

Resume Analyzer Using NLP to Match JOB Description

Shyji Ealias

Assistant Professor/ Dept of CSE
Sri Sairam College of Engineering, Bengaluru, India
shyjealias@gmail.com

<https://orcid.org/0000-0001-8178-7762>

TalaMarla Rajitha, Rakshitha A, Supriya N A, Shreyas R

Student, Dept of CSE

Sri Sairam College of Engineering, Bengaluru, India
sce22cs034@sairamtap.edu.in, sce22cs118@sairamtap.edu.in
sce22cs125@sairamtap.edu.in, sce22cs045@sairamtap.edu.in



Publication History

Manuscript Reference: IRJCS/RS/Vol.12/Issue12/DCCS10082

Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.12/Issue12/DCCS10082

Received:12,November 2025,Revised: 20,November 2025,Accepted: 30 November 2025 Published Online:18 December 2025 <https://www.irjcs.com/volumes/Vol12/iss-12/03.DCCS10082.pdf>

Article Citation:Shyji,TalaMarla,Rakshitha,Supriya,Shreyas(2025),Resume Analyzer Using NLP to Match JOB Description, IRJCS: International Research Journal of Computer Science, Volume 12, Issue 12 of 2025 pages 694-699

Doi:<https://doi.org/10.26562/irjcs.2025.v1212.03>

BibTeX Key Shyji@2025Resume

IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2025, ISSN 2393-9842, <https://doi.org/10.26562/irjcs.2025.v1212.03> The journal's official abbreviation is IRJCS.

Orcid: <https://orcid.org/0009-0004-9398-7488>

Copyright©2025 copyright by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In the changing world of recruitment, organizations must efficiently sift through large numbers of resumes to find candidates that fit specific job roles. This paper introduces a web-based Resume Analyzer System that uses Natural Language Processing (NLP) and machine learning to automate resume evaluation and job matching. The system includes two main modules: Admin (HR/Recruiter) and Job Seeker. The admin module allows recruiters to post jobs, manage candidate applications, and view ranked resumes based on similarity scores. The Job Seeker module enables applicants to register, upload resumes in PDF or DOCX formats, explore job openings, and apply directly through the platform. Using NLP, the system extracts key features such as skills, work experience, and educational background from resumes and job descriptions. A machine learning-based matching algorithm calculates a similarity index to evaluate candidate suitability and highlights missing or recommended skills. Additionally, the platform classifies resumes into relevant categories, assesses document quality, and integrates with online professional profiles. By reducing manual work and bias in the screening process, the proposed Resume Analyzer improves accuracy, fairness, and efficiency in recruitment, benefiting both recruiters and job applicants.

Keywords: Resume Parsing, Natural Language Processing (NLP), Automation, Candidate Screening, Resume Matching, Job Description, Applicant Tracking System, Semantic Similarity.

INTRODUCTION

Manual screening of resumes against different job descriptions is very challenging for both recruiters and job seekers in today's competitive job market. Recruiters must sift through a large variety of resume formats, which makes the assessment process time-consuming, prone to errors, and biased due to personal judgment. In this context, recent advancements in NLP and machine learning offer a helpful solution: resume analyzers automatically extract and match relevant candidate information to specific job requirements with improved objectivity and scalability. NLP-driven resume analyzers use advanced techniques such as text extraction, skill identification, named entity recognition, and semantic similarity computation with models like TF-IDF and cosine similarity to structure resumes and job descriptions. They analyze skills, work experience, education, and other key attributes. These systems measure the fit between candidate profiles and job specifications, often providing ranked match scores or simple predictions like "Good Fit" or "Not a Good Fit" to recruiters. These tools can also identify missing skills, recommend resources for acquiring those skills, and suggest improvements to make resumes more searchable by applicant tracking systems. Various projects and studies conducted in recent years show the effectiveness of NLP-based resume analysis across different platforms. These projects combine large language models, machine learning classifiers, and real-time feedback dashboards. Such systems not only simplify hiring for companies but also empower job seekers by making the application process clearer, more personalized, and better suited to market demands. Automated resume screening technologies are expected to become essential for improving recruitment outcomes and shortening time-to-hire in workforce management.

LITREATURE SURVEY

1. Julian A (2024) proposed a model that sorts resumes and evaluates candidate skills against job requirements by using machine learning techniques to improve the accuracy of candidate short listing.
2. Thangaramya K, Logeswari G. (2024) present an automated resume parsing and ranking system that uses NLP to extract key information from resumes. This system aims to improve recruitment efficiency by ranking candidates based on their qualifications in relation to job descriptions, helping recruiters make better decisions.
3. Pai K, Dhage S (2023) explores resume evaluation using latent Dirichlet allocation and NLP techniques for effective candidate selection. They show how topic modeling can help understand candidate qualifications and fit for specific roles.
4. The work by Khan et al. (2023) discusses the creation of a resume parser and summarizer to simplify recruitment processes. This tool takes unstructured data from resumes and reformats it into structured formats. They explain how natural language processing techniques and machine learning improve data extraction and help organizations, reducing the time recruiters spend identifying information about educational backgrounds, skills, and work experiences. Although existing systems are quite useful, the authors note that they have limitations related to language complexity and ambiguity, which can lead to missing qualified applicants. Their proposed system is trained on diverse datasets and supports multiple document formats for better accuracy in parsing. This work emphasizes the ongoing progress in NLP and AI to boost the efficiency of resume parsing and lessen human error during recruitment.
5. The paper discusses a job recruitment portal that incorporates NLP for resume screening. The system analyzes resumes and job descriptions to identify key information such as skills, education, and work experience. The portal ranks candidates based on how well they fit the job roles, showing how NLP can enhance recruitment processes by providing detailed and accurate analysis, leading to better hiring decisions.
6. In the paper "Resume Parser and Analyzer Using NLP," Patil et al. propose a modern system for automating the recruiting process with the use of NLP. The system they describe downloads resumes in PDF and DOC formats, which it parses to extract key details, including personal information, education, skills, and experience. It uses Named Entity Recognition and Part-of-Speech tagging for field extraction, which are then analyzed to rank resumes and provide suggestions to job seekers. The paper highlights Python libraries that can aid in tokenization and lemmatization, allowing effective preprocessing for unstructured data. The system helps recruiters make decisions by providing visualizations and downloadable reports when it stores parsed data in a database. Thus, this technique increases accuracy and reduces workload, enabling a large number of applications to be handled more easily in the recruitment automation process.

PROBLEM IDENTIFICATION

In today's recruitment process, HR professionals receive hundreds of resumes for a single job posting. Manually reading, short listing, and comparing resumes with job descriptions is:

- Time-consuming.
- Subjective and prone to bias.
- Inefficient when handling many applications.

There is a need for a smart system that can automatically analyze resumes, extract skills using NLP, match those skills with job descriptions, and generate ranking scores. This will improve accuracy and reduce screening time.

OBJECTIVES

The main goal of the Resume Analyzer using NLP to match job descriptions is to automate recruitment by analyzing resumes and aligning them with job descriptions. The system aims to help recruiters quickly find the most suitable candidates and assist job seekers in enhancing their resumes for better job opportunities.

Specific Objectives:

- To extract relevant information from resumes (PDF/DOCX) using NLP techniques.
- To compare candidate skills with job requirements and calculate a match score.
- To recommend suitable job openings to applicants based on their skills and experience.
- To identify missing skills and suggest improvement resources like courses and tutorials.
- To allow admin users (HR/Recruiters) to post job openings and view ranked applicants.
- To analyze GitHub and LinkedIn profiles as extended resumes.
- To use machine learning (clustering) to group similar resumes for faster filtering.

PROPOSED SYSTEM

The proposed Resume Analyzer system will be a web application that uses NLP to match job description requirements. It can automate recruitment by analyzing resumes and matching them with relevant job roles. Key components and features of the system include:

- Secure Admin Module: Recruiters can post jobs, manage applications, and view ranked candidate resumes based on similarity scores and skill matches.
- User Module: This includes features for registration, the ability to upload resumes in PDF or DOCX format, and a personalized dashboard.

PDF Text Extraction:

This application starts by extracting text from PDF files for resumes and job descriptions. Using a PDF parsing library, it reads each page, converts the text to plain text, and combines all the pages into one text string. This ensures the content of any resume or job description is accessible for further processing.

Text Preprocessing:

Preprocessing refines the extracted text data. The following steps are applied:

- Tokenization: Breaking down text into individual words or tokens for precise manipulation.
- Lower casing: Standardizing text by converting all characters to lowercase, ensuring that words like "Python" and "python" are treated the same.
- Removing Stop Words: Eliminating common words that add no real value to the analysis, like "the" or "and," by using a stop word list to reduce noise in the dataset.
- Non-Alphanumeric Filtering: Removing symbols and numbers that are not relevant to skill or similarity analysis, leaving only meaningful keywords. This preprocessing pipeline results in a clean, reduced dataset ready for NLP analysis.

Skill Extraction and Matching:

The core of the analysis lies in matching the resume and job description in terms of skills. The system dynamically identifies and extracts skills from the job description so it can evaluate whether they appear in the candidate's resume.

- Dynamic Skill Extraction: A pre-defined list of skills (like "Python," "SQL") relevant to any job is used. Extracting these skills from the job description ensures a tailored skill set for each role.
- Skill Matching and Scoring: The system calculates the skill match score by identifying and comparing the overlap of these skills with the resume. A higher overlap score indicates a better alignment of the candidate's skills with the job's requirements, while unmatched skills highlight potential gaps.

Application Flow and User Interface

- The user interface allows users to interact with the application easily and provides a natural flow.
- File Upload: Users can upload both resume and job description files in PDF format.
- Automated Processing: After file upload, the application extracts, preprocesses, and analyzes the documents, computing skill match, similarity scores, and final predictions.
- Results Display: The results appear on the screen, showing the skill match score, a list of matched skills, the semantic similarity score, and the job fit prediction.
- The UI also displays helpful error messages for missing files or processing errors, improving the user experience.

RESEARCH METHODOLOGY

The research methodology outlines the clear approach used to design, execute, and assess the Resume Analyzer that uses NLP for Job Description Matching. This methodology includes five key phases: data collection, preprocessing, NLP-driven resume extraction, similarity and ranking calculation, and the generation of final outputs. Both qualitative and quantitative methods are used to ensure proper resume job alignment.

System Architecture Overview:

The methodology follows client-server architecture. The web interface acts as the client, while the backend (NLP and ML processing engine) functions as the server. The admin and user modules connect with a centralized database to store job postings, resumes, and application information.

Data Collection:

The system collects data from various sources:

- Resume files submitted by users (PDF/DOCX formats).
- Job descriptions inputted by the admin.
- Public datasets and external profiles (LinkedIn and GitHub, where allowed). All submitted data is securely stored in the database.

Data Preprocessing:

When a resume is uploaded, preprocessing takes place to convert unstructured text into useful data. This process involves:

- Extracting text using PyPDF2 or python-docx.
- Removing noise: stop words, punctuation, special characters.
- Tokenization: breaking text into words or phrases.
- Lemmatization: changing words to their root form (e.g., "running" becomes "run"). Tools used: spacy NLP model (English - en_core_web_sm).

Feature Extraction Using NLP:

The system extracts important elements from the cleaned text, such as:

- Technical and soft skills.
- Experience keywords.
- Education and degree information.
- Project and job-role keywords.

NLP techniques used:

Technique	Purpose
1. Named Entity Recognition (NER).	To extract Skills, Experience.
2. Keyword Extraction (TF-IDF).	To identify important terms that occur frequently.
3. Word embeddings (spacy vectors).	To calculate similarity with job descriptions.

Resume-Job Matching and Ranking:

To evaluate a resume against a job description:

Both the resume text and the job description are transformed into vector formats.

This results in a match percentage (for example, 85%) that shows how relevant the resume is to the job.

The formula used is:

$$\text{Similarity} = A \cdot B / ||A|| \times ||B||$$

Where:

- A = vector representation of the resume.
- B = vector representation of the job description.

Missing Skills Identification:

The skills extracted from the resume are compared with the required skills listed in the job description.

The system suggests:

- Certification courses (Coursera, Udemy).
- Video tutorials.
- Skill improvement roadmaps.

Machine Learning-based Resume Clustering:

To categorize resumes (for example, Developer, Data Science): Feature vectors are created from the resumes.

Output:

Recruiters can explore candidates by cluster instead of reviewing resumes manually. External Profile Analysis (LinkedIn / GitHub).

An optional module retrieves public profile information if a URL is provided: GitHub → projects, repositories, technology stack.

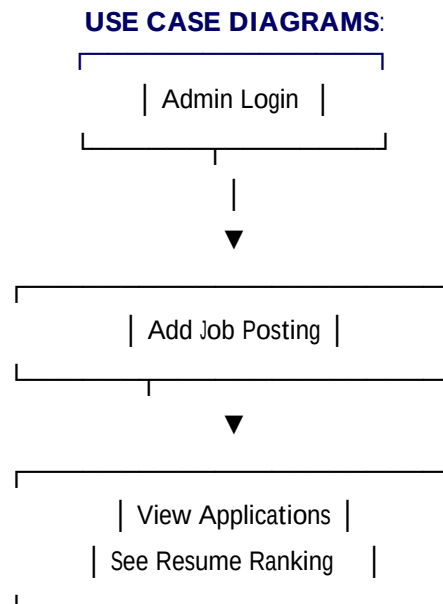
LinkedIn → experience, skills, endorsement count.

System Evaluation:

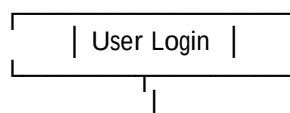
The evaluation is done through:

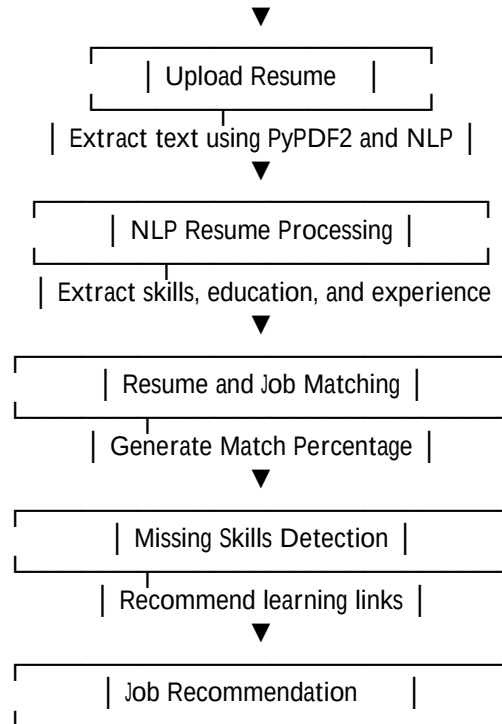
- The accuracy of job matching compared to manual evaluations by recruiters.
- Feedback from users and HR professionals.

Admin login:

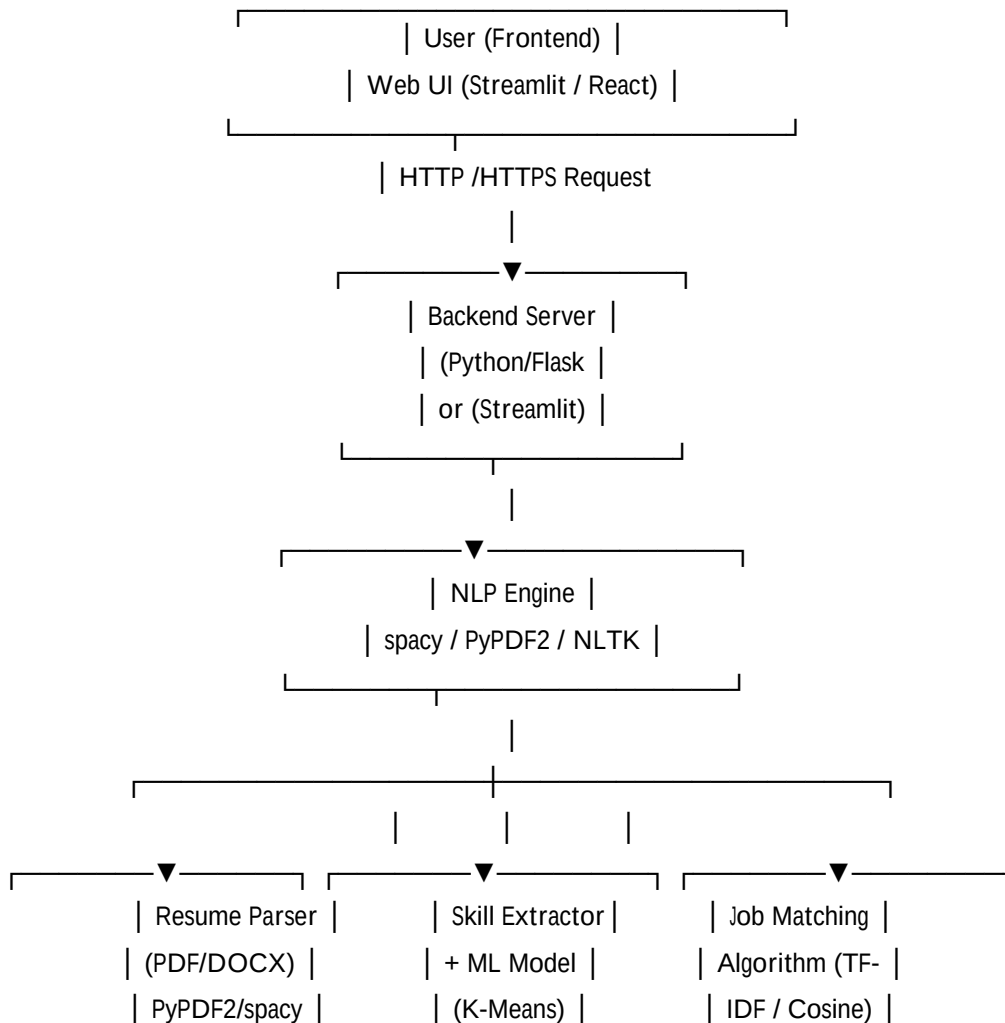


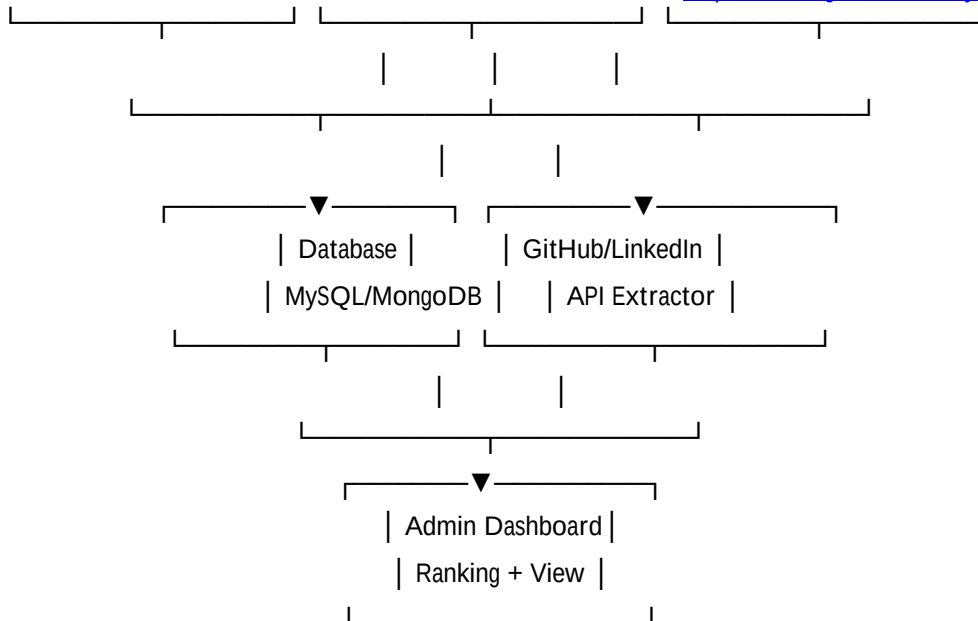
User login:





SYSTEM ARCHITECTURE DIAGRAM (High-Level Design):





CONCLUSION

This research presents an automated resume evaluator that uses NLP and machine learning to improve the efficiency and accuracy of hiring. By automating resume screening, the system cuts down on manual errors and matches candidates more effectively to job descriptions. Unlike traditional keyword-based systems, it employs NLP to understand the context of a candidate's qualifications, offering a deeper and more precise evaluation. As recruitment processes become more digital and the number of applications increases, this system will be crucial for streamlining hiring efforts. By reducing bias and improving candidate feedback, it promotes transparency. With future upgrades like AI integration, multilingual support, and partnerships with other platforms, the system shows potential to change how employers and job seekers approach the recruitment process.

REFERENCES

1. Julian A, Haripriya K. 2024. NLP-based resume analysis and adaptive skill assessment system. In Proceedings of the 3rd International Conference for IEEE.
2. Thangaramya K, Logeswari G. 2024. Automated resume parsing and ranking using natural language processing. In Proceedings of the Internet of IEEE.
3. Jagwani V, Meghani S, Pai K, Dhage S. 2023. Resume evaluation through latent Dirichlet allocation and natural language processing for effective candidate selection. ArXiv preprint arXiv:2307.15752.
4. Khan, Farzana, Hamdan Patel, Arshad Shaikh, Fawzah Sayed, and Abdul Rehman Soorya. "Resume Parser and Summarizer." International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) 3 (2023).
5. IEEE Xplore. (2018). Automated Resume Screener Using Natural Language Processing. Available at: [Accessed 15 Jul. 2024].
6. Patil, N., Yadav, S., & Biradar, V. (2023). Resume Parser and Analyzer Using NLP. International Research Journal of Modernization in Engineering Technology and Science, 5(4), 7210-7216.