International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven Resume Ranking Ensuring Fair and Efficient Hiring Decision.

Sumit Kumar*[1], Mididoddi Keerthana*[2], Thaduri Varuntej*[3], Vangoori Harshavardhan*[4]

*[1] Assistant Professor Of Department Of CSE (AI & ML) Of ACE Engineering College, India.

*[2,3,4] Students Of Department Of CSE (AI & ML) Of ACE Engineering College, India.

ABSTRACT

AI-driven Resume Ranking helps companies deal with the huge number of resumes they receive. The usual way of going through resumes takes a lot of time and can often be unfair. This project focuses on building a machine learning model that can automatically read, sort, and rank resumes using natural language processing (NLP). It pulls out important details like education, work experience, skills, and qualifications using techniques like Named Entity Recognition (NER) and TF-IDF Term Frequency-Inverse Document Frequency. By doing this, the system makes hiring faster, reduces bias, cuts down on manual work, and improves how well candidates match job openings. In the end, each resume is scored and ranked based on how well the candidate fits the role.

Keywords: Named Entity Recognition (NER), Natural Language Processing (NLP),Term Frequency-Inverse Document Frequency, Machine Learning(ML).

INTRODUCTION

With the increasing number of candidates graduating college every year, choosing the right candidates from the huge numbers could be a hectic task for the recruiters. Identifying the individuals with the right skill set for a job and automatically ranking them would aptly ease the process of shortlisting. This project has been designed with an overview of easing recruitment by automatically screening and ranking resumes, helping to weigh down the difficulties of manually shortlisting candidates. In this project the webpage serves as the entry point collecting information from the user, which is sent to the backend for further processing. The final stage comprises ranking the resumes classified by sorting them in terms of their rank. The integration of the technologies like NLP and ML has automized the resume screening and ranking process. It can also be noted that the model may face challenges while dealing with unstructured data from the resumes and to deal with this, a major condition of having the resumes be in a particular structure that compliments the working of the model and overall processing is expected to be maintained

BACKGROUND

Going through resumes one by one takes a lot of time—and with so many people applying for jobs these days, it's easy to feel overwhelmed. On top of that, manual screening can sometimes lead to unfair decisions, just because it's hard to be 100% consistent. Big companies often have fancy software to help with this, but smaller businesses usually don't have those kinds of tools or the budget for them. That's where this project comes in. We're building a system that uses Machine Learning (ML) and Natural Language Processing (NLP) to help automate resume screening. Instead of someone having to read every resume, the system picks out the most important info—like skills, work experience, and education—and checks how well it matches what the job is looking for. The goal is to make hiring simpler, faster, and fairer. By taking care of the repetitive part of screening, this tool helps hiring teams spend less time sorting through resumes and more time talking to the right candidates. It's a smarter, more balanced way to find the best person for the job.

LITERATURE REVIEW

1. S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, and A. Gurjar (2019)

This study presents a web application aimed at automating the process of resume screening. The system streamlines candidate evaluation by reducing manual efforts involved in reviewing resumes. The tool was introduced at the 2019 International Conference on Nascent Technologies in Engineering (ICNTE) in Navi Mumbai. While the system enhances operational efficiency, it does not incorporate advanced machine learning or semantic matching techniques, limiting its adaptability across diverse job descriptions.

2. Pradeep Kumar Roy (2020)

Roy proposed an automated approach to resume classification and matching to accelerate candidate shortlisting. The system employs classifiers for initial categorization and uses content-based recommendation methods—such as cosine similarity and k-NN—to identify resumes that best match the job description. The method increases efficiency in candidate selection but does not explore deep learning or context-aware NLP models that could improve precision.

3. V. Kumar and V. Garg (2017)

This paper discusses a security framework for managing unstructured resume data stored in MongoDB, a NoSQL database. The study highlights potential security threats and presents mitigation strategies to ensure secure storage and access of resume data. While it addresses a critical data protection need, the work does not engage with resume processing or classification models, making it more relevant to database security than document analysis.

4. J. Petersheim, J. Lahey, J. Cherian, A. Pina, G. Alexander, and T. Hammond (2023)

The authors examine how students and recruiters assess computer science resumes differently. Their findings reveal divergent evaluation criteria and highlight the subjective nature of resume screening. The study offers insights into human judgment in resume assessment but does not explore automation or self-supervised learning techniques.

5. D. Pant, D. Pokhrel, and P. Poudyal (2013)

This paper introduces a system for filtering resumes using techniques such as NLP, word matching, character positioning, and regular expressions. The model enhances screening precision for software engineering candidates. However, the reliance on rule-based processing limits scalability and may not generalize well to other domains or more complex resume structures.

6. Yong Luo, Huaizheng Zhang, Yongjie Wang, Yonggang Wen, and Xinwen Zhang (2018)

This work presents ResumeNet, a framework leveraging domain knowledge to assess resume quality automatically. Implemented with knowledge graphs and deep learning, ResumeNet improves objectivity and consistency in resume evaluation. The system offers strong performance but requires comprehensive knowledge base construction, making it resource-intensive for new domains.

7. M. Berdanier, M. McCall, and G. M. Fillenwarth (2021)

The authors propose Domain Discourse Density (DDD) as a metric to evaluate the presence of field-specific language in engineering resumes. Higher DDD scores correlate with stronger resumes. This approach provides a novel linguistic feature for resume analysis but lacks integration with automated screening tools or broader machine learning frameworks.

8. Murugan Anandarajan, Chelsey Hill, and Thomas Nolan (2019)

This study emphasizes text preprocessing—such as tokenization, stop-word removal, stemming, and lemmatization—as foundational for resume analytics. These steps improve model performance in downstream tasks. While crucial, the paper focuses on preprocessing techniques rather than on end-to-end resume evaluation systems.

9. Devaraja G, Krishna Vardhni, Dharshita R, and A. Mahadevan (2023)

The authors conduct a comparative study of classification algorithms (Naive Bayes, SVM, neural networks) applied to textual data, highlighting strengths and weaknesses across models. Although informative for selecting models, the paper does not apply these comparisons specifically to resume data, limiting its practical impact in HR analytics.

10. Patle and Chouhan (2013)

This research evaluates various SVM kernel functions—linear, polynomial, and RBF—for classification tasks, analyzing performance in terms of accuracy and computation. The findings aid model selection but lack direct application to resume classification and do not incorporate recent advances in self-supervised or deep learning.

COMPARISON TABLE

S.No	Title	Author Name	Methodology used	Findings from the Reference paper
1	A Web Application forResume Screening (ICNTE, 2019)	S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, A. Gurjar	Developed a web-based system to automate resume screening using keyword matching and scoring algorithms to assist recruiters in shortlisting candidates.	The tool automates the resume evaluation process, improving efficiency in recruitment by reducing manual workload and enabling faster, more consistent candidate assessments.
2	Resume Classification and Matching	PradeepKumar Roy	Combined classification algorithms with cosine similarity and k-NN for content-based recommendation. Resumes were first categorized and then matched to job descriptions.	The proposed system streamlines candidate shortlisting by automating classification and ranking of resumes, significantly reducing time and effort in hiring processes.
3	Security Analysis in MongoDB for Unstructured Data	J. Kumar, V. Garg	Conducted a security audit and vulnerability analysis for storing and accessing unstructured data (e.g., resumes) in MongoDB, a NoSQL database.	Identified potential threats in unstructured data handling and recommended security measures to ensure safe and private data storage in hiring systems.
4	Student vs Recruiter Resume Evaluation Perspectives	C. Petersheim, J. Lahey, J. Cherian, A. Pina, G. Alexander, T. Hammond	Comparative study involving resume evaluation by both students and recruiters, highlighting the differing criteria and	Found discrepancies in resume assessment standards; recruiters prioritized clarity and relevance, while students focused more on format and length.

			perceptions between both groups.	
5	Automated Resume Screening for Software Engineers	D. Pant,D. Pokhrel, P. Poudyal	Used NLP techniques, word and pattern matching, character position indexing, and regular expressions to extract relevant data from resumes.	Improved screening accuracy by filtering resumes based on keywords and formatting, making it suitable for software engineering role applications.
6	ResumeNet: A Learning-Based Framework for Automatic Resume Quality Assessment (ICDM, 2018)	Yong Luo, Huaizheng Zhang, Yongjie Wang, Yonggang Wen, Xinwen Zhang	Proposed a deep learning framework called ResumeNet that uses NLP and classification models to assess and score resume quality.	Achieved reliable automated evaluation of resumes based on quality metrics; useful for large-scale hiring scenarios to filter high-quality candidates.
7	Characterizing Disciplinarity and Conventions in Engineering Resumes (IEEE, 2021)	C. G. P. Berdanier, M. McCall, G. M. Fillenwarth	Introduced the Disciplinary Discourse Density (DDD) metric to quantify discipline-specific language; qualitative analysis of resume content and style.	Found that resumes with higher DDD scores were more effective and aligned with engineering norms; recommended teaching disciplinary writing conventions to students.
8	Text Preprocessing in Practical Text Analytics (Springer, 2019)	Murugan Anandarajan, Chelsey Hill, Thomas Nolan	Discussed foundational NLP preprocessing steps including tokenization, stop-word removal, stemming, and lemmatization to prepare text for analysis.	Highlighted how effective preprocessing enhances data consistency, model accuracy, and downstream analytics, especially in text classification tasks.
9	Comparative Study of Text Classification	Devaraja G., Krishna Vardhni,	Empirically evaluated multiple classifiers such as Naive Bayes, SVM,	Identified strengths and limitations of each model; provided insights into the best

	Models (OCIT, 2023)	Dharshita R., A. Mahadevan	and deep learning models using precision, recall, F1 score, and visual tools like confusion matrices.	model choices for different types of resume classification problems.
10	SVM Kernel Functions for Classification (ICATE, 2013)	S. Patle, D. S. Chouhan	Compared several SVM kernel functions—linear, polynomial, and RBF—on different datasets to analyze classification performance and computational efficiency.	Found RBF kernel generally provided better accuracy for non-linear data; offered guidance on selecting appropriate kernels based on dataset characteristics.

RESEARCH GAPS IN EXISTING SYSTEMS:

1.Bias and Fairness in Resume Screening

AI models trained on historical recruitment data may perpetuate biases related to gender, ethnicity, age, or university attended. Lack of robust debiasing algorithms and fairness-aware ML techniques tailored specifically to resume ranking.Development of transparent, fair, and explainable systems that minimize discrimination while preserving performance.

2. Insufficient Contextual Understanding

NLP models often fail to capture the nuanced context of job roles, career progression, or transferable skills.Limited ability of existing models to accurately understand implicit experiences or domain-specific terminology.Improved contextual embeddings and knowledge-aware NLP techniques for better role fit prediction.

3. Lack of Publicly Available Benchmark Datasets

Research is hindered by the absence of standardized, labeled, and diverse datasets of resumes and hiring outcomes.No large-scale open dataset exists that supports reproducibility and benchmarking.Creation and release of anonymized resume datasets with metadata (e.g., job titles, hiring decisions).

4. Limited Explainability and Transparency

Most AI systems for resume ranking function as “black boxes,” offering little insight into decision-making Scarce research on interpretable AI models that can explain why a resume is ranked higher or lower. Integration of explainable AI (XAI) frameworks to justify and validate model predictions to both recruiters and applicants.

5. **Static Modeling of Candidate Profiles**

Resumes are treated as static documents, ignoring continuous skill acquisition or experience updates. Existing models do not incorporate dynamic candidate evolution or learning progress.

PROPOSED SYSTEM

To address these gaps, a proposed system could extend self-supervised learning to diverse document analysis tasks (e.g., table detection, OCR) by training on varied datasets beyond RVL-CDIP and Tobacco-3482. It would incorporate an automated hyperparameter optimization pipeline using techniques like Bayesian optimization to maximize SimCLR and Barlow Twins performance. Additionally, the system would develop document-specific augmentations (e.g., text distortion, layout shifts) to enhance representation learning, validated through a standardized augmentation framework. Finally, it would test scalability by applying self-supervised models to smaller, industry-specific datasets, ensuring practical applicability in low-resource settings.

CONCLUSION AND FUTURE SCOPE

This is project shows how AI can make the process of reading and ranking resumes faster and easier. By using machine learning and natural language processing, the system looks at things like a person's skills, work experience, and how well their resume matches the job. This saves time and effort for recruiters and helps make sure every candidate is judged fairly. The use of models like XGBoost gave good results and showed that this kind of system can work well in real hiring situations. In short, this project proves that AI can help companies find the right people more quickly and make better hiring decisions.

REFERENCES

1. S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, A. Gurjar, "Web Application for Screening Resume," 2019.

2. Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia, "A Machine Learning approach for automation of Resume Recommendation system," 2020.

3. J. Kumar and V. Garg, “Security analysis of unstructured data in NOSQL MongoDB database,” 2017 Int. Conf. on Computing and Communication Technologies for Smart Nation (IC3TSN), Gurgaon, India, pp. 300-305, 2017.

4. C. Petersheim, J. Lahey, J. Cherian, A. Pina, G. Alexander, and T. Hammond, "Comparing Student and Recruiter Evaluations of Computer Science Resumes," 2023.

5. D. Pant, D. Pokhrel, and P. Poudyal, "Automatic Software Engineering Position Resume Screening using NLP and Regex," 2022.

6. Yong Luo; Huaizheng Zhang; Yongjie Wang; Yonggang Wen; Xinwen Zhang,” ResumeNet: A Learning-Based Framework for Automatic Resume Quality Assessment”, 2018 IEEE International Conference on Data Mining (ICDM), December 2018

7. C. G. P. Berdanier, M. McCall and G. M. Fillenwarth, "Characterizing Disciplinarity and Conventions in Engineering Resume Profiles," in IEEE Trans. Profess. Comm., vol. 64, no. 4, pp. 390-406, Dec. 2021.

8. Murugan Anandarajan, Chelsey Hill, and Thomas Nolan. "Text preprocessing" in Practical text analytics:Maximizing the value of text data, Switzerland: Springer Nature Switzerland AG, pp. 45-59, 2019.

9. Devaraja G, Krishna Vardhni, Dharshita R and A. Mahadevan, "A Comparative and Analytical Study of Text Classification Models using Various Metrics and Visualizations," 2023 OITS Int. Conf. Information Technology (OCIT), Raipur, India, pp. 1-6, 2023.

10. A. Patle and D. S. Chouhan, "SVM kernel functions for classification," 2013 Int. Conf. Advances in Technology and Engineering (ICATE), Mumbai, India, pp. 1-9, 2013