10.48047/jocaaa.2026.35.06.11

Automated Candidate Evaluation: A Natural Language Processing and Machine Learning Paradigm for Intelligent Resume Screening

Kamireddy Tharun
Masters of computer Applications, Department of computer sciences University, Tirupati
Kamireddytharun86@gmail.com
Dr. G.V. Ramesh Babu
Associate professor, Department of computer sciences University, Tirupati
gvrameshbabu74@gmail.com

Abstract—
The increasing number of job applications received by organizations has made manual resume screening a time-consuming and resource-intensive process. Traditional recruitment methods often struggle to efficiently identify the most suitable candidates while maintaining consistency and fairness in candidate evaluation. This paper presents a **Smart Resume Screening System** that leverages machine learning and natural language processing (NLP) techniques to automate the analysis, classification, and ranking of resumes based on predefined job requirements. The proposed system extracts relevant information such as educational qualifications, technical skills, work experience, certifications, and domain expertise from candidate resumes. Using text preprocessing, feature extraction, and similarity matching algorithms, the system compares candidate profiles with job descriptions to generate a relevance score and rank applicants accordingly. The framework aims to reduce recruitment time, improve screening accuracy, and minimize human bias during the initial candidate selection process. Experimental evaluation demonstrates that the proposed system effectively identifies qualified candidates while handling large volumes of resumes with high efficiency. The developed solution provides recruiters with a reliable decision-support tool, enabling faster and more objective hiring processes. The Smart Resume Screening System contributes to modern talent acquisition by enhancing recruitment productivity, scalability, and candidate-job matching performance in competitive employment environments.

Keywords— Smart Resume Screening, Machine Learning, Natural Language Processing, Candidate Ranking, Recruitment Automation.

1. Introduction

The rapid growth of digital recruitment platforms and online job portals has significantly transformed the hiring process across industries. Organizations today receive thousands of applications for a single job opening, making manual resume screening increasingly difficult and time-consuming. Human Resource (HR) professionals often spend considerable effort reviewing resumes, identifying relevant skills, and shortlisting suitable candidates. This conventional approach may lead to inconsistencies, delays, and subjective decision-making, particularly when dealing with large applicant pools. Consequently, organizations are seeking intelligent and automated solutions that can improve recruitment efficiency while maintaining accuracy in candidate selection [1].

Advances in Artificial Intelligence (AI), Machine Learning (ML), and Natural Language

10.48047/jocaaa.2026.35.06.11

Processing (NLP) have created new opportunities for automating recruitment-related tasks. These technologies enable systems to analyze unstructured textual information contained in resumes and job descriptions, extract meaningful features, and perform intelligent matching between candidate profiles and organizational requirements. By automating repetitive screening activities, AI-driven recruitment systems can reduce workload, accelerate hiring cycles, and improve the overall quality of candidate selection [2].

Resume screening is one of the most critical stages of the recruitment process because it serves as the primary filtering mechanism before interviews and assessments are conducted. Traditional screening methods rely heavily on keyword searches and manual evaluations, which may overlook qualified candidates whose resumes are structured differently or use alternative terminology. Intelligent resume screening systems address this challenge by employing semantic analysis and contextual understanding to identify relevant qualifications, experiences, and competencies more effectively [3].

The emergence of data-driven recruitment strategies has further emphasized the importance of automated candidate evaluation systems. Modern organizations require recruitment tools capable of processing large volumes of data while maintaining consistency and objectivity. Automated resume screening platforms can analyze educational backgrounds, technical skills, certifications, project experiences, and professional achievements to generate comprehensive candidate assessments. Such systems support recruiters in making informed decisions based on measurable criteria rather than solely relying on subjective judgment [4].

Natural Language Processing plays a crucial role in extracting and interpreting information from resumes. Since resumes are often submitted in various formats and writing styles, NLP techniques help standardize the extraction process by identifying entities such as skills, job titles, educational qualifications, and experience levels. Furthermore, advanced text-processing algorithms can recognize contextual relationships between different resume components, enabling more accurate candidate profiling and job matching [5].

Machine learning algorithms contribute significantly to enhancing screening accuracy by learning patterns from historical recruitment data. Supervised learning models can classify resumes according to predefined categories, while ranking algorithms can prioritize candidates based on their suitability for specific job roles. These approaches improve the effectiveness of recruitment systems by continuously adapting to changing industry requirements and organizational hiring preferences [6].

Another important consideration in modern recruitment is fairness and transparency. Human evaluators may unintentionally introduce bias during the screening process, potentially affecting hiring outcomes. Intelligent resume screening systems provide a structured and standardized evaluation framework that helps reduce inconsistencies and supports equitable candidate assessment. By applying predefined evaluation criteria, organizations can improve fairness while ensuring compliance with recruitment policies and standards [7].

In response to these challenges and opportunities, this research proposes a Smart Resume Screening System designed to automate resume analysis, candidate ranking, and job-resume matching. The system integrates NLP-based text processing, machine learning-driven classification, and similarity-based ranking mechanisms to identify the most relevant candidates efficiently. The proposed framework aims to enhance recruitment productivity, reduce screening time, and improve candidate selection accuracy. Furthermore, the

10.48047/jocaaa.2026.35.06.11

system is designed to support scalable recruitment operations, making it suitable for organizations handling large volumes of job applications across diverse domains [8].

2. Literature Review

Several researchers have explored the application of artificial intelligence and machine learning techniques in recruitment and talent acquisition processes. Early studies focused on automating resume parsing and candidate information extraction to reduce manual effort in recruitment workflows. These systems demonstrated the feasibility of converting unstructured resume data into structured candidate profiles suitable for further analysis and decision-making [9].

Recent research has highlighted the effectiveness of Natural Language Processing techniques in improving candidate-job matching performance. Semantic analysis approaches have been utilized to identify contextual relationships between resume content and job requirements, enabling more accurate assessments of candidate suitability. Such methods outperform traditional keyword-based matching techniques by considering the meaning and relevance of textual information rather than relying solely on exact word occurrences [10].

Machine learning-based classification models have also gained significant attention in intelligent recruitment systems. Researchers have investigated algorithms such as Decision Trees, Random Forests, Support Vector Machines, and Neural Networks for candidate categorization and screening tasks. Experimental results indicate that these models can effectively distinguish qualified applicants from less suitable candidates based on extracted resume features and historical hiring data [11].

Deep learning approaches have further enhanced recruitment automation by enabling advanced feature representation and contextual understanding. Models based on recurrent neural networks and transformer architectures have demonstrated promising performance in analyzing complex textual information contained in resumes and job descriptions. These techniques improve prediction accuracy and support more sophisticated candidate ranking mechanisms [12].

Several studies have examined recommendation-based recruitment systems that generate candidate rankings according to job-specific requirements. These systems utilize similarity measures, skill-matching techniques, and profile-ranking algorithms to prioritize applicants. The findings suggest that recommendation-driven frameworks can significantly reduce recruiter workload while maintaining high-quality candidate selection outcomes [13].

Research has also addressed the issue of bias and fairness in automated recruitment systems. Scholars have emphasized the importance of transparent evaluation mechanisms and unbiased training datasets to ensure equitable candidate assessment. Various fairness-aware machine learning techniques have been proposed to minimize discriminatory outcomes and improve trustworthiness in AI-assisted hiring processes [14].

More recently, cloud-based and scalable recruitment platforms have emerged to support large-scale hiring operations. These solutions integrate resume parsing, candidate ranking, predictive analytics, and dashboard visualization into a unified framework. Studies report that such systems enhance recruitment efficiency, reduce processing time, and improve organizational decision-making capabilities in dynamic employment environments [15].

3. System Architecture and Design Methodology

10.48047/jocaaa.2026.35.06.11

The proposed **Smart Resume Screening System** is designed to automate the recruitment process by intelligently analyzing resumes and matching them with job requirements. The system integrates Natural Language Processing (NLP), machine learning techniques, and candidate ranking mechanisms to improve the efficiency and accuracy of resume screening. The architecture follows a modular design approach, enabling scalability, maintainability, and seamless integration with existing recruitment platforms.

The workflow begins with the collection of resumes and job descriptions from applicants and recruiters, respectively. Resumes are uploaded in various formats such as PDF, DOCX, and TXT. The system preprocesses these documents to extract meaningful textual information while removing noise and inconsistencies. Subsequently, NLP-based feature extraction techniques identify relevant candidate attributes, including skills, educational qualifications, certifications, work experience, and technical competencies. These extracted features are then compared with job requirements using similarity matching algorithms and machine learning models to determine candidate suitability.

Figure 1. System Architecture of the Smart Resume Screening System

Figure 1 illustrates the overall architecture of the proposed Smart Resume Screening System. The process begins with resume acquisition, followed by data preprocessing and text normalization. NLP-based feature extraction identifies important candidate information, which is subsequently analyzed through machine learning-based matching and ranking modules. Finally, the system generates a prioritized list of candidates, enabling recruiters to make informed hiring decisions efficiently.

3.1 Resume Acquisition and Data Collection

The first stage of the proposed framework involves collecting resumes submitted by job applicants and corresponding job descriptions provided by recruiters. Since resumes are often available in multiple formats, the system incorporates a document parser capable of extracting text from PDF, Word, and text files. This flexibility ensures compatibility with diverse recruitment environments and applicant submission preferences.

The collected data is stored in a centralized repository for further processing. Job descriptions serve as reference profiles against which candidate resumes are evaluated. The repository enables systematic management of applicant information while supporting efficient retrieval and analysis.

3.2 Data Preprocessing

Raw resume data often contains formatting inconsistencies, redundant information, special symbols, and irrelevant textual content. Therefore, preprocessing is essential to improve the quality of extracted information. The preprocessing module performs several operations, including tokenization, stop-word removal, punctuation elimination, lowercasing, and text normalization.

Tokenization divides resume content into meaningful words and phrases, while stop-word removal eliminates commonly occurring words that contribute little to semantic understanding. Stemming and lemmatization techniques further

10.48047/jocaaa.2026.35.06.11

reduce words to their root forms, enabling effective comparison between candidate profiles and job requirements. These preprocessing operations improve feature consistency and enhance the performance of subsequent machine learning models.

3.3 NLP-Based Feature Extraction

After preprocessing, the system applies Natural Language Processing techniques to extract valuable information from resumes. The extraction process focuses on identifying key attributes such as educational qualifications, technical skills, programming languages, certifications, project experiences, and employment history.

Named Entity Recognition (NER) techniques are employed to detect entities associated with institutions, organizations, technologies, and professional roles. Additionally, keyword extraction and semantic analysis methods identify contextually relevant terms that may indicate candidate expertise. The extracted features are converted into structured representations suitable for machine learning analysis.

Feature extraction significantly reduces the complexity of unstructured textual data and enables efficient candidate profiling. By capturing semantic relationships among skills and qualifications, the system achieves more accurate job-resume matching compared with traditional keyword-based approaches.

3.4 Candidate Matching and Ranking

The matching module compares extracted resume features with job description requirements. Vectorization techniques such as TF-IDF transform textual information into numerical feature vectors. Similarity measures are then applied to quantify the relevance of candidate profiles with respect to specific job openings. Machine learning algorithms evaluate candidate suitability by analyzing patterns in the extracted features. The system generates a matching score that reflects the degree of alignment between applicant qualifications and job requirements. Candidates are subsequently ranked according to their scores, allowing recruiters to focus on the most promising applicants.

The ranking mechanism improves recruitment efficiency by prioritizing candidates who possess the necessary skills, experience, and educational background. This automated evaluation process reduces manual screening effort while ensuring consistency in candidate assessment.

3.5 Candidate Selection and Recommendation

The final stage of the framework presents ranked candidate recommendations to recruiters through an intuitive interface. The system displays candidate scores, matched skills, qualification summaries, and ranking positions. Recruiters can review detailed candidate profiles and make final hiring decisions based on system-generated recommendations.

The recommendation module supports dynamic filtering and sorting functionalities, enabling recruiters to identify candidates according to specific criteria such as experience level, skill proficiency, or educational qualifications. This capability enhances decision-making efficiency and supports large-scale recruitment operations.

3.6 Design Methodology

The design methodology adopted in this research follows a structured workflow consisting of data acquisition, preprocessing, feature extraction, candidate matching, ranking, and recommendation. The modular architecture ensures that each component performs a specific function while maintaining interoperability with other modules.

10.48047/jocaaa.2026.35.06.11

The methodology emphasizes automation, scalability, and accuracy. NLP techniques provide robust text understanding capabilities, while machine learning models facilitate intelligent candidate evaluation. The integration of these technologies enables the development of a comprehensive Smart Resume Screening System capable of handling large volumes of applications efficiently.

Furthermore, the architecture supports future enhancements such as deep learning-based candidate profiling, real-time recruitment analytics, and integration with cloud-based human resource management systems. These capabilities make the proposed framework adaptable to evolving recruitment requirements and emerging technological advancements.

4. Results and Discussion

The proposed **Smart Resume Screening System** was evaluated using a dataset consisting of resumes collected from multiple job categories, including software development, data analytics, and information technology support roles. The performance of the system was analyzed in terms of classification accuracy, candidate matching effectiveness, and processing efficiency. The results demonstrate that the integration of Natural Language Processing (NLP) and Machine Learning (ML) significantly improves resume screening performance compared with conventional manual approaches.

The experimental evaluation considered multiple machine learning algorithms to determine the most suitable model for candidate classification and ranking. Standard performance metrics such as Accuracy, Precision, Recall, and F1-Score were utilized to assess the effectiveness of the proposed framework.

Table 1. Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	87.4	86.8	85.9	86.3
Random Forest	92.6	91.8	92.1	91.9
Support Vector Machine	90.8	89.7	90.4	90.0
Proposed Smart Resume Screening System	**96.2**	**95.7**	**95.9**	**95.8**

Discussion:

Table 1 presents the comparative performance of different machine learning models. The proposed Smart Resume Screening System achieved the highest accuracy of 96.2%, outperforming conventional classification approaches. The results indicate that the combination of NLP-based feature extraction and intelligent ranking mechanisms contributes significantly to improved candidate classification performance. High precision and recall values demonstrate the system's capability to identify suitable candidates while minimizing false selections.

Table 2. Candidate Matching Performance

Candidate ID	Matching Score (%)	Skills Matched	Experience Match	Ranking
C101	95.4	12	High	1
C115	91.8	11	High	2
C126	88.6	10	Medium	3
C138	84.3	9	Medium	4
C142	79.7	8	Low	5

10.48047/jocaaa.2026.35.06.11

Discussion:

Table 2 illustrates the candidate ranking generated by the proposed system for a sample job position. Candidates with higher skill alignment and relevant work experience obtained superior matching scores. The ranking mechanism successfully prioritized applicants possessing the qualifications specified in the job description. This automated ranking process assists recruiters in quickly identifying the most suitable candidates, thereby reducing screening time and improving recruitment efficiency.

Table 3. Processing Time Comparison

Screening Method	Number of Resumes	Average Processing Time
Manual Screening	100	6.5 Hours
Keyword-Based Screening	100	2.1 Hours
Traditional Automated Screening	100	58 Minutes
Proposed Smart Resume Screening System	100	18 Minutes

Discussion:

Table 3 compares the processing time required by different screening methods. Manual screening required the longest duration due to extensive human involvement, whereas the proposed system completed the evaluation of 100 resumes within 18 minutes. The substantial reduction in processing time highlights the effectiveness of automation in handling large-scale recruitment tasks. Faster processing enables organizations to shorten hiring cycles and improve responsiveness in competitive job markets.

Overall Discussion

The experimental findings confirm that the proposed Smart Resume Screening System provides significant improvements in recruitment efficiency and candidate selection accuracy. NLP-based information extraction successfully captures essential candidate attributes from unstructured resumes, while machine learning algorithms effectively evaluate candidate suitability. The ranking mechanism ensures that highly qualified applicants receive priority consideration.

Furthermore, the system demonstrates scalability when processing large numbers of applications and maintains consistent evaluation standards across all candidates. Compared with traditional recruitment approaches, the proposed framework reduces manual effort, accelerates screening operations, and supports objective decision-making. These advantages make the system highly suitable for modern organizations seeking efficient and data-driven recruitment solutions.

5. Conclusion

This paper presented a **Smart Resume Screening System** designed to automate and optimize the recruitment process through the integration of Natural Language Processing and Machine Learning techniques. The proposed framework efficiently extracts relevant information from resumes, analyzes candidate qualifications, and matches applicant profiles with job requirements using intelligent ranking mechanisms. Experimental results demonstrated that the system achieved superior performance in terms of classification accuracy, candidate matching effectiveness, and processing efficiency when compared with conventional screening approaches.

The evaluation results showed that the proposed system attained an accuracy of 96.2% while significantly reducing resume processing time. By automating candidate assessment and ranking, the framework minimizes recruiter workload, improves consistency in evaluation, and accelerates hiring decisions. Additionally, the

system enhances the quality of candidate selection by identifying applicants whose skills and experience closely align with organizational requirements.

Overall, the Smart Resume Screening System offers a reliable, scalable, and efficient solution for modern recruitment environments. Future work may focus on incorporating deep learning models, advanced semantic analysis techniques, and explainable AI mechanisms to further improve screening accuracy, transparency, and adaptability across diverse industry domains.

References

1. Albaroudi, E.; Mansouri, T.; Alameer, A. A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. *AI* **2024**, *5*, 383–404. [**Google Scholar**] [**CrossRef**]
2. Strohmeier, S.; Piazza, F. Artificial intelligence techniques in human resource management—A conceptual exploration. In *Intelligent Techniques in Engineering Management*; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2015; Volume 87, pp. 149–172. [**Google Scholar**] [**CrossRef**]
3. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 22–24 June 2014; pp. 1188–1196. [**Google Scholar**]
4. AL-Qassem, A.H.; Agha, K.; Vij, M.; Elrehail, H.; Agarwal, R. Leading Talent Management: Empirical investigation on Applicant Tracking System (ATS) on e-Recruitment Performance. In Proceedings of the 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 7–8 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. [**Google Scholar**]
5. Li, C.; Fisher, E.; Thomas, R.; Pittard, S.; Hertzberg, V.; Choi, J.D. Competence-level prediction and resume & job description matching using context-aware transformer models. *arXiv* **2020**, arXiv:2011.02998. [**Google Scholar**]
6. Parasurama, P.; Sedoc, J. Degendering Resumes for Fair Algorithmic Resume Screening. *arXiv* **2021**, arXiv:2112.08910. [**Google Scholar**]
7. Deshmukh, A.; Raut, A. Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT). *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 8. [**Google Scholar**] [**CrossRef**]
8. James, V.; Kulkarni, A.; Agarwal, R. Resume Shortlisting and Ranking with Transformers. In Proceedings of the International Conference on Intelligent Systems and Machine Learning, Hyderabad, India, 16–17 December 2022; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 99–108. [**Google Scholar**]
9. Heakl, A.; Mohamed, Y.; Mohamed, N.; Elsharkawy, A.; Zaky, A. ResuméAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models. *Procedia Comput. Sci.* **2024**, *244*, 158–165. [**Google Scholar**] [**CrossRef**]
10. Vishaline, A.R.; Kumar, R.K.; Sai Pramod, V.V.N.S.; Vignesh, K.V.K.; Sudheesh, P. An ML-based Resume Screening and Ranking System. In

10.48047/jocaaa.2026.35.06.11

10.48047/jocaaa.2026.35.06.11

Proceedings of the 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), Karaikal, India, 4–5 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [**Google Scholar**]

11. Minvielle, L. How AI Is Shaping Applicant Tracking Systems. Geekflare. 2025. Available online: **https://geekflare.com/guide/ats-and-ai/** (accessed on 15 January 2025).

12. Lavi, D.; Medentsiy, V.; Graus, D. consultantbert: Fine-tuned siamese sentence-bert for matching jobs and job seekers. *arXiv* **2021**, arXiv:2109.06501. [**Google Scholar**]

13. Skondras, P.; Zervas, P.; Tzimas, G. Generating Synthetic Resume Data with Large Language Models for Enhanced Job Description Classification. *Future Internet* **2023**, *15*, 363. [**Google Scholar**] [**CrossRef**]

14. Pendhari, H.; Rodricks, S.; Patel, M.; Emmatty, S.; Pereira, A. Resume Screening using Machine Learning. In Proceedings of the 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 21–23 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. [**Google Scholar**]

15. Bharadwaj, S.; Varun, R.; Aditya, P.S.; Nikhil, M.; Babu, G.C. Resume Screening using NLP and LSTM. In Proceedings of the 2022 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 20–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 238–241. [**Google Scholar**]