****

**GOVERNMENT POLYTECHNIC KOLHAPUR**

(First Autonomous Institute of Government of Maharashtra)

Department of

**INFORMATION TECHNOLOGY**

Project Based upon Skills Gained in Internship program

Project Name:

**Virtual HR :CV Extraction**

Project By

**Nawaz Bashir Sayyad 216062**

Under the Guidance of

Prof. Sachine Pukale Sir.

In the academic Year 2023-2024

|  |  |  |
| --- | --- | --- |
| S.no | Title | Page |
| 1. | PROBLEM STATEMENT | 1 |
| 2. | INTODUCTION | 2 |
| 3. | OBJECTIVE | 3 |
| 4. | FEATURES | 4 |
| 5. | TECHNOLOGIES USED | 5 |
| 6. | SYSTEM ARCHITECTURE | 7 |
| 7. | FLOW OF SYSTEM | 8 |
| 8. | DEMO SCREEN-SHOTS | 9 |
| 9. | CONCLUSION | 11 |
| 10. | CERTIFICATE | 12 |

INDEX

**PROBLEM STATEMENT**

In today's competitive job market, the importance of resumes cannot be overstated. Every day, companies of all sizes receive a substantial volume of resumes from potential candidates. Managing and sifting through this continuous influx of data within the confines of a single document can be an arduous task. The challenge intensifies when searching for specific individuals or their qualifications, buried amidst a sea of pages.

Imagine a solution where this overwhelming task is transformed into a well-organized process. What if you could effortlessly extract and systematize resume data into an Excel sheet? This would revolutionize the management and search for potential candidates. By harnessing the power of Excel's functions, you would gain the ability to swiftly and accurately locate individuals with the exact skills and qualifications you're seeking.

To meet this pressing need, we've introduced a cutting-edge module that converts not only CSV and Cove files but also various other resume formats into the universally accessible Excel format using JSON. This innovation simplifies the process of managing resumes, making it easier than ever to find specific candidate information.

But our solution doesn't stop there. In addition to this data transformation, we've also pioneered the development of a virtual human resource assistant, or HR assistant, capable of providing real-time insights and information derived from the Excel sheet. This intelligent assistant becomes your ally in resolving the challenges posed by resumes, ensuring that you can swiftly and efficiently locate the best candidates based on your unique requirements.

**INTRODUCTION**

In the era of rapid technological advancements and ever-increasing data volumes, the need for efficient data management and analysis has become paramount. The ability to extract meaningful insights from complex data sources is a critical aspect of decision-making in various fields, from business and finance to healthcare and research. In response to this demand, the development of data extraction tools has gained substantial importance.

This project report delves into one such innovative solution – the conversion of diverse resume formats, including CSV and Cove files, into a structured Excel sheet using JSON. The objective is to simplify the management and analysis of resume data, making it more accessible and streamlined for recruiters and HR professionals.

The proliferation of digital resumes and job applications has led to a deluge of information. This information often remains scattered and unorganized, hindering the swift identification of potential candidates with specific qualifications or skills. Our project addresses this challenge by introducing a systematic approach to data transformation. By converting resumes into a familiar Excel format, we aim to facilitate efficient data management and enhance the search for qualified candidates.

This report not only presents the technical details of the data conversion module but also introduces a virtual human resource assistant designed to work in tandem with the Excel sheet. This AI-powered assistant is geared towards offering real-time insights and information from the transformed data, further enhancing the capability to find and select the best candidates based on specific criteria.

In the subsequent sections, we will delve into the technical aspects of our project, detailing the processes, methods, and results achieved. The project's significance lies not only in its technical execution but also in its potential to revolutionize the way organizations manage and leverage resume data.

**OBJECTIVE**

The primary objective of this project is to provide an innovative solution to address the challenges faced by Human Resources (HR) departments in managing a large influx of resumes and to streamline the hiring process. This project aims to:

1. Develop a data extraction module capable of converting various resume formats, such as CSV and Cove files, into a structured Excel sheet using JSON.

2. Enable HR professionals to efficiently manage and analyze large volumes of resume data, replacing time-consuming manual tasks with automated processes.

3. Facilitate swift and accurate candidate searches based on specific qualifications and skills, reducing the time and effort required to identify potential candidates.

4. Introduce a virtual human resource assistant that can interact with the Excel sheet, providing real-time insights and information from the transformed data, thus enhancing the HR department's decision-making capabilities.

5. Offer an innovative approach to transform the HR workflow, making it more efficient and effective in handling a substantial number of resumes.

In summary, this project seeks to provide an idea-driven solution to revolutionize HR processes, specifically addressing the handling and management of resumes. By automating the extraction and analysis of resume data, our objective is to improve the HR department's efficiency, reduce manual workload, and enhance the overall effectiveness of candidate selection.

**FEATURES**

1. PDF and Text Resume Extraction: This system has the ability to extract data from PDF and text file resumes, transforming the content into a structured Excel sheet.

2. Interactive HTML User Interface: The project provides an intuitive HTML user interface, allowing users to visualize the extracted data. This interface facilitates updates and deletions of the imported information.

3. Data Modification: Users can easily update or delete data entries within the Excel sheet, ensuring accurate and up-to-date information.

4. Efficient Search Functionality: The core feature of this project is the Virtual HR Assistant, capable of answering specific queries related to the extracted data. Users can ask questions like Who is from Kolhapur? to obtain precise information quickly.

5. Saves Time: By automating the data extraction, organization, and retrieval processes, this project significantly reduces the time and effort required to manage and analyze resume data.

6. Versatile Data Input: The system accepts various data sources, including PDF files, text files, and CSV files, making it adaptable to a wide range of resume formats.

7. HR Virtual Assistant: The HR Virtual Assistant serves as the centerpiece of the project. It can answer questions, provide insights, and assist in data-driven decision-making within the HR domain, streamlining the hiring process.

These features collectively make the project a user-friendly, efficient, and versatile tool for HR professionals and organizations looking to optimize their resume management and candidate selection processes.

**TECHNOLGIES USED**

1. Python 3.11:

- Python serves as the project's backbone, offering a wide range of libraries and tools for data manipulation and backend functionality.

- Python's versatility allows for seamless integration with other technologies and enables rapid development.

2. Flask Framework:

- Flask, a micro web framework for Python, plays a pivotal role in developing the project's web interface.

- It provides an efficient and lightweight solution for routing, request handling, and serving web pages.

3. CSV and Pandas:

- Pandas: This powerful data manipulation library is instrumental for data correlation, extraction, and transformation.

- CSV: Working in conjunction with Pandas, CSV handling ensures that data from CSV files is efficiently managed and integrated into the Excel sheet.

4. JQuery:

- JQuery, a JavaScript library, enhances the user interface by enabling dynamic and asynchronous responses.

- This eliminates the need for full page refreshes, resulting in a smoother and more interactive user experience.

5. HTML and CSS:

- HTML (Hypertext Markup Language): HTML is used to structure the project's web pages, creating a well-organized and accessible layout for displaying data and user interaction.

- CSS (Cascading Style Sheets): CSS is employed to style the HTML elements, ensuring a visually appealing and user-friendly interface.

6. OpenAI API:

- The project integrates OpenAI's API for document extraction and natural language processing.

- This technology enables the HR Virtual Assistant to understand and respond to user queries, enhancing the overall user experience.

7. Virtual Environment:

- A virtual environment is used to encapsulate and manage project dependencies, ensuring that the various technologies and libraries work together cohesively.

- This environment guarantees the project's stability and compatibility, even in complex, multi-technology applications.

These combined technologies create a robust and versatile solution for HR professionals. They streamline the resume data extraction, management, and analysis process, offering a user-friendly interface and an intelligent HR Virtual Assistant that significantly improves the efficiency of the hiring workflow.

**SYSTEM ARCHITECTURE**

STAGE 1

MAKE A PROMT WHICH WILL POSTFIX TO THE QUERY . # prompt

prompt\_postfix = <document>

  \n###

  \nExtract the key sections from the resume above into json.

READ THE DATA FROM THE FILE . AND STORE IT INTO THE DOCUMENT AS STR

Document

Stage 2

Note: get\_structured\_data(document, prompt\_postfix, deployment\_id):

is user defined function in the program

PASS THE VALUES TO THE FUNCTION

def get\_structured\_data(document, prompt\_postfix, deployment\_id):

    content = prompt\_postfix.replace('<document>', document)

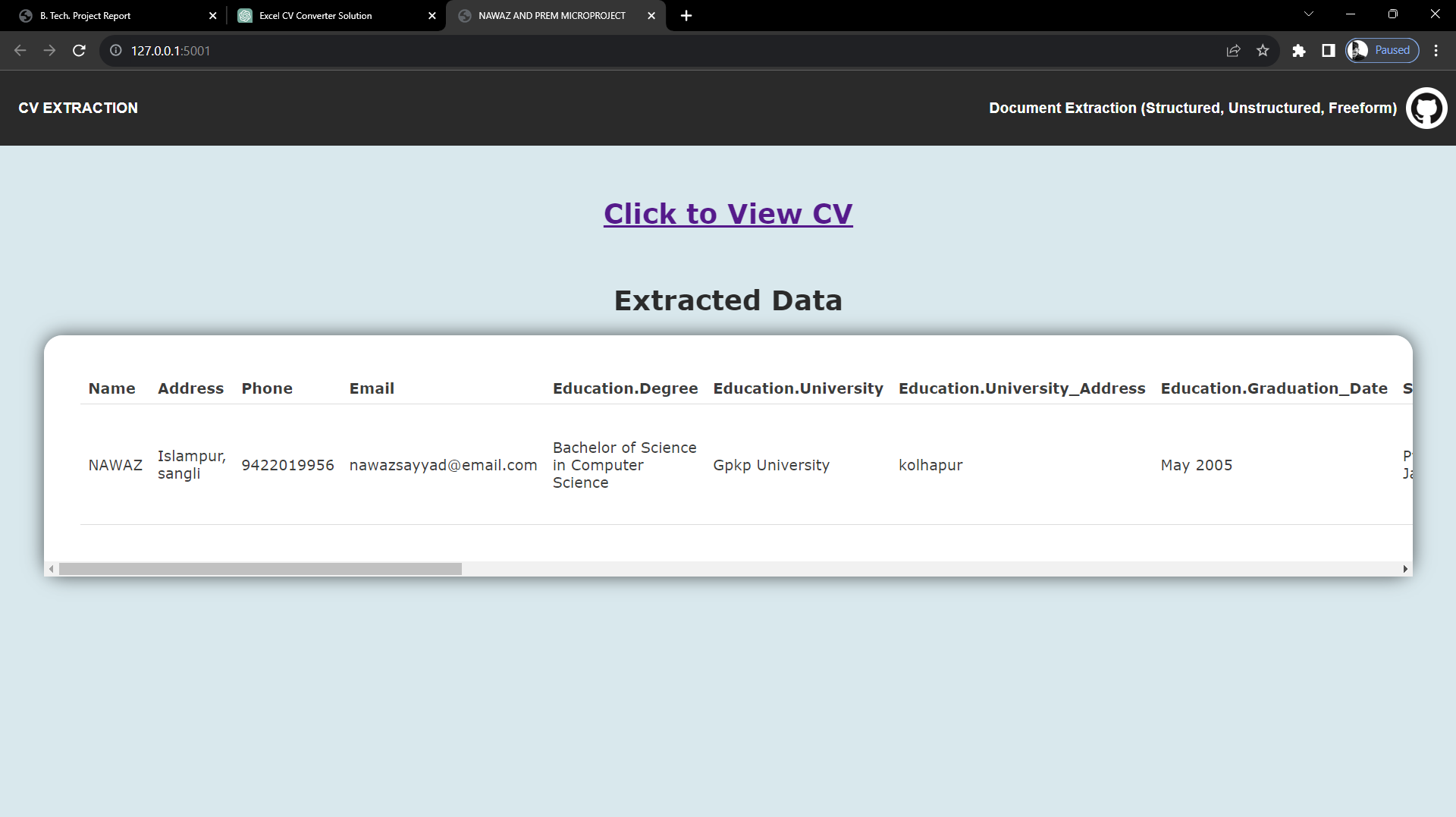
    messages = [{ role : user , content :content},]; #print(messages)

    structured\_data = request\_api(messages, deployment\_id)

    return structured\_data

CSV FILE THAT CONTAINES THE DATA IN THE STURCTURED AND TABULAR FORMAT

CONVERTING THE THE JSON FORMAT INTO THE Dictionary AND SAVING INTO THE OUTPUT\_DIC , AND DUMPING IT TO THE DATA.JSON  
AND PUTTING THAT JSON INTO THE DATZ.JSON FILE



WEBPAGE-------------------->

**VIRTUAL HR:**

**FLOW**

Ask QUESTION

USER

Update

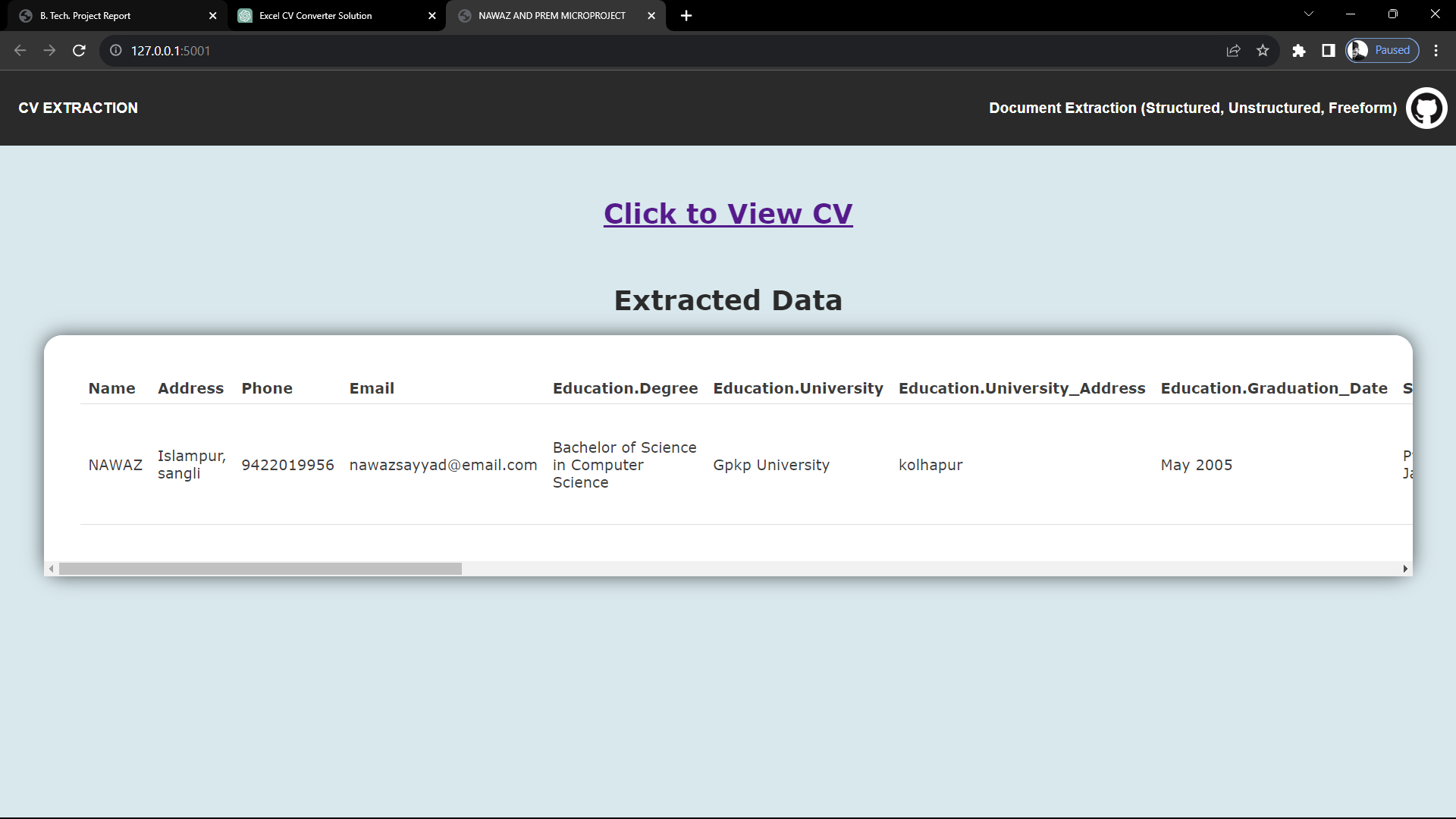
CSV

GET RESPONSE

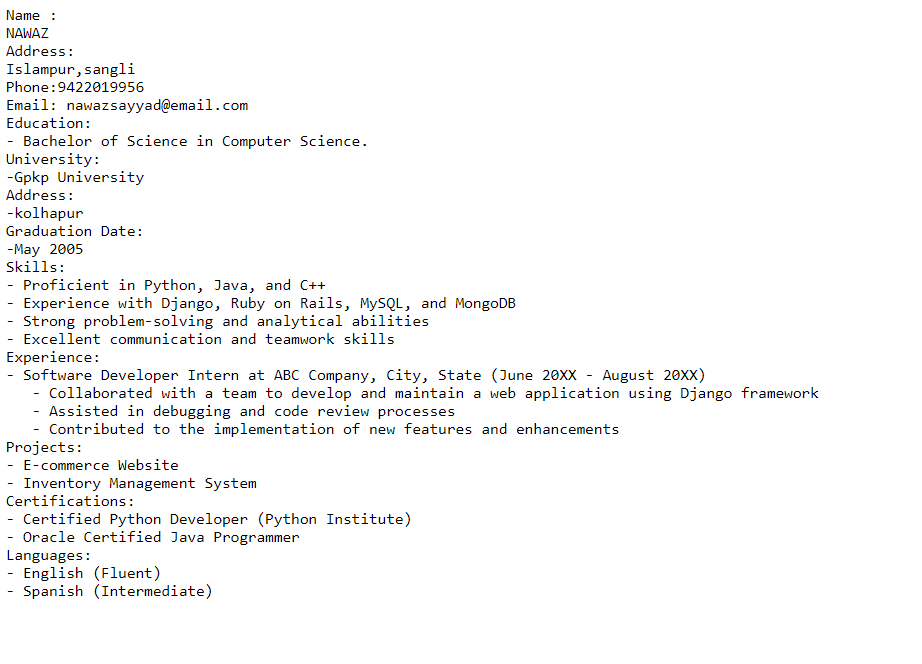
CV

DEMO SCREENSHOTS

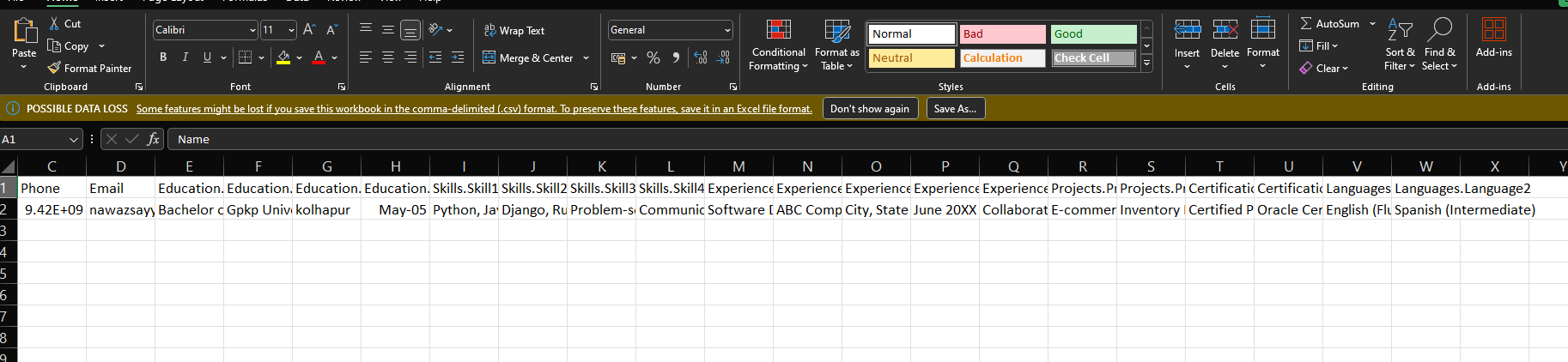
EXTRACTED CV



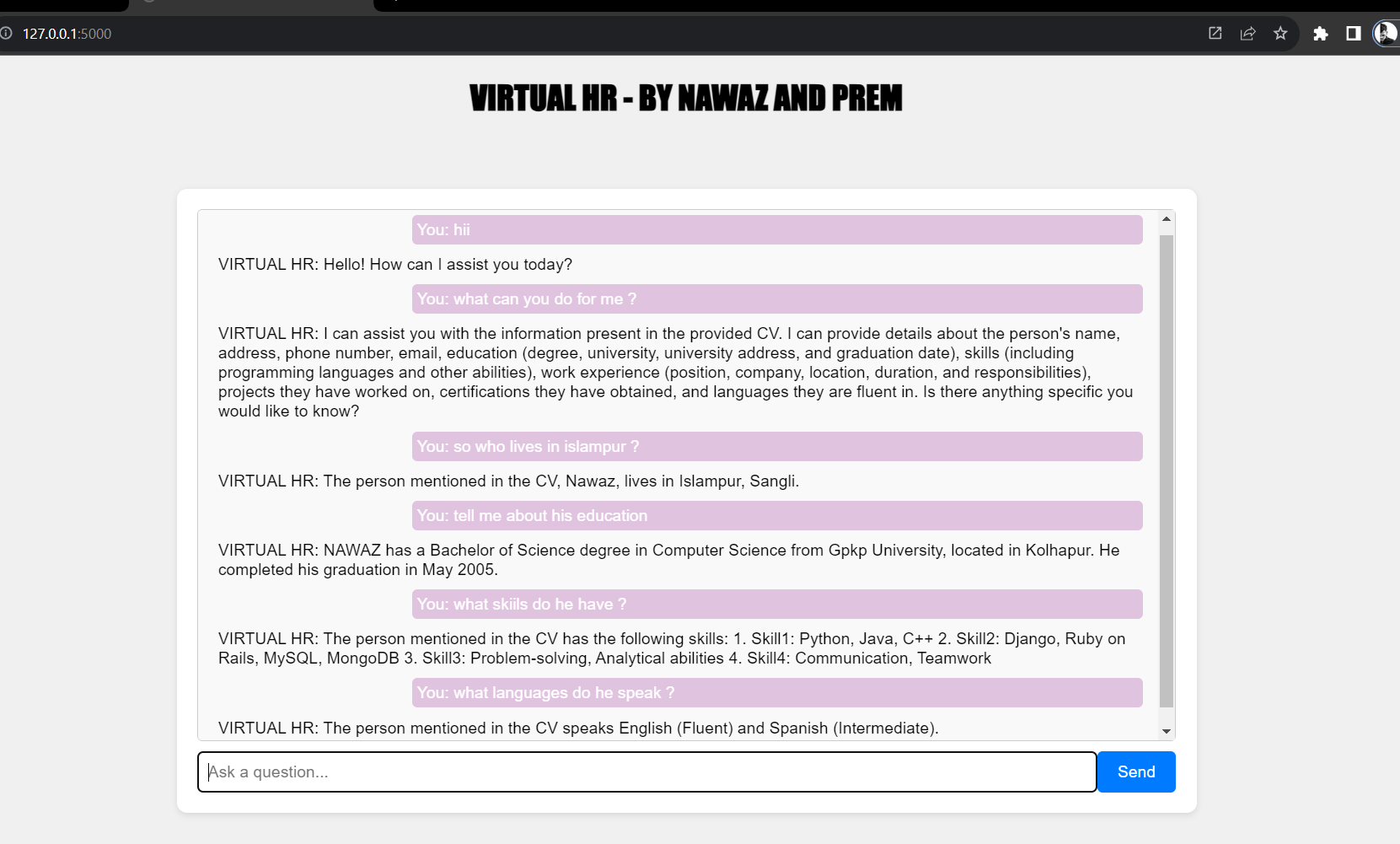
TEXT CV



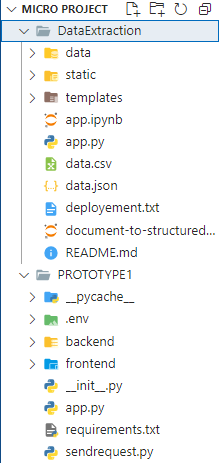
UPDATED CSV



VIRTUAL HR



DIRECTORY



**CONCLUSION**

In the age of digital transformation, where data is at the heart of informed decision-making, our project stands as a beacon of innovation in the realm of HR and resume management. The primary aim of this endeavor was to address the challenges faced by HR departments in dealing with an ever-growing influx of resumes and to provide a solution that streamlines the entire process.

Through the integration of cutting-edge technologies, such as Python, Flask, CSV, Pandas, JQuery, HTML, CSS, and OpenAI, we have successfully developed a comprehensive system that not only extracts and organizes resume data but also offers an interactive user interface and a virtual HR assistant.

The key outcomes of this project are as follows:

1. Efficient Resume Data Management: The ability to extract data from PDFs, text files, and CSVs, and transform it into a structured Excel sheet, has revolutionized the way HR departments manage resumes. This has significantly reduced manual efforts and errors, making the process more efficient.

2. User-Friendly Interface: With the integration of HTML and CSS, we have provided a visually appealing and user-friendly interface. Users can easily visualize and interact with the data, making updates and deletions as needed.

3. Enhanced User Experience: The project leverages JQuery to create an interactive and seamless user experience. Asynchronous responses reduce page refreshes, ensuring a smoother and more efficient user interface.

4. HR Virtual Assistant: The incorporation of OpenAI's API has given rise to a virtual HR assistant capable of understanding and responding to user queries. This assistant enhances decision-making and further simplifies the hiring process.

5. Optimized Workflow: By encapsulating these technologies in a virtual environment, we have ensured seamless compatibility and cohesiveness, offering a robust and efficient solution for HR professionals.

In conclusion, our project not only meets its objectives but also sets a new standard for resume and HR management. It simplifies the traditionally labor-intensive process, making it efficient, accurate, and user-friendly. The fusion of diverse technologies and the introduction of an HR Virtual Assistant herald a new era in HR operations, one that promises to save time, reduce errors, and ultimately optimize the candidate selection process. The future of HR management is here, and it's innovative, streamlined, and user-centric.

****

**GOVERNMENT POLYTECHNIC KOLHAPUR**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CERTIFICATE**

This is to certify that **Nawaz Bashir Sayyad [216062]** has successfully completed the internship program, demonstrating exemplary knowledge and skills in the application of new technologies. Under my guidance, Nawaz undertook and successfully accomplished the project titled "Virtual HR: CV Extraction," addressing a real-world problem.

This achievement highlights Nawaz's proficiency and dedication in the field, making significant contributions to the project's success. The successful completion of this internship program is a testament to Nawaz's commitment to continuous learning and application of IT skills in a real-world context.

Date: [10/11/2023]

**PROF. SHOBHA NADGERI PROF. SACHIN PUKALE SIR**

HEAD OF DEPARTMENT INTERNSHIP GUIDE