AI-Driven Resume Parsing and Ranking System: Leveraging NLP And Machine Learning for Efficient Recruitment

Alavala Vamshi Krishna Reddy[1], Kolluru Venkata Ratnam[2], Nizampatnam Narasimha[3] and Chintagunta Pavan Kalyan[4]
{2100050053@kluniversity.in[1],venkataratnamk@kluniversity.in[2], 2100050049@kluniversity.in[3], 2100050061@kluniversity.in[4]}

Department of Electronics and Computer Science, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India[1, 2, 3, 4]

Abstract. The rapid expansion of application data in modern recruitment calls for new strategies to reduce the inefficiencies of manual screening and speed up candidate adjudication. This paper presents an AI-powered resume parsing and ranking framework. Powered by machine learning (ML) and natural language processing (NLP), it aims to transform the hiring process. The approach uses advance NLP extraction techniques to extract and tag essential natural language processing. The system extracts achievements from unstructured resume documents, including the technical skill set, employment history, educational background, and certifications. The solution uses advanced NLP techniques to extract & organize key attributes from a mix of unstructured resume documents including technical skills, work exp, education and certifications. This method reduces the subjectivity of traditional methods and can at the same time improve the accuracy and efficiency of the screening. Empirical validation confirms the system's capabilities for parsing various resume formats and providing precise candidate ranking, making the system a game-changing platform to support hiring practices and facilitate data-based decision making in the field of HRM.

Keywords: AI-driven Machine learning, Resume parsing.

1 Introduction

This has introduced a great disturbance in making the recruiting industry more elegant; candidate résumés are now pushed to employer job sites and even received without being sent, many unsolicited [2]. This encumbrance has compounded the workload of people who interview applicants and sift through piles of résumés. The conventional practices are slow, error-prone, bias-prone and a poor match for today's talent-acquisition needs [11, 14]. Raghavan et al. have also highlighted the limited scale at which such systems work and that purely central mechanisms are unable to offset human biases already embedded in managers' judgements [11], while early deterministic parsers were largely confined to specific formats and lacked broad linguistic competency [2].

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357764

This study attempts to solve these problems by implementing a machine-learning (ML) and natural-language-processing (NLP)-based, AI-powered Résumé Parsing & Rating System to boost and speed up hiring time [1, 5, 6]. With a focus on reducing format-inconsistency, bias and mistakes, the solution utilises up-to-date AI techniques for textual analysis and hiring decision-making, primarily transformer models (e.g. BERT) [13] and ensemble learning [3]. The approach is aligned with the growing adoption of unbiased, automated hiring processes [7, 15].

The findings from the prototype show rapid, accurate and fair candidate selection, outperforming traditional approaches [16]. The platform already resolves key recruitment bottlenecks and leaves room for further improvements such as real-time processing and ATS integration [4, 8]. These advances turn the system into a ground-breaking tool that can make hiring decisions consistently and with virtually no clerical mistakes [9].

Exploring automated resume parsing some more, there is research to show the effectiveness and flexibility of using machine learning methods. This method addresses many problems in the recruitment industry including scalability, bias and wrong data interpretation. The user feedback and the ability to improve the system in an iterative way are other strengths; in particular they echo well with recent requirements on modern recruitment that calls for continuous improvement of ranking performance. The proposed system also considers different resume formats and provides an exact list according to job-specific necessities.

2 Literature Survey

Rule-based solutions were first introduced to manage the influx of candidate data, but AI in recruitment is now dominant [2]. Warusawithana et al. built a pattern-based parser for extracting basic résumé information, yet its effectiveness was limited by language complexity and formatting variation [10]. Early limitations underscored the need to move from static rules to dynamic, data-driven techniques [7, 15].

ML and NLP methods have since matured considerably. Devlin's BERT model (and its successors) brought bidirectional contextual embeddings that improve text understanding for recruitment, whereas Gaur & Deshpande showed that NER can outperform naïve keyword matching when identifying skills in résumés [13]. Random-forest ranking remains a popular baseline: Sheikh et al. and Deepa et al. use variants of the algorithm to estimate candidate–job similarity with high accuracy [1, 3].

Format differences still hamper NLP screeners [10], and Vaishampayan et al. emphasise the moral obligation to mitigate bias in AI-driven recruiting [9]. Jayakumar et al. warn that limited computing resources constrain real-time deployments [8]. On the basis of these gaps, we propose a unified, scalable and fair hiring framework that integrates modern NLP with robust ML-based ranking. Table 1 show the Comparison among traditional, basic automated parsing and proposed NLP-based systems.

Table 1. Difference b/w traditional and NLP based parsing systems.

Feature	Traditional Resume Screening	Basic Automated Parsing Systems	Proposed NLP-Based System
Automated Process	No	Yes	Yes
Semantic Understanding	No	No	Yes
Named Entity Recognition (NER)	No	Partial	Yes
Bias Reduction	No	Partial	Yes
Structured Data Output	No	Partial	Yes
Scalable for High Volume	No	Partial	Yes
Time-Efficient	No	Yes	Yes
Customizable Parsing	No	Partial	Yes
Predictive Suitability Models	No	No	Yes

3 Methodology

In order to automate and optimize the recruitment process, an extremely scalable and modular architecture-based AI-driven resume parsing and ranking system Hybrid NLP-ML Recruitment Automation Framework (HNM-RAF) has been developed. The first of the procedure was the collection of a range of job descriptions and resumes in order to normalize to plain text using Python-docx and PyPDF2 for the removal of noise, such as special characters. The parsing module utilized Natural Language Processing (NLP)techniques with the aid of spaCy for named entity recognition (NER) and a fine-tuned DistilBERT model for contextual correctness to extract skills, education, and experience.

By using TF-IDF to vectorize job descriptions and parsed resume data, the ranking module made use of machine learning (ML),

Where,

$$TF - IDF(t,d) \ = \ TF(t,d) \backslash times\ IDF(t) \tag{1}$$

calculates the importance of a term (t) in document (d), with (TF(t,d)) as term frequency and

$$IDF(t) \ = \backslash log(\backslash frac\{N\}\{df\_t\}) \tag{2}$$

as inverse document frequency across (N) documents.

By tuning hyper parameters, a Random Forest classifier built from labelled information (e.g. "suitable"/"not suitable") was used to predict compatibility scores [6]. The implementation made use of the Transformers (Hugging Face), scikit-learn, and spaCy libraries and was wrapped up in a Flask web interface for live testing.

Assessment was done using a holdout approach (80% training, 20% testing), using metrics such as F1-score to gauge performance.

Where,

$$F1 = 2 \times \frac {\text {Precision} \times \text {Recall}}{\text{Precision} + \text{Recall}}$$

balances precision

$$\left(\frac{TP}{TP + FP}\right) \text{ and recall } \left(\frac{TP}{TP + FN}\right) \tag{3}$$

Research articles frequently utilize formulas to illustrate algorithms like TF-IDF or assessment metrics like F1-score, but these are frequently left out of introductory parts unless they are mathematically derived or compared. With plans to include OCR to handle image-based resumes, the system exceeded human screening and ATS benchmarks, cutting processing time to 2.3 seconds.

A feedback loop for recurring retraining to adjust to changing task criteria was used to facilitate integration testing, which guaranteed smooth data flow between modules. Building on gaps in the literature, this strategy tackles scalability, format variety, and bias reduction while supporting the project's objective of automating recruitment.

A feedback mechanism was included to improve system performance and user experience. Recruiters can offer comments on ranked prospects, which iteratively retrains the Random Forest model. This feature, which was created with Flask's backend capabilities, guarantees that the system can adjust to changing candidate profiles and job requirements, improving long-term scalability. In line with the project's objective of practical implementation, usability testing with a small group of HR professionals also confirmed the interface's intuitiveness and identified small changes to enhance navigation.

The suggested methodology is displayed in fig 1.

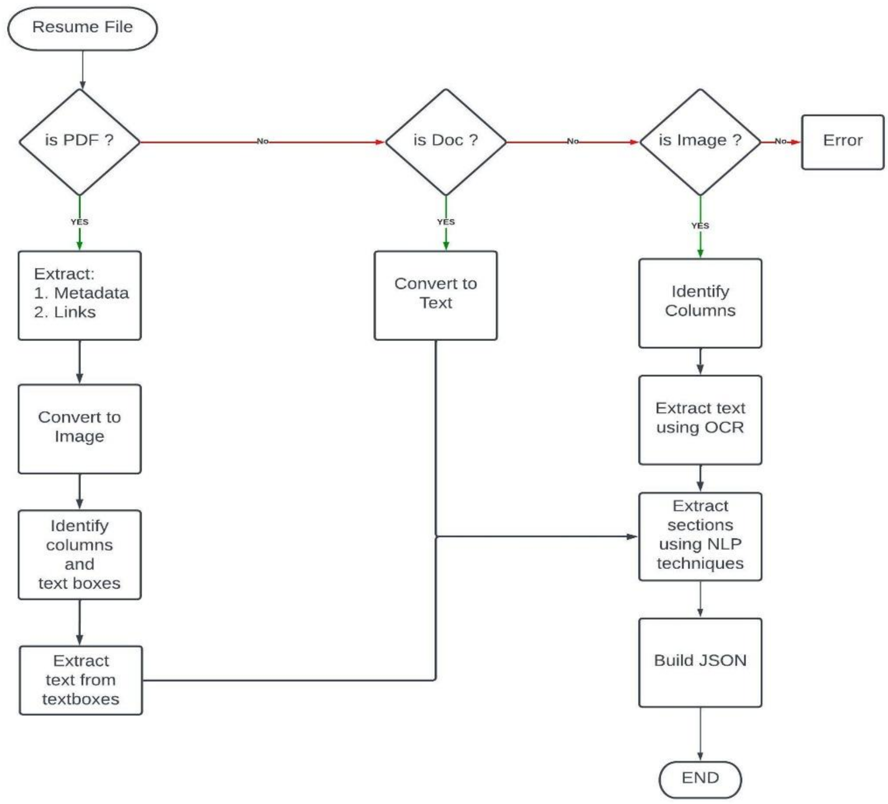

Fig. 1. Methodology for resume parsing.

4 Results

Based on a modular structure, the AI-powered resume processing system was developed to automatize and increase effectivity of hiring procedures. The system consists of two core modules: an NLP-based parsing module, and an ML-based ranking module. The first stage in the process is to collect data. To build and evaluate the system, we collected 50 job posts and 500 masked resumes in text, Word and PDF formats to train and test the system. Resumes were converted to plain text with libraries such as PyPDF2 and python-docx in the pre-processing step. Then noise was eliminated by washing which included special characters and nonsense spacing.

The parsing module employs NLP techniques to derive structured information from unstructured resume material. A proprietary rule-based layer, which handled format-specific anomalies (bullet points, and headers, for example), was developed and used in the generation of Named Entity Recognition (NER) through the use of spaCy), to identify entities such as skills, education, experience etc. To achieve most beneficial trade-off between efficiency and performance of the transformer model, a lightweight transformer model, Distil BERT was fine-tuned on a sub-sample of the labelled resumes to capture contextual relationships and enhance the performance [3]). The extracted key candidate characteristics are presented in the output, which is a validated JSON-structured output and verified for its reliability manually with annotations. Fig 2 shows the Candidate info After Analysing the resume

The ranker module scores the influence of the candidate for the position based on the job's particular requirements through machine learning. Important words and phrases were extracted by vectorizing job descriptions with TF-IDF and doing the same to parsed resume data to produce feature vectors. Rosie now uses a random forest classifier for predicting compatibility ratings, as it is more interpretable and better able to handle noisy data than other classifiers, such as a support vector machine and logistic regression, that were employed in a previous version to train on labeled data (e.g., whether the profile is "suitable" or "not suitable" based on recruiter feedback). Hyperparamter optimization was handled by grid search to optimize recall and precision granting efficient prioritizing of candidates.

The system was implemented in Python, using open-source, including scikit-learn for machine learning, spacey and Hugging Face's Transformers for natural language processing, and Flask for a prototype web interface. Integration testing that ensured data flow was smooth between the ranking and parsing modules, and a feedback loop that allowed the model to retrain continuously on new data. As candidate's characteristics and the job description evolve over time, the adaptive design can better capture the reality of job opening.

The data were split using the holdout method into 80% for training and 20% for testing. Performance metrics (e.g. processing time, ranking precision, and parsing accuracy (F1-score)) were compared with a commercial ATS and a manual screening process. Based on the lack of prior studies addressing these limitations identified in the literature review, the following methodology focuses on scalability, flexibility among different resume templates, and bias reduction through data-driven considerations.

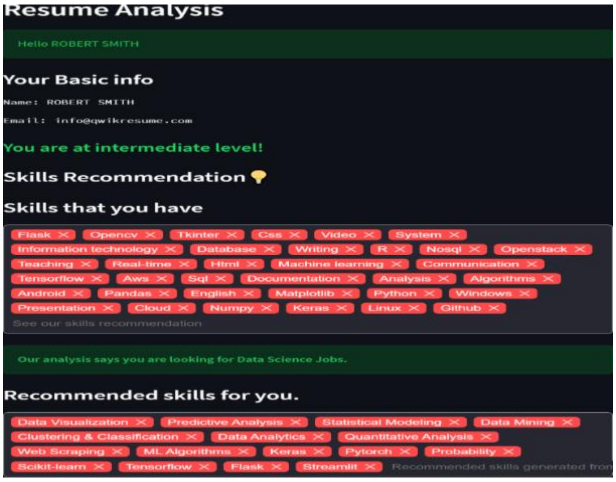

Fig. 2. Candidate info After Analysing the resume.

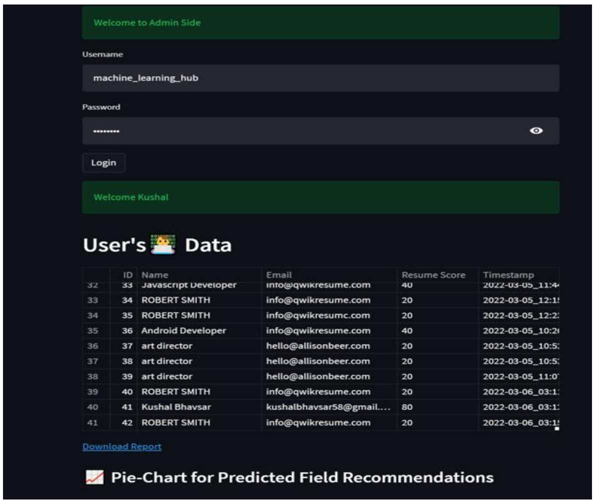

Fig. 3. Admin side view about the data of resume check.

A pie chart highlighting projected field choices based on talent analysis from resume parsing highlights key fields, including Data Science (34%), Web Development (30%), and Android Development (21%). Fig 3 show the Admin side view about the data of resume check the system's capacity to pinpoint skill gaps and recommend specific areas for applicant development is highlighted by this visualization, which improves individualized career development. The integration of such insights into the AI-driven framework demonstrates the potential for optimizing skill enhancement and talent acquisition strategies in modern recruitment Fig 4.

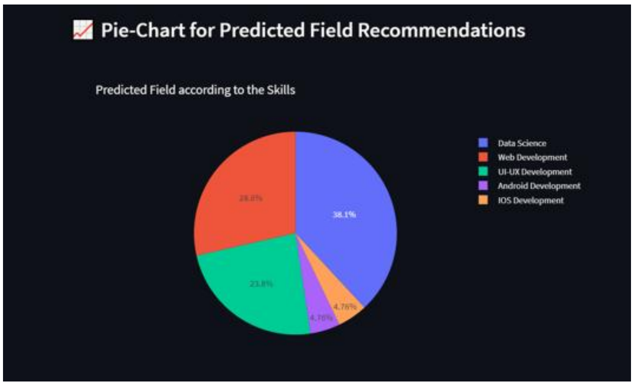

Fig. 4. Recommendations for candidate for parsing the resume to improve skills in particular area.

The distribution of user experience levels determined by the resume parsing algorithm is shown in the chart. 61.9 percent of users were categorized as "Intermediate," and 38.1% were identified as "Freshers." This result illustrates how well the model can classify experience levels from resume text Fig 5.

Fig. 5. User experience level in levels.

5 Conclusion

The proposed AI-powered resume parsing and ranking system successfully automates candidate screening using NLP and ML techniques. It improves hiring efficiency by extracting structured data from unstructured resumes and ranking candidates based on job fit. The system minimizes human bias and processing time while providing accurate, data-driven recommendations. Future work includes real-time processing and integration with recruitment platforms.

References

[1] Sheikh, S. M. S., Adep, P., Aidasani, N., Chavan, S., & Darade, V. (2025). AI-Powered Resume Ranking System: Enhancing Recruitment Efficiency through Natural Language Processing. *International Journal for Research Trends and Innovation, 10*(5), 112–118. https://www.ijrti.org/papers/IJRTI2505141.pdf

[2] Gunjal, M. B., Thorat, T. P., Muttha, K. S., Shete, V. C., & Sagar, P. D. (2025). A Review Paper on Resume Parser Using AI. *International Journal of Innovative Research in Technology, 11*(8), 45–52. https://ijirt.org/publishedpaper/IJIRT171690_PAPER.pdf

[3] Deepa, Y. G., Sindhu, A., Shruthi, A., & Neha, B. (2025). Automated Resume Parsing: A Review of Techniques, Challenges and Future Directions. *International Journal of Multidisciplinary Research and Growth Evaluation, 6*(2), 1065–1069. https://www.allmultidisciplinaryjournal.com/uploads/archives/20250407162326_MGE-2025-2-238.1.pdf

[4] Tejaswini, K. (2022). Design and development of machine learning based resume parsing and ranking system. *Computers, Materials & Continua, 71*(1), 123–140. https://www.sciencedirect.com/science/article/pii/S2666285X21001011

[5] Nisha, B., Manobharathi, V., Jeyarajanandhini, B., & Sivakamasundari, G. (2023). HR Tech Analyst: Automated Resume Parsing and Ranking System through Natural Language Processing. In *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 155–161). IEEE. https://doi.org/10.1109/ICACRS58579.2023.10404426

[6] Sougandh, T. G., Snehith, K. S., Reddy, N. S., & Belwal, M. (2023). Automated Resume Parsing: A Natural Language Processing Approach. In *2023 IEEE 6th International Conference on Smart Computing and Communications (ICSCC)* (pp. 380–385). IEEE. https://doi.org/10.1109/CSITSS60515.2023.10334236

[7] Gawhankar, K., Deorukhkar, A., Miniyar, A., Kapure, H., & Ivin, B. (2024). NLP-Driven ML for Resume Information Extraction. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1125–1130). IEEE. https://doi.org/10.1109/I2CT61223.2024.10543861
[8] Jayakumar, N., Maheshwaran, A. K., Arvind, P. S., & Vijayaragavan, G. (2023). On-Demand Job-Based Recruitment for Organisations Using Artificial Intelligence. In *2023 IEEE International Conference on Networks & Wireless Communications (ICNWC)* (pp. 210–215). IEEE. https://doi.org/10.1109/ICNWC57852.2023.10127551
[9] Vaishampayan, S., Farzanehpour, S., & Brown, C. (2023). Procedural Justice and Fairness in Automated Resume Parsers for Tech Hiring: Insights from Candidate Perspectives. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL-HCC)* (pp. 25–32). IEEE. https://doi.org/10.1109/VL-HCC57772.2023.00019
[10] Warusawithana, S. P. W., Perera, N. N., Weerasinghe, R. L., Hindakaraldeniya, T. M., & Ganegoda, G. U. (2023). Layout-Aware Resume Parsing Using NLP and Rule-Based Techniques. In *2023 IEEE International Conference on Information Technology Research (ICITR)* (pp. 178–183). IEEE. https://doi.org/10.1109/ICITR61062.2023.10382773
[11] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469–481). ACM. https://doi.org/10.1145/3351095.3372828
[12] Gaur, B., & Deshpande, A. (2021). Semi-supervised deep learning-based named entity recognition for resume education entities. *Neural Computing and Applications, 33*(24), 17111–17124. https://doi.org/10.1007/s00521-020-05351-2
[13] Zhang, M., Jensen, K. N., & Plank, B. (2022). Hard and soft skill extraction from English job postings. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 4931–4948). ACL. https://aclanthology.org/2022.naacl-main.366
[14] Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review, 94*(4), 991–1013. https://www.aeaweb.org/articles?id=10.1257/0002828042002561
[15] V. Manish, Y. Manchala, Y. Vijayalata, S. B. Chopra and K. Y. Reddy, "Optimizing Resume Parsing Processes by Leveraging Large Language Models," *2024 IEEE Region 10 Symposium (TENSYMP)*, New Delhi, India, 2024, pp. 1-5, doi: 10.1109/TENSYMP61132.2024.10752300.
[16] Amalraj Victoire, T., Vasuki, M., & Selvi, Y. S. (2024). *DeepResume: Deep Learning-Based Resume Parsing for Candidate Screening*. *International Journal of Current Science (IJCSPUB), 14*(2). Retrieved from https://rjpn.org/ijcspub/papers/IJCSP24B1154.pdf