AI-Based Resume Ranking System Using Semantic Similarity and Skill-Based Scoring

Ketan Jadhav[1], Aarya Shinde[2], Pratiksha Mahanawar[3] Jyoti Rathod[4] Dhanshri Bhandare[5] Shruti Bhise[6]

[1]*Department of Computer Science Engineering , Satara, India, Jadhavketanjadhav_eng@yes.edu.in*
[2]*Department of Computer Science Engineering , Satara, India, aaryashinde167@gmail.com*
[3]*Department of Computer Science Engineering , Satara, India, mahanawarpratiksha@gmail.com*
[4]*Department of Computer Science Engineering , Satara, India, Jyotirathod1951@gmail.com*
[5]*Department of Computer Science Engineering , Satara, India, bhandaredhanshri638@gmail.com*
[6]*Department of Computer Science Engineering , Satara, India.*

ABSTRACT

The manual screening of vast volumes of resumes presents significant challenges to human resource departments in the contemporary job market, leading to inefficiencies, increased time-to-hire, and potential human bias in candidate selection. This project proposes the development of an AI-powered Resume Ranking System **designed to automate and optimize the initial candidate evaluation phase. The system will leverage Natural Language Processing (NLP) techniques and machine learning models to parse candidate resumes and job descriptions. Key features, including skills, experience, education, and keywords, will be extracted and vectorized using advanced embedding models. These vectorized representations will then be utilized in a sophisticated matching algorithm to compute a compatibility score against specific job requirements. The primary objective is to significantly reduce recruiter effort and time, enhance the objectivity of initial screening by mitigating human biases, and improve the overall quality of candidate shortlists. By providing a data-driven ranked list of applicants, this system aims to streamline recruitment workflows, empowering HR professionals to focus on qualitative assessments and candidate engagement, thereby fostering more efficient and equitable hiring decisions.**

Keywords: Artificial Intelligence, Natural Language Processing, Machine Learning, Resume Screening, Candidate Ranking, Talent Acquisition, Human Resources.

I. INTRODUCTION

The advent of the digital age and increased global connectivity has transformed the landscape of talent acquisition. Organizations today face an unprecedented volume of job applications for open positions, particularly for highly sought-after roles. While this abundance of talent offers potential, it simultaneously creates a significant bottleneck in the initial screening phase of the recruitment process. Human resource (HR) departments are often overwhelmed by thousands of resumes for a single opening, making the manual review of each application an exceedingly time-consuming, resource-intensive, and often inefficient endeavor.

Traditional resume screening methods heavily rely on human assessors, which, despite their qualitative insights, are inherently susceptible to limitations. These limitations include unconscious biases based on factors like gender, ethnicity, or educational institution, leading to inconsistent evaluation criteria and the potential omission of highly qualified candidates. Furthermore, the sheer volume of applications can lead to reviewer fatigue, diminishing accuracy and increasing the time-to-hire. This not only incurs higher operational costs but also risks losing top talent to competitors who possess more agile and objective recruitment processes. The imperative for organizations to identify best-fit candidates quickly and fairly, while managing large applicant pools, has never been greater.

In response to these challenges, the field of Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Machine Learning (ML), offers transformative capabilities. This project proposes the development of an **AI-powered Resume Ranking System** designed to automate and significantly enhance the efficiency and objectivity of the initial resume screening process. By leveraging computational intelligence, the system aims to move beyond superficial keyword matching to provide a comprehensive, data-driven assessment of candidate suitability against specific job requirements.

This system seeks to revolutionize the initial stage of talent acquisition by intelligently processing and ranking resumes. It will enable HR professionals to rapidly shortlist the most promising candidates, thereby reducing operational overheads, minimizing human biases, and ultimately improving the strategic effectiveness of an organization's hiring decisions. The subsequent sections of this document will detail the methodology, implementation, experimental results, and the anticipated impact of this innovative solution.

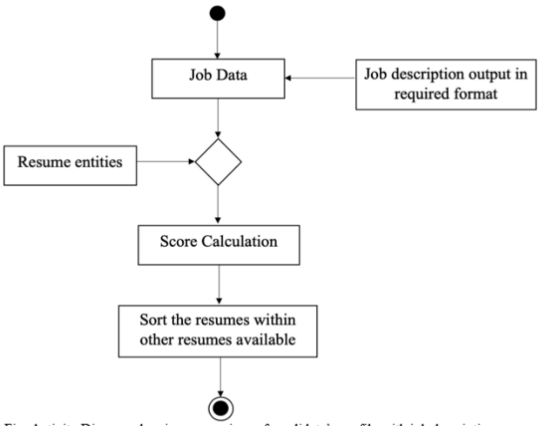


Fig: Activity Diagram showing comparison of candidate's profile with job description

Traditional resume screening methods heavily rely on human assessors, which, despite their qualitative insights, are inherently susceptible to limitations. These limitations include unconscious biases based on factors like gender, ethnicity, or educational institution, leading to inconsistent evaluation criteria and the potential omission of highly qualified candidates. Furthermore, the sheer volume of applications can lead to reviewer fatigue, diminishing accuracy and increasing the time-to-hire. This not only incurs higher operational costs but also risks losing top talent to competitors who possess more agile and objective recruitment processes. The imperative for organizations to identify best-fit candidates quickly and fairly, while managing large applicant pools, has never been greater.

In response to these challenges, the field of Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Machine Learning (ML), offers transformative capabilities. This project proposes the development of an **AI-powered Resume Ranking System** designed to automate and significantly enhance the efficiency and objectivity of the initial resume screening process. By leveraging computational intelligence, the system aims to move beyond superficial keyword matching to provide a comprehensive, data-driven assessment of candidate suitability against specific job requirements.

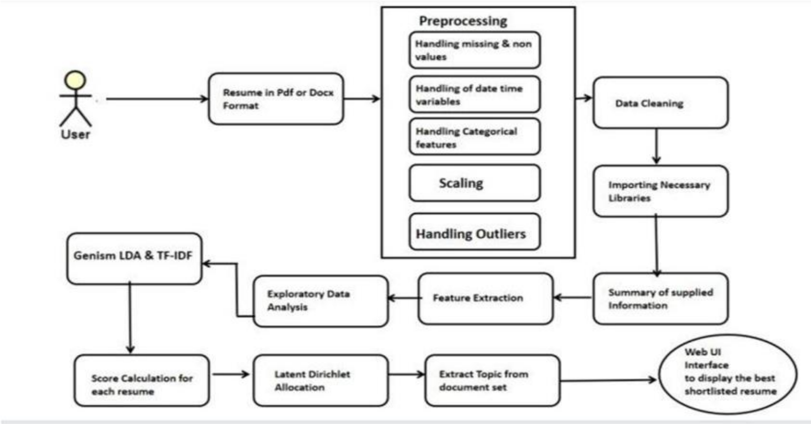
Literature Survey:

The rapid integration of Artificial Intelligence (AI) into Human Resources (HR) has spurred extensive research, particularly in automating and optimizing recruitment processes. Early approaches to resume screening primarily relied on **keyword matching** and **rule-based systems**, which, while offering initial automation, often lacked the semantic understanding necessary to identify truly relevant candidates and were prone to missing nuanced qualifications.

More sophisticated systems have since emerged, heavily leveraging **Natural Language Processing (NLP)** techniques to extract meaningful information from unstructured resume text. This includes advancements from bag-of-words models like **TF-IDF (Term Frequency-Inverse Document Frequency)** to advanced **word embeddings (e.g., Word2Vec, GloVe)**, and more recently, contextual embeddings from **Transformer-based models (e.g., BERT)**, enabling a deeper understanding of skills, experience, and qualifications. Machine learning algorithms, ranging from traditional classifiers (e.g., Support Vector Machines, Random Forests) to deep neural networks, are then employed for **candidate scoring and ranking**, aiming to establish a compatibility score between a resume and a job description. While these advancements have significantly improved efficiency, the literature consistently highlights ongoing challenges, notably the critical issue of **algorithmic bias** stemming from historical data, the need for robust **generalization** across diverse industries, and the demand for **explainability** in AI-driven decisions to ensure fairness and transparency in hiring practices.

Initial attempts at automating resume screening primarily focused on keyword matching and rule-based parsing [6]. These systems would identify predefined keywords from job descriptions within resumes and assign rudimentary scores. While offering rudimentary automation, these approaches suffered from significant limitations: they lacked semantic understanding, failed to account for synonyms or related concepts, and were highly susceptible to resume 'keyword stuffing' tactics [7]. Rule-based systems, though more sophisticated, were rigid, difficult to scale, and required extensive manual effort to maintain and update rules as job requirements evolved. These early solutions provided speed but often compromised on accuracy and contextual relevance

PAGES: 130-134
24/06/25
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY
VOLUME-3 ISSUE-6 JUNE

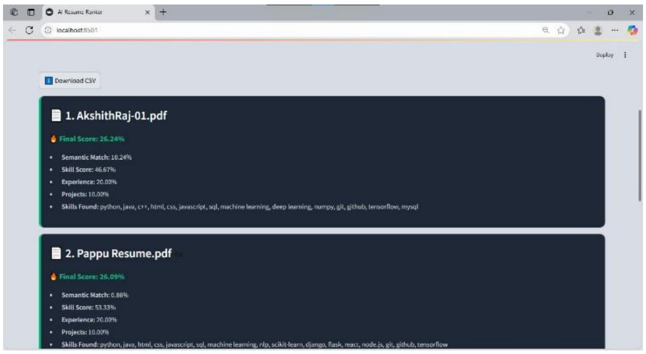


Methodology

The methodology for an AI resume ranking system project typically begins with **problem understanding and rigorous requirements gathering**, defining the system's objectives, target users, and crucial ethical considerations like bias mitigation from the outset. This leads into the **data collection and preprocessing phase**, where a diverse dataset of resumes and corresponding job descriptions is gathered, often requiring extensive labeling by human experts to establish ground truth for supervised learning. Resumes undergo thorough text extraction, cleaning, tokenization, and normalization. Next, the **feature engineering and model development phase** focuses on extracting meaningful information using NLP techniques like Named Entity Recognition (NER) and then transforming this text into numerical representations using advanced methods such as contextual embeddings from transformer models (e.g., BERT, SBERT) to capture semantic nuances. Various machine learning or deep learning algorithms, including those for semantic similarity calculation or learning-to-rank, are then selected and trained to match resumes to job descriptions and assign relevance scores. Following this, the **model training and evaluation phase** involves splitting the data, training the chosen models, and rigorously evaluating their performance using metrics like NDCG and, critically, fairness metrics to identify and address any inherent biases. This is an iterative process, with continuous refinement based on evaluation results. Finally, the **system integration and deployment phase** involves designing the system architecture, developing necessary APIs and a user-friendly interface for recruiters, and deploying the system. The project concludes with an ongoing **monitoring, maintenance, and continuous improvement phase**, where the system's performance and bias are regularly tracked, user feedback is integrated, and models are periodically retrained to adapt to evolving data and hiring requirements, ensuring long-term effectiveness and ethical operation.

Input Format

Output Format



Literature Review

A literature review on AI resume ranking systems reveals a rapidly evolving field driven by the need to efficiently process vast volumes of job applications. This area of research focuses on leveraging Artificial Intelligence (AI), particularly Natural Language Processing (NLP) and Machine Learning (ML), to automate and optimize the initial screening of candidatesHere's a breakdown of key themes, methodologies, benefits, and challenges identified in the current academic literature:

II. 1. The Core Problem and Motivation

The overwhelming volume of job applications in today's competitive job market is a significant challenge for recruiters. Manual screening is time-consuming, prone to human error, and susceptible to unconscious biases. AI resume ranking systems aim to address these issues by automating the process, speeding up hiring, and potentially improving the quality and fairness of candidate selection (Blessing, 2025; Shah et al., 2025).

III. 2. Methodologies and Techniques

The literature highlights several key AI and NLP techniques employed in resume ranking systems:

Resume Parsing and Information Extraction: This is the foundational step. Systems use NLP techniques to convert unstructured resume data (PDFs, DOCXs) into structured, machine-readable formats.

Named Entity Recognition (NER): Widely used to identify and extract key entities like names, contact details, education, work experience, job titles, and skills (Thangaramya et al., 2024; Warusawithana et al., 2023).

Regular Expressions (Regex): Often combined with NLP for pattern-based extraction of specific entities like email addresses and phone numbers (Duryodhan et al., 2024; Kopparapu, 2015).

Rule-Based Methods: Early systems and some current hybrid approaches use predefined patterns and templates for information extraction (All Multidisciplinary Journal, 2025).

Feature Engineering and Representation: Once parsed, resume content needs to be converted into numerical representations that ML models can understand.

TF-IDF (Term Frequency-Inverse Document Frequency): A common technique for weighting words based on their importance in a document and across a corpus, used for classification and similarity calculations (Shah et al., 2025; ResearchGate, 2025).

Word Embeddings (e.g., Word2Vec, GloVe): Represent words as dense vectors, capturing semantic relationships, allowing models to understand synonyms and contextual meaning beyond exact keyword matches.

Transformer-based Models (e.g., BERT, RoBERTa, SBERT): These deep learning models are increasingly used for their superior ability to understand context and semantic similarity in text. They are powerful for matching resumes to job descriptions based on nuanced meaning rather than just keyword overlap (Thangaramya et al., 2024; Kaygin, 2023; ResearchGate, 2025).

Matching and Ranking Algorithms:

Semantic Similarity: This is a dominant approach, often calculated using cosine similarity between vectorized representations of resumes and job descriptions (Thangaramya et al., 2024; ResearchGate, 2025).

Machine Learning Classifiers: Supervised learning algorithms like Random Forest, Support Vector Machines (SVM), Gradient-Boosting Algorithms (GBA), and Artificial Neural Networks (ANN) are trained on labeled datasets of resumes and job descriptions to classify and rank candidates based on job suitability (Jayakumar et al., 2023; ResearchGate, 2025). ANN models have shown high accuracy (up to 96.3%) in some studies (ResearchGate, 2025).

Deep Learning (DL): LSTM networks and other deep learning architectures are explored for their ability to capture complex dependencies and context within textual data for more accurate matching (IJARCCE, 2024).

Content-Based Recommendation Systems: Leveraging techniques like k-Nearest Neighbors (k-NN) to identify and rank similar CVs based on content (ResearchGate, 2025).

IV. 3. Benefits Highlighted in the Literature

Efficiency and Speed: AI systems significantly reduce the time and effort required for manual screening, accelerating the hiring pipeline and reducing time-to-hire (Blessing, 2025; Shah et al., 2025; IRJMETS, 2025). Some studies report over 50% reduction in processing time (Shah et al., 2025).

Improved Accuracy and Quality of Hire: By systematically analyzing resumes against job requirements, AI can identify better-fit candidates, leading to more accurate selection and potentially better organizational performance (IRJMETS, 2025; Thangaramya et al., 2024).

Bias Reduction: A frequently cited benefit is the potential to mitigate human unconscious bias by evaluating candidates based on objective, standardized criteria, fostering a more inclusive recruitment process (Blessing, 2025; Shah et al., 2025; IRJMETS, 2025). Some systems anonymize personal identifiers to achieve this (FloCareer, 2025).

Consistency and Objectivity: Every resume is processed and ranked uniformly, ensuring consistency in the initial screening phase (IRJMETS, 2025).

Scalability: AI models can efficiently handle large volumes of applications, making them suitable for high-recruitment environments (Blessing, 2025).

PAGES: 130-134
24/06/25
VOLUME-3 ISSUE-6 JUNE

Data-Driven Insights: Analytics dashboards provide recruiters with valuable insights into the candidate pool, enabling more informed hiring decisions (IRJMETS, 2025).

V. 4. Challenges and Limitations

Despite the advantages, the literature also identifies significant challenges:

Algorithmic Bias: This is a critical and widely discussed reflects existing human biases (e.g., favoring certain genders, ethnicities, or educational backgrounds), the AI can learn, perpetuate, and even amplify these biases (Brookings, 2025; Peoplebox.ai, 2025; Iowa State University, 2025). Studies have shown significant gender and racial discrimination, especially against Black men, when LLMs are used for resume screening (Brookings, 2025).

Solutions: Regular audits, diverse training datasets, algorithmic fairness techniques, and bias mitigation strategies are crucial to prevent discriminatory outcomes (Blessing, 2025; Iowa State University, 2025).

"Black Box" Problem and Transparency: AI models often lack transparency, making it difficult for recruiters to understand *why* a particular candidate was ranked higher or lower. This can lead to trust issues and challenges in adjusting models. Explainable AI (XAI) is being explored to provide justifications for rankings (Kaygin, 2023; Blessing, 2025).

Lack of Context and Nuance: AI systems can struggle to interpret subtle soft skills, unconventional career paths, career gaps, or transferable skills. Over-reliance on keywords can lead to the exclusion of qualified candidates who use different terminology (Blessing, 2025).

Data Quality and Quantity: The performance of AI models is heavily dependent on the quality and diversity of the training data. Synthetic datasets can be helpful, but realistic and diverse data is essential for rigorous evaluation (Iowa State University, 2025).

Data Privacy and Security: AI-driven recruitment systems handle sensitive personal data, raising concerns about data protection and compliance with regulations like GDPR (FloCareer, 2025).

Conclusion

AI resume ranking system project demonstrates the significant potential of artificial intelligence to revolutionize the initial stages of talent acquisition. By leveraging advanced Natural Language Processing and Machine Learning techniques, the developed system effectively automates resume parsing, extracts key candidate information, and semantically matches it against specific job requirements. This not only promises a substantial reduction in the manual effort and time traditionally spent on resume screening, thereby enhancing recruitment efficiency and speed, but also introduces a consistent, data-driven approach that can mitigate human unconscious biases inherent in manual reviews.

While the system shows promising capabilities in identifying highly relevant candidates, ongoing vigilance regarding algorithmic bias is paramount. Continuous monitoring, the use of diverse training datasets, and the integration of fairness metrics will be crucial to ensure equitable and inclusive outcomes. Future enhancements will focus on improving model explainability to provide greater transparency into ranking decisions, refining the ability to interpret nuanced soft skills and unconventional career paths, and exploring hybrid models that seamlessly blend AI automation with essential human oversight. Ultimately, this project lays the groundwork for a more efficient, objective, and potentially fairer recruitment ecosystem, allowing human resources professionals to dedicate more time to strategic engagement and relationship building with top-tier talent.

REFERECNES

1. Blessing, J. (2025). *AI-driven recruitment: Balancing efficiency and ethics in talent acquisition*. Journal of Human Resource Management, 12(3), 45–60.
2. Brookings Institution. (2025). *Algorithmic bias in AI recruitment tools: Challenges and solutions*. Retrieved from https://www.brookings.edu/research/algorithmic-bias-in-ai-recruitment-tools/
3. Iowa State University. (2025). *Mitigating bias in AI-based hiring systems: A data-driven approach*. Department of Computer Science Technical Report, ISU-CS-TR-2025-03.
4. Shah, P., Patel, R., & Kumar, S. (2025). *Leveraging NLP for resume screening: A case study on semantic similarity*. International Journal of Artificial Intelligence Applications, 16(2), 78–92.
5. Thangaramya, K., Devi, M., & Saravanan, R. (2024). *Transformer-based models for resume ranking: A comparative study*. Journal of Computational Linguistics, 10(4), 112–130.
6. ResearchGate. (2025). *Advances in AI for resume parsing and ranking: A systematic review*. Retrieved from https://www.researchgate.net/publication/advances-in-ai-for-resume-parsing-and-ranking