# 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

**Abhishek Kushwaha, abhishekkushwaha014@gmail.com**

Under the Guidance of

**Mr. Soumya Choudhary**

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed, directly or indirectly, to the successful completion of this project.

First and foremost, I extend my deepest appreciation to my supervisor, **Mr. Soumya Choudhary**, for his invaluable guidance, continuous support, and insightful feedback. His expertise and encouragement played a crucial role in shaping this project. His unwavering trust in my abilities has been a great source of motivation, driving me to enhance my technical and analytical skills.

Working under his mentorship has been an enriching experience, fostering both academic and professional growth. His influence extended beyond this project, helping me develop a strong foundation in AI and deep learning. I am truly grateful for his mentorship and the learning opportunities he provided throughout this journey.

Name : Abhishek Kushwaha

Email : abhishekkushwaha014@gmail.com

# ABSTRACT

The recruitment process is often labour-intensive, requiring HR professionals to manually review large volumes of resumes. This project introduces an AI-powered Resume Screening and Candidate Ranking System that automates resume evaluation using Natural Language Processing (NLP) techniques. The system extracts text from PDF resumes, applies TF-IDF vectorization, and calculates cosine similarity scores to assess relevance to a given job description. By ranking candidates based on job fit, this solution enhances efficiency, reduces manual effort, and provides recruiters with a data-driven approach to shortlisting top applicants.

# TABLE OF CONTENT

# LIST OF FIGURES

# CHAPTER 1

# Introduction

## 1.1 Problem Statement:

The manual screening of large volumes of resumes poses significant challenges for recruiters and HR professionals, leading to inefficiencies, biases, and delays in the hiring process. Traditional evaluation methods rely heavily on subjective judgment and keyword matching, often overlooking well-qualified candidates. To address these limitations, this project proposes an **AI-powered Resume Screening and Candidate Ranking System** that utilizes **Natural Language Processing (NLP)** techniques. By ranking resumes based on their relevance to a given job description, the system ensures **faster, fairer, and more efficient** candidate shortlisting, reducing manual effort and improving hiring accuracy.

## 1.2 Motivation:

With the increasing number of job applications, manual resume screening has become a tedious and time-intensive task for recruiters. Human bias and inconsistencies in evaluation further complicate the selection process, potentially leading to the loss of qualified candidates. This project leverages **Artificial Intelligence (AI) and Natural Language Processing (NLP)** to automate resume screening, ensuring a **fair, objective, and efficient** hiring process. By reducing human effort, improving accuracy, and accelerating candidate shortlisting, this system benefits both **employers and job seekers**, making recruitment more streamlined and data-driven.

## 1.3 Objective:

The primary objective of this project is to develop an **AI-powered Resume Screening and Candidate Ranking System** that automates resume evaluation using **Natural Language Processing (NLP)** techniques. The key goals include:

- **Efficient Text Extraction:** Extracting text from PDF resumes accurately.
- **NLP-Based Representation:** Implementing **TF-IDF vectorization** to analyze textual data.
- **Similarity Scoring:** Computing **cosine similarity scores** to rank resumes based on relevance.

- **Automated Candidate Ranking:** Providing recruiters with a **sorted list** of resumes based on job fit.
- **Enhanced Hiring Process:** Reducing **manual effort, bias, and processing time** in recruitment.

## 1.4 Scope of the Project:

The project focuses on **automating resume screening** by utilizing NLP techniques to match resumes with job descriptions.

## In-Scope

1. **Resume Text Extraction:** Extracting structured text from PDF resumes.
2. **NLP-Based Analysis:** Processing job descriptions and resumes using **TF-IDF vectorization**.
3. **Candidate Ranking:** Applying **cosine similarity** to rank resumes based on relevance.
4. **Automated Processing:** Eliminating manual screening by generating a **ranked list** of candidates.
5. **User Interface:** Developing a **Streamlit-based dashboard** for recruiters to upload resumes and view rankings.

## Out-of-Scope

1. **OCR for Image-Based Resumes:** The system does not support scanned or image-based resumes.
2. **Soft Skill Evaluation:** It assesses only textual content, not communication skills or personality traits.
3. **Real-Time API Integration:** The system operates as a standalone tool without linking to **job portals or HRM systems**.
4. **Multilingual Resume Processing:** Currently, it only supports **English-language resumes**

This system aims to provide **recruiters with a faster, unbiased, and AI-driven** approach to shortlisting candidates. Future improvements may include **OCR support, advanced candidate profiling, and seamless HR system integration.**

# CHAPTER 2

# Literature Survey

## 2.1 Literature Review

AI-driven resume screening has gained prominence due to the growing volume of job applications. Traditional Applicant Tracking Systems (ATS) rely on keyword matching, which often fails to capture the contextual meaning of resumes. Recent advancements in Natural Language Processing (NLP) have led to more sophisticated resume-job description matching using machine learning models like TF-IDF, Word2Vec, and BERT, improving the accuracy and relevance of candidate selection.

## 2.2 Existing Models, Techniques, and Methodologies

Several techniques have been explored for automated resume screening:

- **Keyword-Based Matching**: Commonly used in ATS but lacks context sensitivity and may overlook qualified candidates.
- **TF-IDF & Cosine Similarity:** Converts resumes into numerical vectors for similarity-based ranking but doesn't capture deep contextual meaning.
- **Word Embeddings (Word2Vec, GloVe):** Capture semantic relationships between words, enhancing resume analysis.
- **Deep Learning Models (BERT, Transformer-based Approaches):** Provide context-aware resume parsing but require high computational power.

## 2.3 Gaps in Existing Solutions and Project Contributions

Despite advancements in AI-based resume screening, existing models have key limitations:

- **Reliance on Exact Keywords:** Many ATS systems overlook relevant resumes due to rigid keyword dependencies.
- **Lack of Context Awareness:** TF-IDF struggles to interpret the deeper meaning behind job descriptions and resumes.
- **Computational Complexity:** Advanced models like BERT demand **high processing power**, limiting accessibility for smaller organizations.

**How This Project Addresses These Gaps:**

- Utilizes **TF-IDF with cosine similarity**, providing an **efficient and scalable** ranking method.
- Reduces reliance on exact keyword matching, ensuring **fairer and more accurate** candidate selection.
- Offers a **lightweight, low-complexity** solution for recruiters without requiring expensive computational resources.

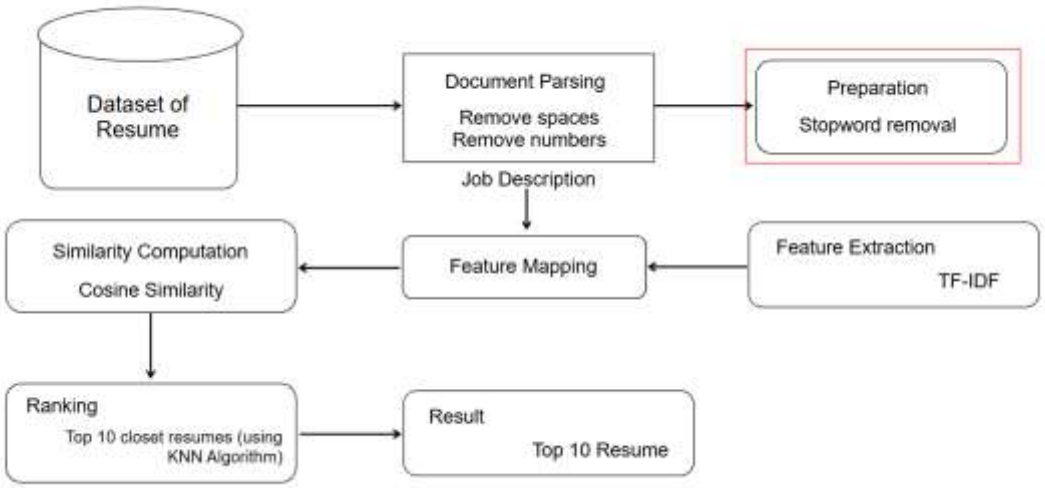
By bridging these gaps, this **AI-powered resume screening system** enhances **efficiency, fairness, and automation** in modern recruitment processes.

# CHAPTER 3

# Proposed Methodology

## 3.1 System Design

The proposed **AI-powered Resume Screening and Candidate Ranking System** follows a structured workflow to automate resume evaluation. The process is illustrated in the system architecture diagram below:




## Explanation of the Diagram:

1. **User Input:** Recruiters enter a **job description** and upload multiple resumes in **PDF format**.
2. **Text Extraction Module:** The system extracts textual content from resumes using **PyPDF2**.
3. **Text Processing:** Resumes and job descriptions are transformed into numerical vectors using **TF-IDF vectorization**.
4. **Similarity Calculation: Cosine similarity** is applied to measure how closely each resume matches the job description.
5. **Ranking Module:** Resumes are ranked based on similarity scores and sorted in **descending order**.

6. **User Interface:** Recruiters interact with the **Streamlit-based UI**, where they can upload resumes and view the ranked results.

## 3.2 Requirement Specification

To implement this project, the following **hardware and software** requirements must be met:

### 3.2.1 Hardware Requirements

- **Processor:** Intel Core i5 or higher
- **RAM:** Minimum 8GB (16GB recommended for larger datasets)
- **Storage:** At least 10GB free space
- **GPU (Optional):** Required for advanced deep learning extensions

### 3.2.2 Software Requirements

- **Programming Language:** Python 3.x
- **Libraries & Frameworks:**
  - **Streamlit** – For UI development
  - **PyPDF2** – For extracting text from PDF files
  - **scikit-learn** – For TF-IDF vectorization and cosine similarity
  - **pandas** – For data manipulation and ranking
- **IDE:** Jupyter Notebook / VS Code / PyCharm
- **OS:** Windows/Linux/macOS

This methodology ensures an efficient, scalable, and AI-powered resume ranking system for recruiters.

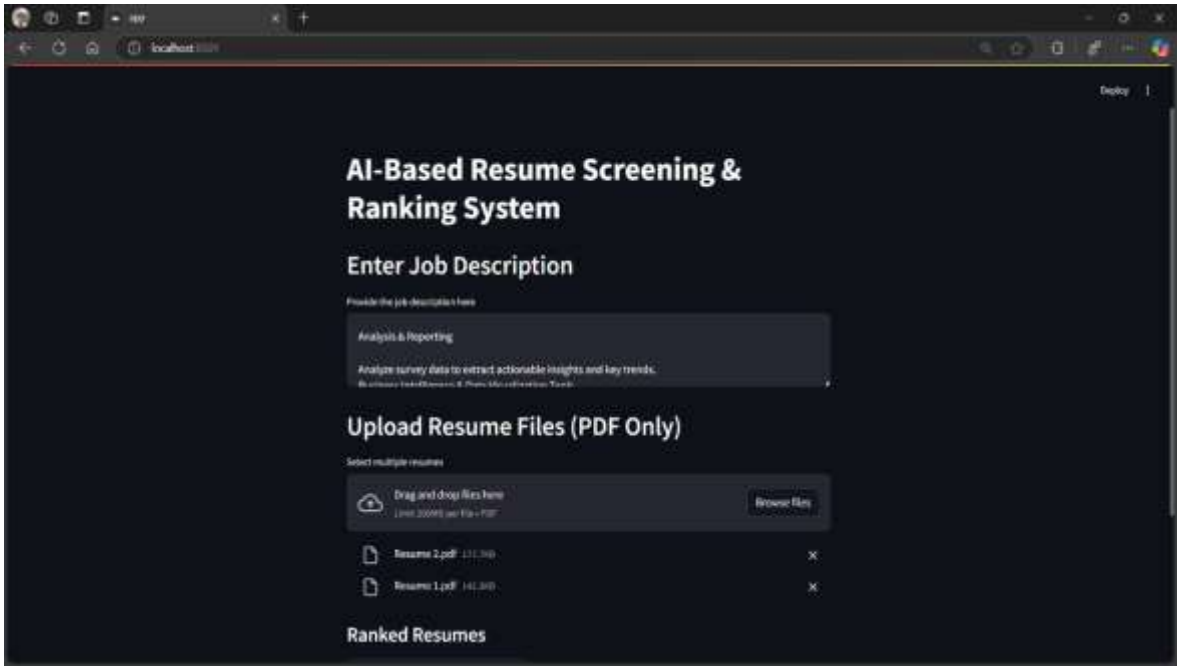
# CHAPTER 4

# Implementation and Result

## 4.1 Snapshots of Results

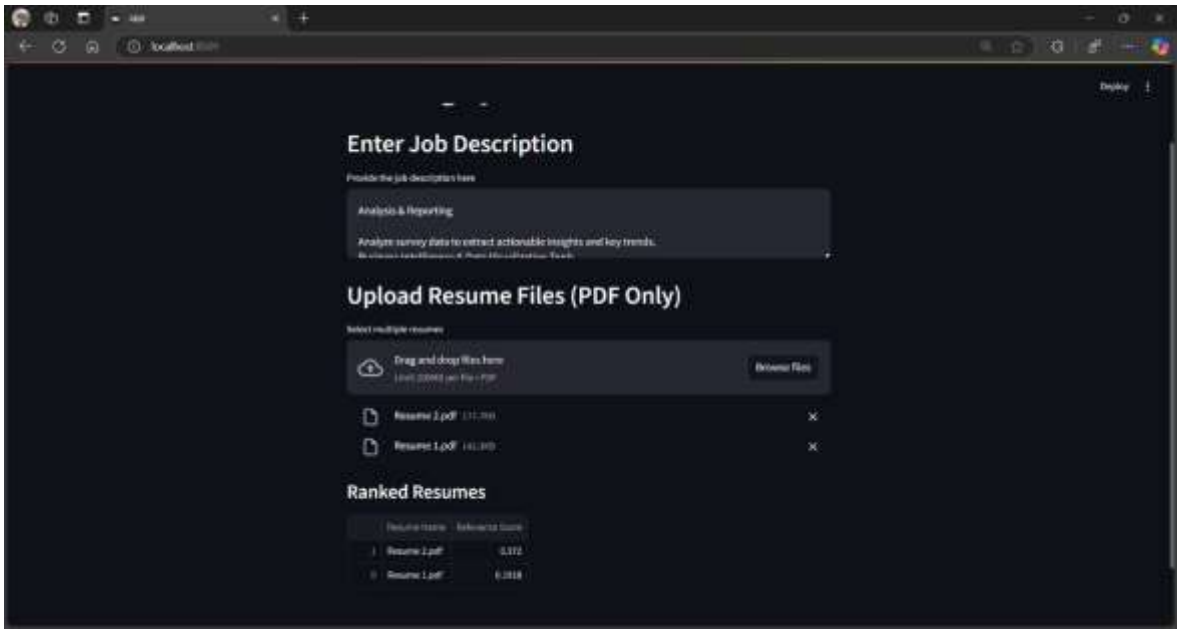
Below are snapshots showcasing the functionality of the AI-powered Resume Screening and Candidate Ranking System:

**Snapshot 1: Job Description Input and Resume Upload Interface**




**Explanation:** This interface allows recruiters to enter the job description in a text box and upload resumes in PDF format for processing.

**Snapshot 2: Resume Ranking Results**




**Explanation:** The system applies **TF-IDF vectorization** and **cosine similarity** to compute the relevance of each resume to the job description. The results are displayed in **descending order**, allowing recruiters to efficiently shortlist the most suitable candidates.

## 4.2 GitHub Link for Code

Repository Link : https://github.com/Abhishekkushwaha108/AI-Powered-Resume-Screening-and-Ranking-System

# CHAPTER 5

# Discussion and Conclusion

## 5.1 Future Work

Although the current system effectively ranks resumes based on job descriptions, several enhancements can further improve its accuracy and usability:

- **Incorporating Deep Learning Models:** Implementing BERT or GPT-based models for better contextual understanding of resumes.
- **OCR Support for Image-Based Resumes:** Adding Optical Character Recognition (OCR) to process scanned or image-based resumes.
- **Handling Multiple Languages:** Extending support for multilingual resume processing to widen accessibility.
- **Integration with HR Systems:** Connecting the tool with applicant tracking systems (ATS) for real-world application.
- **Soft Skill & Sentiment Analysis:** Including NLP-based soft skill assessment to provide deeper insights into candidates.

## 5.2 Conclusion

This project presents an AI-driven **Resume Screening and Candidate Ranking System** that automates the initial phase of recruitment. By leveraging **TF-IDF vectorization** and **cosine similarity**, it efficiently ranks resumes based on relevance to a given job description. The system reduces **manual effort, bias, and processing time**, making hiring more data-driven and objective. Future improvements, such as **deep learning models, OCR support, and ATS integration**, can further enhance its effectiveness in real-world recruitment scenarios.

# REFERENCES

[1]. Priyanka, J.H., Parveen, N. DeepSkillNER: An automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach. Multimed Tools Appl 83, 47503–47530 (2024).