

## **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

**Adityan Nair, [adityannair007@gmail.com](mailto:adityannair007@gmail.com)**

Under the Guidance of

**Soumya Chaudhury for P1**

## 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 my supervisor, Soumya Chaudhury for being a great mentor and the best adviser I could ever have. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional.

We would also like to express our sincere appreciation to Microsoft and SAP's joint CSR initiative, TechSaksham, for providing us with the platform, resources, and learning opportunities that have greatly enriched our knowledge and practical skills. The initiative has guided us in equipping with industry-relevant technical expertise, which has been instrumental in shaping this project. The exposure, training sessions, and mentorship provided through TechSaksham have empowered us with valuable insights and hands-on experience to bridge the gap between academia and industry.

## ABSTRACT

---

This project presents an AI-powered resume screening and ranking system designed to streamline the recruitment process by automating the evaluation of resumes against job descriptions. Traditional resume screening is time-consuming and often subjective, leading to inefficiencies in hiring. Our solution leverages **Natural Language Processing (NLP)** techniques, specifically **TF-IDF** and **BERT-based embeddings**, to compute the similarity between resumes and job descriptions. The system extracts text from resumes in various formats (PDF, DOCX, TXT, CSV) and compares their relevance using **cosine similarity**. The tool, developed with **Streamlit**, enables recruiters to upload multiple resumes and receive ranked results.

Key results show that **BERT embeddings** outperform TF-IDF in capturing the context and meaning of words, leading to better ranking accuracy. Future improvements include **OCR integration for scanned resumes** and **industry-specific keyword matching**. This project bridges the gap between manual and AI-driven resume screening, increasing efficiency and reducing hiring bias.

## TABLE OF CONTENT

---

<b>Abstract</b>	<b>I</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Problem Statement	1
1.2 Motivation	1
1.3 Objectives	1
1.4. Scope of the Project	2
<b>Chapter 2. Literature Survey</b>	<b>3</b>
<b>Chapter 3. Proposed Methodology</b>	<b>6</b>
<b>Chapter 4. Implementation and Results</b>	<b>7</b>
<b>Chapter 5. Discussion and Conclusion</b>	<b>9</b>
<b>References</b>	<b>10</b>

## LIST OF FIGURES

<b>Figure No.</b>	<b>Figure Caption</b>	<b>Page No.</b>
<b>Figure 1</b>	Main Screen	<b>7</b>
<b>Figure 2</b>	Job Description inserted and Resumes imported	<b>8</b>
<b>Figure 3</b>	ATS score generated for resumes	<b>8</b>
<b>Figure 4</b>		
<b>Figure 5</b>		
<b>Figure 6</b>		
<b>Figure 7</b>		
<b>Figure 8</b>		
<b>Figure 9</b>		

## LIST OF TABLES

[illegible]

# CHAPTER 1

## Introduction

### 1.1 Problem Statement:

The hiring process is often burdened by the manual screening of large volumes of resumes, which is time-consuming and susceptible to human biases. Traditional keyword-based filtering methods fail to capture the contextual relevance of skills and experience, leading to suboptimal hiring decisions. An AI-powered solution can **automate resume screening**, improve accuracy, and enhance recruiter productivity.

### 1.2 Motivation:

The demand for AI-driven HR solutions is increasing, especially in large organizations that receive thousands of resumes per job opening. Implementing an intelligent ranking system reduces recruitment costs and ensures fair candidate evaluation. The use of NLP-based models like **BERT** helps in understanding resume content beyond simple keyword matching, offering better hiring decisions.

### 1.3 Objective:

- To develop an AI-powered resume screening system that ranks resumes based on relevance to a job description.
- To implement **TF-IDF and BERT-based models** for text similarity comparison.
- To provide a **user-friendly interface using Streamlit** for easy resume upload and ranking visualization.
- To explore **OCR-based text extraction** for scanned resumes.

## 1.4 Scope of the Project:

Define the scope and limitations.

- The system supports **PDF, DOCX, TXT, and CSV formats**.
- Focuses on **semantic similarity** rather than just keyword matching.
- Uses **Streamlit** for easy deployment and recruiter interaction.
- Currently, the model is **generalized across industries**, with future work including industry-specific tuning.



## CHAPTER 2

### Literature Survey

#### 2.1 Review relevant literature or previous work in this domain.

##### Traditional Resume Screening Methods

**Historically, resume screening has been a manual process, requiring HR professionals to read and evaluate each resume individually. This approach is not only time-consuming but also prone to biases and errors. Some traditional methods include:**

1. **Keyword Matching:** Early automated systems relied on simple keyword searches to filter resumes, which often led to false positives and false negatives due to a lack of contextual understanding.
2. **Boolean Search Techniques:** Recruiters used logical operators (AND, OR, NOT) to refine search queries, but these methods still lacked the ability to assess semantic meaning.
3. **Rule-Based Expert Systems:** Some companies implemented rule-based filters that classified resumes based on predefined rules, but these systems were rigid and required frequent updates.

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

##### Existing Resume Screening Systems

Several AI-driven resume screening tools exist in the market today, including:

- **LinkedIn Recruiter:** Uses AI algorithms to match job descriptions with candidate profiles.
- **HireVue AI:** Integrates NLP and ML techniques to analyze resume content.
- **IBM Watson Recruitment:** Leverages AI and cognitive computing to screen resumes and predict candidate suitability.

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

**Despite advancements in AI-powered screening, several gaps remain:**

- Existing systems rely heavily on keyword matching, leading to poor ranking accuracy.
- Many tools struggle with scanned PDFs, requiring OCR capabilities.
- Lack of industry-specific tuning makes general-purpose AI less effective for niche job roles.
- Bias in AI models remains an issue, requiring ongoing improvements in fair recruitment techniques.

**Gaps in existing solutions:**

- Traditional keyword-based filtering lacks contextual meaning.
- Many systems do not support multiple file formats.
- Scanned resumes require OCR-based extraction.

**Our project addresses these gaps by:**

- Combining TF-IDF with BERT embeddings to capture both keyword relevance and contextual meaning.
- Integrating OCR capabilities to extract text from scanned resumes.
- Exploring industry-specific keyword matching for better accuracy in specialized job domains.
- Enhancing ranking logic to reduce bias and improve fair evaluation.

**Several NLP-based approaches have been proposed for resume ranking:**

- TF-IDF (Term Frequency-Inverse Document Frequency): Used for text vectorization but lacks contextual understanding.
- Word2Vec & FastText: Capture word relationships but are limited in sentence-level comprehension.
- BERT (Bidirectional Encoder Representations from Transformers): Provides deep contextualized embeddings, significantly improving NLP tasks.

**Our system addresses these gaps by integrating TF-IDF for basic matching and BERT for contextual ranking, along with OCR support for scanned PDFs.**

---

## CHAPTER 3

### Proposed Methodology

#### System Design

The system consists of the following key components:

1. **Resume Extraction Module** – Uses pdfplumber, mammoth, and pandas to extract text.
2. **Preprocessing Module** – Removes stopwords and tokenizes text.
3. **Similarity Computation Module** – Uses TF-IDF and BERT embeddings to compute similarity.
4. **Ranking & Display Module** – Uses cosine similarity to rank resumes and display them via Streamlit.

id	title	description
integer primary key auto-incremented	text not null	text not null

Table 1. Job description

name	extracted_text	job_id
text not null	text not null	integer

Table 2. Resumes

#### Requirement Specification

##### Hardware Requirements

- **CPU/GPU-enabled system (MPS for Mac, CUDA for Nvidia GPUs)**
- **Minimum 8GB RAM**
- **Storage: ~5GB for model storage and logs**

### **Software Requirements**

- **Python 3.8+**
- **Libraries: streamlit, pandas, pdfplumber, mammoth, sentence-transformers, sklearn**

## **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.

Deploy

## AI-Powered Resume Screening System

### Add Job Description

Job Title

Enter the Job Description

Save Job Description

### Select Job Description for Screening

Select a Job

No options to select.

### Upload Resumes

Main Screen

Deploy

## AI-Powered Resume Screening System

### Add Job Description

Job Title

Senior Software Developer

Enter the Job Description

✓ Excellent communication and teamwork skills.

Preferred Skills

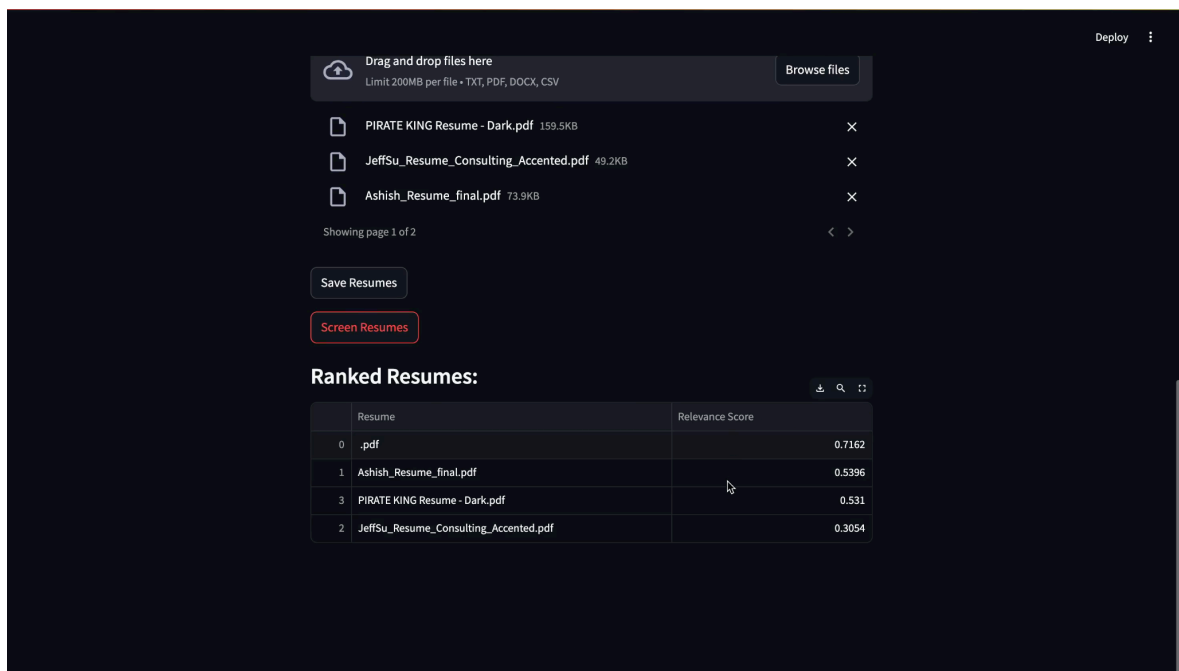
- Experience with TypeScript and GraphQL.
- Knowledge of containerization and microservices architecture.
- Familiarity with automated testing frameworks (Jest, Cypress).

Benefits

- 🏆 Competitive salary + annual performance bonuses
- 🏠 Comprehensive health, dental, and vision insurance
- 🌳 Generous PTO + paid holidays
- 🏡 Flexible work environment (Hybrid/Remote options)
- 📚 Professional development & learning opportunities

Save Job Description

Job Description inserted and Resumes imported



The ATS Score generated for the resumes

## 4.2 GitHub Link for Code:

<https://github.com/adityannair007/AI-powered-Resume-Screening-and-Ranking-System/tree/main>

# CHAPTER 5

## Discussion and Conclusion

### 5.1 Future Work:

While the current system successfully automates resume screening and ranking, several enhancements can be made to improve accuracy and usability:

- Integration of OCR for scanned resumes: The current system struggles with scanned documents. Adding Optical Character Recognition (OCR) using pytesseract will enable text extraction from image-based resumes.
- Improved Industry-Specific Customization: Fine-tuning models for specific industries (e.g., IT, healthcare, finance) can improve relevance.
- Bias Mitigation Strategies: Implementing techniques to detect and reduce bias in AI-driven decision-making.
- Enhanced UI and Filtering: Allow recruiters to filter candidates based on experience level, skills, and education.
- Real-Time API for ATS Integration: Developing an API to integrate with Applicant Tracking Systems (ATS) for seamless HR workflow.

## 5.2 Conclusion:

This project presents an AI-powered resume screening and ranking system that enhances recruitment efficiency by automating candidate evaluation. The use of TF-IDF for keyword relevance and BERT embeddings for contextual ranking provides a more effective method of shortlisting resumes. The tool reduces manual effort, minimizes hiring biases, and improves decision-making for recruiters. Future improvements, such as OCR, industry-specific tuning, and bias reduction, will further refine the system and enhance its practical applicability in real-world hiring scenarios.

## REFERENCES



- [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.