AI Resume Screening & Candidate Ranking System

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

Mudgolwad Ravi , mudgolwadravi@gmail.com

Under the Guidance of

Saomya Chawdary

ACKNOWLEDGEMENT

ABSTRACT

This project, **"AI Resume Screening & Candidate Ranking System"**, aims to automate the process of resume screening by leveraging **Natural Language Processing (NLP) and Machine Learning techniques**. The system extracts text from resumes and compares them with job descriptions using **TF-IDF vectorization and cosine similarity** to rank candidates based on relevance.

The key objectives of this project include reducing manual effort in hiring, improving candidate-job matching, and enhancing efficiency in the recruitment process. By leveraging AI-driven techniques, our model automates resume screening, minimizing bias and ensuring a fairer evaluation of candidates. The system extracts text from resumes, preprocesses it using Natural Language Processing (NLP), and applies **TF-IDF vectorization and cosine similarity** to measure relevance to job descriptions. The results demonstrate that our model effectively ranks resumes, providing recruiters with a data-driven, scalable, and efficient approach to candidate shortlisting. Additionally, the system allows for real-time ranking, helping hiring teams focus on the most suitable candidates without extensive manual review.

TABLE OF CONTENT

CHAPTER 1

Introduction

1.1 Problem Statement:

Recruiters spend significant time manually reviewing resumes, leading to inefficiencies and potential biases in hiring. This project aims to automate resume screening by ranking candidates based on their suitability for a job description using **AI and NLP techniques**.

1.2 Motivation:

With the increasing number of job applications, automating the initial screening process helps recruiters focus on the most relevant candidates. AI-powered resume screening reduces hiring time and enhances fairness in candidate selection.

1.3 Objective:

- Develop an AI system to **analyze and rank resumes** based on job descriptions.

- Utilize **TF-IDF vectorization** and **cosine similarity** for text comparison.

- Enhance **efficiency and accuracy** in resume screening.

1.4 Scope of the Project:

This project focuses on automating the **resume ranking process** for recruiters. It does not assess candidate personality traits or soft skills but provides a data-driven approach for shortlisting resumes based on textual content.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

AI-driven resume screening has been widely explored in recent years. Various techniques, including **Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning**, have been utilized to automate recruitment processes. Studies have shown that text similarity metrics such as **TF-IDF, word embeddings, and transformer-based models (e.g., BERT)** are effective in resume ranking and job matching.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Existing Models, Techniques, and Methodologies
Several existing models have been used for resume screening, including:

- TF-IDF & Cosine Similarity: Traditional text similarity measurement techniques used for ranking resumes.
- Word2Vec & Doc2Vec: Word embedding models for capturing semantic meaning in resumes.
- BERT & Transformer Models: Deep learning-based NLP models that improve contextual understanding of job descriptions and resumes.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

Despite advancements, existing solutions have several limitations:

- Lack of Context Awareness: Traditional techniques like TF-IDF do not consider contextual relationships between words.
- Bias in Training Data: Many AI models inherit biases from historical hiring data, leading to unfair candidate evaluation.
- Limited Customization: Most solutions do not allow recruiters to tweak ranking criteria based on specific job requirements.

Our project addresses these limitations by providing an interactive AI-powered resume ranking system with customizable ranking parameters and real-time updates.

CHAPTER 3

Proposed Methodology

3.1   System Design
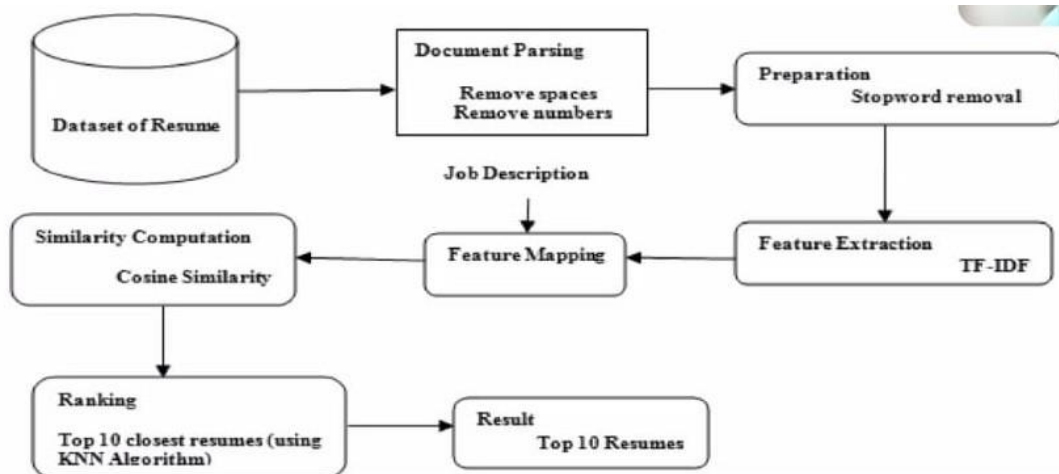
The flowchart represents the process of AI-powered resume screening and ranking using **TF-IDF and Cosine Similarity**.

1. **Dataset of Resumes** – The system starts with a collection of resumes.

2. **Document Parsing** – Resumes are cleaned by **removing spaces and numbers** to enhance text processing.

3. **Preparation** – **Stopwords are removed** to retain only meaningful words.

4. **Feature Extraction (TF-IDF)** – Resumes and job descriptions are converted into numerical representations using **TF-IDF vectorization**.

5. **Feature Mapping** – The job description is also vectorized to compare with resumes.

6. **Similarity Computation (Cosine Similarity)** – The system calculates similarity scores between resumes and the job description.

7. **Ranking (KNN Algorithm)** – The **top 10 closest resumes** are identified using the **K-Nearest Neighbors (KNN) algorithm** based on similarity scores.

8. **Result** – The system outputs the **top 10 most relevant resumes** for further evaluation.

This automated approach enhances **hiring efficiency, reduces manual effort, and minimizes bias**, ensuring fair and accurate resume ranking for recruiters.

3.2   Requirement Specification

3.2.1   Hardware Requirements:

Processor: Intel i5 or higher

RAM: 8GB or more

Storage: 20GB free space

3.2.2   Software Requirements:

Python

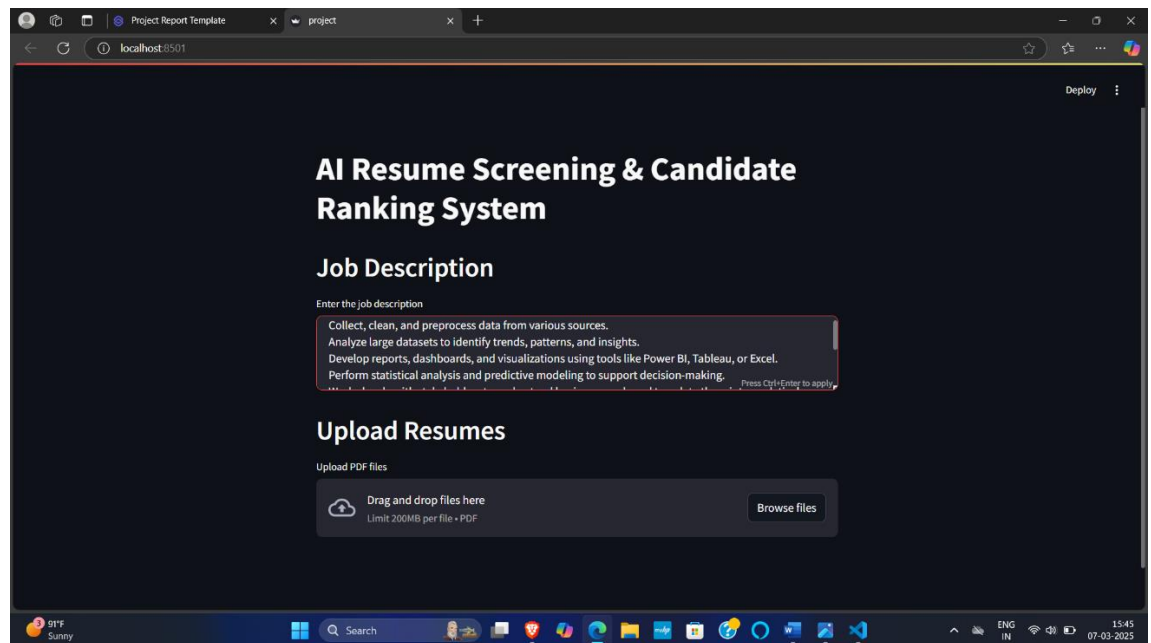Streamlit

Scikit-learn

PyPDF2

Pandas

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Kindly provide 2-3 Snapshots which showcase the results and output of your project and after keeping each snap explain the snapshot that what it is representing.
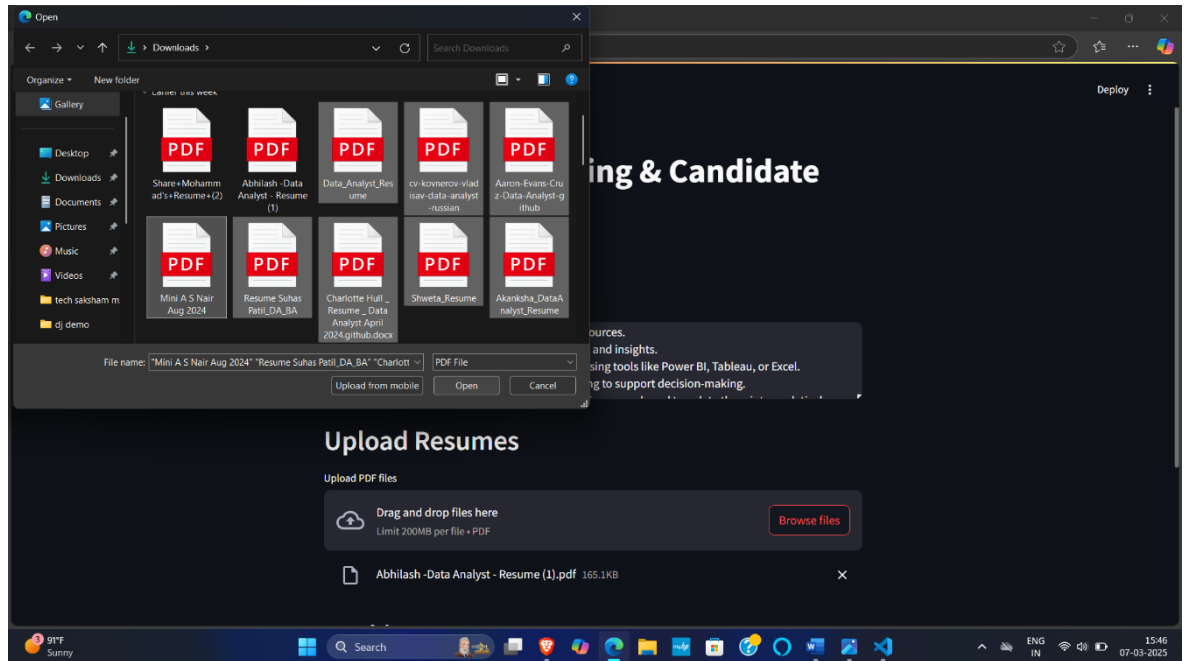
1. **Job Description Input Interface:**
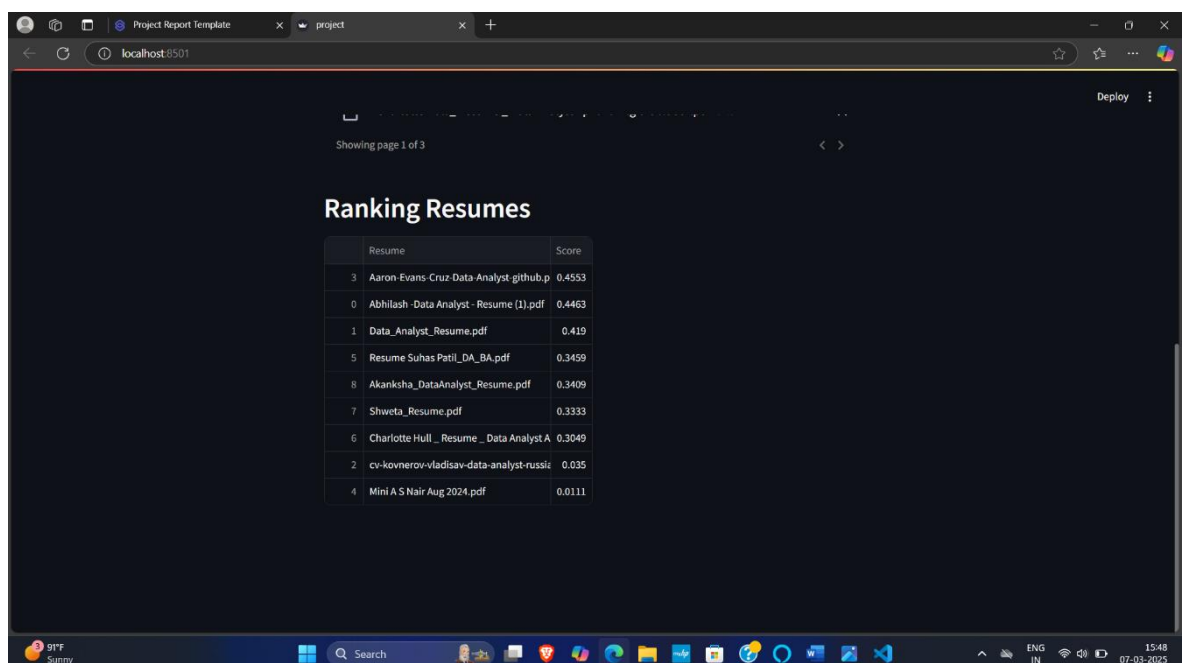   o Users enter the job description in the Streamlit app.

2. **Resume Upload Interface:**
   - o Users upload multiple PDF resumes.

3. **Ranked Resume Output:**
   - o The system displays ranked resumes with their similarity scores.

4.2 GitHub Link for Code:

https://github.com/mudgolwadravi/AI-powered-Resume-Screening-and-Ranking-System.git

CHAPTER 5

Discussion and Conclusion

5.1   Future Work:

- Implement deep learning techniques such as BERT for improved accuracy.
- Expand the system to include skill extraction and keyword analysis.
- Integrate bias reduction mechanisms to ensure fair candidate evaluation.

5.2   Conclusion:

- **Functional Solution:** This project introduces an AI-powered resume screening and ranking system that automates candidate shortlisting using **machine learning and NLP techniques**. By leveraging **TF-IDF vectorization and cosine similarity**, the system efficiently matches resumes to job descriptions, reducing manual effort and enhancing the hiring process.

- **User-Friendly Interface:** The system is designed with an intuitive interface that allows users to **upload resumes, enter job descriptions, and view ranked candidates**. Its **responsive design and error-handling mechanisms** ensure a smooth user experience for recruiters.

- **Data Processing and Ranking:** The system **extracts and preprocesses** textual data from resumes and job descriptions, applying **TF-IDF vectorization** to generate numerical features. Using **cosine similarity and KNN-based ranking**, it identifies the most relevant resumes, enabling recruiters to make data-driven hiring decisions.

- **Future Enhancements:** Future improvements may include **advanced NLP models (such as BERT or GPT-based embeddings)** for deeper semantic understanding, **customizable ranking criteria**, and an **expanded dataset** to improve accuracy and adaptability. Integrating **real-time applicant tracking system (ATS) support** and **automated feedback mechanisms** can further refine the hiring process.

REFERENCES

Gangoda, Nikethani, et al. "Resume Ranker: AI-Based Skill Analysis and Skill Matching System." *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2024.

Shukla, Anant. "Resume Screening and Ranking with spaCy." *Turkish Online Journal of Qualitative Inquiry* 12.7 (2021).

Amro, Belal, Arwa Najjar, and Mario Macido. "An Intelligent Decision Support System For Recruitment: Resumes Screening and Applicants Ranking." (2022).