INTELLIGENT RESUME SHORTLISTING USING NLP AND SEMANTIC MATCHING

Sushant Dev Tanwar[1], Harsh Mahajan[1], Pallavi Sakalley [2], Ajitesh Ghoshl[2]

[1]*Department of Computer Science and Engineering Parul University, Gujarat, India*
[2] *Department of Computer Science and Engineering Parul University, Gujarat, India*
[1]*sushanttanwar78@gmail.com,*
[1]*harshmahajan9892@gmail.com,*
[2]*pallavi.sakalley40382@paruluniersity.ac.in,*
[2]*ajitesh.ghosh36908@paruluniversity.ac.in,*

Keywords: **ARTIFICIAL INTELLIGENCE, NATURAL LANGUAGE PROCESSING, AUTOMATED RESUME SCREENING, EXPLAINABLE MACHINE LEARNING, SEMANTIC SIMILARITY**

Abstract

The recruitment process often requires evaluating a large number of resumes, making manual screening inefficient and prone to inconsistencies. Traditional Applicant Tracking Systems (ATS) rely heavily on keyword-based filtering techniques, which frequently fail to capture the contextual meaning of candidate profiles in relation to job requirements. This research proposes an AI-driven resume shortlisting framework that utilizes Natural Language Processing (NLP), semantic similarity analysis, and explainable scoring techniques to automate candidate evaluation. The proposed system extracts structured information from unstructured resume documents and converts both resumes and job descriptions into semantic embeddings. A weighted ranking model evaluates candidates based on skill relevance, professional experience, project similarity, and educational qualifications. To ensure fairness, bias-sensitive attributes are masked during evaluation. Furthermore, an explainable AI component provides interpretable insights into candidate rankings. Experimental evaluation demonstrates that the proposed framework improves matching accuracy and significantly reduces resume screening time compared to conventional keyword-based approaches. The system provides a scalable, transparent, and ethical solution that can be integrated into modern recruitment platforms.

1. Introduction

Recruitment is one of the most critical processes in any organization, as it directly affects workforce quality and productivity. With the rapid growth of online job applications, recruiters often receive hundreds or even thousands of resumes for a single job opening. Reviewing such a large volume of applications manually is both time-consuming and inefficient.

Most existing Applicant Tracking Systems mainly use keyword-based filtering to process resumes. However, this approach often cannot understand the contextual meaning or semantic relationship between a candidate's resume and the job description. Advancements in Artificial Intelligence and Natural Language Processing have made it possible to analyse textual information more effectively.

This study proposes an AI-based resume shortlisting system that integrates semantic matching, explainable scoring, and bias-aware evaluation. The key contributions of this work are:

- A semantic framework resume–job alignment
- A transparent and weighted candidate scoring model
- A bias-reduction strategy for fair evaluation

2. Methodology

2.1 Resume Preprocessing

Resumes uploaded by candidates are first converted into plain text. The system performs several preprocessing steps to prepare the data for analysis, including tokenization, removal of stop words, and lemmatization. These steps help improve the quality of textual data and remove irrelevant information.

2.2 Semantic Representation

Both resumes and job descriptions are converted into vector representations through sentence-level embeddings. This enables the system to identify contextual similarities instead of depending solely on exact keyword matches.

2.3 Candidate Scoring Model

The system calculates candidate relevance using a weighted scoring mechanism that considers multiple factors.

- Skill Relevance: 45%
- Experience Alignment: 30%
- Project Similarity: 15%
- Educational Qualification: 10%

The final candidate score is computed as the weighted sum of these components.

2.4 Bias Mitigation Strategy

To ensure fair evaluation, the system removes sensitive attributes such as candidate names, gender indicators, and institution names before analysis. This prevents the model from being influenced by non-technical factors.

2.5 Explainable Scoring Module

The explainable AI component provides detailed reasoning for candidate rankings. The module generates:

- A detailed breakdown of candidate scores
- Lists of matched and missing skills
- Explanations for candidate ranking

3. LITERATURE REVIEW

The rapid increase in job applications has encouraged researchers to explore automated approaches for resume screening in order to improve recruitment efficiency. Early Applicant Tracking Systems mainly relied on keyword-based filtering techniques, where resumes were shortlisted based on the presence of predefined terms. Although useful for basic filtering, such approaches often fail to capture the deeper contextual meaning between candidate qualifications and job requirements.

To address these limitations, researchers began applying Natural Language Processing (NLP) techniques to extract structured information from unstructured resumes. Methods such as tokenization, part-of-speech tagging, and named entity recognition have been used to identify important attributes including skills, professional experience, and educational qualifications. However, traditional NLP approaches based on frequency counts and rule-based methods often struggle with synonym detection and variations in domain terminology.

To enhance the effectiveness of automated screening, machine learning-based approaches were later introduced. Various algorithms including logistic regression, support vector machines, and decision trees have been utilized to classify or rank resumes according to their relevance for a particular job role. Although these models improve the automation of candidate selection, they generally require labelled training data and often provide limited transparency regarding their decision-making process.

Recent advancements in deep learning have enabled the use of semantic embedding techniques such as Word2Vec, GloVe, and transformer-based models like BERT. These models allow systems to analyse contextual relationships between words and sentences, making it possible to compare resumes and job descriptions more effectively. These approaches have improved matching performance but also raise concerns regarding fairness and model transparency.

To mitigate such issues, several studies have explored bias-aware recruitment frameworks that conceal sensitive attributes like name, gender, or educational institution during evaluation. In addition, explainable AI methods have been incorporated to clarify how automated systems reach their decisions, thereby increasing trust and accountability in AI-assisted recruitment tools.

Despite these improvements, concerns related to transparency and bias still remain. To address these issues, researchers have introduced bias-aware recruitment systems that remove sensitive attributes during evaluation. Additionally, explainable AI techniques are increasingly used to provide insights into model decisions and improve trust in automated recruitment systems

4. RESULTS

The performance of the proposed AI-based resume shortlisting system was evaluated using a dataset consisting of resumes from different technical job roles. The system's performance was compared with traditional keyword-based screening methods to analyse improvements in accuracy, efficiency, and transparency.

4.1 Matching Accuracy

The semantic matching approach demonstrated higher accuracy in identifying relevant candidates compared to traditional keyword filtering. Even when exact keywords were not present, candidates whose experience and skills matched the job context were ranked higher. This confirms that semantic embeddings provide a more meaningful comparison between resumes and job descriptions.

4.2 Reduction in Screening Time

The automated system significantly reduced the time required for resume screening. Unlike manual review or keyword filtering methods, the proposed framework processes and ranks resumes automatically, allowing recruiters to shortlist candidates within seconds.

This demonstrates the system's ability to handle large volumes of applications efficiently.

4.3 Explainability Evaluation

The explainable scoring component provided a detailed breakdown of candidate rankings. Recruiters were able to view factors such as skill match percentage, relevance of experience, and missing competencies.

This level of transparency helps overcome one of the major limitations of traditional black-box AI systems.

4.4 Bias Mitigation Assessment

By masking sensitive attributes such as candidate names and institution identifiers, the system ensured that evaluation decisions were based strictly on job-related factors. This approach contributes to a more objective and fair recruitment process.

Overall, the experimental findings indicate that the proposed system enhances accuracy, efficiency, and fairness compared to conventional resume screening approaches.

5. CONCLUSION

This research presented an AI-driven resume shortlisting framework that integrates Natural Language Processing, semantic similarity analysis, explainable scoring mechanisms, and bias-aware evaluation techniques to automate the recruitment screening process. Unlike traditional keyword-based methods, the proposed system captures contextual relationships between resumes and job descriptions, resulting in more accurate candidate ranking.

The experimental results demonstrate improvements in shortlisting accuracy, substantial reduction in screening time, and increased transparency through explainable decision outputs. The inclusion of bias-mitigation strategies further ensures that candidate evaluation remains fair and focused on job-relevant qualifications.

Overall, the framework provides a scalable and interpretable solution that can be integrated into modern Applicant Tracking Systems to improve recruitment efficiency. Future research may focus on supporting multilingual resumes, implementing dynamic scoring weights based on job roles, and integrating automated interview analysis modules to create a fully intelligent recruitment pipeline.

6. ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to the **Department of Computer Science and Engineering, Parul University**, for providing the academic environment, infrastructure, and resources required to conduct this research work. The authors are also grateful to their faculty mentors and academic supervisors for their valuable guidance, constructive feedback, and continuous encouragement throughout the development of this project. Their insights and suggestions played an important role in improving the quality of this research study.

7. REFERENCES

[1] Smith, T., Jones, M.: 'Resume screening using natural language processing', *IET Systems Biology*, 2022, 16, (4), pp. 210–218

[2] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.

[3] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543.

[4] L. Brown, H. Thomas, and C. James, "Artificial intelligence in recruitment systems," *IET Communications*, vol. 15, no. 6, pp. 455–463, 2021.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, Minneapolis, USA, 2019, pp. 4171–4186.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *International Conference on Learning Representations (ICLR)*, 2013.

[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *KDD*, 2016.

[8] Python Software Foundation.: 'Python 3.13.2 documentation', available at https://docs.python.org/3/, accessed January 2026

[9] GeeksforGeeks.: 'Step-by-step guide to calculating RMSE using scikit-learn', *GeeksforGeeks*, 2 November 2024, available at https://www.geeksforgeeks.org/step-by-step-guide-to-calculating-rmse-using-scikit-learn/, accessed January 2026

[10] Hugging Face.: 'Hugging Face – The AI community building the future', available at https://huggingface.co/, accessed January 2026

[11] Fiori, A.: 'AI: prompt engineering to reduce hallucinations (Part 1)', *Flowygo Blog*, 17 June 2024, available at https://flowygo.com/en/blog/ai-prompt-engineering-to-reduce-hallucinations-part-1/, accessed January 2026

[12] Improving retrieval for RAG based question answering models on financial documents, *arXiv preprint*, available at https://arxiv.org/html/2404.07221v2, accessed January 2026

[13] Hugging Face.: 'Transformers model documentation', available at https://huggingface.co/docs/transformers/main_classes/model#transformers.AutoModel , accessed January 2026