

# **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

**Arpita.S.Biradar**

<mailto:arpitabiradar30@gmail.com>

Under the Guidance of

**Prof.Choudhary**

## ACKNOWLEDGEMENT

---

We are incredibly grateful to everyone who supported us, both directly and indirectly, throughout this thesis journey.

First and foremost, we extend our heartfelt thanks to our supervisor Choudhary Sir, for being an exceptional mentor and guide. His unwavering support, insightful advice, and constructive feedback have been instrumental in shaping this project. His confidence in me has been a constant source of motivation, pushing me to think creatively and stay committed to my work. Over the past year, it has been an absolute privilege to learn from him—not just about this project, but also about the broader aspects of our field. His guidance has not only helped me navigate challenges but has also shaped me into a more capable and responsible professional.

## ABSTRACT

---

Recruiting the right candidates is a time-consuming and challenging task, especially when dealing with a large volume of resumes. Traditional screening methods often require significant manual effort, leading to inefficiencies and potential biases. To address this, we developed an AI-powered Resume Screening and Ranking System—a web-based application that automates the process of categorizing resumes based on their content.

This system utilizes machine learning to analyze text extracted from resumes in various formats, including PDF, DOCX, and TXT. Through advanced text preprocessing and classification techniques, it predicts the most suitable job category for each resume. Built using Python and Streamlit, the application offers a user-friendly interface for recruiters and HR professionals, helping them streamline their hiring process.

Our methodology involves text extraction, data cleaning, feature extraction using techniques like TF-IDF, and model training with machine learning algorithms. The system was evaluated using a diverse dataset, and the results demonstrate its effectiveness in accurately classifying resumes into relevant job categories.

In conclusion, this AI-powered tool significantly reduces the manual effort involved in resume screening, enhances efficiency, and minimizes human biases in the hiring process. By automating initial resume filtering, recruiters can focus on the most promising candidates, leading to a more efficient and fair hiring experience.

## 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	2
1.3 Objectives	3
1.4 Scope of the Project	4
<b>Chapter 2. Literature Survey</b>	<b>5</b>
<b>Chapter 3. Proposed Methodology</b>	<b>7</b>
<b>Chapter 4. Implementation and Results</b>	<b>10</b>
<b>Chapter 5. Discussion and Conclusion</b>	<b>12</b>
<b>References</b>	<b>14</b>

## LIST OF FIGURES

Figure No.	Figure Caption	Page No.
<b>Figure 1</b>	Process Flow/System Design	<b>07</b>
<b>Figure 2</b>	Application User-Interface	<b>10</b>
<b>Figure 3</b>	Resume Ranking	<b>10</b>
<b>Figure 4</b>	Resume Word Cloud and Ranking Report	<b>11</b>

## CHAPTER 1

### Introduction

#### 1.1 Problem Statement:

Hiring the right candidate is challenging due to the large volume of resumes organizations receive. Manually screening them is time-consuming, inefficient, and prone to bias. AI and machine learning can automate this process, making it faster and more accurate. Using NLP, resumes can be analyzed and categorized effectively. This improves recruitment efficiency, helping companies find the best talent quickly.

#### Problem Being Addressed

Recruiters and HR professionals receive a massive influx of resumes for every job opening, making it difficult to manually review and shortlist the best candidates. Traditional resume screening methods are not only time-consuming but also prone to human bias, inconsistency, and inefficiencies. As a result, qualified candidates may be overlooked, and the hiring process becomes slower and less effective.

#### Significance of the Problem

1. **Time-Consuming Process** – Reviewing hundreds or even thousands of resumes manually takes significant time and effort, delaying the hiring process.
2. **Human Bias & Inconsistencies** – Personal biases and subjective judgment can affect candidate selection, leading to unfair hiring decisions.
3. **High Hiring Costs** – Inefficient screening increases recruitment costs due to extended hiring cycles and misallocation of HR resources.
4. **Missed Opportunities** – Qualified candidates may be overlooked if their resumes do not immediately stand out to human reviewers.
5. **Scalability Issues** – As organizations grow, manually screening resumes becomes unsustainable, especially for companies with high hiring needs.

## 1.2 Motivation:

The motivation behind this project stems from the growing challenges in modern recruitment. With companies receiving thousands of applications for a single job posting, manually screening resumes is inefficient, time-consuming, and prone to bias. To address this, AI and machine learning can revolutionize the hiring process by automating resume screening, ensuring fairer and faster candidate selection.

### Potential Applications

- **Automated Resume Screening** – Streamlines the hiring process by ranking resumes based on relevance.
- **Bias Reduction in Hiring** – Ensures fair selection by focusing on skills and qualifications rather than human biases.
- **Recruitment for Large Organizations** – Helps HR teams efficiently manage high application volumes.
- **Career Portals & Job Matching Platforms** – Enhances job recommendation systems by intelligently matching candidates with roles.
- **Talent Acquisition for Startups & SMEs** – Reduces hiring costs for small businesses by minimizing manual effort.

### Impact

- **Time & Cost Efficiency** – Significantly reduces the time spent on initial resume screening.
- **Improved Candidate Experience** – Faster processing leads to quicker responses and better engagement.
- **Data-Driven Hiring Decisions** – Provides insights based on data rather than subjective judgment.
- **Scalability for Future Growth** – Can be adapted to various industries and hiring needs.

### 1.3 Objective:

The primary objective of this project is to develop an **AI-powered Resume Screening and Ranking System** that automates and enhances the recruitment process by efficiently analyzing and shortlisting resumes. The system leverages **machine learning and natural language processing (NLP)** to improve hiring efficiency, reduce bias, and ensure fair candidate selection.

#### Specific Objectives:

1. **Automate Resume Screening** – Develop an AI-based system that can extract, analyze, and classify resumes based on job relevance.
2. **Improve Hiring Efficiency** – Reduce the time and effort required for manual resume screening, allowing recruiters to focus on interviewing top candidates.
3. **Enhance Fairness & Reduce Bias** – Use data-driven decision-making to minimize human bias in candidate selection.
4. **Develop a User-Friendly Interface** – Build a simple and intuitive platform that HR professionals can use easily.
5. **Ensure Scalability & Accuracy** – Train machine learning models to handle large datasets while maintaining high accuracy in ranking resumes.
6. **Integrate NLP for Resume Analysis** – Utilize natural language processing techniques to extract meaningful insights from resumes and match candidates effectively.
7. **Provide Data-Driven Insights** – Offer recruiters valuable analytics on candidate qualifications and job-market trends.

By achieving these objectives, the project aims to transform the recruitment process, making it faster, more efficient, and more reliable for organizations.



## 1.4 Scope of the Project:

This project focuses on developing an **AI-powered Resume Screening and Ranking System** that automates the initial stages of recruitment. It aims to enhance efficiency, reduce human bias, and provide data-driven insights for candidate selection.

### Scope:

1. **Resume Parsing & Extraction** – Supports multiple formats (PDF, DOCX, TXT) for extracting candidate information.
2. **Machine Learning-Based Resume Classification** – Uses NLP and AI models to categorize resumes based on job relevance.
3. **Ranking System** – Assigns scores to resumes based on qualifications, skills, and experience.
4. **User-Friendly Web Interface** – Provides recruiters with an easy-to-use platform for uploading resumes and reviewing ranked results.
5. **Bias Reduction** – Uses objective criteria to ensure fair candidate evaluation.
6. **Scalability** – Capable of handling a large number of resumes efficiently.

### Limitations:

1. **Limited to Text-Based Resumes** – Cannot process image-based or handwritten resumes.
2. **Dependent on Training Data** – Accuracy depends on the quality and diversity of the training dataset.
3. **No Real-Time Candidate Interaction** – Does not include interview scheduling or communication with applicants.
4. **Context Understanding Limitations** – May struggle with complex career transitions, non-standard resume formats, or ambiguous job descriptions.
5. **Requires Periodic Updates** – AI models need regular retraining to stay up-to-date with evolving industry trends and job market requirements.

Despite these limitations, the system provides a **significant improvement** over manual screening methods, making recruitment **faster, fairer, and more efficient** for organizations.

## CHAPTER 2

### Literature Survey

A literature survey is essential to understand the existing research, methodologies, and technologies in AI-powered resume screening. It helps identify current solutions, their limitations, and how this project improves upon them.

#### 2.1 Review of Relevant Literature and Previous Work

Several studies and projects have explored AI-based hiring solutions, focusing on resume screening, candidate ranking, and job matching. Research in **Natural Language Processing (NLP)**, **Machine Learning (ML)**, and **Deep Learning (DL)** has significantly contributed to automating recruitment processes.

- **Automated Resume Screening Using NLP & ML** – Prior studies have leveraged NLP techniques such as **TF-IDF**, **Word2Vec**, and **BERT** to analyze resume content and classify candidates based on job requirements.
- **Job-Candidate Matching Algorithms** – Research has examined how AI models predict the best-fit candidate by comparing resume data with job descriptions.
- **Bias Detection in Hiring AI** – Some studies highlight the risk of AI models inheriting biases from training data and propose fairness-improving techniques.

#### 2.2 Existing Models, Techniques, and Methodologies

Several AI-based resume screening systems have been developed, utilizing different methodologies:

- **Rule-Based Systems** – Some early models relied on predefined keyword matching, which lacked flexibility and context understanding.
- **Machine Learning Algorithms** – Popular models like **Naïve Bayes**, **Random Forest**, and **SVM** have been used for resume classification but often require feature engineering.
- **Deep Learning Models** – More recent approaches employ **LSTMs**, **BERT**, and **Transformer models** for advanced resume understanding and classification.
- **ATS (Applicant Tracking Systems)** – Widely used in HR, ATS solutions scan resumes but often struggle with non-traditional formats or nuanced experience descriptions.

## 2.3 Gaps and Limitations in Existing Solutions

Despite advancements, current AI-driven recruitment tools face several limitations:

- **Keyword Dependency** – Many systems rely heavily on keyword matching, leading to the rejection of qualified candidates who use different terminology.
- **Lack of Context Understanding** – Existing models may misinterpret career transitions, skills, or job role variations.
- **Bias in AI Models** – If trained on biased datasets, AI models can inherit discrimination, impacting fair candidate selection.
- **Limited Resume Formats** – Some ATS and AI models struggle with parsing complex or non-standard resume formats.
- **Scalability and Processing Speed** – High volumes of resumes can slow down AI models if not optimized properly.

## How This Project Addresses These Gaps

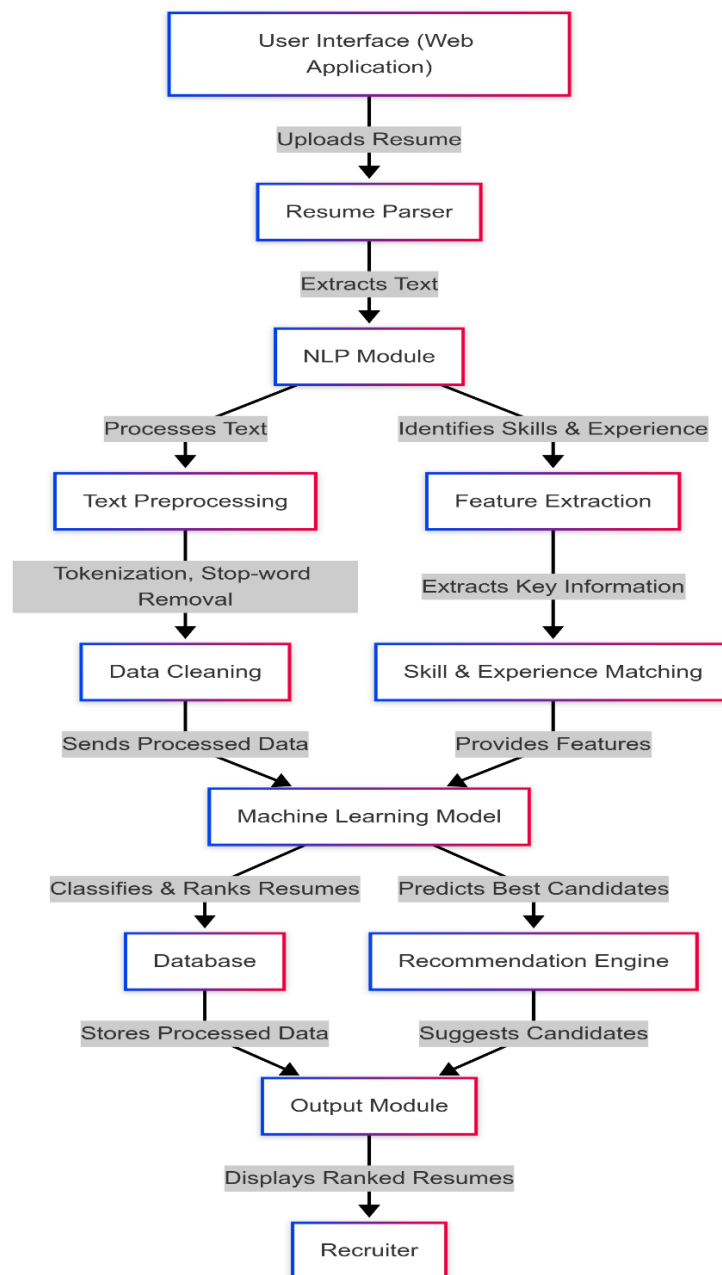
- **Advanced NLP Models** – Uses contextual embeddings (e.g., BERT) for better resume understanding.
- **Bias Reduction Techniques** – Incorporates fairness-aware AI to minimize bias in hiring decisions.
- **Enhanced Resume Parsing** – Supports multiple formats and intelligently extracts key information.
- **Improved Job-Candidate Matching** – Utilizes machine learning to match candidates beyond just keyword comparison.
- **Scalable System** – Designed to handle large datasets efficiently without compromising accuracy.

By addressing these challenges, this project aims to create a **more effective, unbiased, and intelligent AI-powered resume screening system**, improving the recruitment process for organizations.

## CHAPTER 3

### Proposed Methodology

#### 3.1 System Design



**Fig1:Process Flow/System Design**

### 3.2 System Design Explanation

#### 1. User Interface (Web Application)

- Recruiters can upload resumes in multiple formats (PDF, DOCX, TXT).
- The interface allows users to view and manage processed resumes.

#### 2. Resume Parser

- Extracts text data from uploaded resumes.
- Uses Optical Character Recognition (OCR) for scanned documents if necessary.
- Cleans and structures extracted information for further processing.

#### 3. Natural Language Processing (NLP) Module

- Performs text preprocessing (tokenization, stop-word removal, stemming, and lemmatization).
- Identifies and extracts key skills, experience, and relevant information.

#### 4. Machine Learning Model

- Classifies resumes based on job relevance using AI/ML algorithms (e.g., Random Forest, BERT, or TF-IDF with SVM).
- Scores and ranks resumes according to job descriptions.

#### 5. Database (Storage System)

- Stores processed resumes and extracted candidate profiles.
- Maintains a structured dataset for future reference and analytics.

#### 6. Output Module (Ranked Resume Display)

- Presents ranked resumes to recruiters in an easy-to-interpret format.
- Allows sorting, filtering, and downloading candidate details.

This structured system ensures **efficient, unbiased, and accurate** resume screening, reducing manual effort in hiring.

### 3.3 Requirement Specification

This section outlines the necessary tools and technologies required to implement the **AI-powered Resume Screening and Ranking System** effectively.

#### 3.2.1 Hardware Requirements:

- **Processor:** Intel Core i5 or higher (or equivalent AMD Ryzen)
- **RAM:** Minimum 8GB (16GB recommended for better performance)
- **Storage:** Minimum 256GB SSD (Recommended: 512GB SSD for faster processing)
- **GPU (Optional):** NVIDIA GPU (for deep learning-based NLP models, if required)
- **Peripherals:** Standard keyboard, mouse, and monitor
- **Server (if cloud-based):** AWS, Google Cloud, or Azure instance with at least 4 vCPUs and 16GB RAM

#### 3.2.2 Software Requirements:

- **Operating System:** Windows 10/11, macOS, or Linux (Ubuntu recommended)
- **Programming Languages:**
  - Python (Primary Language)
  - JavaScript (for Web Application)
- **Frameworks & Libraries:**
  - Flask/Django (Backend API)
  - React.js or Angular (Frontend)
  - Pandas, NumPy (Data Processing)
  - NLTK, SpaCy, or BERT (Natural Language Processing)
  - Scikit-learn, TensorFlow, or PyTorch (Machine Learning)
- **Database:**
  - PostgreSQL / MySQL (Relational Database)
  - MongoDB (NoSQL for unstructured resume data)
- **Development Tools:**
  - VS Code / PyCharm (Code Editor)
  - Postman (API Testing)
  - Git & GitHub (Version Control)
  - Docker (Containerization)
  - Jupyter Notebook (For ML model development)

## CHAPTER 4

### Implementation and Result

#### 4.1 Snap Shots of Result:

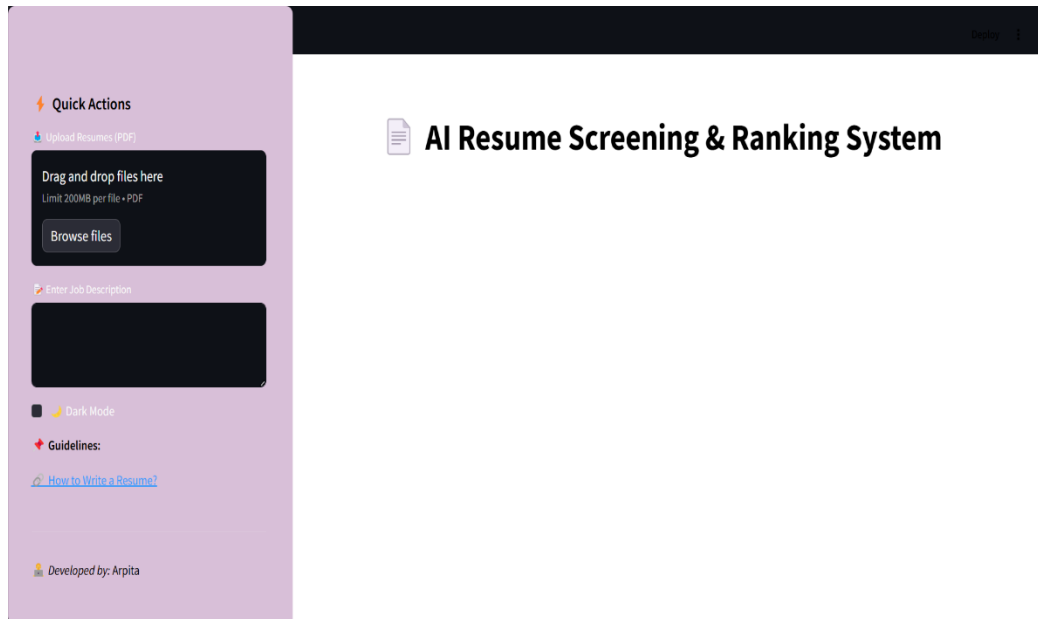


Fig2:Application User-Interface

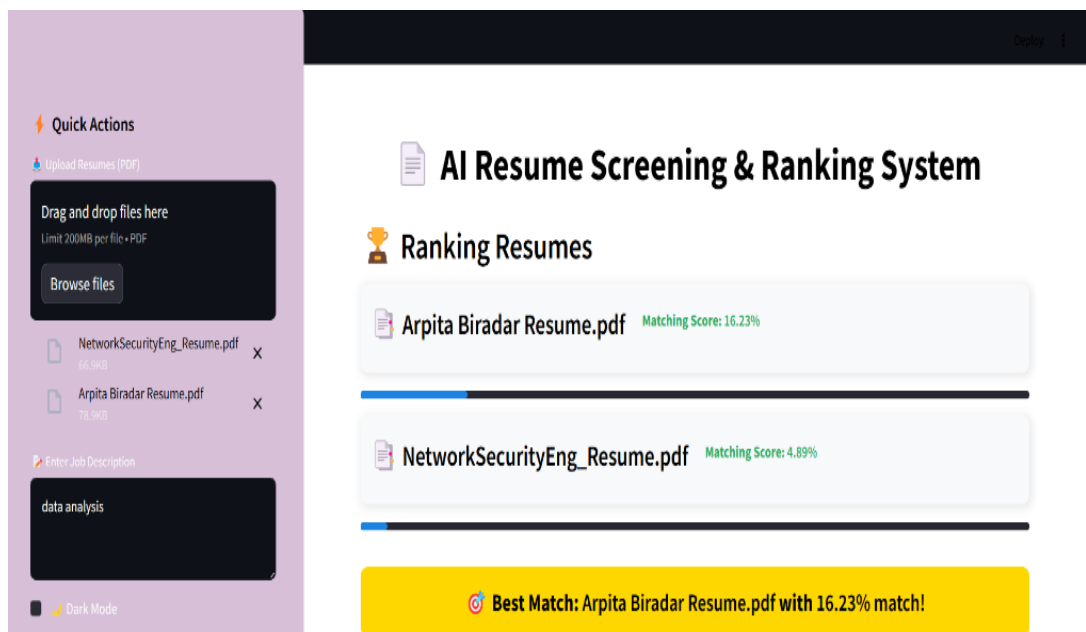
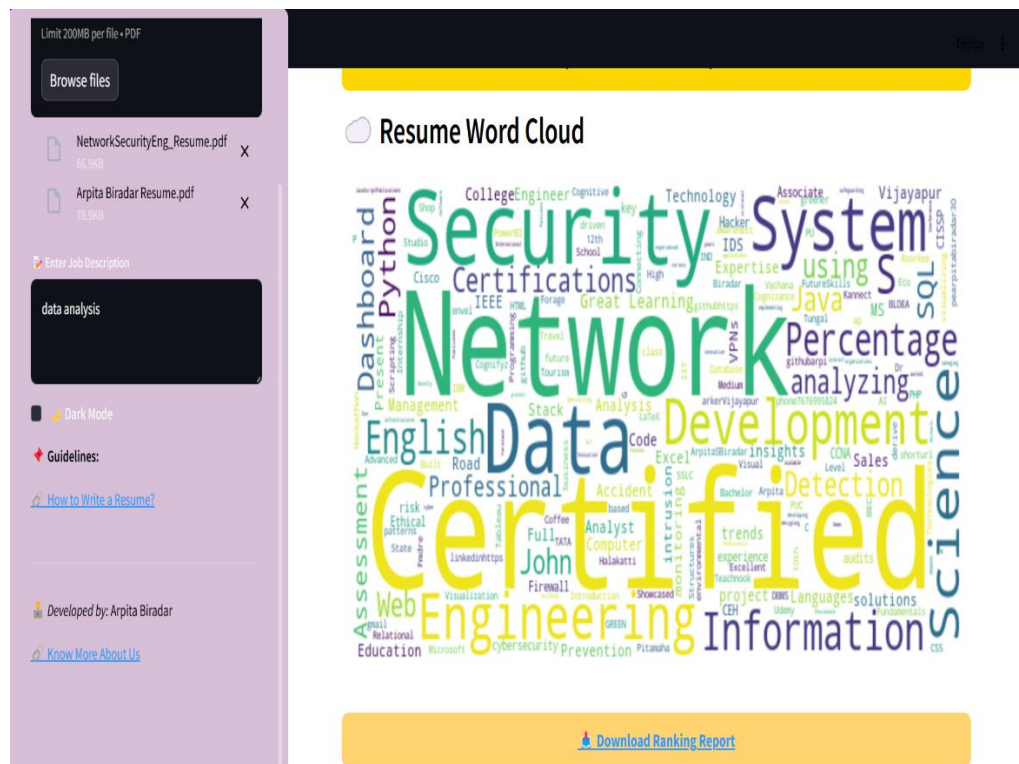


Fig2:Resume Ranking



### Fig3:Resume Word Cloud and Ranking Report

## 4.2 GitHub Link for Code:

[https://github.com/ArpitaSBiradar/AI\\_Resume\\_Screening\\_and\\_Resume\\_Ranking.git](https://github.com/ArpitaSBiradar/AI_Resume_Screening_and_Resume_Ranking.git)



## CHAPTER 5

### Discussion and Conclusion

#### 5.1 Future Work:

While the **AI-powered Resume Screening and Ranking System** provides an efficient way to automate recruitment, there is always room for improvement. Below are some areas for future enhancements:

##### 1. Enhancing Model Accuracy and Bias Reduction

Implement more advanced **Deep Learning models** (e.g., Transformer-based models like GPT or BERT fine-tuning) for better contextual understanding.

Improve **bias detection and mitigation techniques** to ensure fairness in candidate selection.

##### 2. Expanding Data Sources and Resume Formats

Support additional resume formats such as **LinkedIn profiles, online portfolios, and JSON-based resume data**.

Integrate with **ATS (Applicant Tracking Systems)** for seamless resume parsing.

##### 3. Real-time Candidate Matching and Feedback

Develop a **real-time candidate matching feature** where recruiters can instantly see the most relevant candidates for a job.

Provide **feedback loops** where recruiters can rate AI-generated rankings to improve future results.

##### 4. Multi-language Support

Extend **NLP capabilities** to support **multiple languages**, making the system applicable to global hiring.

##### 5. Cloud-Based and Scalable Architecture

Deploy the system using **serverless architecture** (AWS Lambda, Google Cloud Functions) for **scalability and cost-efficiency**.

Implement **distributed computing** for handling large volumes of resumes efficiently.

## 6. Candidate Profile Enrichment

Use **web scraping and social media data analysis** to **enrich candidate profiles** with additional information such as GitHub contributions, Kaggle projects, and publications.

## 7. AI-driven Interview Assistance

Develop an **AI-based interview assistant** to conduct **preliminary interviews** using **chatbots and voice recognition AI**.

## 5.2 Conclusion:

The **AI-powered Resume Screening and Ranking System** significantly enhances the recruitment process by automating **resume analysis, classification, and ranking**. By leveraging **Natural Language Processing (NLP) and Machine Learning (ML)**, the system ensures **efficient, accurate, and unbiased candidate selection**, reducing the time and effort required for manual screening.

This project contributes to **faster hiring, improved decision-making, and fairer recruitment** by eliminating human biases and inconsistencies. Its ability to **integrate with existing HR systems** makes it highly scalable and adaptable for various industries.

Overall, this system **revolutionizes talent acquisition**, enabling recruiters to **focus on the best candidates quickly and effectively**, ultimately **streamlining and improving the hiring process**.

## REFERENCES

### 1. Books & Research Papers:

- Goldberg, Y. (2017). **Neural Network Methods for Natural Language Processing**. Morgan & Claypool Publishers.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. arXiv preprint arXiv:1810.04805.

### 2. Machine Learning & NLP Resources:

- Pedregosa, F., et al. (2011). **Scikit-learn: Machine Learning in Python**. Journal of Machine Learning Research, 12, 2825–2830.
- Bird, S., Klein, E., & Loper, E. (2009). **Natural Language Processing with Python**. O'Reilly Media.

### 3. Technical Reports & White Papers:

- LinkedIn Economic Graph Research. (2020). **AI in Hiring: Reducing Bias and Improving Candidate Matching**.

### 4. GitHub Repository:

- Dhingra, K. (n.d.). *AI-powered Resume Screening and Ranking System* [GitHub Repository]. Retrieved from <https://github.com/Kalpanadhingra/AI-powered-Resume-Screening-and-Ranking-System.git>