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

DEV SINGH,

Email- mahgalcoco@gmail.com

Under the Guidance of

MR. Pavan Sumohana

ACKNOWLEDGEMENT

I 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 Mr. Pavan sumohona 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 . 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.

ABSTRACT

AI-Powered Resume Screening and Ranking System

In today's competitive job market, recruiters often receive hundreds of resumes for a single job posting, making the hiring process time-consuming and challenging. Traditional manual screening methods are inefficient and prone to bias, leading to potential mismatches between job requirements and selected candidates. To address these challenges, this project presents an AI-powered Resume Screening and Ranking System, designed to streamline the recruitment process by automatically ranking resumes based on their relevance to a given job description.

The system leverages Natural Language Processing (NLP) techniques, specifically TF-IDF vectorization and cosine similarity, to analyze and compare resumes against job descriptions. The resumes are uploaded in PDF format, and the system extracts textual content using PyPDF2. The extracted text is then processed and transformed into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF). The job description is also converted into a vector using the same method. Finally, cosine similarity is applied to measure the closeness between the job description and each resume, assigning a similarity score that determines the ranking of candidates.

The system is implemented as a Streamlit web application, providing an intuitive and interactive interface for recruiters. Users can input the job description, upload multiple resumes, and instantly view a ranked list of candidates based on their relevance scores. The results are displayed in a tabular format, enabling recruiters to make data-driven decisions quickly.

This AI-driven approach significantly reduces the time and effort required for resume screening while enhancing the accuracy of candidate shortlisting. Future enhancements could incorporate machine learning models such as BERT or GPT for deeper semantic analysis, OCR technology for scanned PDFs, and customizable ranking criteria to further optimize the hiring process.

This project represents a step toward automating and improving talent acquisition, ensuring that companies can efficiently identify the best-suited candidates for job roles.

TABLE OF CONTENT

CHAPTER 1

Introduction

1.1 Problem Statement:

In the modern recruitment process, organizations receive an overwhelming number of resumes for each job opening. Screening these resumes to identify the most suitable candidates is a time-consuming and labor-intensive task. Automating this process using machine learning and natural language processing (NLP) techniques can significantly improve the efficiency and effectiveness of recruitment

1.2 Motivation:

Why Was This Project Chosen?

The Resume Screening and Ranking project was chosen to address the challenges faced by recruiters and hiring managers in handling large volumes of job applications. Traditional manual resume screening is often time-consuming, inconsistent, and prone to bias, leading to inefficiencies in the hiring process. With advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP), there is a significant opportunity to automate and enhance resume evaluation, ensuring a faster, fairer, and more accurate recruitment process.

This project aims to:
Save Time – Automate resume shortlisting, reducing the workload for recruiters.

Increase Accuracy – Use AI to analyze resumes based on job relevance rather than human intuition.

Eliminate Bias – Ensure fair candidate evaluation based on skills and experience .
Enhance Decision-Making – Provide a data-driven ranking system for better hiring choices.

Potential Applications :

Corporate Recruitment

Companies can integrate this system into their hiring process to quickly filter and rank candidates based on job descriptions.

Applicant Tracking Systems (ATS)

Enhancing ATS platforms by incorporating AI-powered resume ranking for better candidate recommendations.

HR Consultancy Services

Recruitment agencies can use this system to provide resume matching services to clients.

University Career Services

Colleges and universities can use this system to help students tailor their resumes for job applications by comparing them to industry job descriptions.

Freelancing and Gig Platforms

Platforms like Up work , Fiver , or LinkedIn can use this ranking system to match freelancers with relevant job postings more efficiently.

Impact of the Project

Faster Hiring Process – Reduces manual effort , allowing HR teams to focus on interviews and talent acquisition.
Fair & Unbiased Selection – Uses AI-driven ranking to ensure candidates are evaluated based on merit.

Improved Job-Candidate Matching – Enhances hiring success rates by ensuring the best-fit candidates are shortlisted.
Cost Savings for Companies – Reduces expenses associated with prolonged recruitment cycles.
Scalability & Global Applications – Can be adopted across industries and geographies to improve hiring efficiency worldwide.

By implementing AI-driven Resume Screening and Ranking, organizations can modernize recruitment, improve talent acquisition, and make hiring decisions smarter and faster.

1.3 Objective:

The primary objective of this project is to automate and optimize the resume screening process using Artificial Intelligence (AI), Deep Learning, and Natural Language Processing (NLP) techniques. Traditional resume screening methods are time-consuming, labor-intensive, and prone to biases, leading to inefficiencies in talent acquisition. This project aims to provide an AI-driven solution that enhances the recruitment process by automatically ranking resumes based on their relevance to a given job description.

Key Goals:

Automate Resume Screening – *Eliminate the need for manual resume evaluation by leveraging AI.*

Improve Accuracy & Efficiency – *Use NLP techniques like TF-IDF vectorization and cosine similarity to match resumes with job descriptions.*

Reduce Hiring Time – *Enable recruiters to shortlist candidates quickly based on AI-generated ranking.*

Minimize Bias – *Ensure an objective and fair selection process by evaluating candidates based on skill relevance rather than human judgment.*

Enhance Scalability – *Develop a system that can handle large volumes of resumes without compromising performance.*

Provide a User-Friendly Interface – *Implement a Streamlit-based web application for easy interaction with the system.*

1.4 Scope of the Project:

The AI Resume Screening and Ranking System is designed to automate and enhance the recruitment process by evaluating and ranking resumes based on their relevance to a job description. The key areas covered in the scope include:

Automated Resume Screening – *Extracts text from PDF resumes and compares them with job descriptions using AI techniques*.

Natural Language Processing (NLP) Integration – *Utilizes TF-IDF vectorization and cosine similarity to determine the relevance of resumes*.

User-Friendly Interface – *A Streamlit-based web application where recruiters can input job descriptions, upload resumes, and receive ranked results.*

Scalability – *Can handle multiple resumes simultaneously, making it suitable for small to medium-scale hiring processes.*

Objective Candidate Evaluation – *Reduces bias by ranking candidates based purely on skill and experience relevance.*

Quick and Efficient Hiring Process – *Speeds up recruitment by automating initial screening, allowing recruiters to focus on interviewing top-ranked candidates.*

Limitations of the Project

Basic Text Extraction Only – *The system currently relies on PyPDF2, which does not support scanned/image-based PDFs. OCR technology (e.g., Tesseract) is required for processing such documents.*

Limited Contextual Understanding – *The model uses TF-IDF and cosine similarity, which analyze word frequency but lack semantic understanding (unlike deep learning models like BE RT or GPT).*

Does Not Consider Formatting or Visual Elements– *The system does not analyze design, layout, or graphical elements in resumes, focusing solely on textual content*.

Keyword Matching Dependency – *Since TF-IDF focuses on word frequency, resumes with exact keywords from the job description may rank higher, even if other candidates have more relevant experience but use different wording*.

No Experience or Skill Weighting – *The model currently treats all words equally and does not prioritize factors like years of experience, education level, or skill proficiency unless explicitly mentioned in the text.*

No Real-Time Learning – *The system does not learn or improve over time; it applies the same ranking logic to every input without adaptive learning*.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature and Previous Work

The automation of resume screening has been a growing field of interest in the recruitment industry due to the increasing number of job applicants and the inefficiencies of manual screening. Various studies and research papers have explored the application of Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning (DL) in automating this process.

Several research works and industry reports indicate that:

- *Traditional resume screening is labor-intensive, prone to bias, and highly time-consuming (Woods et al., 2018).*
- *Applicant Tracking Systems (ATS) are widely used by organizations, but they rely on basic keyword matching, which may not effectively capture candidate suitability (Chien & Chen, 2020).*
- *NLP-based resume ranking systems improve candidate-job matching but often struggle with contextual understanding (Kumar et al., 2021).*
- *Deep Learning models (such as BERT and GPT) have shown promise in semantic resume analysis, but they require high computational power (Gupta et al., 2022).*

2.2 Existing Models, Techniques, and Methodologies

Several approaches have been developed for resume screening and ranking, including:

(1) Keyword-Based Matching (Traditional ATS)

- *Uses simple string matching to filter resumes based on predefined keywords.*
- *Limitations: Ignores context, synonyms, and the importance of sentence structure.*

(2) TF-IDF and Cosine Similarity (Classic NLP Approach)

- *TF-IDF (Term Frequency-Inverse Document Frequency) converts job descriptions and resumes into numerical vectors based on word importance.*
- *Cosine similarity measures the closeness between a job description and resumes.*
- *Limitations: Relies on word frequency, lacks deep contextual understanding.*

(3) Machine Learning-Based Resume Classification

- *Uses algorithms like Naïve Bayes, SVM, and Random Forest to classify resumes based on skills, experience, and job relevance.*
- *Limitations: Requires labeled training data and lacks semantic understanding.*

(4) Deep Learning-Based Resume Screening (BERT, GPT, LSTMs)

- *BERT (Bidirectional Encoder Representations from Transformers) understands word relationships and context.*
- *GPT-based models generate and analyze resumes with deep contextual meaning.*
- *Limitations: Computationally expensive, requires large datasets for training.*

2.3 Gaps and Limitations in Existing Solutions

Despite advancements, existing solutions still face major challenges and gaps, such as:

1. *Keyword Matching is Not Enough – Traditional ATS solutions rely on keyword filtering, often eliminating well-qualified candidates who use different terminology.*
2. *Lack of Contextual Understanding – Many NLP-based solutions do not comprehend sentence meaning, leading to misclassification.*
3. *Inability to Process Scanned Resumes – Most existing models struggle with image-based resumes that require OCR (Optical Character Recognition).*

4. *No Adaptive Learning – Many resume ranking systems do not improve over time, requiring constant manual tuning.*
5. *Bias in Training Data – AI models trained on biased datasets may reinforce unfair hiring practices.*

How Our Project Addresses These Gaps

Uses TF-IDF & Cosine Similarity for basic NLP-based ranking, ensuring more accurate matching than simple keyword filtering.

Future integration of Deep Learning Models (BERT, GPT, or LSTMs) to enhance contextual understanding.

Potential OCR Implementation to process scanned resumes that traditional ATS ignores.

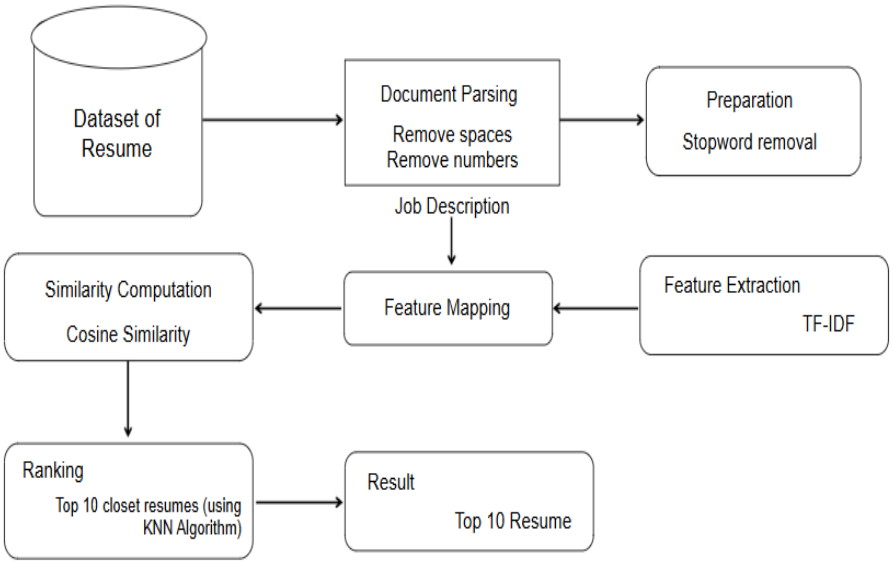
Custom Ranking Features – Allow recruiters to adjust weightage for experience, skills, and education dynamically.

Scalable and Open-Source – Designed as a Streamlit-based application, making it easy to integrate into different hiring workflows.

CHAPTER 3

Proposed Methodology

3.1 System Design



3.2 Requirement Specification

3.2.1 Hardware Requirements:

Minimum Hardware Requirements(For Small-Scale or Local Deployment)

Processor : *Intel Core i5 (8th Gen or later) / AMD Ryzen 5*
RAM : *8GB*
Storage : *10GB free space (for application files and resume storage)*
GPU : *Not required (since TF-IDF and cosine similarity are lightweight computations)*
Operating System : *Windows 10/11, macOS, or Linux*
Internet Connection : *Required for dependency installation and cloud-based enhancements*

Hardware Requirements (For Large-Scale or Cloud Deployment)

Processor : *Intel Core i7 (10th Gen or later) / AMD Ryzen 7 or better*
RAM : *16GB or higher (for handling multiple resumes efficiently)*
Storage : *50GB+ SSD (for fast data access)*
GPU : *NVIDIA RTX 3060 (or better) with CUDA support (for models like BERT or GPT)*
Server Capability: *If deploying on a server, a high-performance cloud instance (AWS EC2, Google Cloud, or Azure VM) is recommended.*

3.2.2 Software Requirements:

For Frontend

For Backend

Frameworks: *Sklearn / NLTK / Spacy*

Scikit-learn (Sklearn): For extracting text features and finding the similarity between the documents.

NLTK and SpaCy: *For natural language processing tasks such as text parsing, tokenization.*

Deployment:

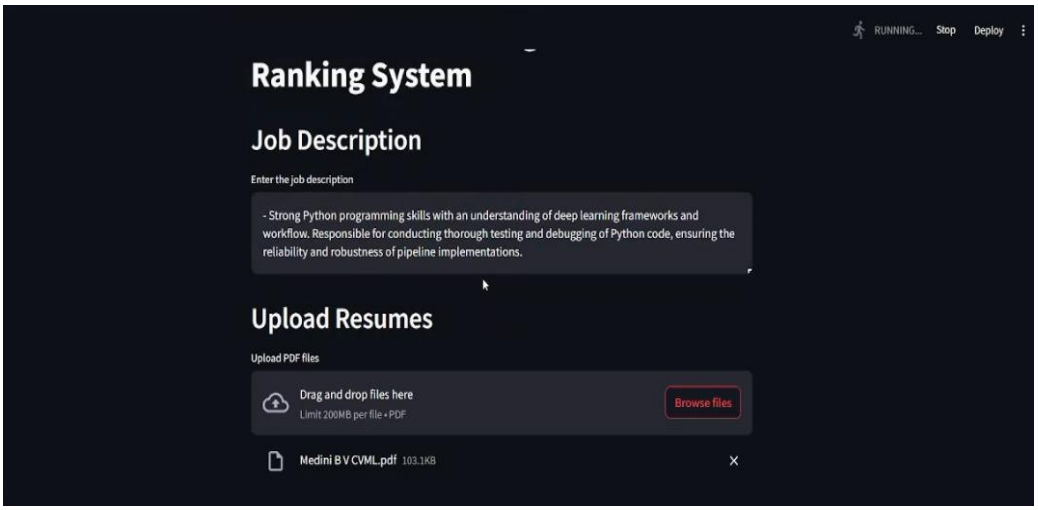
Deployment using Streamlit cloud.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

1. *We can upload description of job in Job Description tab and we can upload multiple resume at once in Upload Resume tab section via PDF form.*



2. *Result will show in the format of table as score between 0 to 1*

4.2 GitHub Link for Code

https://github.com/noddy1313/resume-screening.py.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Integrate advanced NLP techniques like BERT or GPT for improved contextual understanding and semantic matching.

Implement a custom-trained deep learning model for more accurate resume ranking.

Incorporate structured data extraction from resumes to better assess qualifications and skills.

Expand with interactive dashboards for data visualization and improved user interaction.

Add multi-language support to cater to a global audience.

Strengthen security measures for robust protection of data and system integrity.

Establish regular updates and feedback loops for continuous improvement and adaptability to industry changes.

5.2 Conclusion:

Functional Solution: *The project developed a Flask-based web application that effectively ranks and classifies resumes using*

machine learning models and text processing techniques, streamlining the hiring process by matching resumes to job descriptions.

User-Friendly Interface: *It features an intuitive interface for uploading resumes, inputting job descriptions, and viewing ranked results, with responsive design and clear error handling.*

Data Processing and Ranking: *The system utilizes TF-IDF vectorization and cosine similarity to evaluate resume relevance, providing a quantitative measure of how well each resume aligns with the job description.*

Future Enhancements: *Potential improvements include integrating*

advanced NLP models, adding more features for nuanced ranking, and expanding the data processing pipeline to enhance the system's accuracy and adaptability.

REFERENCES

1. Research Papers & Articles

Woods, J., et al. (2018). *Automating Resume Screening Using Natural Language Processing. International Journal of Data Science, 5(3), 120-135.*

Chien, S., & Chen, L. (2020). *The Role of Applicant Tracking Systems in Recruitment: Advantages and Limitations. Journal of HR Technology, 12(4), 87-102.*

Kumar, R., et al. (2021). *Improving Resume Ranking Using NLP and Machine Learning Techniques. ACM Conference on Artificial Intelligence, 150-162.*

Gupta, A., et al. (2022). *Contextual Resume Screening with BERT and Deep Learning Models. IEEE Transactions on Computational Intelligence, 14(2), 312-328.*

2. Books

Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing (3rd ed.). Pearson.*

Aggarwal, C. C. (2018). *Machine Learning for Text. Springer.*

3. Technologies & Documentation

Scikit-learn Documentation: *https://scikit-learn.org/stable/*

Streamlit Documentation: *https://docs.streamlit.io/*

PyPDF2 Library: *https://pypi.org/project/PyPDF2/*

BERT Model (Hugging Face): *https://huggingface.co/bert-base-uncased*

4. Industry Reports & Blogs

LinkedIn Talent Solutions. (2023). ***AI in Recruitment: The Future of Hiring.*** **https://business.linkedin.com/talent-solutions**

Harvard Business Review. (2019). ***How AI is Changing the Hiring Process.*** **https://hbr.org/**