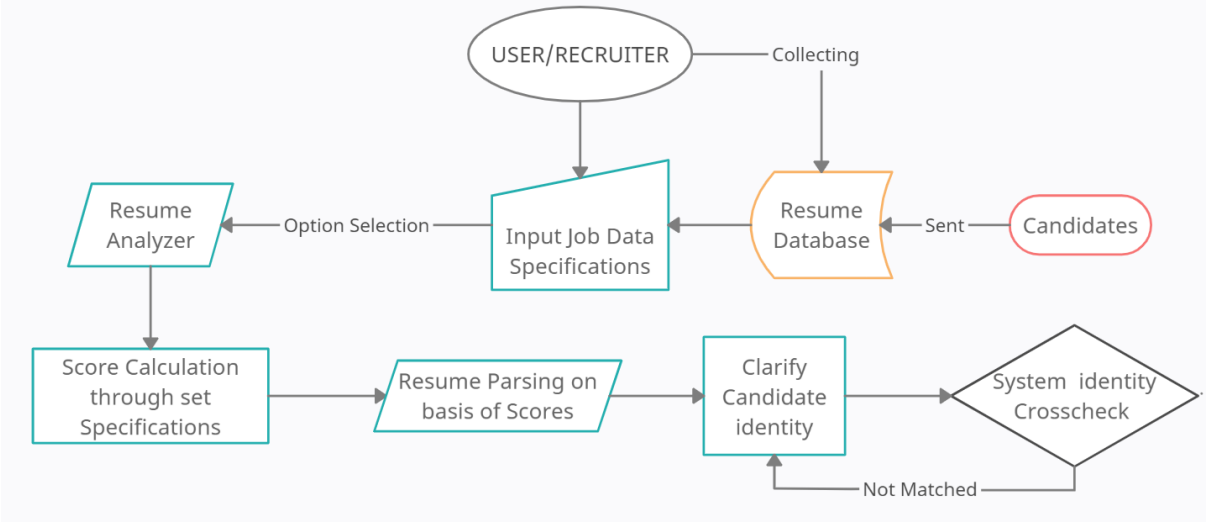

Fig 4.2.1 Resume Parser Model

Explanation:

Resume parser does include different jobs so it can score the resume of all the candidates on the basis of specific job description. The process uses a computer application which collect the information of the candidates with their resume (resume can be in doc, word and pdf format). After entering the resume and job data, resume analyser calculates the score of each resume . In the end it selects the potential candidates with high score according to the requirements of the company.

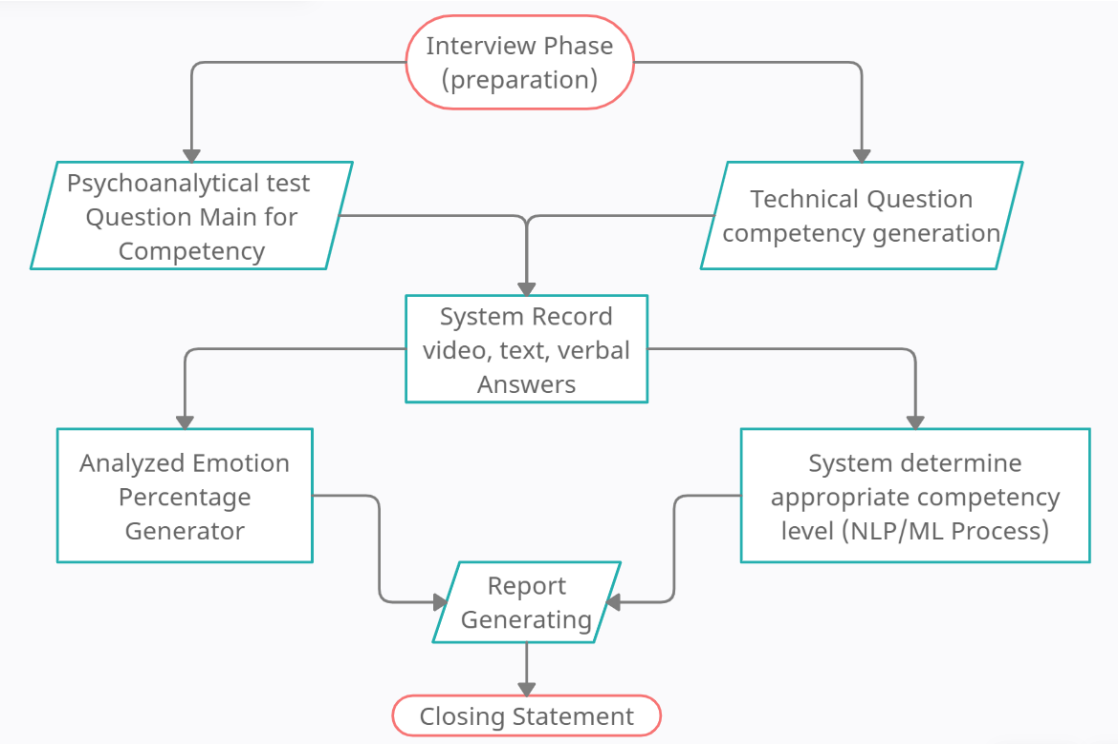

Fig 4.2.2 Interview Process

An interview bot helps to limit how much the interviewer can evaluate you based on your clothes and non-verbal behaviour instead of your work. At the same time, the lack of face-to-face interaction means that you can have your answers prepared and written to best highlight your accomplishments.

Front End Design

Fig 4.2.3 Front End Design : Home page

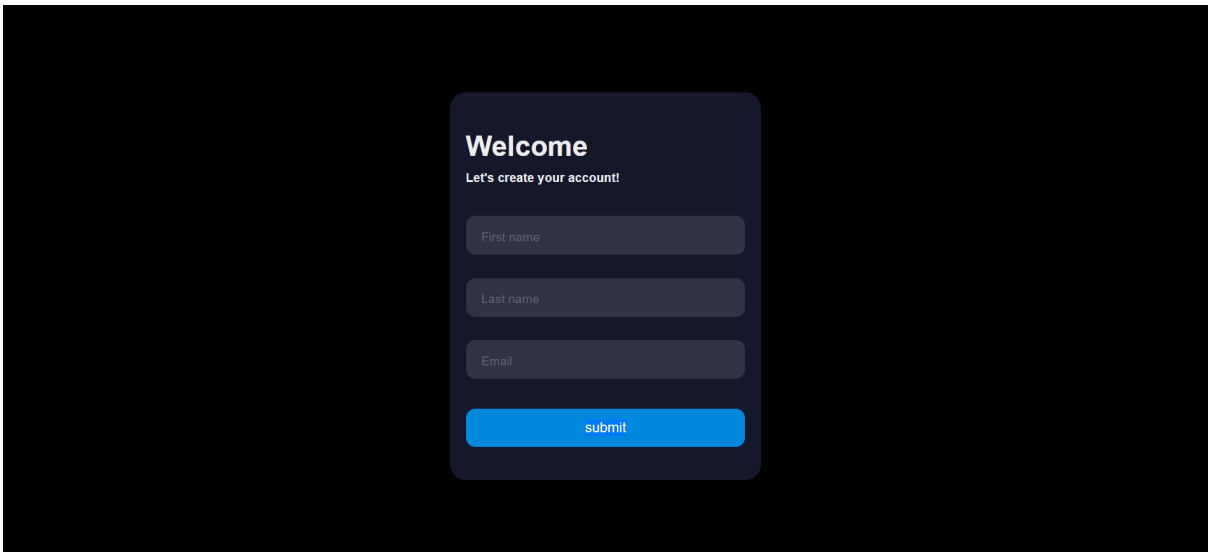

Fig 4.2.4 Front End Design : Registration page

CHAPTER - 5
Implementation Details

5.1 Assumptions and Dependencies

5.1.1 Assumptions

The following Assumption was taken into consideration:

• The Parser parses the resumes and convert them into the text file and then that text file is read to get the attributes of the candidate and store them in structured form in the json file.
• This json file contains the attributes of the student in ranked format and then it is read to show the out put to the student/employer.

5.1.2 Dependencies

The dependencies are as follows:
• For User interface Django web framework is used.
• Python programing language is used to parse the document and store the information in structured form.

5.2 Implementation Methodologies

5.2.1 Modular Description of Project

Resume parser does include different jobs so it can score the resume of all the candidates on the basis of specific job description. The process uses a computer application which collect the information of the candidates with their resume (resume can be in doc, word and pdf format). After entering the resume and job data, resume analyser calculates the score of each resume . In the end it selects the potential candidates with high score according to the requirements of the company.

5.3 Detailed Analysis and Description of Project

PyPDF2 : It is a python library used for performing major tasks on PDF files such as extracting the document-specific information, merging the PDF files, splitting the

pages of a PDF file. In this project PyPDF2 is used to extract text from the resume so it can be further analysed.

re : A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression. re is used to remove unwanted characters like numbers, commas etc from the text which is extracted from the resume.

counter : After extracting the text from resume and specific job description, counter is used to convert the text in form of vector.

	job	data
2	software developer	B.Tech or M.Tech in EE/ECE/CSE from a reputed engineering college Programming knowledge of C/C++ for HLS is required. SystemC is a plus Proficiency in C++ is mandatory Knowledge of one of the scripting languages like Perl, Tcl. Experience with LINUX platforms. Basics of Verilog/VHDL and experience in Logic Synthesis/High-Level Synthesis is plus Sound knowledge of windows technologies Good debugging and investigation skills
3	software engineer	A background in software engineering, software design or architecture and an understanding of how your Along with at least four years of experience in software development streams Have implementation knowledge of PySpark and Data warehousing, and good SQL query skills Experience in implementing programming best practice, especially around scalability, automation, virtuali An understanding of cloud, ideally AWS Experience in working with code repositories, bug tracking tools and wikis Coding experience in multiple programming languages Experience with DevOps and Agile methodology and associated toolsets and methodologies A background in solving highly complex, analytical and numerical problems
4	data scientist	B.Tech/B.E./MSc./MBA in engineering, maths, or statistics from a top college Minimum 1 year of experience with data science, machine learning, and programming Expertise in Python, object-oriented programming, and machine learning libraries Knowledge of statistics, machine learning, AI, and NLP tools and techniques Knowledge of Git, CI, Spark, AWS, and data pipelines Knowledge of Excel, PowerPoint, and Shell is desirable Good written and verbal communication skills Strong problem-solving, and quantitative skills

Fig 5.3.1 Specific Job Description

Cosine Similarity : Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.

DF (Data Frame): Pandas DataFrame is a widely used data structure which works with a two-dimensional array with labelled axes (rows and columns). DataFrame is

defined as a standard way to store data that has two different indexes, i.e., row index and column index

Select the potential candidates based on their resume score .

```
df.sort_values('score',ascending=False).head(10)    #top 10 candidates
```

	name	job	resumeData	score
2	Robert	software developer	software developer ii robert smithphone ...	0.465833
3	Smith	software developer	robert smith embedded software developer phon...	0.446500
0	Vivek	software developer	vivek seeking a position to utilize my skills ...	0.316749
1	Rajesh	software developer	vivek phone no email kv gmailcom seeking a p...	0.244288

Fig 5.3.2 Selecting potential resume

CHAPTER - 6
Result and Discussion

6.1 Test cases and Results

When the candidate submits his/her resume it is ranked and stored in the database and is later retrieved when required. We have tested our web application by considering following test case:

6.1.1 Unit Testing

We are firstly parsing the resumes and transforming them to text file and then reading them and parsing and storing the info in json format. After this we are ranking these resumes and storing them in our database. Later this data is shown to the user in web user interface.

Fig 6.1.2 Home page

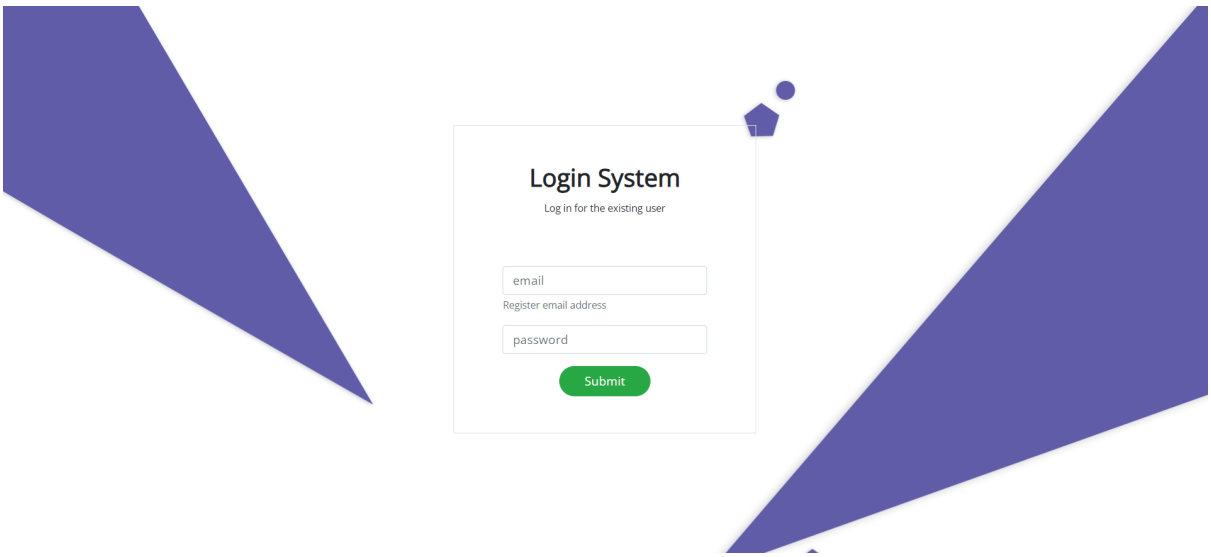

Fig 6.1.3 Login Page

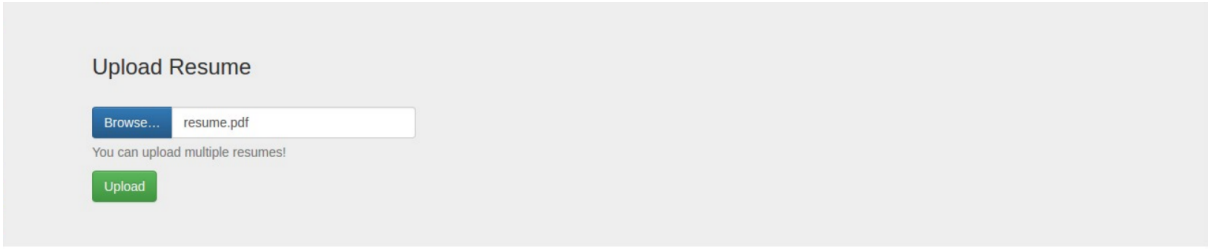

Fig 6.1.4 Upload resume

CHAPTER - 7
Conclusion and Future Scope

7.1 Conclusion

Our system will provide better and efficient solution to current hiring process. This will provide potential candidate to the organisation and the candidate will be successfully be placed in an organisation which appreciate his/her skillset and ability.

7.2 Future Scope

The application can be extended further to other domains like Telecom, Healthcare, Ecommerce and public sector jobs.

CHAPTER - 8
References

[1] **Kaggle** - Resume Screening using Machine Learning
https://www.kaggle.com/code/gauravduttakiit/resume-screening-using-machine-learning

[2] **tds -** Resume Screening with Python
https://towardsdatascience.com/resume-screening-with-python-1dea360be49b

[3] Match Your Resume to a job Description using python
https://randerson112358.medium.com/resume-scanner-2c30f5baf92c

[4] **w3schools -** Django Introduction
https://www.w3schools.com/django/django_intro.php

[5] **Analytic Vidya -** Complete guide to build your AI chatbot with NLP in python
https://www.analyticsvidhya.com/blog/2021/10/complete-guide-to-build-your-ai-chatbot-with-nlp-in-python/

[6] **Javatpoint -** Django File upload
https://www.javatpoint.com/django-file-upload#:~:text=Django%20provides%20built%2Din%20library,%2Fform%2Ddata%22%20property.

[7] **FreeCodeCamp -** Backend Web Development with python
https://www.freecodecamp.org/news/backend-web-development-with-python-full-course/

[8] **IRJET** - Interactive interview Chatbot
https://www.irjet.net/archives/V6/i4/IRJET-V6I4584.pdf