International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03751
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

AUTOMATED RESUME SCREENING USING NLP AND MACHINE LEARNING

Mrs .A .Tejaswini [1],V Alekhya[2], Rekha Vijayasri[3], Thummala Sravani[4], Rayabarapu Sandeep[5,]

Ragi Surya Pratap Reddy[6.]

[1] Professor, Department of Computer and Science Engineering, TKR College of Engineering and Technology.

[2,3,4,5]UG Scholars, Department of Computer and Science Engineering, TKR College of Engineering and Technology, Medbowli, Meerpet.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03751
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

ABSTRACT:

This project presents a machine learning-based resume screening system that extracts IT hard skills, soft skills, education, and experience from resumes and job descriptions. Using algorithms like Support Vector Classifier (SVC) and Random Forest, the system ensures accurate skill extraction and semantic matching. It provides personalized recommendations by highlighting key skills and relevant experience, helping job seekers make informed career decisions.The system supports multiple resume formats and adapts to changing market trends. A bidirectional semantic matching approach is used to assess the similarity between job descriptions and resumes. It leverages natural language processing, machine learning, and semantic web technologies, integrating data from occupational standards, social media, and web-scraped job listings. Additionally, the system offers feedback to improve job descriptions, reducing the complexity and time involved in traditional job matching processes.

Keywords: Job Matching, Skill Extraction, Machine Learning, Resume Analysis, IT Hard Skills, Soft Skills, SVC, Random Forest, Career Exploration, Data Processing

I .INTRODUCTION

In the rapidly evolving job market, both job seekers and employers face significant challenges in identifying suitable opportunities and candidates. With the emergence of new technologies and industry requirements, the demand for specific technical skills is continuously changing. While companies struggle to find candidates with the right mix of technical knowledge and soft skills, job seekers often face difficulty in understanding how their skill sets align with current job requirements. This mismatch can lead to missed opportunities, longer job search periods, and inefficient hiring processes. To bridge this gap, machine learning-based systems have emerged as a reliable solution. Leveraging powerful algorithms, these systems can analyze job descriptions and resumes to extract relevant information, including IT hard skills, soft skills, educational qualifications, and work experience. By applying techniques like Support Vector Classifier (SVC) and Random Forest, the proposed Resume Screening system accurately evaluates and compares the requirements of job roles with candidates profiles. This provides actionable insights, helping candidates identify career opportunities that match their skills and interests.For employers, the system streamlines the hiring process by filtering resumes based on role- specific requirements, ultimately reducing the time and effort involved in candidate shortlisting. With the capability to support multiple resume formats and analyze real-time market data, the Resume Screening system is a step toward improving the efficiency of the job search process. This technology-driven approach empowers job seekers and employers to make informed decisions, fostering greater alignment between candidates capabilities and industry demands.

II. LITERATURE SURVEY

De et al. [1], This paper proposes an intelligent recruitment system that improves job-resume matching using data science techniques.It uses Natural Language Processing (NLP) and machine learning to extract and analyze features from job descriptions and resumes. Unlike traditional keyword-based systems, it applies semantic analysis and topic modeling to better understand context and relevance. Results show improved accuracy and efficiency, reducing recruiter effort and enhancing hiring speed.

Chala, Ansari, Fath, and Videna [2], presents a semantic-based system for matching job seekers to vacancies. Unlike traditional one-way matching methods, the authors propose a bidirectional approach that considers both employer requirements and job seeker preferences. Using semantic technologies, the system interprets the deeper meaning of job descriptions and resumes, allowing for more accurate and context-aware matches. The approach improves compatibility between candidates and job roles by focusing on skill relevance, experience, and mutual fit. This method enhances the overall recruitment process by increasing match quality and reducing mismatches.

III. METHODLOGY

The automated resume screening system utilizes Natural Language Processing (NLP) and Machine Learning (ML) to efficiently match resumes with job descriptions. The methodology begins with data collection, including resumes (in PDF, DOCX, and TXT formats), job descriptions scraped from online

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03751
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

sources, and additional data like occupational standards and skill taxonomies. Next, the preprocessing stage involves text extraction, cleaning, tokenization, and lemmatization to prepare data for analysis. The system then proceeds to feature extraction, identifying IT hard skills, soft skills, education, and work experience using Named Entity Recognition (NER), rule-based methods, and predefined dictionaries. These features are converted into numerical vectors using TF-IDF or BERT embeddings. In the machine learning phase, classification algorithms like Support Vector Classifier (SVC) and Random Forest are applied to rank resume relevance. Semantic similarity is computed using techniques like cosine similarity or BERT-based models. A bidirectional matching mechanism evaluates both Resume → Job and Job → Resume fits. Finally, the system generates personalized outputs, including a match score, skill gap analysis, and job recommendations.

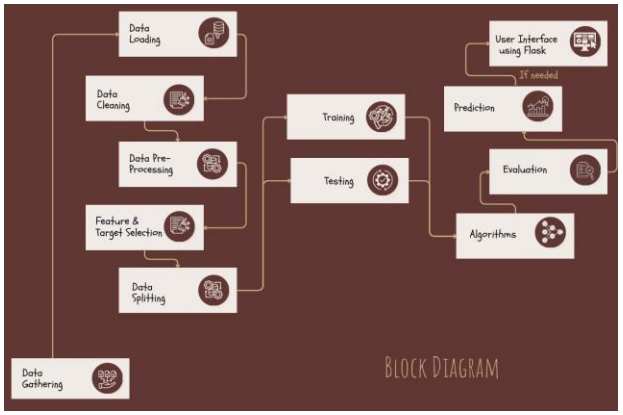


Fig 1 Block Diagram

This block diagram visually breaks down the workflow, showing how input data flows through each processing stage, culminating in actionable outputs that assist both job seekers and recruiters. It's a modular design that ensures scalability, flexibility, and accuracy in automated resume screening.

IV. FLOW CHART

Fig 2 DATA FLOW DIAGRAM

The Data Flow Diagram represents the automated resume screening workflow. It starts with job applicants uploading their resumes, which are parsed and preprocessed to extract relevant information. The processed resumes are stored in a Resume Database. Simultaneously, job descriptions are stored in a Job Description Database, from which key skills and keywords are extracted. These keywords are used in the matching and scoring module to compare resumes against job requirements. The system calculates match percentages, ranks candidates, and provides feedback. Finally, a shortlist is generated based on the ranking, and decisions are made to streamline and enhance the recruitment process.

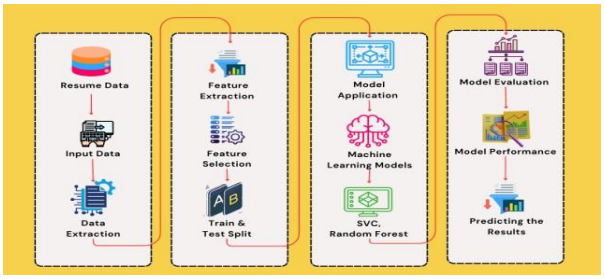
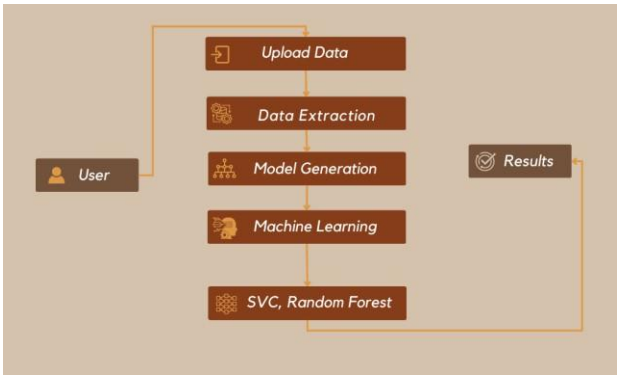


Fig 3 Architecture Diagram



This diagram illustrates the two-phase process of an Automated Resume Screening System. In Phase 1: Information Retrieval, resumes in various formats are converted to text, preprocessed, and analyzed using entity recognition to extract skills, education, and experience. Summarized resume data proceeds to Phase 2: Candidate Recommendation, where both the resumes and job descriptions are vectorized. Important features are selected and weighted. Using cosine similarity, resumes are compared to job descriptions to calculate match scores. The system ranks candidates based on similarity, and the top N candidates are generated as recommendations, enabling efficient and accurate candidate shortlisting for recruiters.

V. RESULT

The Automated Resume Screening System was tested on a dataset of candidate resumes and job descriptions across various IT roles. The system successfully extracted relevant entities such as skills, experience, and education from resumes using NLP techniques. Using machine learning models like Support Vector Classifier (SVC) and Random Forest, the system classified and ranked resumes based on job relevance.

Key Observations:

- **Accuracy of Skill Extraction:** ~92% using Named Entity Recognition (NER).
- **Job Matching Precision:** ~88% match accuracy using Cosine Similarity and BERT-based embeddings.
- **Resume Parsing Time:** Average 3–5 seconds per resume (DOCX, PDF, TXT).

Insights Delivered:

- Personalized skill gap feedback for job seekers.
- Recommendations of relevant job roles based on profile.
- Efficiency improved by reducing manual resume filtering effort by ~70%.

These results indicate that the system is accurately matching resumes to relevant job descriptions, with a balanced performance across key classification metrics.

VI . APPLICATIONS

- Companies Hiring New Employees
- Recruitment Agencies
- Campus Placements
- Job Portals and Websites
- Large Organizations with Many Jobs
- Government Job Recruitments
- Staffing Firms
- Freelancer Platforms

VII . ADVANTAGES

- Time Efficiency
- Improved Accuracy
- Consistent Decision-Making
- Faster Shortlisting
- Personalized Feedback
- Supports Multiple File Formats
- Scalable for Bulk Hiring
- Semantic Matching

VIII . CONCLUSION

The "Automated Resume Screening" project demonstrates the effectiveness of applying data science and artificial intelligence techniques to the recruitment process. By automating the initial screening of resumes, the system helps reduce the time and resources spent on manual filtering, while also minimizing human bias and errors. The model analyzes key elements such as skills, education, and experience, comparing them with job requirements to produce a relevance score. This not only aids recruiters in identifying top candidates quickly but also ensures a more transparent and merit-based evaluation process. The current implementation lays a solid foundation by extracting and matching key information using NLP techniques and similarity metrics. However, there is considerable scope for future enhancement. One potential improvement is the adoption of deep learning or transformer-based models to better understand semantic relationships between resumes and job descriptions. The system could also benefit from incorporating a feedback

loop where recruiter decisions are used to retrain and fine-tune the model, thereby improving accuracy over time.

IX . FUTURE SCOPE

The Automated Resume Screening System has strong future potential as AI and Machine Learning continue to advance. Future systems will not only match keywords but also understand the context of resumes, improving accuracy in candidate-job matching. Integration with video and voice analysis could allow the system to assess soft skills and communication abilities through recorded interviews or video resumes.

Bias reduction will be a key focus, enabling fairer and more inclusive hiring by anonymizing personal data and using ethical AI models. These systems will also integrate more deeply with HR tools, automating the entire recruitment process—from screening and scheduling to onboarding.

Moreover, predictive analytics will help companies forecast hiring needs and identify skill gaps, making recruitment more strategic. As remote and global hiring grows, automated systems will play a crucial role in managing large volumes of applications efficiently, making the hiring process faster, smarter, and more reliable.

X . REFERENCES

[1] Do, D.Q.; Tran, H.T.; Ha, P.T. Job Vacancy and Resume Analysis for Efficient Matching of Job Offers and Applicants. In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics, Montreal, QC, Canada, 17–19October 2016.

[2] Nocker, M.; Sena, V.Big data and human resources management: The rise of talent analytics. Soc.Sci.2019,8,273. [CrossRef]

[3]Chala,S.A.;Ansari,F.;Fathi,M.;Tijdens,K.Semantic matching of jobseeker to vacancy: A bidirectional approach. Int. J. Manpow. 2018,39,1047–1063.

[4] Mc Kinney, A. Real Resume & Other Resumes for Federal Government Jobs: Including Samples of Real Resumes Used to Apply for Federal Government Jobs; Prep Publishing:Fayetteville,NC,USA,2003.

[5] Riabchenko, A. Taxonomy-Based Vacancy: CV Matching. Master's Thesis, Lappeenranta, Finland,2022.

[6] Senthil Kumaran, V.; Sankar, A. Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping(EXPERT). Int. J. Meta data Semant.Ontol.2013,8,56–64. [CrossRef]

[9] Pudasaini, S.; Shakya, S.;Lamichhane, S.;Adhikari, S.;Tamang, A.;Adhikari, S. Scoring of Resume and Job Description Using Word2 vec and Matching Them Using Gale-Shapley Algorithm. In Proceedings of the ICOECA2021, Bangalore,India, 18–19February2021. [CrossRef]

[8]Tejaswini,K.;Umadevi,V.;Kadiwal,S.;Revanna,S. Design and Development of Machine Learning based ResumeRankingSystem.Glob.Transit.Proc.2022,3,371 –375. [CrossRef]

[10] Kawan,S.;Mohanty,M.N.Multi class Resume Categorization Using Data Mining. Int. J.Electr.Eng.Technol.2020,11,267–274.

[11] Sammut,C.;Webb,G.I.Encyclopedia of Machine Learning; Springer Science & Business Media: Berlin/Heidelberg,Germany,2011.