**Project Title**

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

**Suhas Martha, suhasmartha@gmail.com**

Under the Guidance of

**Saomya Chaudhury**

**ACKNOWLEDGEMENT**

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.   
  
Firstly, we would like to thank our supervisor, Saomya Chaudhary, for being a great mentor and providing valuable insights throughout this project. His guidance, encouragement, and constructive feedback were instrumental in shaping the development of this project. His confidence in our abilities served as a great motivation.  
  
We would also like to extend our gratitude to AICTE and TechSaksham for providing this internship opportunity, which enabled us to explore and implement AI-based solutions in real-world applications. This project would not have been possible without their support.

#### **ABSTRACT**

In the modern recruitment process, organizations receive an overwhelming number of resumes for each job opening. Manually screening these resumes to identify the most suitable candidates is time-consuming and inefficient. This project aims to automate resume screening and ranking using AI and NLP techniques.  
  
The proposed solution involves developing a web-based system that extracts relevant information from resumes, compares them with job descriptions, and ranks candidates based on their relevance. We use BERT-based Semantic Matching and Named Entity Recognition (NER) for structured data extraction.  
  
Additionally, interactive visualizations help recruiters analyze candidate profiles more effectively. The system enhances efficiency in hiring processes by reducing manual effort while ensuring accurate candidate selection.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 1

1.4. Scope of the Project 1

**Chapter 2.**  **Literature Survey 2**

2.1 Relevant Literature 2

2.2 Existing Models 2

2.3 Limitations 2

**Chapter 3.**  **Proposed Methodology 4**

3.1 System Design 4

3.2 Requirement Specification 4

**Chapter 4.**  **Implementation and Results 5**

4.1 Snapshots of the Results 5

4.2 GitHub Link 7

**Chapter 5. Discussion and Conclusion 8**

**References 9**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | Code for proposed project | 5 |
|  | Compiling the code | 5 |
|  | Deployed project interface | 6 |
|  | Uploading resume into platform | 6 |
|  | Results for single uploaded resume | 7 |
|  | Results for multiple uploaded resumes | 7 |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

Screening a large number of resumes for job openings is a complex and time-intensive task. Recruiters often spend hours filtering resumes manually, leading to inefficiencies and possible biases in candidate selection. Automating this process using AI and Natural Language Processing (NLP) can significantly improve recruitment efficiency and accuracy.

* 1. **Motivation:**

The motivation behind this project is to develop a system that can:

* Reduce the time taken for resume screening.
* Improve accuracy in candidate selection.
* Enable recruiters to make data-driven hiring decisions.
* Provide a scalable and automated hiring process for organizations.
  1. **Objective:**

The objectives of this project are:

* Automate the resume screening process using AI.
* Extract and structure information from resumes for ranking.
* Implement BERT-based Semantic Matching to compare resumes with job descriptions.
* Develop an intuitive web-based application using Streamlit.
* Provide interactive data visualizations for better insights.
  1. **Scope of the Project:**

The project focuses on automating resume screening using NLP and deep learning. It supports PDF and DOCX resume formats and extracts candidate details, skills, and experience using NER and semantic similarity techniques. The system is designed for deployment on a cloud-based platform and integrates interactive analytics for recruiters.

**CHAPTER 2**

**Literature Survey**

**2.1 Review of Relevant Literature**

The process of resume screening has been extensively studied in AI and HR technology domains. Traditional keyword-based filtering techniques have been widely used but suffer from lack of contextual understanding. Recent advancements in Natural Language Processing (NLP) have introduced machine learning models, such as TF-IDF, Word2Vec, and BERT, that improve the accuracy of resume parsing and ranking.

**2.2 Existing Models, Techniques, and Methodologies**

Several models have been proposed for automated resume screening:

* **TF-IDF & Cosine Similarity:** Used for basic resume-job description matching but lacks semantic understanding.
* **Word2Vec & FastText:** Improves upon keyword matching by incorporating word relationships but still struggles with long-form document comparison.
* **BERT & Transformer-based Models:** Provides deep contextual understanding, making it effective for **resume ranking and skill matching**.
* **Named Entity Recognition (NER):** Extracts structured information such as names, education, skills, and experience from resumes.
* **Deep Learning-based Ranking Models:** Incorporates **LSTMs and CNNs** for more advanced resume classification.

**2.3 Gaps and Limitations in Existing Solutions**

* **Lack of Contextual Understanding:** Many traditional models fail to capture the true relevance of candidate skills in context.
* **Limited Adaptability to Different Resume Formats:** Many systems do not handle **PDF, DOCX, and scanned documents efficiently**.
* **No Interactive Insights for Recruiters:** Existing solutions often lack **data visualization** to help recruiters quickly assess the candidate pool.
* **Security & Privacy Issues:** Resume data needs to be securely processed and stored to comply with **data protection regulations**.

**How Our Project Addresses These Gaps**

* **Uses BERT for Contextual Matching:** Improves accuracy over keyword-based models.
* **Supports Multiple File Formats:** Handles **PDF and DOCX** with **OCR-based extension possibility**.
* **Enhances User Experience with Analytics:** Provides **interactive dashboards** for skill and experience analysis.
* **Implements Secure Resume Processing:** Uses **encryption techniques** to ensure candidate data protection.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

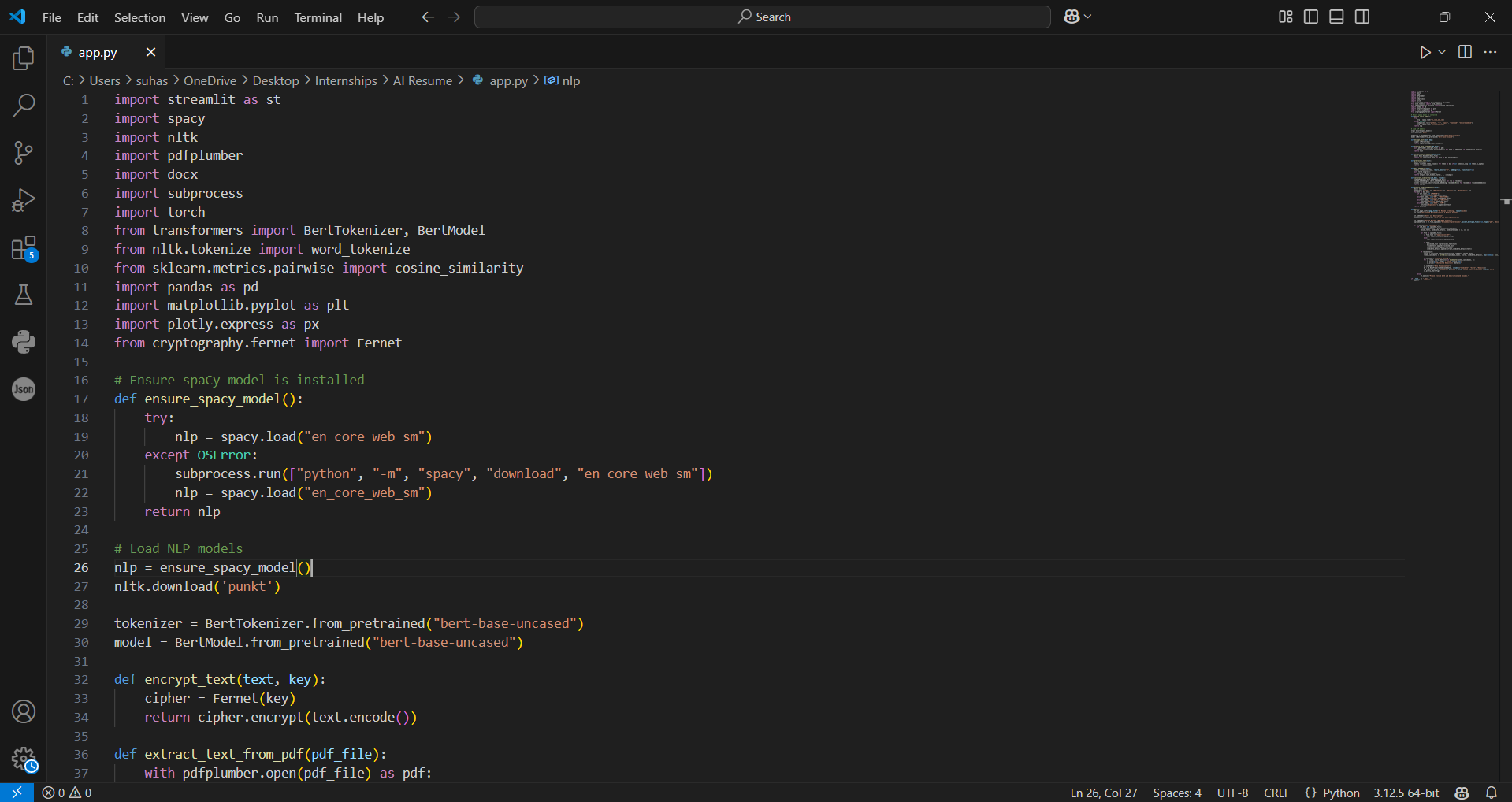
The system consists of the following components:

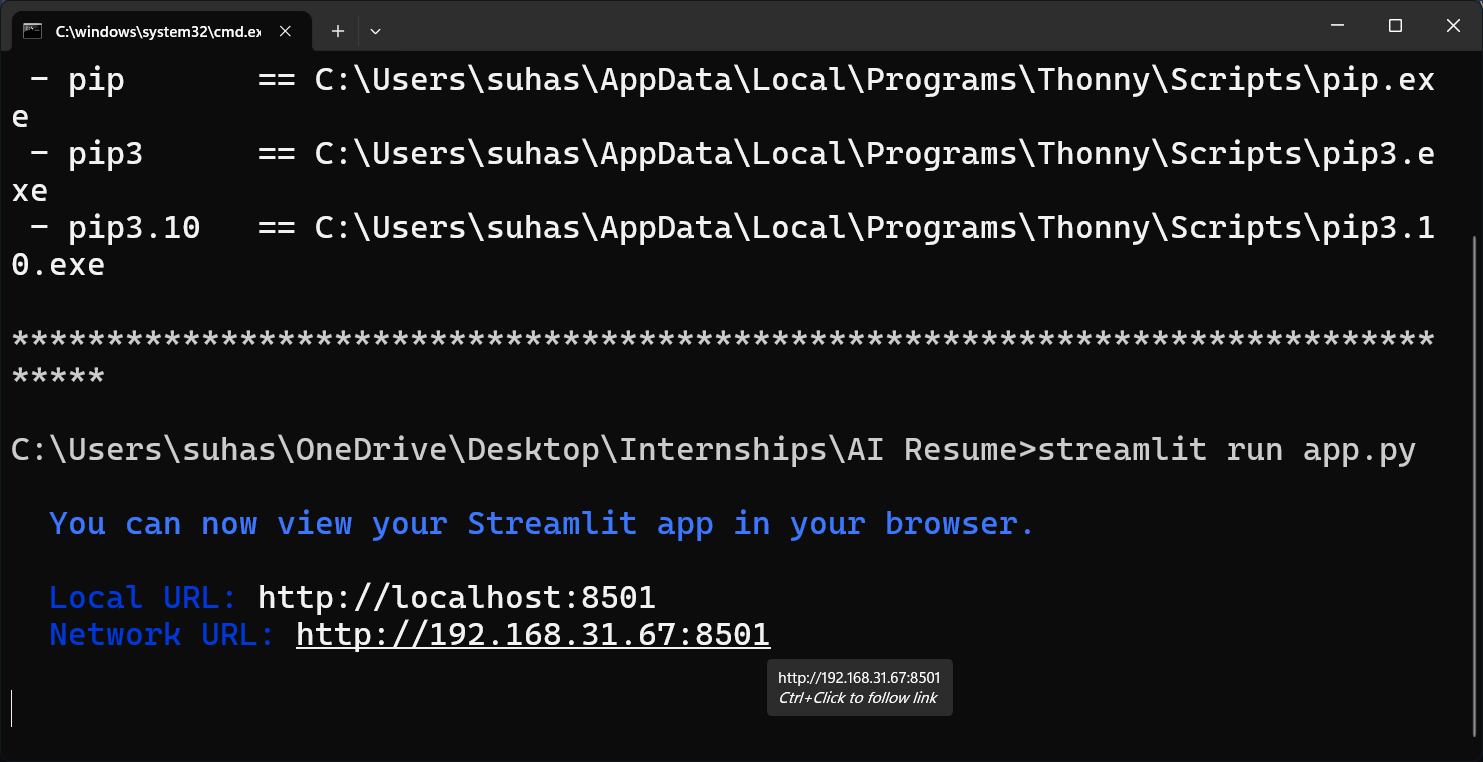
* **Resume Parsing:** Extracts text from PDF and DOCX files.
* **Preprocessing:** Tokenization, stop-word removal, and lemmatization.
* **NER Model:** Identifies candidate details such as name, education, skills, and experience.
* **BERT-based Matching:** Compares resumes with job descriptions.
* **Ranking Algorithm:** Assigns scores based on **semantic similarity** and extracted features.
* **Web Interface:** Provides an interactive dashboard for recruiters.
  1. **Requirement Specification**
     1. **Hardware Requirements:**
* Python, Streamlit, SpaCy, Transformers (Hugging Face), Scikit-learn, Matplotlib, Plotly
* BERT-based Pretrained Model
  + 1. **Software Requirements:**
* Minimum 8GB RAM, 4-core CPU
* GPU acceleration for faster model inference (optional)

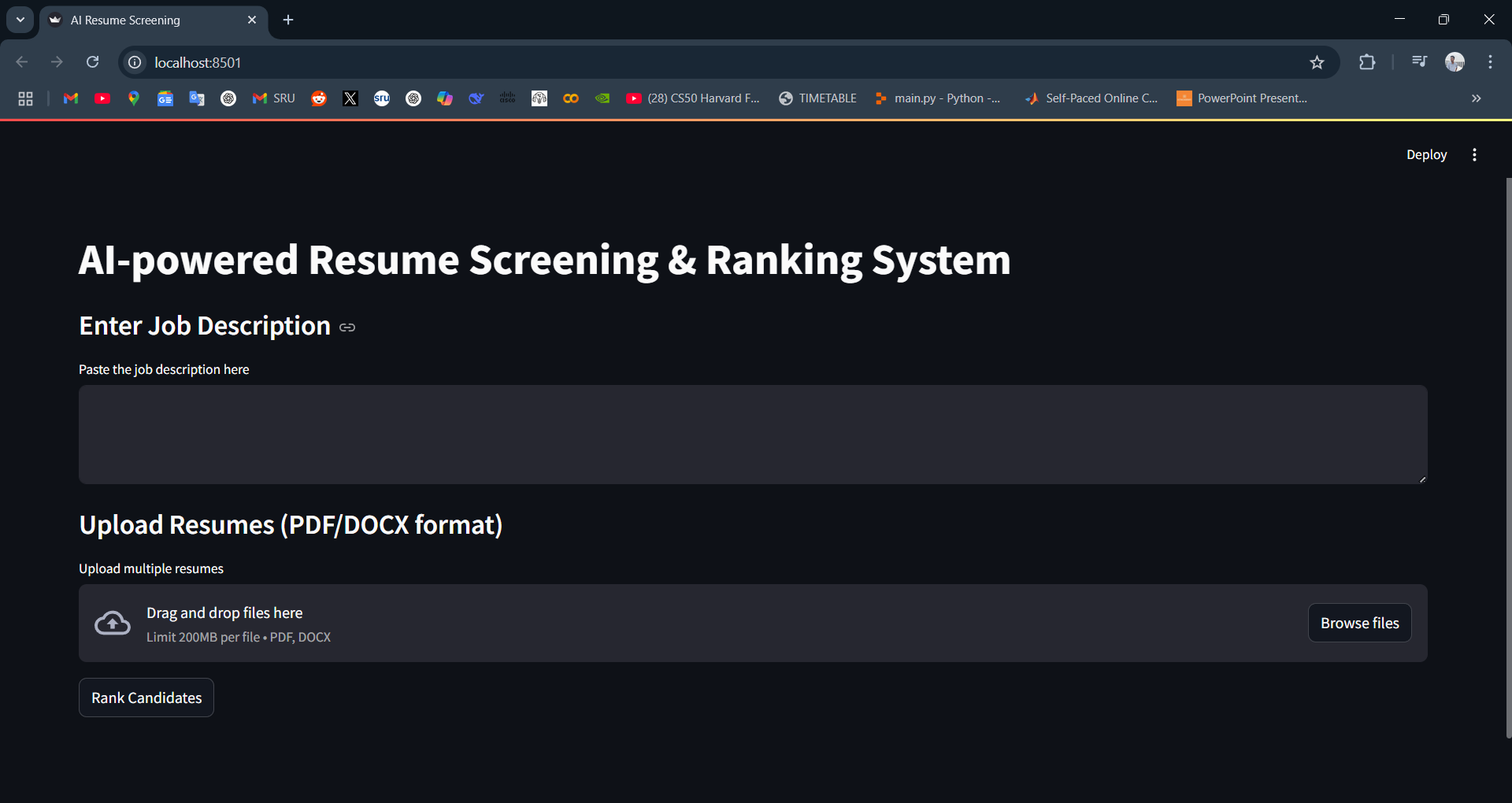
**CHAPTER 4**

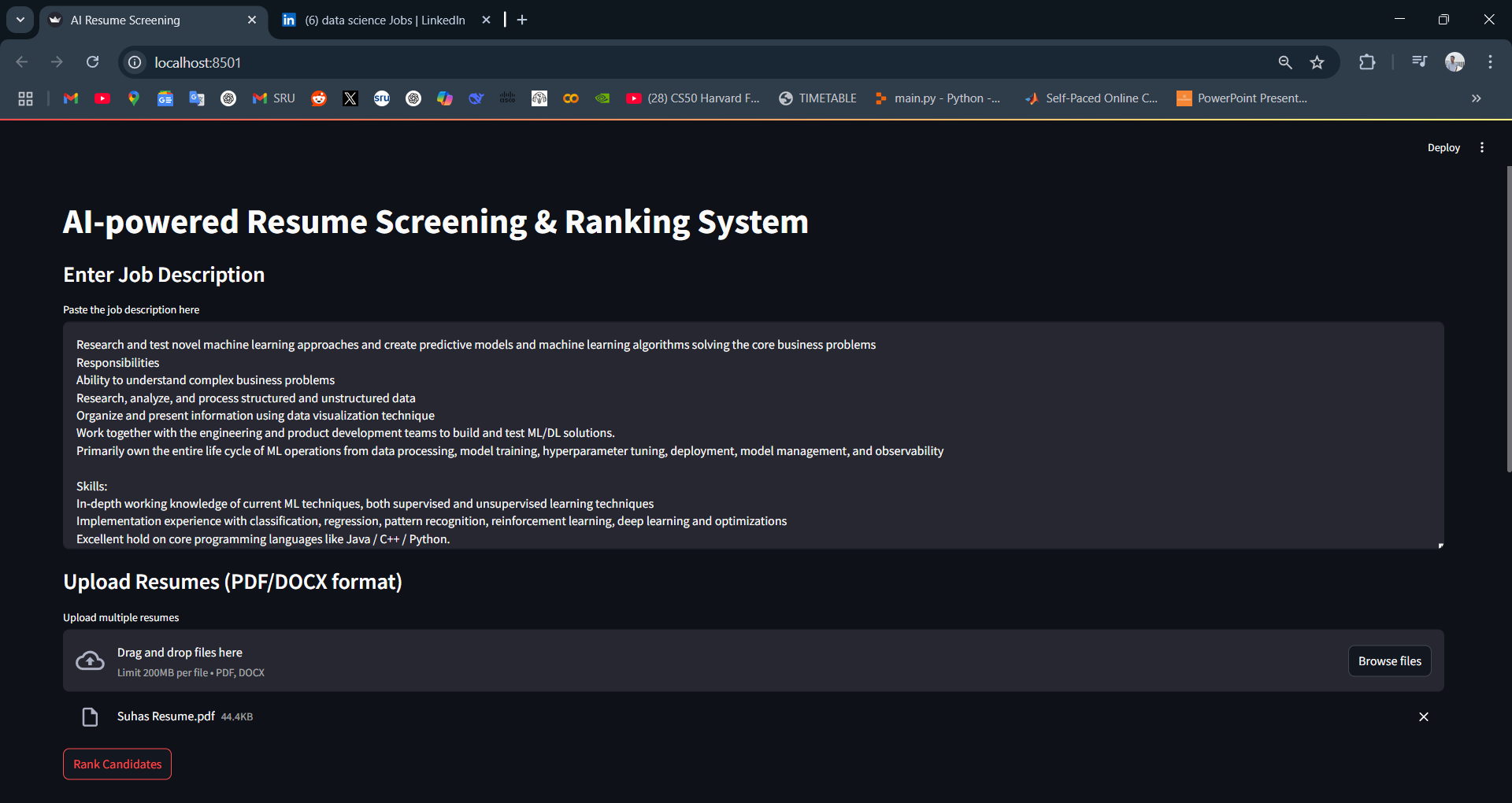
**Implementation and Result**

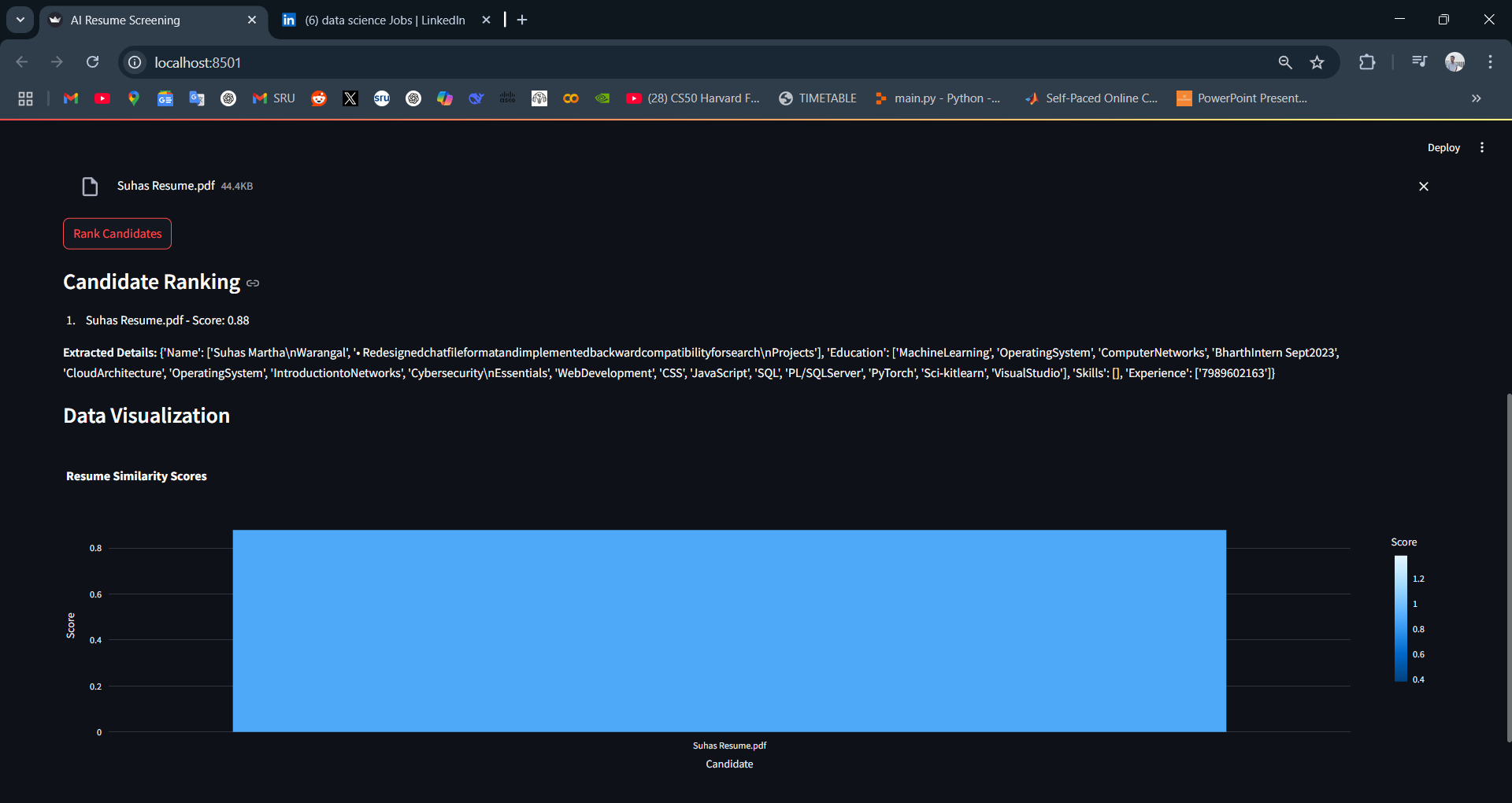
* 1. **Snap Shots of Result:**

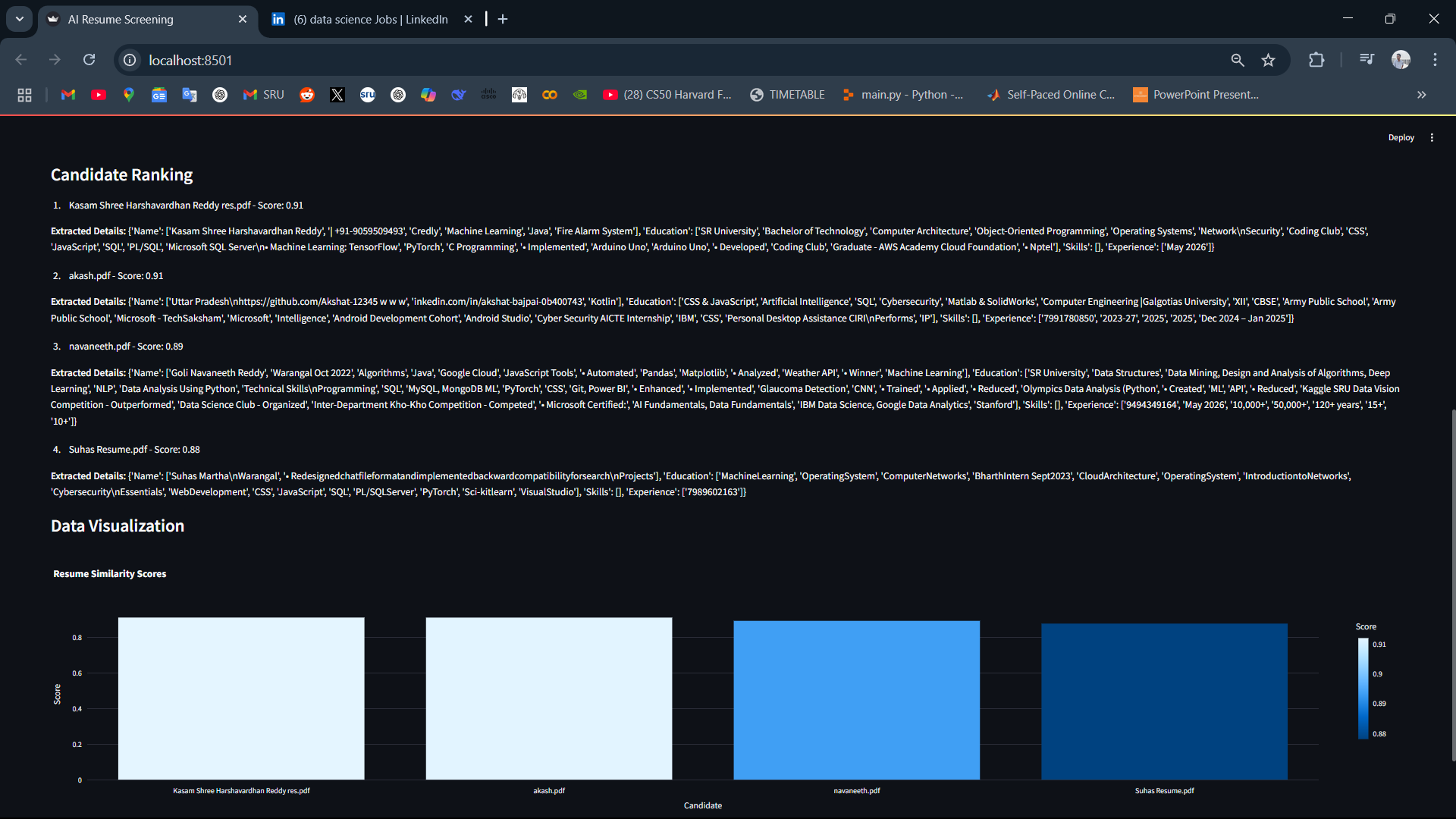












* 1. **GitHub Link for Code:**

<https://github.com/SuhasMartha/AI-Resume-Screening-System>

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**
* **Enhance NLP Techniques:** Integrate more advanced AI models like GPT for better contextual understanding.
* **Expand Multi-Language Support:** Enable the system to process resumes in multiple languages.
* **Improve Security Measures:** Implement encryption for **secure resume storage and processing**.
* **ATS Integration:** Connect with existing **Applicant Tracking Systems (ATS)** for seamless hiring.
  1. **Conclusion:**

This project successfully demonstrates the automation of resume screening using BERT-based Semantic Matching and NLP techniques. It reduces recruiter workload, improves hiring accuracy, and provides data-driven decision-making insights. Future enhancements will further refine the system for industry-level deployment.

**REFERENCES**

1. Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805, 2019.
2. Matthew Honnibal and Ines Montani, "spaCy 2: Natural Language Understanding with Bloom Embeddings," 2017.
3. AICTE & TechSaksham Internship Guidelines, 2024.