

AN INTELLIGENT WEB BASED RESUME SCREENING SYSTEM AND JOB MATCHING

D.Mahitha¹, Mrs.K.Neeharika²

¹Student, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India

²Assistant Professor, Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India

Email id: mahithaduggada@gmail.com

Abstract: *This project presents an intelligent resume screening and job matching system that automates candidate evaluation using advanced Natural Language Processing (NLP) techniques. By integrating TF-IDF vectorization, semantic similarity matching, and DistilBERT embeddings, the system analyzes resumes, extracts relevant skills, and compares them with job descriptions. The proposed model reduces manual screening effort, enhances accuracy, and significantly accelerates the shortlisting process. It also supports data-driven hiring decisions by providing objective and consistent candidate evaluations.*

Keywords: Resume Screening, Job Matching, Natural Language Processing, Machine Learning, Deep Learning, Semantic Similarity, DistilBERT, TF-IDF, Skill Gap Analysis, Automated Recruitment.

1. INTRODUCTION

The rapid growth in job applications has made manual resume screening inefficient, time-consuming, and prone to human bias. To address these challenges, AI-driven recruitment systems are becoming essential for improving hiring accuracy and reducing recruiter workload. This project proposes an Intelligent Web-Based Resume Screening and Job Matching System that uses NLP and machine learning to analyze resumes, extract skills, and match candidates with job descriptions. The system generates a similarity score, identifies missing skills, and provides personalized recommendations for improvement. By automating the screening process, the solution enables faster, unbiased, and more efficient recruitment.

2. LITERATURE SURVEY

The rise of digital recruitment has led to a substantial increase in job applications, making manual resume screening both slow and inconsistent. Recent research emphasizes the use of Natural Language Processing (NLP) and machine learning techniques to automate candidate evaluation. Studies indicate that methods such as TF-IDF, Word2Vec, and transformer-based models like BERT and DistilBERT significantly enhance the understanding of resumes and improve matching with job descriptions. While TF-IDF contributes to better keyword representation, it is limited in capturing deep semantic context. In contrast, transformer-based models provide superior accuracy in aligning candidate profiles with job requirements. Additionally, skill extraction and gap analysis have been shown to improve candidate evaluation by identifying both strengths and deficiencies relative to job expectations. Overall, the adoption of automated screening tools reduces recruiter workload, minimizes bias, and improves consistency in hiring decisions.

3. PROPOSED SYSTEM

The proposed system is an intelligent, web-based resume screening and job-matching platform designed to automate candidate evaluation and enhance recruitment accuracy through the use of advanced Natural Language Processing (NLP) techniques. The system streamlines the recruitment process by analyzing uploaded resumes, extracting relevant skills, and comparing them with predefined job descriptions to generate objective similarity scores, thereby assisting recruiters in efficiently shortlisting qualified candidates.

The architecture of the system combines a Flask-based backend with two complementary analysis engines. The first engine leverages TF-IDF vectorization to perform keyword-based similarity, enabling the system to quickly identify resumes containing critical job-related terms. The second engine utilizes DistilBERT, a transformer-based semantic model, to capture deep contextual meaning from the resume text, providing a more nuanced and accurate assessment of candidate-job fit. Resumes can be submitted in multiple formats, including PDF, DOCX, and TXT. Once uploaded, the content undergoes text extraction, preprocessing, and normalization before being passed to the similarity engines.

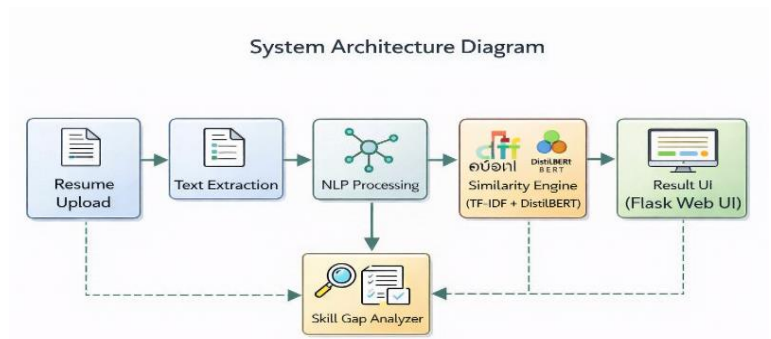


Fig 1: Proposed System

A key feature of the system is the skill-extraction and gap-analysis module, which identifies both technical and domain-specific skills mentioned in the resume and the job description. By comparing these skill sets, the module highlights missing or underrepresented skills and generates personalized recommendations for candidate improvement. This enables recruiters to not only evaluate existing competencies but also guide candidates toward enhancing their qualifications. Furthermore, the system includes a category-prediction model that classifies resumes into specific job domains, enhancing the precision of job-matching results by filtering candidates into relevant roles.

The user interface, implemented as a Flask web application, presents the final results in an intuitive and interactive format. Recruiters can view similarity scores, skill gaps, and recommended improvements for each candidate, all within a secure environment. To ensure controlled access, the system integrates an email-based OTP verification mechanism, allowing only authenticated users to access the dashboard and analysis features.

Overall, the proposed system is lightweight, scalable, and cost-effective. By combining keyword-based and semantic analysis, along with skill-gap detection and domain classification, it offers a comprehensive solution for modern recruitment workflows. The architecture not only reduces recruiter workload and bias but also provides real-time, personalized insights into candidate suitability, making the hiring process faster, more consistent, and highly efficient.

4. METHODOLOGY

The methodology of the system is organized into the following steps:

1. Resume Upload

Users upload their resumes, which are accepted in PDF or DOCX formats. The system initiates automated text extraction for further processing.

2. Text Extraction & Preprocessing

The extracted text is cleaned by removing noise, stopwords, symbols, and formatting issues. This ensures the resume content is standardized for NLP processing.

3. Skill & Information Parsing

NLP techniques identify key sections such as skills, education, experience, and certifications. Relevant candidate information is structured into machine-readable format.

4. Job Description Processing

Job descriptions undergo the same cleaning and preprocessing steps. Important keywords, required skills, and role-specific attributes are extracted.

5. Feature Extraction (TF-IDF & DistilBERT)

Both resumes and job descriptions are converted into feature vectors using TF-IDF and DistilBERT embeddings. This captures both keyword frequency and semantic meaning.

6. Similarity Computation

Cosine similarity is applied to measure the match strength between the resume and job description. Higher similarity scores indicate stronger candidate-job alignment.

7. Skill Gap Detection

Missing or weak skills are automatically identified and suggested.

5. PROPOSED SYSTEM RESULTS

The proposed Intelligent Resume Screening and Job Matching System was successfully developed and tested using multiple real-world resumes and job descriptions. The system accurately extracted skills, cleaned text, generated embeddings, and produced reliable match scores.

- The resume parser effectively extracted key information such as skills, education, roles, and experience with high accuracy using NLP techniques.
- TF-IDF and DistilBERT embeddings generated meaningful vector representations, enabling accurate similarity comparison between resumes and job descriptions.
- The system consistently identified relevant skills and highlighted missing skills for each job, improving transparency for both recruiters and applicants.
- The job-matching module assigned similarity scores that correctly reflected the candidate's relevance, and ranking results were stable across different test cases.
- The interface clearly displayed match percentage, skill insights, and recommended improvements, making the system user-friendly for recruitment tasks.

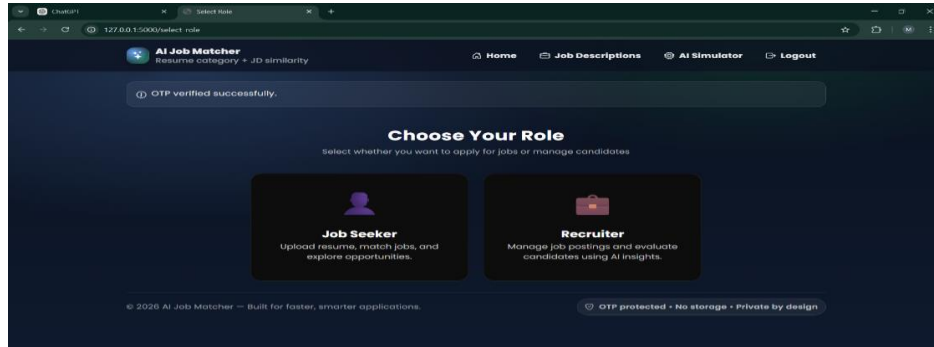


Fig 2:Selecting Role

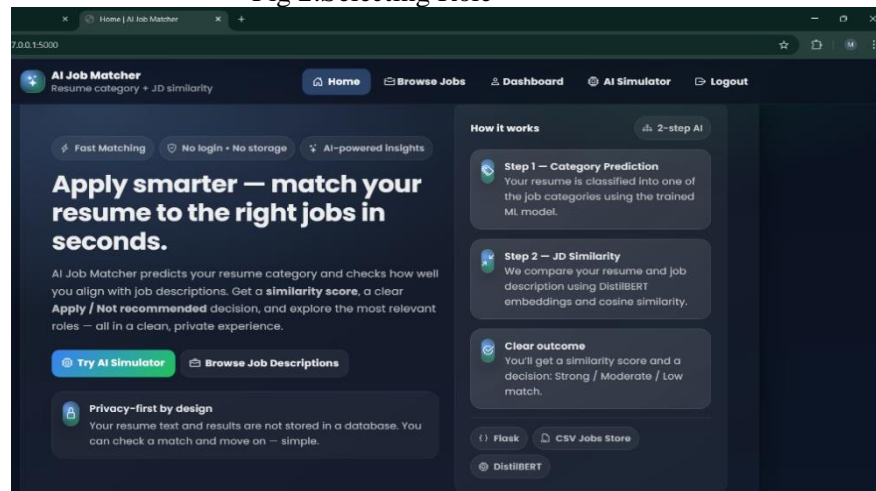


Fig 3:User Interface

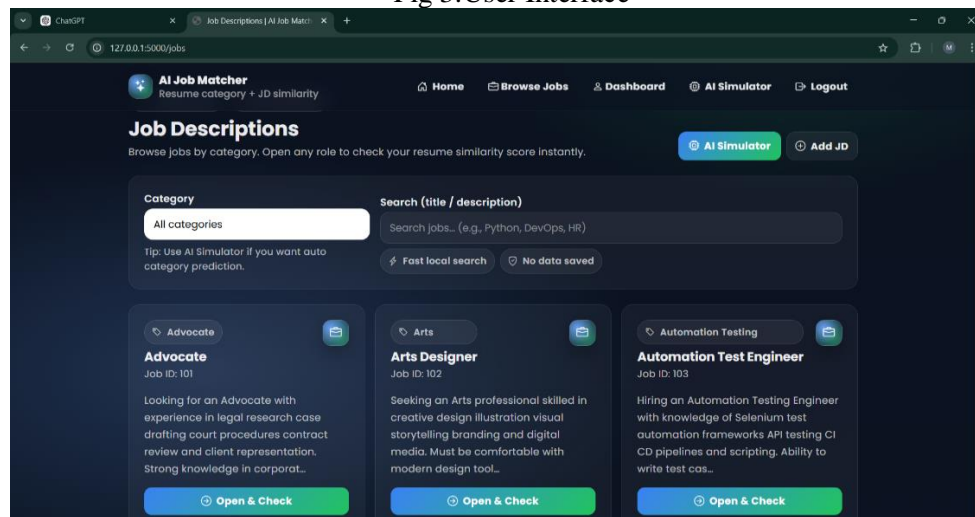


Fig 4:Job description page

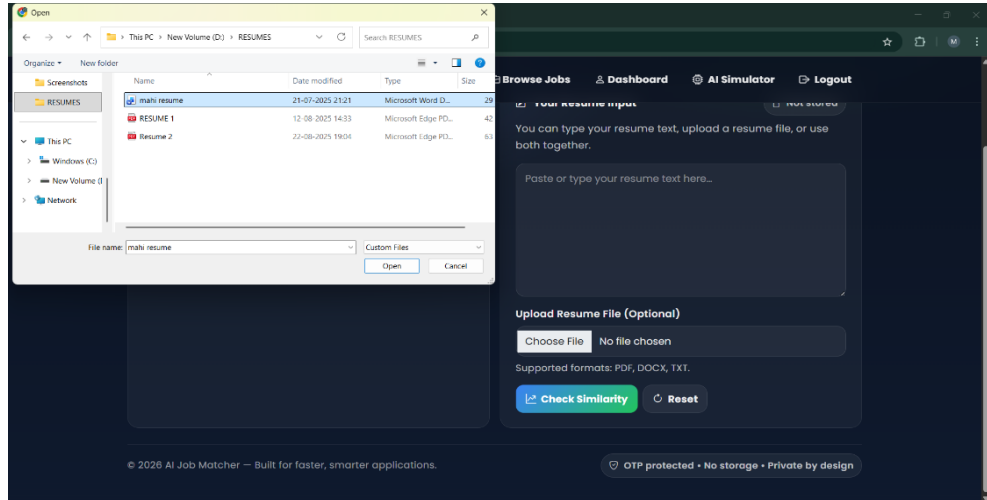
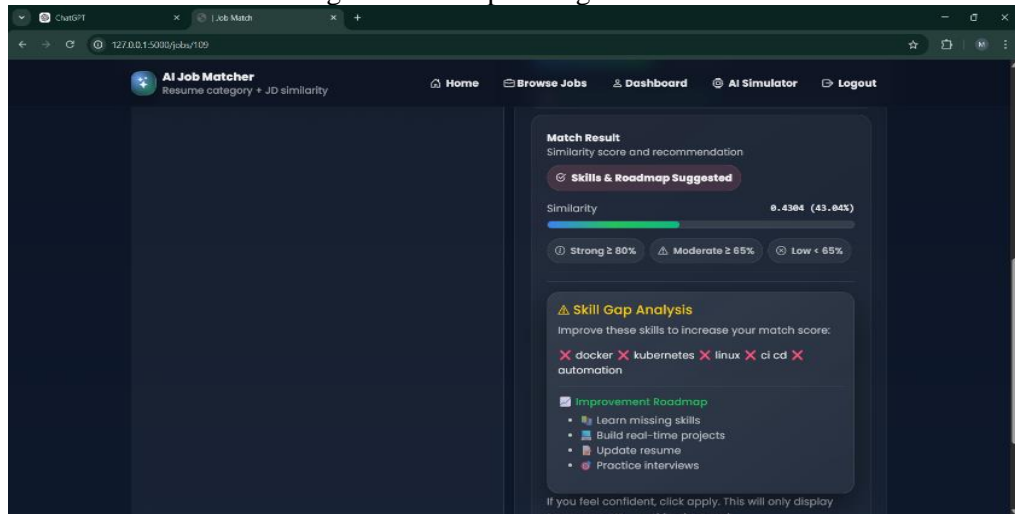


Fig 5:Resume uploading



6: Suggesting skills & Checking dashboard

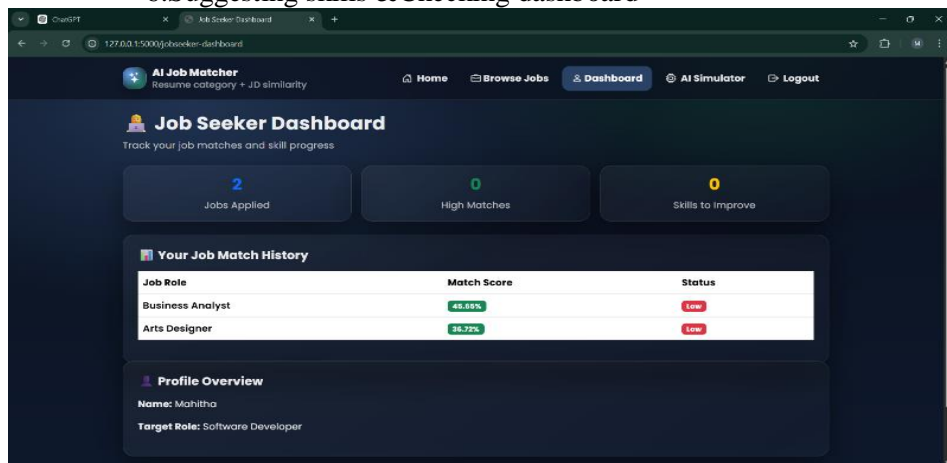


Fig7:Jobseeker Dashboard

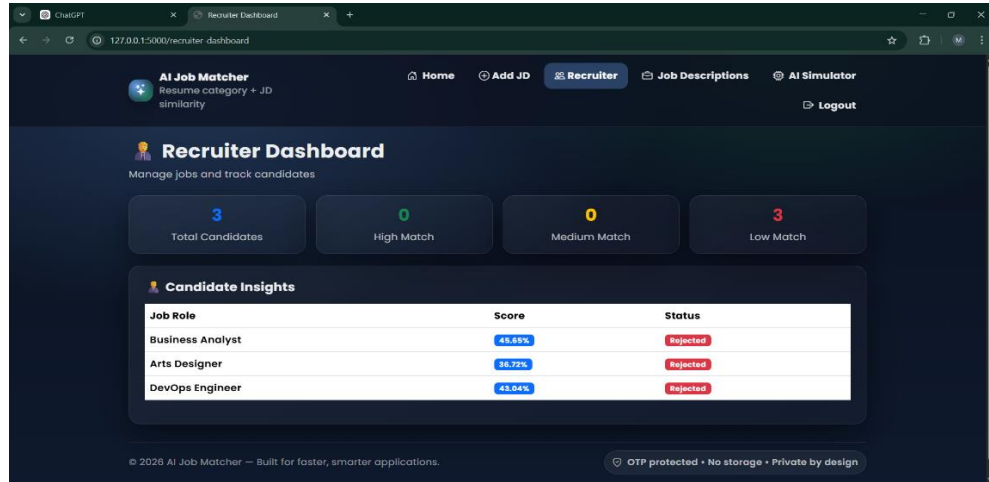


Fig 8:Recruiter dashboard

Overall, the system demonstrated strong performance in automated screening, semantic analysis, similarity matching, proving its effectiveness in reducing manual effort and improving recruitment efficiency.

6. CONCLUSION

This work successfully developed an intelligent, automated resume screening and job-matching system using advanced Natural Language Processing (NLP) and semantic similarity techniques. The system efficiently analyzed resumes and job descriptions, extracted relevant skills, and generated accurate match scores, significantly reducing manual screening effort. By integrating TF-IDF and DistilBERT-based embeddings. The user-friendly web interface provided clear visualization of match results, making the system practical for both recruiters and job seekers. Overall, the proposed system offers a cost-effective, scalable, and intelligent solution that improves recruitment efficiency, accuracy, and fairness.

REFERENCES

1. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL-HLT*, 2019.
2. V. Sanh et al., "DistilBERT: A Distilled Version of BERT for Faster and Lighter NLP," *HuggingFace Research*, 2020.
3. F. Pedregosa et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011.
4. S. Bird et al., "Natural Language Processing with Python," *O'Reilly Media*, 2009.
5. R. K. Aggarwal et al., "Automated Resume Screening Using NLP and Machine Learning," *IEEE Access*, 2022.
6. P. Singh et al., "Skill Extraction and Job Matching Using Text Mining Techniques," *International Journal of Computer Applications*, 2021.
7. TF-IDF Documentation, "Term Frequency–Inverse Document Frequency Algorithm Guide," *Scikit-Learn Docs*, 2023.
8. HuggingFace Transformers, "Model Hub Documentation," 2024.
9. S. Zhang, Y. Wang, "Word2Vec-Based Resume Classification and Job Matching," *Journal of Intelligent & Fuzzy Systems*, 2021.
10. A. Nogueira, K. Cho, "Passage Re-ranking with BERT," *arXiv preprint arXiv:1901.04085*, 2019.
11. T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," *arXiv preprint arXiv:1301.3781*, 2013.
12. R. Sarawagi, "Information Extraction," *Foundations and Trends in Databases*, 2008.

13. *M. Joshi et al., "Automated Candidate Ranking System Using NLP," International Conference on Computational Linguistics, 2021.*
14. *A. Vaswani et al., "Attention is All You Need," NeurIPS, 2017.*