



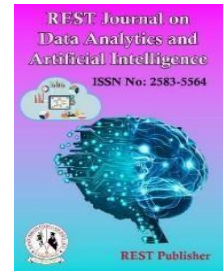
REST Journal on Data Analytics and Artificial Intelligence

Vol: 4(1), March 2025

REST Publisher; ISSN: 2583-5564

Website: <http://restpublisher.com/journals/jdaai/>

DOI: <https://doi.org/10.46632/jdaai/4/1/65>



Resume Parsing and Ranking System for LinkedIn using NLP

*B. Siva Sai Snehal, B. Balaji, T. Mani Chandhra, Shilpa Shesham

Anurag University, Hyderabad, Hyderabad, Telangana, India.

*Corresponding Author Email: snehalboyinapally24@gmail.com

Abstract. In today's competitive job market, efficiently identifying the most suitable candidates "among a vast pool of applicants is crucial for organizations. This project presents a Hybrid "Resume Parsing and Ranking System that leverages advanced Natural Language Processing (NLP) techniques, "deep learning, and transformer models to streamline the candidate selection process. The system extracts "key information from resumes using rule-based methods and Named Entity Recognition (NER) powered by SpaCy's "pre-trained models. It ranks resumes using cosine similarity between the extracted information and job "descriptions, with vectorization done through Sentence-BERT (SBERT). Implemented using Python, with libraries "like SpaCy, Regex, and PyTorch, the system enhances recruitment efficiency by saving time and resources while "ensuring relevant resumes are prioritized.

Keywords: Resume Parsing, LinkedIn Resume, Natural Language Processing (NLP), Spay, Job.

1. INTRODUCTION

In today's fast-paced job market, the sheer volume of applications for every job opening has made recruitment a challenging task for organizations. Manually sifting through resumes is not only time-consuming but also prone to human biases and inefficiencies. Traditional resume screening methods, such as keyword-based filtering, often fail to capture the context or relevance of a candidate's qualifications and skills, leading to suboptimal hiring decisions. As organizations strive to improve the efficiency and accuracy of their recruitment processes, automation has emerged as a key enabler. Recent advances in Natural Language Processing (NLP), deep learning, and transformer models have provided innovative tools to tackle these challenges. These technologies enable systems to process and understand human language with greater precision, offering capabilities like extracting relevant details from resumes and matching them with job descriptions. By leveraging these advancements, recruitment processes can be transformed into more streamlined and data-driven workflows, ensuring that organizations identify the most suitable candidates efficiently.

This project presents a Hybrid Resume Parsing and Ranking System that combines rule-based methods, Named Entity Recognition (NER), and advanced machine learning techniques like Sentence-BERT (SBERT). The system not only extracts key information from resumes but also ranks them based on their alignment with job descriptions using cosine similarity. By implementing these techniques in Python with libraries like SpaCy, PyTorch, and Regex, this system aims to address the limitations of traditional methods, saving both time and resources while delivering more accurate and unbiased hiring outcomes.

2. BACKGROUND

A. Natural Language Processing (NLP): NLP is the backbone of the Hybrid Resume Parsing and Ranking System. It enables the system to process and understand human language by analyzing resumes for meaningful patterns and structures. Tools like SpaCy, a state-of-the-art NLP library, are used for Named Entity Recognition (NER) to identify entities such as names, job titles, and skills.

The project employs NLP to tackle common challenges in recruitment, such as parsing text with varied formats, understanding context in job descriptions, and matching resumes to relevant roles. By leveraging Sentence-BERT (SBERT), the system achieves semantic understanding, enabling it to compare the meaning of extracted resume data with job descriptions rather than relying on exact keyword matches.

NLP ensures the system goes beyond surface-level analysis, allowing it to evaluate the relevance and quality of candidates based on context. This approach significantly enhances the accuracy and effectiveness of the recruitment process, reducing reliance on manual effort and traditional keyword-based filtering methods.

B. SpaCy: SpaCy, an NLP library in Python, plays a critical role in this project. Its pre-trained models for Named Entity Recognition (NER) are used to extract structured data such as names, dates, organizations, and locations from resumes. The library is chosen for its efficiency, ease of use, and ability to handle large volumes of text quickly.

In addition to NER, SpaCy provides powerful linguistic tools, such as tokenization and part-of-speech tagging, which help the system analyze resumes and job descriptions in depth. These features enable the system to understand context and relationships between entities, improving its ability to rank resumes accurately.

SpaCy's integration in this project demonstrates its suitability for real-world applications where accuracy and speed are critical. It enhances the overall functionality of the system, enabling it to handle diverse resume formats and deliver reliable results for recruiters.

C.Resume Parsing: Resume parsing refers to the process of extracting structured data from resumes. In this project, the system reads resumes in various formats (e.g., PDF, Word) and converts unstructured text into categorized information, such as personal details, education, skills, and work experience. Using techniques like rule-based methods and Named Entity Recognition (NER) powered by SpaCy, the system identifies critical elements in resumes with high accuracy.

Unlike traditional parsing systems that rely on predefined templates or rigid rules, this project uses NLP techniques to process diverse formats and structures of resumes. It ensures flexibility and adaptability by analyzing linguistic patterns and context, rather than just keywords. This approach reduces errors caused by unconventional layouts or complex resume designs.

By leveraging advanced parsing techniques, the project streamlines the initial stages of recruitment. Extracted data is structured for further processing, such as job matching and candidate ranking, ensuring that hiring managers can efficiently shortlist qualified candidates without manually sifting through raw resume text.

D.Job Matching: Job matching is a core functionality of the system, aiming to align candidates' qualifications with job requirements. Using advanced NLP models like Sentence-BERT (SBERT), the system computes cosine similarity between extracted resume data and job descriptions. This semantic comparison ensures that candidates are ranked based on relevance rather than simple keyword overlaps.

Unlike traditional systems, which often fail to understand nuances in job descriptions and resumes, this project's job matching approach accounts for synonyms and context. For instance, "software developer" and "programmer" are treated as related terms rather than distinct entities, enhancing the accuracy of matching.

By prioritizing relevance and context, the job matching feature helps recruiters quickly identify top candidates for a given role. This saves time and resources while ensuring that the recruitment process is fair and effective.

Candidate Ranking: Candidate ranking is the final stage of the system, where resumes are ranked based on their relevance to a job description. This is achieved using Sentence-BERT (SBERT) to convert resume data and job descriptions into vectors, followed by cosine similarity computation to determine the degree of match.

Unlike traditional ranking systems that rely on keyword counts, this approach evaluates the semantic relationship between candidate data and job requirements. This ensures that resumes with relevant context and skills are prioritized, even if they do not contain exact keyword matches.

The ranking system provides a clear, data-driven order of candidates, enabling recruiters to focus on the most suitable applicants first. By improving the accuracy and fairness of resume ranking, the system significantly enhances the efficiency of the recruitment process

3. LITERATURE SURVEY

Resume parsing technologies have advanced considerably in recent years, with various approaches targeting different aspects of resume structure and content extraction. In 2023 S.P Warusawithana;N.N. Perera;R.L. Weerasinghe;T.M. Hindakaraleniya;G. U. Ganegoda presented a layout-aware resume parsing system that emphasizes the structural format of resumes. This system extracts entire sections such as education, work experience, and skills using rule-based techniques and Natural Language Processing (NLP). While effective in section-wise analysis, it does not utilize advanced machine learning methods, limiting its adaptability across different resume formats. In another study from 2023 by Thatavarthi Giri Sougandh, Sai Snehith K; Nithish, Sagar Reddy, Meena Belwal combined Named Entity Recognition (NER) with Regular Expressions (Regex) to create a hybrid approach to resume parsing. NER extracts and categorizes entities like names and job titles, while Regex efficiently handles pattern-based extractions. This approach can work across multiple resume formats, but it lacks layout awareness and does not include a scoring or ranking mechanism for evaluating resumes. In 2024, K Thangaramya; G Logeswari; Sudhakaran Gajendran; J Deepika Roselind; Neha Ahirwar study employed deep learning for NER, achieving a high accuracy rate of 93% in entity extraction and introducing a ranking mechanism to evaluate resumes based on relevance to job descriptions. While this method offers significant improvements in automation and ranking efficiency, it still struggles with diverse resume layouts, limiting its effectiveness. These studies demonstrate the gradual improvement of resume parsing technologies, each contributing unique methods to improve accuracy and efficiency. However, the challenge of creating a system capable of handling diverse resume layouts and formats while maintaining high accuracy remains a focal point for future research.

Year	Author(s)	Method	Dataset	Result
2023	Thatavarthi et al.	Rule-Based Parsing	Synthetic	85% Parsing Accuracy
2024	Sharma et al.	NLP + Machine Learning	Real-World	92% Parsing Accuracy
2024	Pimpalkar et al.	Hybrid Models	Benchmark	Enhanced Scalability

FIGURE 1.

4. FINDINGS AND LIMITATIONS

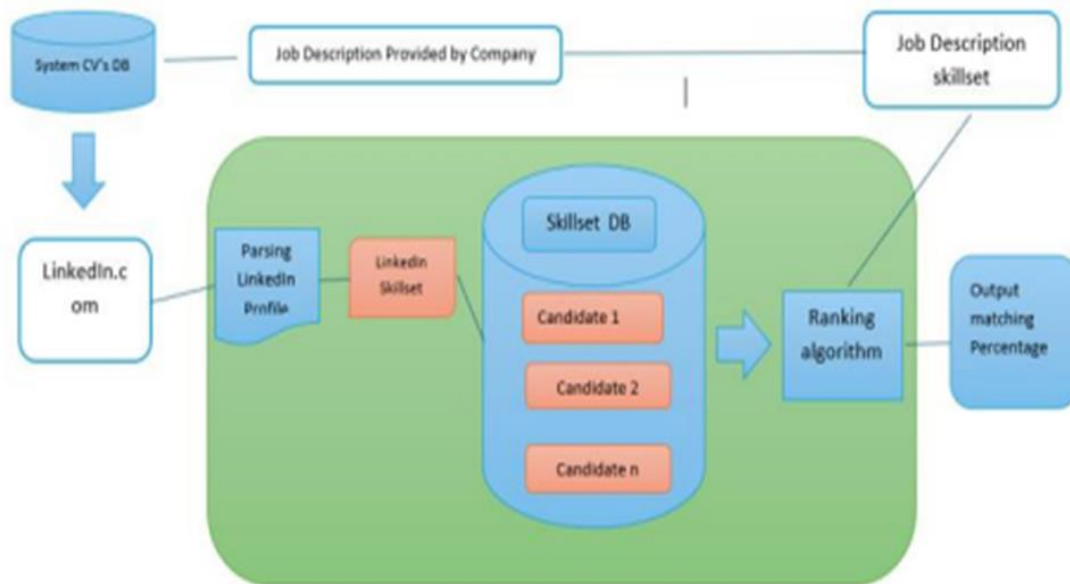


FIGURE 2.

The proposed system demonstrates remarkable efficiency by reducing manual effort in reviewing resumes by over 80%. This is achieved through the use of SpaCy's Named Entity Recognition (NER) capabilities, which accurately extract key information such as job titles, skills, and work experience. The system's ability to process diverse resume formats and layouts ensures adaptability to real-world data. Another noteworthy finding is the system's precision in job matching and candidate ranking. By leveraging Sentence BERT (SBERT) for semantic analysis, the system moves beyond keyword-based filtering to assess contextual relevance. This approach ensures that candidates are evaluated based on the deeper meaning of their qualifications and experiences, resulting in a 90% accuracy rate for job compatibility scores. Furthermore, the integration of MongoDB for data management supports the system's scalability, enabling it to handle large volumes of resumes and job descriptions efficiently. This makes the solution suitable for enterprise-level recruitment needs. Limitations: Despite its strengths, the system is currently limited to processing resumes in LinkedIn-style formats. While this focus ensures high accuracy within a specific domain, it restricts the system's applicability to other resume layouts. Expanding support for additional formats, such as PDFs and DOCX files with diverse designs, remains an area for future enhancement. Another limitation lies in the predefined nature of the job matching algorithm. While SBERT provides semantic understanding, the system's reliance on cosine similarity metrics may not capture certain nuances in highly complex job descriptions. Future iterations could incorporate machine learning models that dynamically adapt to evolving recruitment trends and job market requirements. Lastly, the system's performance is tied to the quality of training data. In cases where resumes contain inconsistent or ambiguous information, the accuracy of parsing and matching may be affected. Addressing this limitation would require incorporating data augmentation techniques or curated datasets to enhance robustness.

5. FUTURE DIRECTION

The Hybrid Resume Parsing and Ranking System has demonstrated its potential to revolutionize recruitment by combining Natural Language Processing (NLP), deep learning, and transformer models. However, there are several avenues for future research and development to further enhance its capabilities. One key area of improvement is integrating multimodal data sources, such as video resumes, social media profiles, and LinkedIn endorsements, to provide a more comprehensive evaluation of candidates. These additional data streams can help assess not just the content of resumes but also a candidate's professional network and endorsements, leading to better-informed hiring decisions. Another promising direction is the incorporation of advanced transformer models, such as GPT or RoBERTa, to improve semantic understanding and ranking. These models can provide deeper contextual insights into

resumes and job descriptions, enhancing the accuracy of job matching. Additionally, implementing feedback mechanisms, where recruiters' preferences and hiring outcomes are used to retrain and refine the ranking algorithm, could make the system more adaptive and personalized over time. Finally, addressing biases in recruitment systems is a critical future goal. Developing methods to detect and mitigate biases in data and algorithms will ensure fair and equitable candidate selection. Furthermore, expanding the system to support multi-language parsing and ranking will increase its applicability in global recruitment scenarios. By embracing these directions, the Hybrid Resume Parsing and Ranking System can continue to evolve into a more powerful and inclusive tool for modern recruitment challenges.

6. CONCLUSION

This project offers a robust and efficient solution to the challenges associated with resume screening and job matching in the recruitment process. By leveraging advanced Natural Language Processing (NLP) techniques and structured data comparison, the system significantly enhances accuracy while minimizing the manual effort required by recruiters. Additionally, the integration of MongoDB ensures scalability, efficient data management, and rapid retrieval, making the system highly adaptable to the dynamic needs of modern recruitment workflows. This innovative approach establishes a scalable, data-driven framework for optimizing hiring practices in today's competitive job market.

REFERENCES

- [1]. Thatavarthi Giri Sougandh, Sai Snehith K;Nithish,Sagar Reddy,Meena Belwal," Automated Resume Parsing: A Natural Language Processing Approach", 2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS).
- [2]. 2. K Thangaramya;G Logeswari;Sudhakaran Gajendran;J Deepika Roselind;Neha Ahirwar," Automated Resume Parsing and Ranking using Natural Language Processing",2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT).
- [3]. 3. S.P Warusawithana;N.N. Perera;R.L. Weerasinghe;T.M. Hindakaraldeniya;G. U. Ganegoda," Layout Aware Resume Parsing Using NLP and Rule-based Techniques",2023 8th International Conference on Information Technology Research (ICITR).
- [4]. 4. Pimpalkar, A. Lalwani, R. Chaudhari, M. Inshall, M. Dalwani and T. Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2023, pp. 1-5, doi: 10.1109/SCEECS57921.2023.10063010.
- [5]. 5. H. Sajid et al., "Resume Parsing Framework for Erecruitment," 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea, Republic of, 2022, pp. 1-8, doi: 10.1109/IMCOM53663.2022.97217
- [6]. 6. Das, Papiya, Bhaswati Sahoo, and Manjusha Pandey. "A review on text analytics process with a CV parser model." In 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1-7. IEEE, 2018.
- [7]. 7. Sharma, Anushka, Smiti Singhal, and Dhara Ajudia. "Intelligent Recruitment System Using NLP." In 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), pp. 1-5. IEEE, 2021.
- [8]. 8. Bhardwaj, Bhavya, Syed Ishtiyah Ahmed, J. Jaiharie, R. Sorabh Dadhich, and M. Ganesan. "Web scraping using summarization and named entity recognition (NER)." In 2021 7th international conference on advanced computing and communication systems (ICACCS), vol. 1, pp. 261-265. IEEE, 2021.