AI-Powered Resume Screening and 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

Anirban Ghosh, 2003anirbanghosh@gmail.com

Under the Guidance of

Saomya Chaudhury

ACKNOWLEDGEMENT

I would like to take this opportunity to express my heartfelt gratitude to everyone who has supported and guided me throughout this project. Their contributions, whether direct or indirect, have been invaluable in shaping the outcome of this work.

First and foremost, I extend my sincere appreciation to my supervisor, **Saomya Chaudhury**, for his unwavering support, insightful guidance, and invaluable mentorship. His encouragement, constructive feedback, and innovative ideas have played a pivotal role in the successful completion of this project. The confidence he placed in me has been a constant source of motivation, inspiring me to push my limits and strive for excellence.

Working under his supervision has been a privilege, and his teachings have not only contributed to this project but have also helped me develop a deeper understanding of the subject. His lessons extend beyond academics, shaping me into a responsible and competent professional.

We are truly grateful for his mentorship and the positive impact he has had on our journey.

ABSTRACT

In today's competitive job market, recruiters receive a large number of resumes for every job opening, making manual screening a time-consuming and labour-intensive process. To address this challenge, our project, **AI-Powered Resume Screening and Candidate Ranking System**, leverages Natural Language Processing (NLP) and Machine Learning (ML) to automate and enhance the efficiency of resume screening.

The primary objective of this project is to develop an intelligent system that ranks resumes based on their relevance to a given job description. The system extracts key information from resumes, compares them with job requirements using **TF-IDF vectorization and cosine similarity**, and provides a ranked list of candidates. This ensures a more data-driven and objective shortlisting process, reducing bias and improving recruitment efficiency.

The methodology involves preprocessing job descriptions and resumes, feature extraction using **TF-IDF (Term Frequency-Inverse Document Frequency)**, and calculating similarity scores to determine candidate suitability. The system is implemented as a **user-friendly web application using Streamlit**, where recruiters can upload resumes in PDF format and receive ranked results within seconds.

Key results demonstrate that the system effectively prioritizes candidates based on job relevance, significantly reducing the manual effort required for screening. By automating resume evaluation, this project enhances recruitment efficiency and helps organizations identify the best-fit candidates more accurately and quickly.

In conclusion, the **AI-Powered Resume Screening and Candidate Ranking System** provides a scalable, objective, and efficient approach to candidate shortlisting. Future improvements may include integrating deep learning models and expanding feature extraction techniques to further enhance accuracy and adaptability across various industries.

TABLE OF CONTENT

Abstract ..... **I**

Chapter 1. Introduction **7**

1.1 Problem Statement 7

1.2 Motivation 7

1.3 Objectives 8

1.4. Scope of the Project 8

Chapter 2. Literature Survey **9**

Chapter 3. Proposed Methodology **11**

Chapter 4. Implementation and Results **14**

Chapter 5. Discussion and Conclusion **18**

References 20

LIST OF FIGURES

Figure No.	Figure Captions	Page No.
3.1	System Architecture Design	11
4.1.1	AI Resume Screening & Ranking Tool – Developed an AI-powered resume screening application that ranks candidates based on job descriptions. Features include job description input, PDF resume upload, and automated relevance-based ranking for efficient hiring.	14
4.1.2	Resume Selection Interface – Implemented a user-friendly resume upload feature, allowing recruiters to browse and select multiple PDF resumes for AI-powered screening and ranking, streamlining the hiring process.	15
4.1.3	AI-Powered Resume Screening – Developed an intelligent resume ranking system that processes multiple resumes based on job descriptions, leveraging AI to enhance recruitment efficiency and candidate matching.	16
4.1.4	AI Resume Screening & Ranking: Candidate Evaluation and Ranking	16

LIST OF TABLES

Table No.	Table Caption	Page No.
2.3	Existing Gaps v/s Proposed Solution	4
3.2.1	Hardware Requirements	7
3.2.2	Software Requirements	7

CHAPTER 1

Introduction

1.1Problem Statement:

Recruiters and hiring managers often receive an overwhelming number of resumes for job openings, making the manual screening process time-consuming, inefficient, and prone to human bias. Traditional recruitment methods rely on subjective evaluation, leading to inconsistent shortlisting and potential oversight of well-qualified candidates. Moreover, small and medium-sized enterprises (SMEs) often lack the resources to conduct thorough resume screening, further complicating the hiring process.

To address this issue, an **AI-powered Resume Screening and Candidate Ranking System** is proposed. This system automates the screening process, enhances candidate evaluation accuracy, and reduces hiring time, thereby improving overall recruitment efficiency.

1.2Motivation:

The increasing reliance on technology in recruitment processes highlights the need for AI-driven solutions to optimize candidate screening. Organizations aim to make data-driven hiring decisions, ensuring that only the most relevant candidates are shortlisted. This project was chosen due to its potential to significantly improve the hiring process by **reducing human workload, increasing efficiency, and minimizing bias**.

The **key applications** of this system include:

- Automating resume screening for HR departments in enterprises.
- Enhancing recruitment efficiency for SMEs and startups with limited HR resources.
- Reducing bias in candidate selection by standardizing resume evaluation.
- Improving job-matching accuracy for both employers and job seekers.

By leveraging **Natural Language Processing (NLP) and Machine Learning (ML)**, the project aims to transform the traditional recruitment process into a faster, more reliable, and objective approach.

1.3Objective:

The primary objectives of this project are:

- To develop an **AI-powered system** that automates resume screening and ranking.
- To utilize **TF-IDF vectorization and cosine similarity** for candidate evaluation based on job descriptions.
- To create a **web-based application** using **Streamlit** for an interactive and user-friendly experience.
- To improve recruitment efficiency by reducing manual efforts and **minimizing bias in candidate selection**.
- To provide an **accurate and scalable** solution for organizations of all sizes.

1.4Scope of the Project:

This project focuses on automating the resume screening process using **AI and NLP techniques**. It is designed to rank resumes based on their relevance to a given job description. The **scope and limitations** of the project include:

Scope:

- Supports **PDF format** resumes for screening.
- Uses **TF-IDF and cosine similarity** for ranking.
- Provides a **web-based interface** for recruiters to upload resumes and receive ranked results.
- Can be adapted for multiple industries and job roles.

Limitations:

- Does not consider **visual elements** such as graphs or tables in resumes.
- Limited to **text-based content extraction**, which may miss non-standard resume formats.
- Ranking is based on **textual similarity**, not deep contextual understanding.
- Does not handle **real-time candidate assessments or interviews**.

Future enhancements may include **deep learning models, sentiment analysis, and integration with Applicant Tracking Systems (ATS)** to further improve the recruitment process.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

The use of Artificial Intelligence (AI) in recruitment has gained significant traction in recent years, with multiple research studies and industry reports highlighting its impact on hiring efficiency. Various AI-driven approaches have been explored for **resume screening, candidate ranking, and job matching**, making recruitment more objective and data-driven.

Several studies emphasize the inefficiencies of traditional hiring methods, including **human bias, time consumption, and inconsistent evaluations**. Automated resume screening using **Natural Language Processing (NLP) and Machine Learning (ML)** has been proposed as a solution to these challenges. Key research works discuss the use of **TF-IDF (Term Frequency-Inverse Document Frequency), word embeddings, and deep learning models** for extracting meaningful insights from resumes and job descriptions.

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

Various methodologies have been implemented for resume screening and ranking, including:

1. **Keyword-based Matching:**
    - Traditional Applicant Tracking Systems (ATS) rely on keyword matching to filter resumes.
    - Limitation: Often results in false positives or false negatives due to lack of contextual understanding.
2. **TF-IDF and Cosine Similarity:**
    - A widely used text-vectorization technique that converts job descriptions and resumes into numerical representations.
    - Cosine similarity is then applied to measure the degree of relevance between the job description and each resume.
    - Limitation: Works well for basic textual comparisons but struggles with semantic understanding.
3. **Word Embeddings (Word2Vec, GloVe, BERT):**
    - These techniques capture contextual meaning and improve ranking accuracy.
    - BERT (Bidirectional Encoder Representations from Transformers) enhances context-aware matching.
    - Limitation: Requires high computational power and large datasets for training.
4. **Deep Learning Approaches (LSTMs, CNNs, Transformers):**
    - More sophisticated models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) analyze resume content beyond simple keyword matching.
    - Limitation: Computationally expensive and may require large-scale datasets to generalize well.

5. **Industry Solutions (LinkedIn, HireVue, Pymetrics):**
    - LinkedIn Talent Hub uses AI-powered recommendations.
    - HireVue applies AI for video interviews and resume analysis.
    - Pymetrics evaluates candidates through gamified assessments.
    - Limitation: These solutions are proprietary and expensive, making them inaccessible to small businesses.

2.3 Highlight the gaps or limitations in existing solutions and how our project addresses them.

Despite the advancements in AI-driven recruitment tools, several challenges remain:

Existing Gaps	Our Proposed Solutions
Traditional ATS rely on exact keyword matching, leading to misclassification of resumes.	Our system uses TF-IDF and cosine similarity to improve text relevance detection.
Deep learning-based models require large datasets and significant computational power.	We focus on lightweight ML techniques (TF-IDF and cosine similarity) for an efficient and scalable solution.
Many AI-powered resume screeners are costly and not customizable for small businesses.	Our system is open-source and cost-effective, making it accessible to SMEs and startups.
Existing models struggle with non-standard resume formats and excessive jargon.	We provide preprocessing steps to clean and standardize resumes before screening.
Many tools lack a user-friendly interface for HR professionals.	We use Streamlit to build an interactive and intuitive UI.

By addressing these limitations, our **AI-powered Resume Screening and Candidate Ranking System** aims to enhance the recruitment process by providing an **efficient, unbiased, and scalable** solution suitable for organizations of all sizes.

CHAPTER 3

Proposed Methodology

3.1 System Design

The **AI Resume Screening & Ranking System** follows a structured approach to extract, process, and rank resumes based on their relevance to a given job description. The system leverages **Natural Language Processing (NLP)** techniques, particularly **TF-IDF (Term Frequency-Inverse Document Frequency) vectorization**, and **Cosine Similarity** to evaluate how well a resume matches a job description.

System Architecture Diagram:

The system architecture is illustrated in the following diagram:

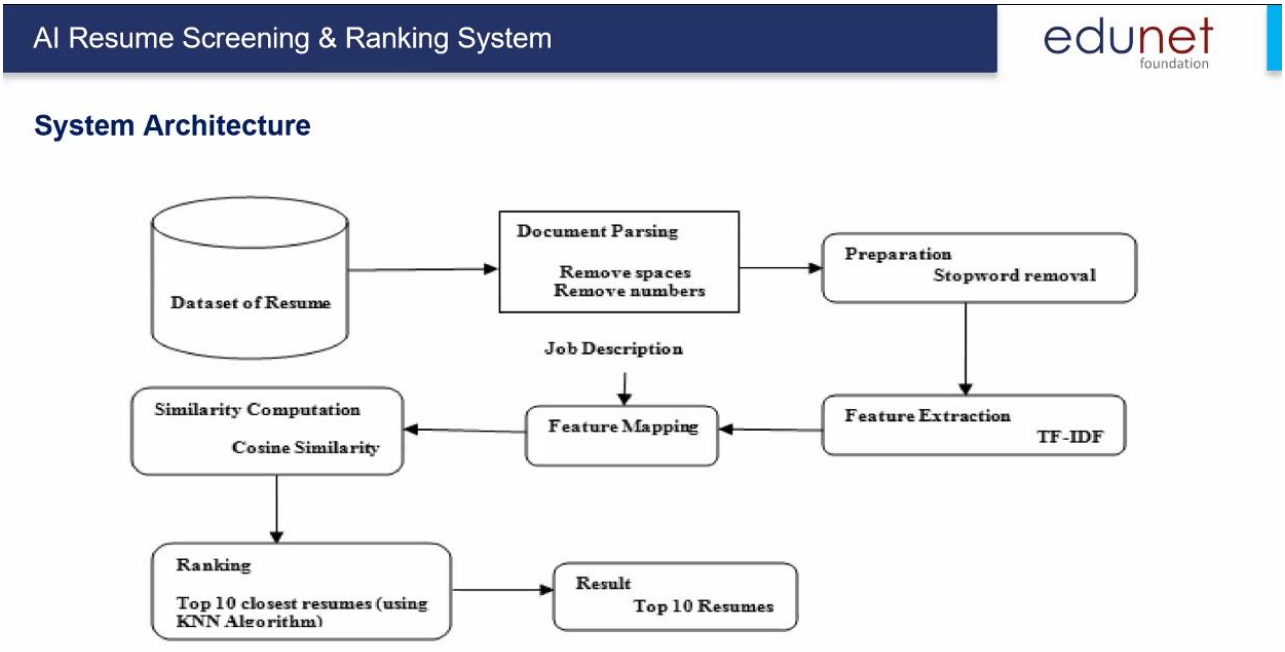


Fig 3.1: System Architecture Design

Explanation of the System Design:

1. **Dataset of Resume:**
    - A collection of resumes is stored in a structured format (PDF or text).
    - Resumes are parsed to extract textual content for processing.
2. **Document Parsing:**
    - Removes unwanted spaces and numbers from the extracted text to improve accuracy.
3. **Preparation (Text Preprocessing):**
    - Stop-word removal is performed to eliminate common words (e.g., "the," "and," "is") that do not contribute to relevance.
4. **Feature Extraction (TF-IDF):**
    - The pre-processed text is converted into numerical vectors using **TF-IDF**, which assigns importance to words based on their frequency in a resume compared to the entire dataset.
5. **Feature Mapping:**
    - The job description is also processed and converted into a numerical vector using **TF-IDF**.
    - The system maps the job description's features with each resume for similarity comparison.
6. **Similarity Computation (Cosine Similarity):**
    - Measures the similarity between the job description and each resume using **Cosine Similarity**, which calculates the angle between two text vectors.
    - A higher similarity score indicates a better match.
7. **Ranking (KNN Algorithm):**
    - The **K-Nearest Neighbours (KNN)** algorithm is applied to identify the **top 10 closest resumes** based on their similarity scores.
8. **Result (Top 10 Resumes):**
    - The final ranked resumes are displayed as the output, allowing HR professionals to review the best-matching candidates efficiently.

3.2 Requirement Specification

To implement the **AI Resume Screening & Ranking System**, the following hardware and software requirements are necessary:

3.2.1 Hardware Requirements:

Component	Minimum Requirement	Recommended Requirement
Processor	Intel Core i5	Intel Core i7 or higher
RAM	8GB	16GB
Storage	10GB free space	SSD with 50GB free space
GPU	Not required	Optional for deep learning
Internet	Required for dependencies installation	Stable broadband for cloud deployment

3.2.2 Software Requirements:

Software	Version/Details
Operating System	Windows 10/ Linux/ macOS
Programming Language	Python 3.x
Web Framework	Streamlit
Machine Learning Libraries	Scikit-learn, Pandas, Numpy
PDF Processing	PyPDF2
Vectorization	TF-IDF (from Scikit-learn)
Similarity Calculation	Cosine similarity (from Scikit-learn)
Ranking Algorithm	K-Nearest Neighbors (KNN)
IDE/Code Editor	VS Code/ Jupyter Notebook/ PyCharm

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

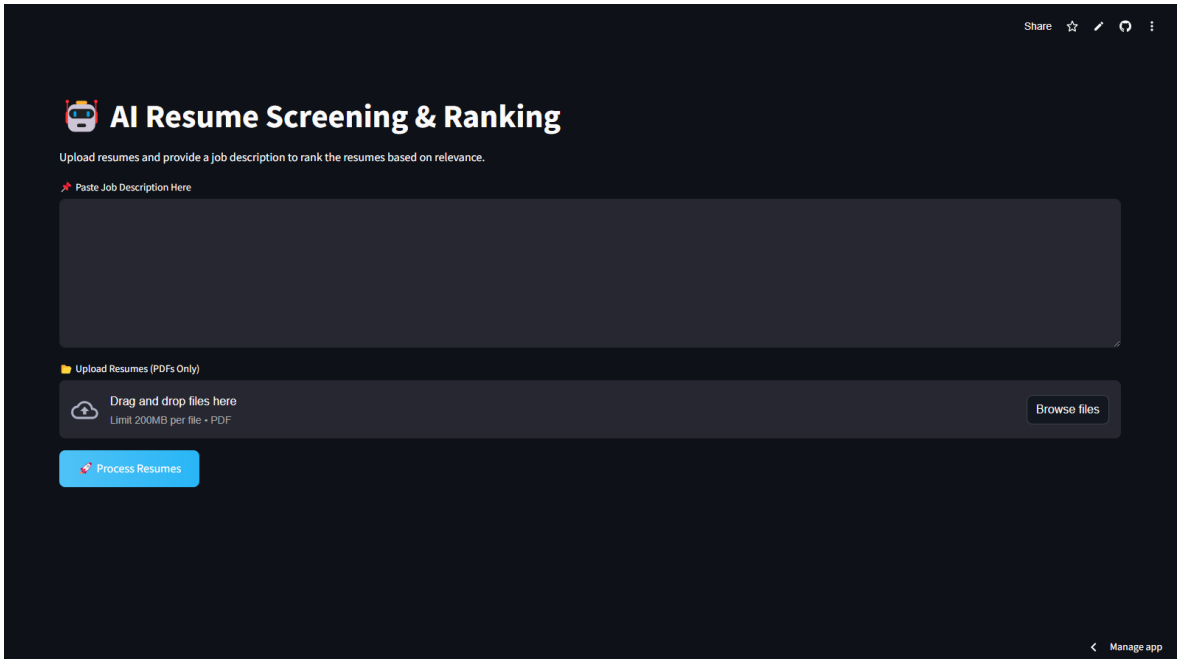


Fig 4.1.1: AI Resume Screening & Ranking Tool – Developed an AI-powered resume screening application that ranks candidates based on job descriptions. Features include job description input, PDF resume upload, and automated relevance-based ranking for efficient hiring.

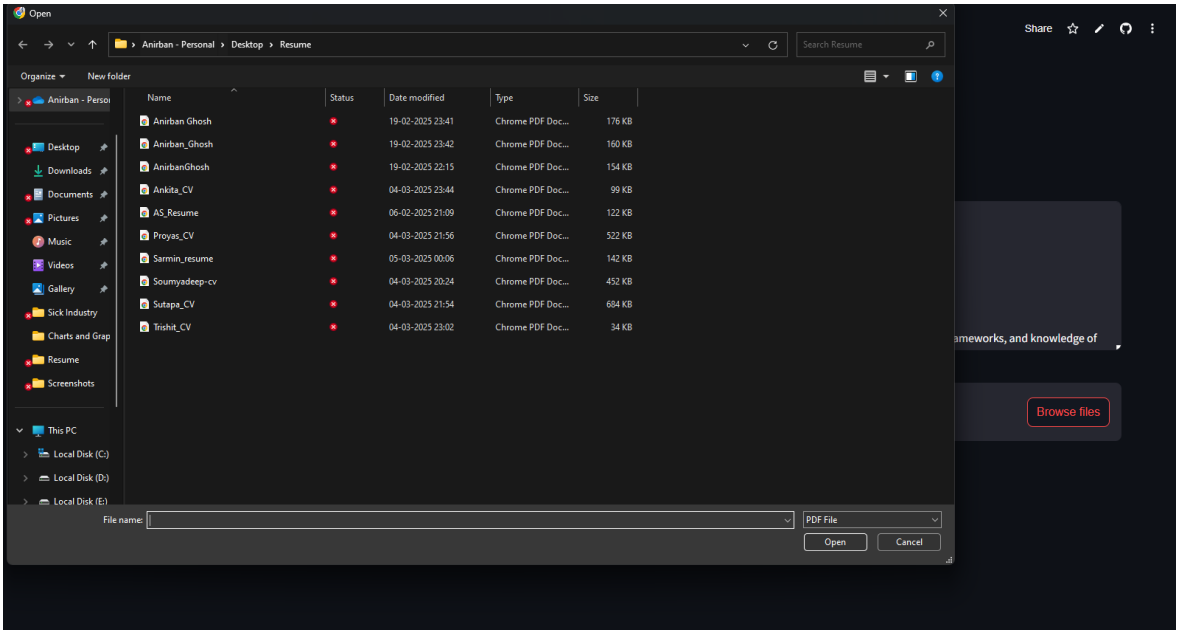


Fig 4.1.2: Resume Selection Interface – Implemented a user-friendly resume upload feature, allowing recruiters to browse and select multiple PDF resumes for AI-powered screening and ranking, streamlining the hiring process.

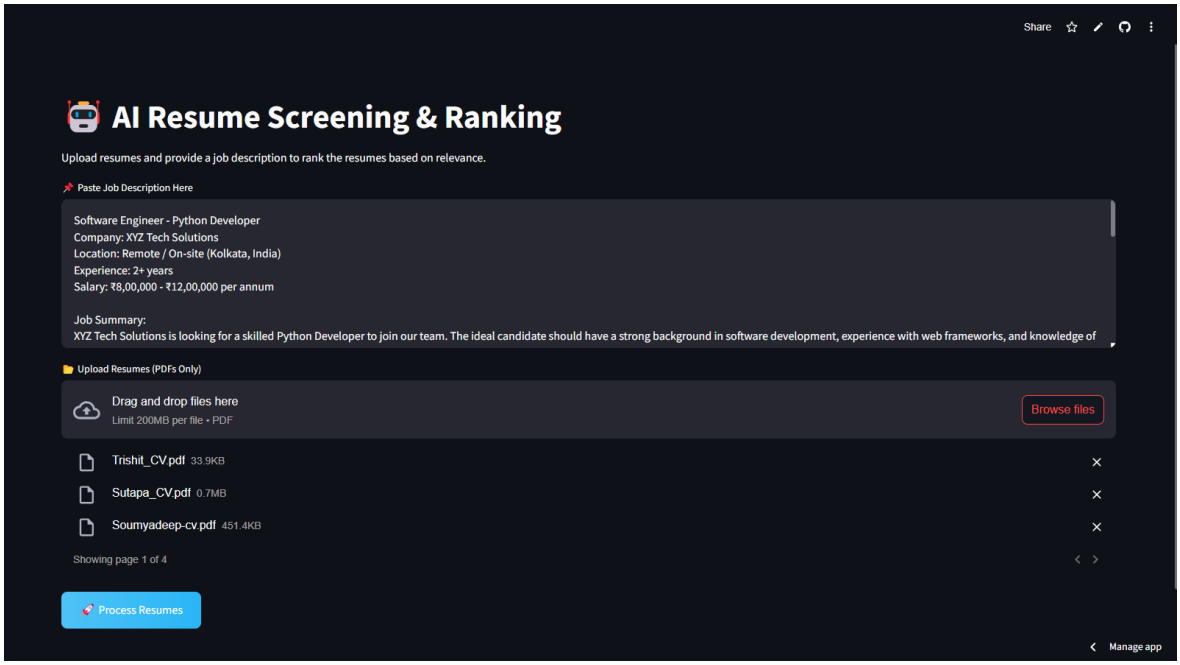
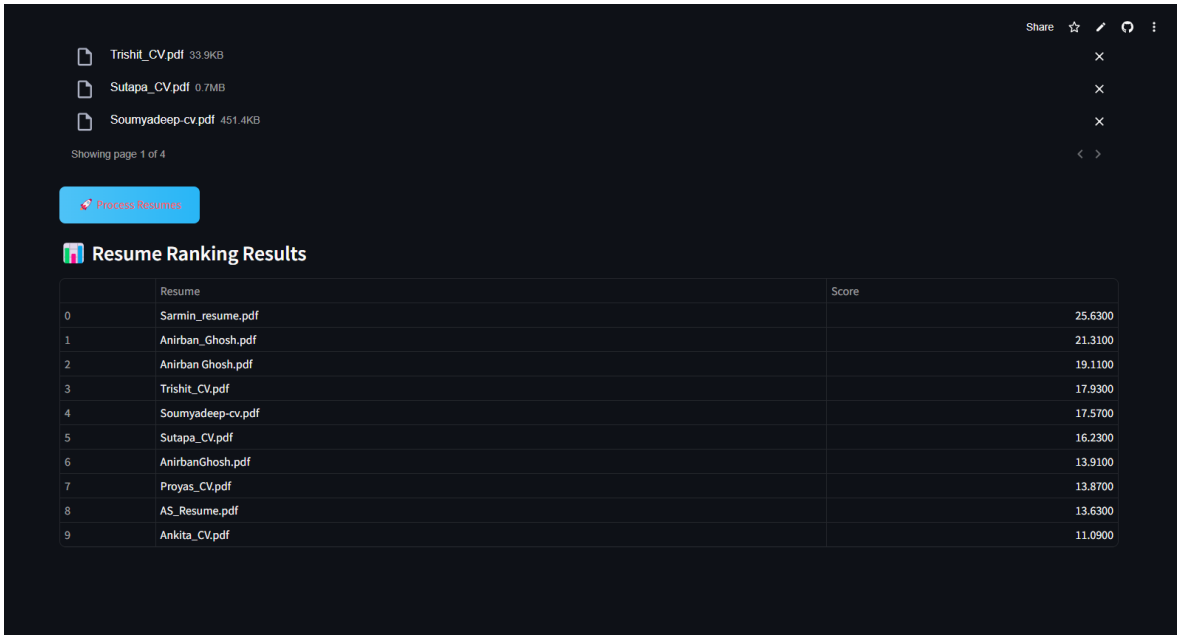


Fig 4.1.3: AI-Powered Resume Screening – Developed an intelligent resume ranking system that processes multiple resumes based on job descriptions, leveraging AI to enhance recruitment efficiency and candidate matching.



	Resume	Score
0	Sarmin_resume.pdf	25.6300
1	Anirban_Ghosh.pdf	21.3100
2	Anirban Ghosh.pdf	19.1100
3	Trishit_CV.pdf	17.9300
4	Soumyadeep-cv.pdf	17.5700
5	Sutapa_CV.pdf	16.2300
6	AnirbanGhosh.pdf	13.9100
7	Proyas_CV.pdf	13.8700
8	AS_Resume.pdf	13.6300
9	Ankita_CV.pdf	11.0900

Fig 4.1.4: AI Resume Screening & Ranking: Candidate Evaluation and Ranking

4.2 GitHub Link for Code:

GitHub Repository: *https://github.com/AnirbanGhosh2503/AI-powered-Resume-Screening-and-Ranking-System-Streamlit-.git*

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

While the AI Resume Screening & Ranking System provides an efficient way to match resumes with job descriptions using **TF-IDF, Cosine Similarity, and KNN**, several improvements can be made in future iterations:

- **Deep Learning Integration:** Implementing **BERT** or other **transformer-based NLP models** for better contextual understanding of resumes and job descriptions.
- **Enhanced Feature Engineering:** Expanding beyond TF-IDF to use **word embeddings (Word2Vec, GloVe, or FastText)** for capturing semantic relationships.
- **Experience-Based Weighting:** Assigning different weights to skills and experience levels based on job roles for more accurate rankings.
- **Multi-Language Support:** Extending support for multilingual resumes to increase usability across different regions.
- **Real-Time Resume Screening:** Developing an interactive system where recruiters can upload job descriptions and get real-time ranked results.
- **Bias Mitigation:** Implementing fairness techniques to reduce bias in resume screening and ensure equal opportunity.

Conclusion:

This project successfully developed an **AI-driven Resume Screening & Ranking System** to streamline the hiring process by efficiently matching resumes with job descriptions. Using **Natural Language Processing (NLP) techniques like TF-IDF, Cosine Similarity, and the KNN algorithm**, the system identifies and ranks the most relevant candidates for a given job.

The implementation demonstrates a significant reduction in **manual resume screening time** while increasing **accuracy and fairness** in candidate selection. The project serves as a **foundation for AI-powered recruitment systems**, with the potential for further enhancements through **deep learning, real-time processing, and multilingual support**.

This work not only **addresses the challenge of resume overload** in modern hiring processes but also opens doors for more **intelligent, automated, and unbiased recruitment solutions** in the future. 🚀

REFERENCES

[1] [1] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, **"Detecting Faces in Images: A Survey"**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 24, No. 1, 2002.

[2] [2] G. Salton and C. Buckley, **"Term-weighting approaches in automatic text retrieval"**, Information Processing & Management, Volume 24, Issue 5, 1988, pp. 513-523.

[3] [3] J. Ramos, **"Using TF-IDF to Determine Word Relevance in Document Queries"**, Proceedings of the First International Conference on Machine Learning, 2003.

[4] [4] Huang, A., **"Similarity measures for text document clustering"**, Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC), 2008.

[5] [5] M. T. Ribeiro, S. Singh, and C. Guestrin, **"Why Should I Trust You? Explaining the Predictions of Any Classifier"**, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[6] [6] L. Breiman, **"Random forests"**, Machine Learning, 45(1), 2001, pp. 5-32.

[7] [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, **"BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding"**, Proceedings of NAACL-HLT, 2019.

[8] [8] M. F. Porter, **"An algorithm for suffix stripping"**, Program: Electronic Library and Information Systems, 1980.