

AI-BASED INTELLIGENT SYSTEM FOR RESUME SHORTLISTING USING NLP.

**PROJECT : Computer Science Project
(CSEMCSPCSP01)**

In

M.Sc. in Computer Science

By

Yash Manohar Chaudhari

ID: 4252695

Supervisor

Mughda Kashyap

ACKNOWLEDGEMENT

I truly thank our Project: Computer Science lecturer for giving key insights and engaging lessons that significantly improved my understanding of the concepts in Computer Science. The course is well structured, containing comprehensive discussions from basic to advanced topics. This helped me in applying concepts to real-world use cases.

I am particularly thankful for the professor's devotion and readiness to address queries throughout the course, which played a crucial part in my learning journey. The abilities from this course were instrumental in effectively completing this computer science project.

YASH MANOHAR CHAUDHARI

ABSTRACT

In today's job market, companies receive hundreds of resumes for a single vacant position, which makes the manual shortlisting process very time consuming and tedious. Recruiters sometimes may without intension, overlook good qualified candidates due to human bias or fatigue. In order to overcome this challenge, this project suggest an AI-based intelligent system for resume shortlisting using Natural Language Processing(NLP) to automate the hiring process and boost candidate evaluation.

Using NLP the project automatically gathers information from the resumes, including qualifications, work experience, abilities and compare them with the job description and its requirements. This system uses semantic similarity model like Sentence-BERT(S-BERT), ranking applications based on meaningful similarity rather than just looking for keywords.

HR staff from the companies can upload resumes to compare them with the job description and can view ranked candidates through a React.js built interactive web interface. The desired result result of this project is a scalable, effective and user-friendly tool that streamlines the hiring process by minimizing human participation, lowering errors, and obtaining accurate and fair candidate shortlisting.

1. INTRODUCTION:

The role of recruiters in a organization is very crucial in building sucessful and effective organization, yet the traditional techniques for shortlisting candidates are ineffective and subjective. Due to exhaustion of analyzing numerous resumes for a long period of time, recruiters may overlook qualified candidates. In today's world with growth in technology such as AI and NLP it is feasible to create systems that automate this process more consistently and fairly.

The goal of this project is to create an AI-driven system that uses semantic similarity matching to comapre the resumes with job description and rank the candidates based on the results and automate the hiring process. Instead on solely depending on keyword matching, the algorithm ranks candidates based on contextual relevance by extracting and evaluating content from resumes and job description. This approach seeks to assist Human Resource(HR) Department by lowering manual labour, enhancing selection quality and guranteeing more equitable hiring decisions.

2. RELATED WORK:

AI-driven hiring and NLP-based text analysis have been the subject of numerous studies. Artificial Intelligence algorithms are used by Indeed recruiters and LinkedIn Talent Acquisition team, however they are frequently restricted and confusing. Researcher like Gupta et al (2021) classified resumes into skill groups using TF-IDF and Naïve Bayes classifiers, and also Poonam Tijare and Mohammed Waseem have used AI and ML techniques and also used BERT model for just name extraction to provide gender neutral classification. Another researcher named Khan et al. '(2022) in his research used BERT model to match job postings with resume. Nevertheless, the majority of these solutions either ignore explainability and bias issues or rely on keyword-based similarity.

This project sets itself apart by using Sentence-BERT(S-BERT) for deep semantic comprehension, FastAPI for a reliable backend and React.js for a interactive frontend. It also places an intense focus on dataset handling, bias mitigation and result accesibilty all of this are critical and crucial in the recruitment process.

3. TECHNICAL BACKGROUND:

- 1) Natural Language Processing:- NLP makes it easy for the machines to fully understand and analyze human language. It serves as a foundation for this project, which helps in the extraction and processing of textual data from resumes and job description.
- 2) Text Embeddings and Semantic Similarity:- Embeddings use numerical vectors in a multidimensional space to represent words or phrases. Cosine similarity is used to calculate the semantic similarity, to determine how well the texts in the resume and the job description agree to each other.
- 3) Sentence-BERT:- S-BERT is a NLP model that has been improved over years for tasks involving sentence similarity. It generates outstanding language embeddings and makes it possible to compare candidates profile and job description effectively.
- 4) Python Libraries:- Pandas for data management, sentence transformers for embeddings, FastAPI for backend and SpaCy for text preparation are the useful python libraries used.

4. METHODOLOGY:

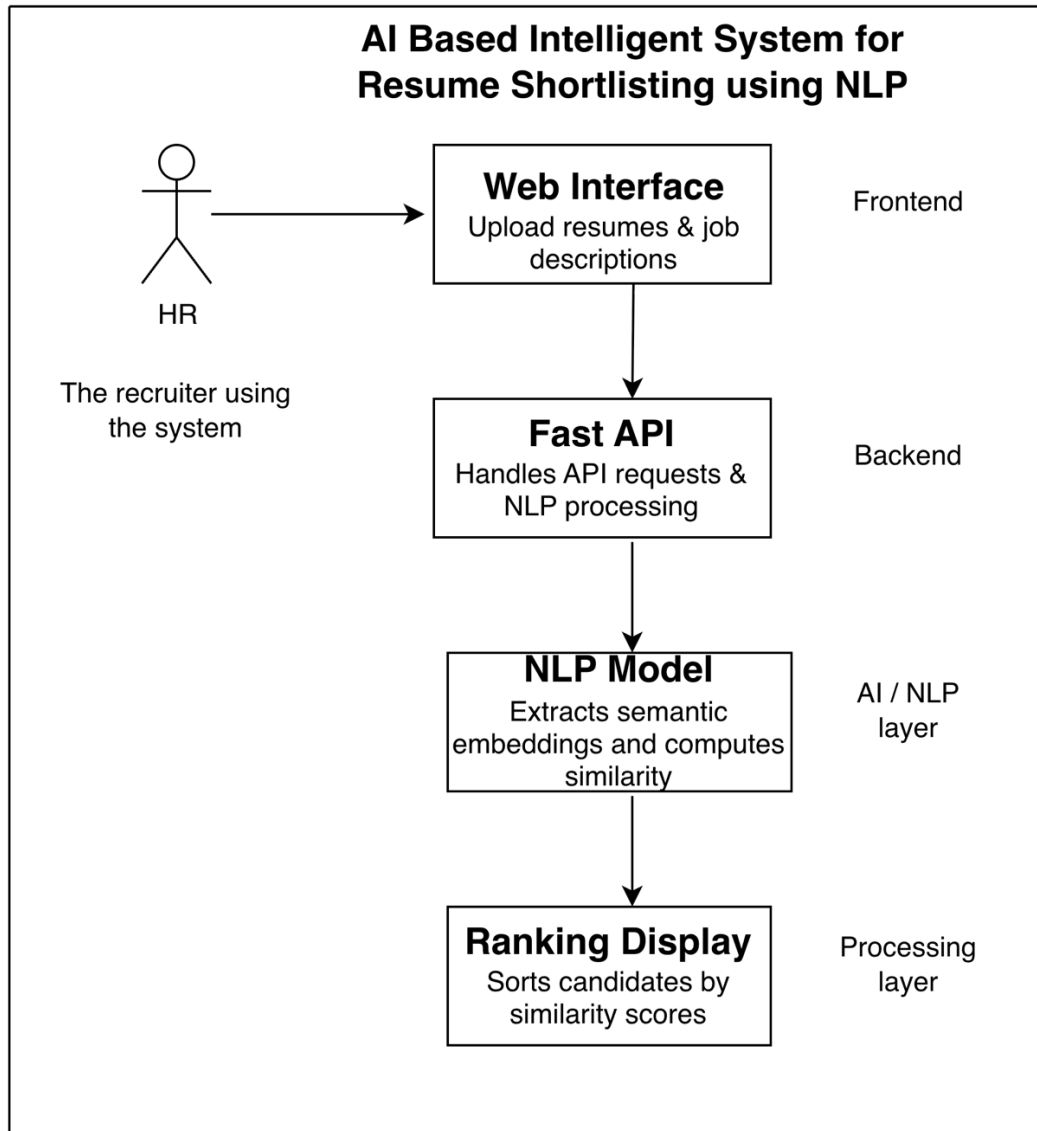
In order to improve and automate the hiring process, the suggested intelligent system combines Artificial Intelligence and Natural Language Processing. A Sentence-BERT model for semantic matching, a FASTAPI backend and React.js frontend makes the systems architecture.

The dataset included in this project are solely used for training and optimizing the NLP model. The HR department can use this website to upload candidates resume and enter the job description to get the candidate rankings. This deployed systems runs only on real-time inputs.

After receiving these inputs, the backend uses spaCy to extract and clean text before using the S-BERT model to create vector embeddings. The model determines the semantic similarity of each resume uploaded by calculating the cosine similarity between it and job description. After ranking the similarity scores of the resumes, a structures ranking list is sent back to frontend.

Multiple file upload, job description entries and a dynamic display of ranked candidates are all made possible by the React.js frontend's user-friendly interface. It uses RESTful API to interact with FASTAPI backend.

This methods ensures that the systems stays data-driven and user-focused. It enables precise and transperant candidate shortlisting with minimizing human interaction with pre-trained NLP intelligence and real-time data processing.



5. IMPLEMENTATION:

The frontend interface, communication layer, and backend AI model are all integrated during the implementation process. File parsing, text preparation, embeddings creation and semantic similarity are all handled by the backend. Pandas, PyPDF2 and docx2txt are used to process resumes. S-BERT is used to convert cleaned text into embeddings, and candidates are ranked using cosine similarity.

The user interface is provided via React.js. Resumes and job description are uploaded by users supplied to FASTAPI using Axios request, the frontend then shows the ranked candidates with similarity scores in an orderly manner after receiving the response.

6. TESTING:

To make sure the AI based shortlisting system is working properly, smoothly and continually in a number of scenarios , testing is an important part of this project. User-level authentication for full capability , integration evaluation of the whole system pipeline, and functional testing of particular modules is part of testing process. The main objectives of testing is to verify that resume reading is correct and done properly, ensure that the semantic similarity analysis produces accurate outcome, and verify's that frontend and backend connectivity is fast and clean.

Testing will be performed in Phase 3.

7. CONCLUSION:

This website demonstrates how online technology, artificial intelligence, and natural language processing can be merged to upgrade or modernise the Hiring Process. By using Sentence-BERT (SBERT) for semantic comprehension, the AI-based resume shortlisting website will consistently and efficiently automate the time-consuming process of manually reviewing resumes. This helps recruiters with an easier, quicker, and more relevant way to evaluate the candidate's resume by mentioning the job description in a preferred context, that finds significant connections beyond the simple keyword matching.

The React.js used for the frontend provides user friendly and easy to use Interface for the users, while the FastAPI developed backend ensures scalable API endpoints and effective processing of the data. By giving job descriptions as input along with the resumes and datasets, Users will be able to check the list of candidates filtered through ratings based on the semantic relevance. Through this website we can see how web frameworks and machine learning can be implemented together to close the gap between technical automation and human decision-making processes.

Furthermore, one of the main ethical issues with AI-driven employment systems is addressed by the incorporation of explainability and bias reduction strategies, which support fairness and transparency. This project can be expanded in multiple ways, such as incorporating Named Entity Recognition for extracting particular and required skills and enhancing model fine-tuning with domain-specific datasets. Also, this entire system can be migrated to cloud platforms for enterprise-scale operations. Through these improvements, we can turn the hiring process into a reliable and production-ready system.

8. REFERENCES:

Tijare, Poonam & Waseem, Mohammed & Sherani, Mohd & Sampath Kumargari, Sai Krishna & Kavitha, P.. (2023). A smart resume screening tool for customized shortlisting. ITM Web of Conferences.56.10.1051/itmconf/20235604001. <https://doi.org/10.1051/itmconf/20235604001>

Blessing, Moses. (2025). AI-Powered Resume Screening: Benefits and Challenges. https://www.researchgate.net/publication/388688179_AIPowered_Resume_Screening_Benefits_and_Challenges

Harsha, Tumula & Moukthika, Gangaraju & Sai, Dudipalli & Pravallika, Mannuru & Anamalamudi, Satish & Enduri, MuraliKrishna. (2022). Automated Resume Screener using Natural Language Processing(NLP). 1772-1777. [10.1109/ICOEI53556.2022.9777194](https://doi.org/10.1109/ICOEI53556.2022.9777194).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. arXiv. <https://arxiv.org/abs/1908.10084>

Job Description Dataset: <https://www.kaggle.com/datasets/kshitizregmi/jobs-and-job-description>

Resume Dataset: <https://www.kaggle.com/datasets/gauravduttakiit/resume-dataset>