

AI-powered Resume Screening and 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

Samrat Sinha, samratsinha7318@gmail.com

Under the Guidance of

Pavan Kumar

Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to mr. Pavan Kumar for offering the insightful and enriching **Transformative Learning in AI** course. This one-month internship provided me with a comprehensive understanding of AI concepts, practical applications, and emerging trends in the field.

I extend my heartfelt appreciation to our mentors, instructors, and coordinators for their guidance, engaging sessions, and valuable feedback throughout the program. Their expertise and support have significantly enhanced my knowledge and problem-solving skills in AI.

I am also grateful to my fellow participants for their collaboration, discussions, and shared learning experiences. This program has not only deepened my technical proficiency but also encouraged critical thinking and innovative approaches to AI-driven solutions.

This transformative journey has been instrumental in shaping my perspective on AI, and I look forward to applying these insights in future endeavours.

Thank you!

ABSTRACT

In the modern recruitment process, organizations receive a large volume of resumes for each job opening, making manual screening a time-consuming and labor-intensive task. To address this challenge, this project presents an **AI Resume Screening & Candidate Ranking System**, leveraging **machine learning** and **natural language processing (NLP)** techniques to automate the screening process efficiently.

The objective of the project is to develop a **Flask-based web application** that ranks and classifies resumes based on job descriptions. The system uses **TF-IDF vectorization** and **cosine similarity** to measure the relevance of resumes, ensuring a more objective and efficient hiring process. The web interface, built using **Streamlit**, allows users to upload resumes, input job descriptions, and view ranked results.

The methodology consists of **data preprocessing**, including **document parsing, stopwords removal, and feature extraction**. The system maps resumes to job descriptions, computes similarity scores, and ranks resumes using the **K-Nearest Neighbors (KNN) algorithm**. The backend is developed in **Python**, utilizing frameworks such as **NLTK, SpaCy, Scikit-learn, NumPy, Pandas, Matplotlib, Eborn, pdf2, and python-docx**. The final deployment is carried out using **Streamlit Cloud**.

Key results demonstrate that the system effectively automates resume ranking, reducing the time required for initial screening. The user-friendly interface enhances recruiter experience, providing clear and accurate candidate rankings.

In conclusion, the **AI Resume Screening & Candidate Ranking System** streamlines the recruitment process by integrating **AI-driven automation**. Future enhancements include **advanced NLP models** like **BERT/GPT**, **structured data extraction**, **interactive dashboards**, **multi-language support**, and **stronger security measures** to improve adaptability and performance.

TABLE OF CONTENT

Abstract	I
Chapter 1. Introduction.....		1
1.1 Problem Statement		1
1.2 Motivation.....		1-2
1.3 Objectives.....		2
1.4 Scope of the Project.....		3
Chapter 2. Literature Survey.....		4-6
Chapter 3. Proposed Methodology.....		7-10
Chapter 4. Implementation and Results		11
Chapter 5. Discussion and Conclusion		12-13
References.....		14

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	The above image shows the block diagram of the entire project	7
Figure 2	The above image shows the web page where users can upload resumes and job descriptions.	11
Figure 3	This above image will shows resume rankings, marks score and AI suggestion of uploaded resume.	11

LIST OF TABLES

Table. No.	Table Caption	Page No.
1	Hardware Requirements	10
2	Software Requirements	10

CHAPTER 1

Introduction

1.1 Problem Statement:

Problem Description and Significance

The problem being addressed is the **manual screening of resumes in the recruitment process**, which is time-consuming and inefficient. Organizations receive an overwhelming number of resumes for each job opening, making it difficult to identify the most suitable candidates quickly. The manual process is **labor-intensive** and **prone to human bias**, leading to potential inefficiencies in hiring the right talent.

This problem is significant because inefficient resume screening can **delay hiring decisions, increase recruitment costs, and reduce productivity**. Automating this process using **machine learning (ML) and natural language processing (NLP)** techniques can enhance the efficiency and effectiveness of recruitment. By leveraging AI-based resume ranking systems, organizations can streamline candidate selection, **improve hiring accuracy**, and **save valuable time and resources**.

The AI Resume Screening & Candidate Ranking System aims to address this challenge by providing an **automated, accurate, and unbiased** approach to resume evaluation, ensuring that recruiters focus on the most relevant candidates.

1.2 Motivation:

In today's fast-paced job market, recruiters face the overwhelming task of manually screening thousands of resumes, often leading to delays, inefficiencies, and overlooked talent. This project was chosen to **revolutionize hiring** by leveraging **AI, NLP, and machine learning** to automate resume screening and ranking. By integrating **TF-IDF and cosine similarity**, the system ensures **accurate, unbiased, and data-driven** candidate selection.

With applications in **corporate hiring, job portals, universities, and large-scale recruitment**, this technology **empowers organizations** to find the best talent faster and more efficiently. It enhances **fairness, scalability, and decision-making** while **reducing hiring costs and bias**.

This AI-driven system is not just a tool—it's a **game-changer** that **transforms recruitment into a smarter, faster, and more effective process**, allowing recruiters to focus on what truly matters: **hiring the right people to drive success!**

1.3 Objective:

The primary objective of this project is to **develop an AI-powered resume screening and candidate ranking system** that enhances the efficiency and accuracy of the recruitment process. The key objectives include:

1. **Automate Resume Screening** – Reduce the time and effort spent manually reviewing resumes by implementing **machine learning and NLP techniques**.
2. **Improve Candidate Ranking** – Utilize **TF-IDF and cosine similarity** to accurately assess and rank resumes based on job descriptions.
3. **Enhance User Experience** – Design a **streamlined and interactive web-based interface** for easy resume uploads, job description inputs, and result visualization.
4. **Ensure Scalability & Adaptability** – Implement advanced **NLP models (BERT/GPT) and structured data extraction** for improved contextual understanding.
5. **Increase Fairness & Accuracy** – Minimize human bias and errors in recruitment by using **data-driven decision-making**.

This project aims to **empower recruiters with AI-driven insights, making hiring faster, fairer, and more efficient**.

1.4 Scope of the Project:

Scope and Limitations

Scope:

This project focuses on **automating the resume screening and candidate ranking process** using AI techniques to improve hiring efficiency. The system will:

- Utilize **machine learning (TF-IDF, cosine similarity) and NLP** to analyze and rank resumes.
- Provide a **web-based platform using Streamlit** for easy interaction.
- Support **multi-language processing** and integration of **advanced models (BERT/GPT)** for contextual understanding.
- Implement **structured data extraction** to assess candidates based on qualifications and skills.
- Ensure **data security** and **scalability** for wider adoption in recruitment processes.

Limitations:

- Accuracy depends on the **quality and structure of resumes** provided.
- **Advanced NLP models require high computational power**, impacting response time.
- **Customization is limited**—specific industry/job role requirements may need manual tuning.
- **Initial deployment on Streamlit Cloud** may have **scalability constraints** for large organizations.

Despite these limitations, the project offers a **powerful AI-driven solution** to streamline hiring processes efficiently.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

Review of Relevant Literature and Previous Work

The AI Resume Screening & Candidate Ranking System builds on prior advancements in machine learning (ML), natural language processing (NLP), and automated recruitment technologies. Previous work in this domain includes:

1. TF-IDF and Cosine Similarity for Text Matching
 - The project leverages TF-IDF (Term Frequency-Inverse Document Frequency), a widely used NLP technique for text relevance scoring.
 - Cosine similarity is commonly employed in information retrieval and recommendation systems to measure text similarity, which is crucial for ranking resumes based on job descriptions.
2. Deep Learning & NLP in Recruitment
 - Research in deep learning models like BERT and GPT has shown significant improvements in semantic understanding and context-based ranking, making AI-driven resume screening more accurate.
 - Many applicant tracking systems (ATS) rely on keyword-based filtering, but modern AI techniques enable more nuanced and intelligent matching.
3. Existing Automated Hiring Systems
 - Companies like LinkedIn, Indeed, and HireVue integrate AI to screen resumes, predict candidate success, and automate hiring decisions.
 - However, challenges such as bias in AI models, data privacy concerns, and interpretability remain key areas for improvement.

How This Project Stands Out

Unlike traditional ATS, this project enhances ranking accuracy using ML & NLP, integrates a user-friendly Streamlit interface, and explores future improvements with advanced AI models. This ensures efficient, fair, and scalable hiring solutions for modern recruitment challenges.

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

Existing Models, Techniques, and Methodologies

The AI Resume Screening & Candidate Ranking System leverages various existing models and techniques in machine learning (ML) and natural language processing (NLP) to enhance recruitment efficiency. Some of the key methodologies include:

1. TF-IDF (Term Frequency-Inverse Document Frequency)
 - A well-established NLP technique used to determine the importance of words in a document relative to a collection of documents.
 - Helps in ranking resumes based on the relevance of keywords compared to job descriptions.
2. Cosine Similarity
 - A mathematical approach to measure the similarity between two text documents.
 - Used to compare resumes with job descriptions and rank candidates accordingly.
3. Word Embeddings (Word2Vec, GloVe, BERT)
 - Advanced NLP models such as Word2Vec and GloVe capture contextual meanings of words.
 - BERT (Bidirectional Encoder Representations from Transformers) enhances deep understanding of job descriptions and resumes.
4. Applicant Tracking Systems (ATS)
 - Many ATS platforms rely on rule-based keyword matching, which often leads to false positives/negatives.
 - Modern AI-powered ATS integrate machine learning algorithms for better candidate ranking.
5. Machine Learning Classification & Ranking Models
 - Logistic Regression, Random Forest, and Neural Networks are widely used for candidate scoring and selection.
 - AI-based ranking systems learn from historical hiring data to refine recommendations.

This project combines these methodologies to create a more intelligent, automated, and fair resume screening system, reducing manual effort and improving hiring accuracy.

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

Gaps in Existing Solutions & How This Project Addresses Them

Current resume screening solutions, including traditional Applicant Tracking Systems (ATS), have significant limitations:

1. **Keyword Dependency & False Positives**
 - Many ATS rely on simple keyword matching, often rejecting qualified candidates due to missing exact keywords or ranking irrelevant ones higher.
 - Our project leverages NLP and semantic analysis to understand the true meaning of resumes beyond just keywords.
2. **Lack of Contextual Understanding**
 - Existing models struggle with understanding industry-specific jargon, synonyms, and nuanced experience levels.
 - Our AI-powered system incorporates deep learning (BERT, Word2Vec) to comprehend context and improve candidate relevance.
3. **Bias & Fairness Issues**
 - Traditional systems may unintentionally favour certain demographics or specific resume formats.
 - Our model is designed to mitigate biases by using unbiased training data and explainable AI techniques.
4. **Manual Effort & Time-Consumption**
 - HR teams still spend hours reviewing shortlisted resumes manually due to inaccuracies in screening tools.
 - Our solution automates ranking and provides human-like scoring, significantly reducing hiring time.

By addressing these gaps, our project enhances efficiency, accuracy, and fairness in AI-driven resume screening, transforming modern recruitment.

CHAPTER 3

Proposed Methodology

3.1 System Design

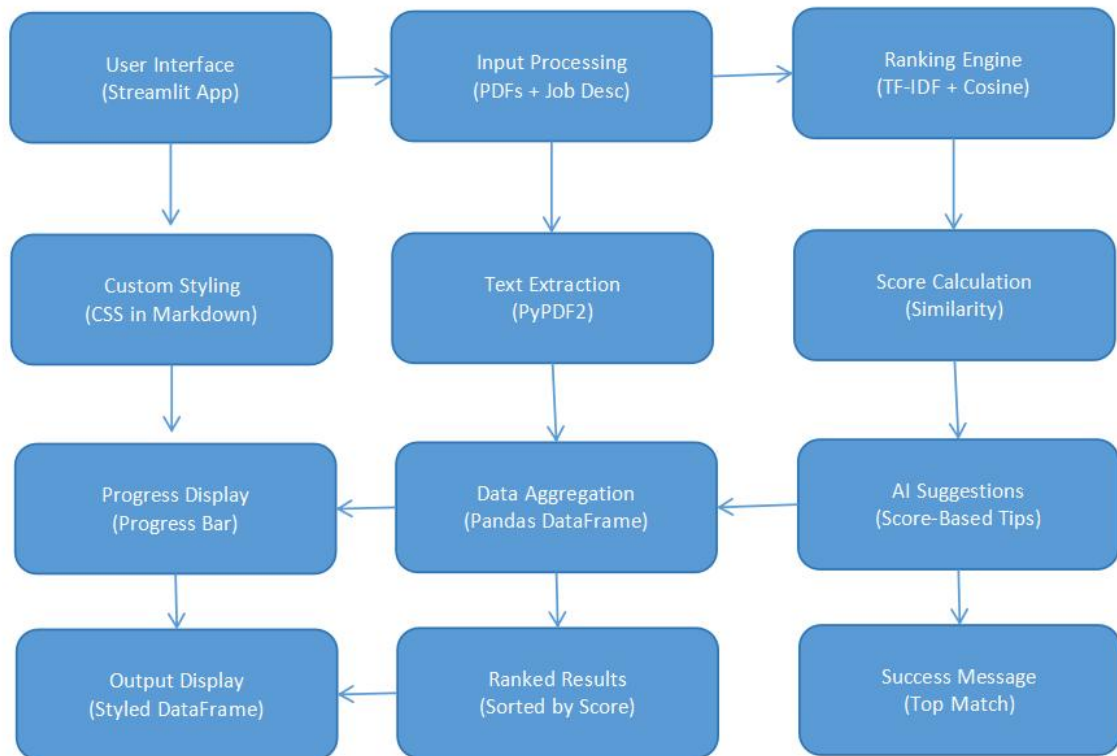


Figure1 : The above image shows the block diagram of the entire project

Detailed Explanation of the Diagram

1. User Interface (Streamlit App)

- Purpose: This is the entry point where users interact with the system.
- Components:
 - ❖ Upload Resumes: A file uploader (via `st.file_uploader`) allows users to upload multiple PDF resumes.
 - ❖ Job Description Input: A text area (via `st.text_area`) collects the job description.

- Flow: Users provide inputs (PDFs and job description), which are passed to the next stage.

- Code Reference: `uploaded_files` and `job_description` variables in the `col1` and `col2` layout.

2. Custom Styling (CSS in Markdown)

- Purpose: Enhances the visual appeal of the UI.
- Details: Custom CSS is injected via `st.markdown` with styles for background, buttons, text areas, and data frames (e.g., green shadows, rounded corners).
- Flow: Applied to the UI elements for a polished look; doesn't affect data processing but improves user experience.

3. Input Processing (PDFs + Job Description)

- Purpose: Prepares the raw inputs for analysis.
- Components:
 - ❖ Text Extraction (PyPDF2): The `extract_text_from_pdf` function reads each uploaded PDF and extracts text using `PdfReader`.
 - ❖ Job Description: The text entered by the user is kept as-is.
- Flow: PDFs are converted to text strings, forming a list of resumes (resumes), alongside the job description.

4. Ranking Engine (TF-IDF + Cosine Similarity)

- Purpose: Computes how well each resume matches the job description.
- Components:
 - ❖ Score Calculation (Similarity): The `rank_resumes` function uses `TfidfVectorizer` to convert text into vectors and `cosine_similarity` to calculate match scores between the job description and resumes.
- Flow: Combines the job description and resumes into a document list, vectorizes them, and computes similarity scores (scores).

5. Progress Display (Progress Bar)

- Purpose: Provides visual feedback during processing.
- Details: A progress bar (`st.progress`) simulates processing time with a 2-second delay (100 steps).
- Flow: Runs after inputs are validated but before results are displayed.

6. Data Aggregation (Pandas DataFrame)

- Purpose: Organizes results into a structured format.
- Details: Creates a `results_df` DataFrame with columns for resume names, match scores (as percentages), and AI suggestions.
- Flow: Scores are multiplied by 100, rounded, and sorted in descending order.

7. AI Suggestions (Score-Based Tips)

- Purpose: Offers improvement tips based on match scores.
- Details: The `generate_resume_tips` function evaluates scores and returns tailored suggestions (e.g., “Excellent match” for >80%).
- Flow: Applied to each score and included in the DataFrame.

8. Output Display (Styled DataFrame)

- Purpose: Presents the ranked results to the user.
- Details: The DataFrame is styled with formatting (e.g., two decimal places for scores) and displayed using `st.dataframe`.
- Flow: Receives aggregated data and renders it with CSS styling from earlier.

9. Success Message (Top Match)

Purpose: Highlights the best-matching resume.

Details: A success message (`st.success`) shows the top resume’s filename.

Flow: Displayed after ranking, pulling the top entry from `results_df`.

3.2 Requirement Specification

To implement the ai-powered resume screening and ranking system, the following tools and technologies are required. These are divided into Hardware Requirements and Software Requirements:

3.2.1 Hardware Requirements:

Table 1

Component	Specifications
Processor	Intel i5 or higher (or equivalent AMD processor)
RAM	Minimum 8 GB (16 GB recommended for faster training)
Storage	SSD with at least 256 GB of free space
GPU	NVIDIA GTX 1060 or higher with CUDA support (recommended for deep learning)
Cloud GPU (Optional)	Google Co lab, AWS EC2, or other cloud-based GPU services

3.2.2 Software Requirements:

Table 2

Category	Tools/Technologies
Programming Languages	Python
PDF Processing	PyPDF2
Data Manipulation	Numpy, Pandas
Data Visualization	Matplotlib
Web Frameworks	Streamlit
Development Tools	Jupyter Notebook, PyCharm, VS Code, Git/Github

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

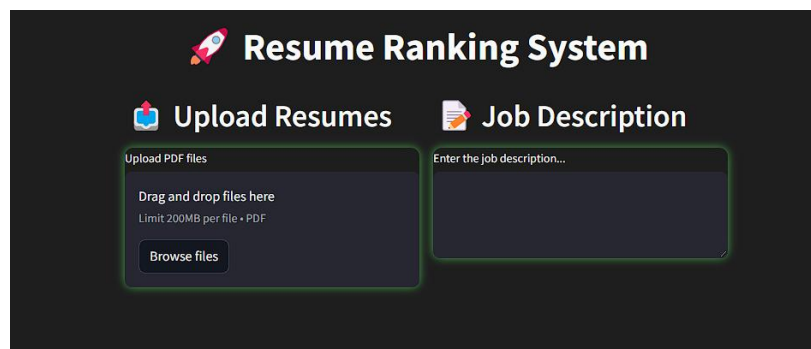


Figure2 : The above image shows the web page where users can upload resumes and job descriptions.

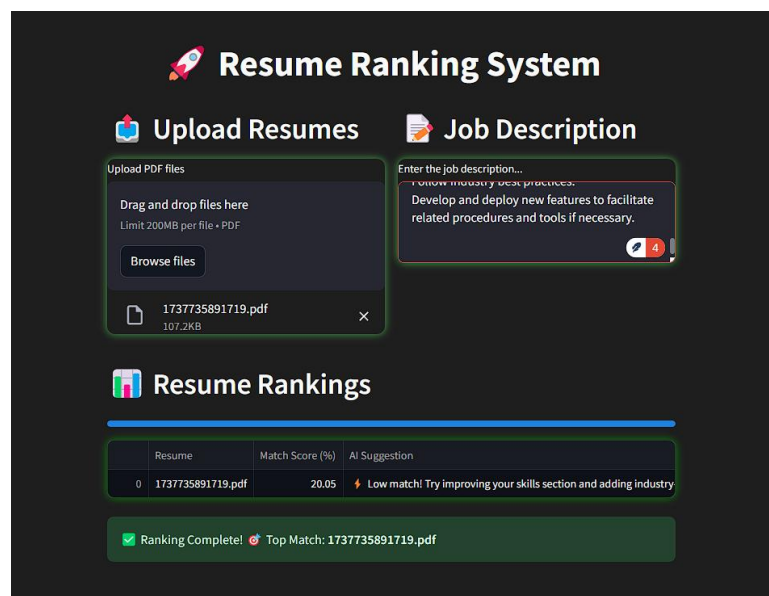


Figure3 : This above image will shows resume rankings, marks score and AI suggestion of uploaded resume.

4.2 GitHub Link for Code: [Github Link](#)

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- **Model:** Upgrade to contextual embeddings (e.g., BERT) and weight key skills.
- **UX:** Add detailed AI tips, visualizations, and a realistic progress bar.
- **Robustness:** Handle scanned PDFs and empty inputs.
- **Scalability:** Batch processing and caching for large datasets.
- **Future Work:** Multi-language support, formatting insights, and external data integration.

5.2 Conclusion:

The Resume Ranking System built with Streamlit, as presented in your code, represents a practical and innovative tool designed to streamline the hiring process by automating resume evaluation. Below is a summary of its overall impact and contribution:

Overall Impact

➤ Efficiency in Recruitment:

The system significantly reduces the time and effort required to manually screen resumes by providing an automated, data-driven ranking based on relevance to a job description. Recruiters can quickly identify top candidates without sifting through large volumes of applications.

➤ Objective Decision-Making:

By leveraging TF-IDF and cosine similarity, the system introduces a level of objectivity to resume evaluation, minimizing human biases that might arise from subjective interpretations of candidate qualifications.

➤ Accessibility for Non-Technical Users:

The intuitive Streamlit interface, enhanced with custom CSS for a polished UI, makes the tool accessible to HR professionals or hiring managers without requiring technical expertise, broadening its usability.

➤ Candidate Support:

The AI-generated suggestions offer candidates actionable feedback to improve their resumes, empowering them to align their applications more closely with job requirements.

➤ Scalability Potential:

While currently suited for small-to-medium resume batches, the system lays the groundwork for scalable recruitment solutions, with potential to handle larger datasets or integrate with enterprise hiring platforms.

The Resume Ranking System delivers a meaningful contribution to the hiring ecosystem by automating and optimizing resume screening with a user-centric design. Its impact is felt in time savings, improved objectivity, and candidate empowerment, while its modular design ensures it can evolve into a more robust tool with advanced features. This project exemplifies how AI and data science can be practically applied to enhance human workflows, particularly in HR and talent acquisition.

REFERENCES

1. <https://medium.com/@bhatadithya54764118/day-96-ai-powered-resume-screening-system-automating-hiring-with-machine-learning-f061803ccc70>
2. <https://ieeexplore.ieee.org/document/10627825>
3. https://ijirt.org/publishedpaper/IJIRT164606_PAPER.pdf