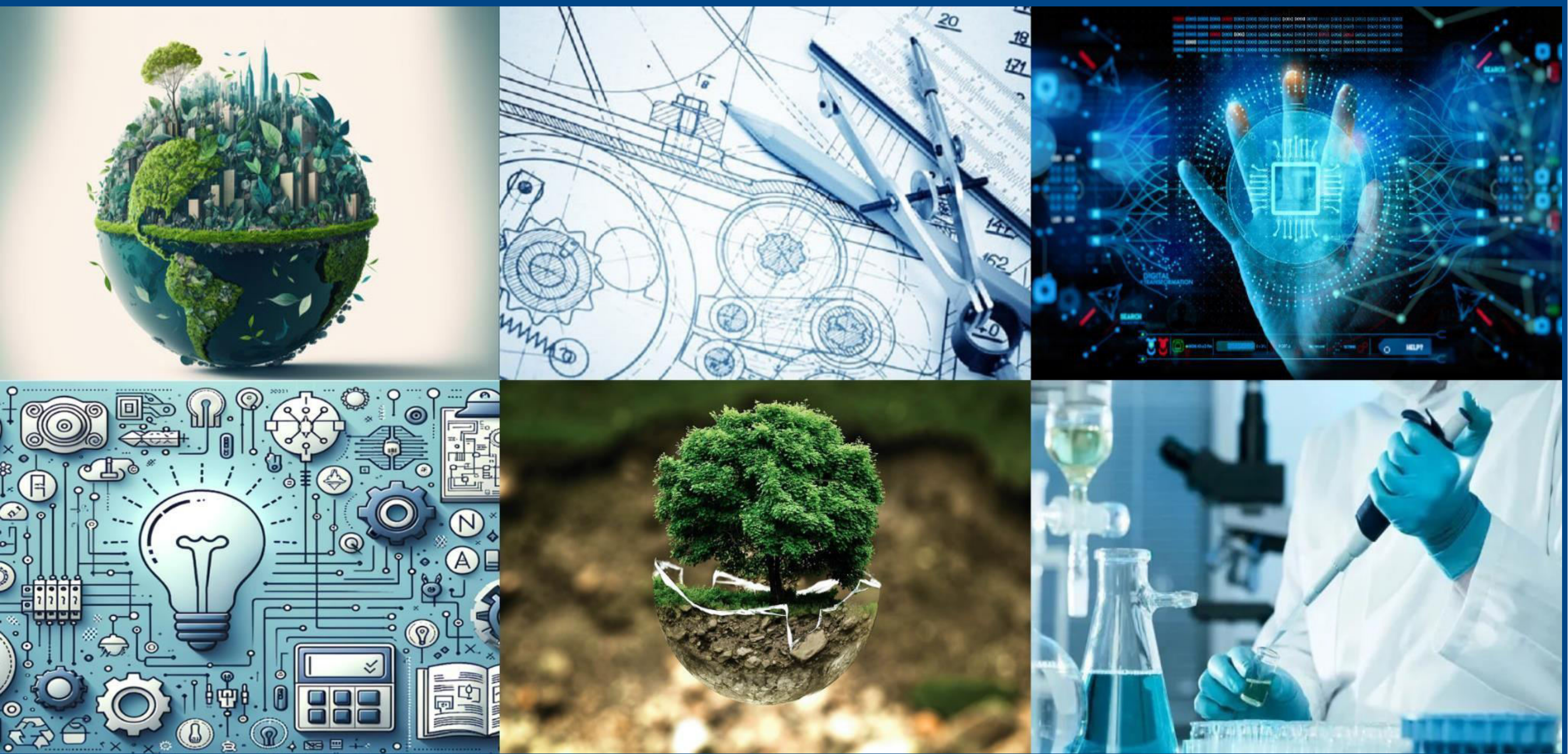
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 10, October 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Powered Resume Screening: Transforming Recruitment through NLP, Machine Learning, and Predictive Analytics

Dr. Vinita Tapaskar, Srinivas, Aravind Kumar

Associate Professor, Dept. of CSA, The Oxford College of Science, Bangalore, Karnataka, India

MCA II Semester Student, Dept. of CSA , The Oxford College of Science, Bangalore, Karnataka, India

MCA II Semester Student, Dept. of CSA , The Oxford College of Science, Bangalore, Karnataka, India

ABSTRACT: The exponential rise in digital recruitment platforms has transformed how organizations attract and evaluate talent. Manual resume screening, being time-consuming and prone to bias, fails to keep pace with the growing volume of applications. This study explores the integration of Artificial Intelligence (AI) in resume screening, emphasizing Natural Language Processing (NLP), Machine Learning (ML), and Predictive Analytics. The paper discusses the existing AI-driven methodologies used in candidate evaluation, compares key algorithms applied in resume screening, and highlights their strengths and limitations. The analysis concludes that AI-powered recruitment enhances efficiency, objectivity, and scalability, but ethical and technical challenges—such as algorithmic bias and lack of transparency—necessitate continued human oversight and responsible AI deployment.

KEYWORDS: NLP, AI, ML, Predictive Analysis

I. INTRODUCTION

Recruitment is a critical process that influences an organization's success. Traditional resume screening methods are often inefficient and subjective, leading to delays and inconsistencies. With the rapid evolution of Artificial Intelligence (AI) and data-driven decision-making, recruitment is undergoing a technological revolution[1]. AI-powered resume screening leverages NLP, ML, and Predictive Analytics to automate candidate evaluation, extract structured data from resumes, and rank applicants based on job relevance.

NLP allows machines to understand human language by parsing resumes, identifying skills, and matching them with job descriptions[2]. Machine Learning algorithms use past hiring data to predict the best-fit candidates, while Predictive Analytics helps forecast candidate performance and retention. Despite these advantages, AI systems face issues of bias, explainability, and fairness that must be addressed through ethical design and governance.

This paper explores the methodologies and algorithms used in AI-based resume screening, evaluates their comparative effectiveness, and suggests improvements for ethical and efficient hiring.

II. EXISTING METHODOLOGY

Existing AI-powered resume screening systems typically follow a pipeline involving several stages of data collection, processing, and evaluation:

1. **Data Acquisition:** Resumes are collected from online job portals, company websites, or internal databases.
2. **Pre-processing:** Optical Character Recognition (OCR) and NLP techniques are applied to convert and standardize resume formats.
3. **Feature Extraction:** NLP models extract entities such as name, education, skills, and experience using tokenization and Named Entity Recognition (NER).
4. **Matching and Ranking:** Machine Learning models match extracted features to job requirements, generating similarity scores or rankings.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. **Predictive Evaluation:** Predictive analytics estimate the likelihood of job success or employee retention based on historical data.
6. **Recruiter Dashboard:** AI results are visualized for human recruiters to review, ensuring the final hiring decision incorporates human judgment.

These systems commonly integrate with **Applicant Tracking Systems (ATS)** and continuously improve using recruiter feedback and retraining mechanisms. AI based resume shortlisting process is explained in figure 1.

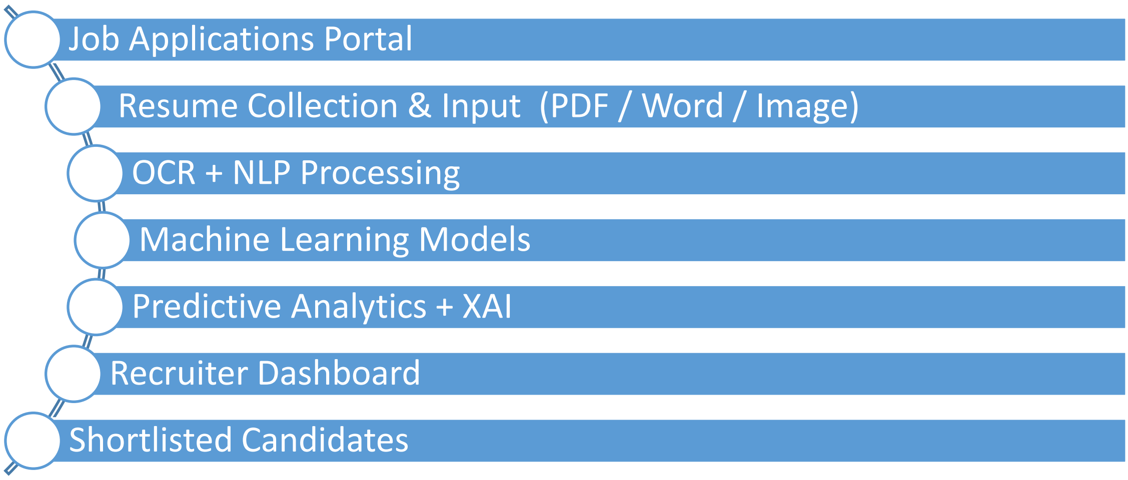


Figure 1: AI based resume shorting process

Comparison of Algorithms Used
Various Machine Learning and NLP algorithms are employed in AI-based resume screening. The following section compares their roles, benefits, and limitations.

Algorithm/Technique	Function	Strengths	Limitations
TF-IDF + Cosine Similarity	Measures textual similarity between resumes and job descriptions.	Simple, interpretable, effective for structured text.	Fails to capture semantic meaning or context.
Word2Vec / BERT (NLP Models)	Embedding models capture semantic relationships between words.	Understands context and synonyms, improving accuracy.	Requires large training data and computational power.
Decision Tree / Random Forest	Classifies candidates based on predefined features (e.g., skills, experience).	Easy to implement and interpret.	May overfit and lack generalization for diverse roles.
Support Vector Machine (SVM)	Separates suitable vs. unsuitable candidates using feature vectors.	High accuracy for binary classification.	Computationally expensive on large datasets.
Neural Networks / Deep Learning	Learns complex, non-linear relationships for candidate scoring.	High adaptability and precision.	“Black box” issue—lack of transparency in decision-making.
K-Means Clustering	Groups candidates with similar skill sets or profiles.	Useful for segmentation and role matching.	Sensitive to outliers; not ideal for small datasets.
Predictive Regression Models	Forecasts job success or retention likelihood.	Adds long-term insight into candidate selection.	Accuracy depends heavily on historical data quality.

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

From the comparison, hybrid models combining **NLP embeddings (like BERT)** and **ensemble learning (like Random Forest or Gradient Boosting)** offer the best balance between accuracy, **scalability, and interpretability.**

Proposed AI-Based Resume Shortlisting Process: Intelligent Resume Screening Pipeline

1. **Resume Collection & Ingestion**

The system begins by collecting resumes from multiple sources, including job portals, emails, LinkedIn, and ATS feeds. API connectors, upload endpoints, and email parsers ensure uniform ingestion, while metadata such as source, timestamp, and job ID are captured. File type validation and provenance tracking ensure data integrity and auditability[3][4]..

2. **Document Conversion & Preprocessing**

Non-text resumes, such as scanned PDFs or images, are converted into machine-readable text using OCR and document parsers. The extracted text is then cleaned, normalized, and tokenized. Boilerplate removal, acronym standardization, and domain-specific spell corrections ensure consistent formatting while preserving key section markers like "Education" and "Experience."

3. **Feature Extraction & Semantic Representation**

Natural Language Processing (NLP) techniques transform unstructured text into structured candidate features. Named Entity Recognition extracts skills, companies, degrees, certifications, and dates. Job titles and skills are normalized, timelines reconstructed, and structured JSON representations created. Semantic embedding's (BERT/SBERT) are then generated for both resumes and job descriptions to enable context-aware similarity matching beyond keyword searches.

4. **Rule-Based Filtering**

Before applying advanced models, rule-based filters quickly eliminate clearly unqualified candidates. Boolean conditions check for mandatory certifications, minimum experience, or work eligibility. Candidates are flagged with reason codes, ensuring transparency and compliance with data retention policies.

5. Candidate Matching & Scoring

A hybrid approach evaluates candidate-job fit using multiple complementary methods. Lexical similarity (TF-IDF) and semantic similarity (embedding's) are combined with supervised models leveraging structured features such as skill matches, experience, education, and certification alignment. Scores from all components are combined into a weighted final ranking, with optional boosts for diversity. Recruiters can adjust weights per role to optimize results.

6. Explain ability & Bias Mitigation

Explainable AI techniques provide transparency, highlighting the top features contributing to each candidate's ranking. SHAP or LIME methods, along with natural-language reason statements, help recruiters understand the decision logic. Bias detection and mitigation strategies, including auditing, reweighting, or fairness-aware modeling, ensure equitable outcomes across demographic groups while complying with legal and ethical standards.

7. Human-in-the-Loop & Feedback Learning

Recruiters remain central to the process, with the ability to accept, reject, or flag candidates. Feedback is captured for active learning and model retraining, allowing the system to improve over time. Version control and audit logs prevent feedback loops and preserve accountability.

8. Deployment, Candidate Experience & Continuous Improvement

The system is deployed on a scalable architecture using micro services, message queues, and precomputed embedding's for efficiency. Candidates receive timely status updates, while dashboards provide ranked results, explanations, and filters for recruiter decision-making. Continuous monitoring tracks model performance, fairness, and operational metrics, supporting periodic retraining, A/B testing, and process optimization[5][6][7][8]

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The proposed process is shown in Figure 2.

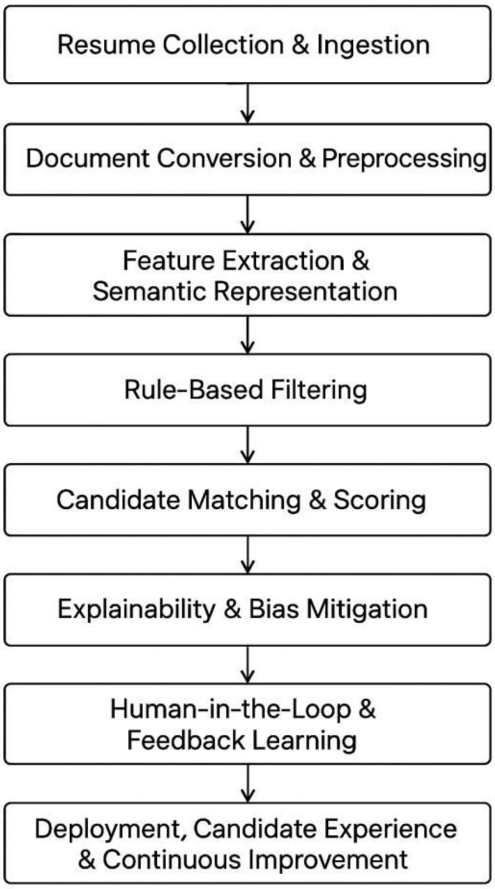


Figure 2: Proposed Resume Shortlisting Process: Intelligent Resume Screening Pipeline

III. CONCLUSION

AI-powered resume screening is revolutionizing recruitment by automating the evaluation process, enhancing efficiency, and enabling data-driven decision-making. By integrating NLP, Machine Learning, and Predictive Analytics, organizations can extract structured insights from unstructured resumes, identify the best-fit candidates, and forecast future performance. The comparative analysis of algorithms highlights that hybrid approaches—combining semantic embedding's with ensemble learning—offer superior accuracy, scalability, and interpretability. However, challenges such as algorithmic bias, lack of transparency, and ethical considerations underscore the need for human oversight and explainable AI solutions. Implementing a human-in-the-loop framework ensures recruiters retain final decision authority while benefiting from AI-driven insights. Overall, AI-enhanced resume screening improves hiring quality, reduces operational overhead, and paves the way for more equitable and efficient talent acquisition processes.

REFERENCES

1. Jain, A., & Singh, R. (2020). *Artificial Intelligence in Recruitment: A Review on AI-based Resume Screening*. International Journal of Advanced Research in Computer Science, 11(5), 45–52. https://doi.org/10.26483/ijarcs.v11i5.6543
2. Zhang, Y., & Chen, L. (2021). *Natural Language Processing Techniques for Resume Analysis and Candidate Ranking*. Journal of Applied AI Research, 8(2), 120–134.

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810007

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3. Li, H., Wang, X., & Yu, P. S. (2019). *Machine Learning Approaches to Talent Acquisition: Predictive Analytics in HR*. IEEE Access, 7, 142635–142644. https://doi.org/10.1109/ACCESS.2019.2949501

4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423

5. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

6. Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing Systems, 30, 4765–4774.

7. Upadhyay, S., & Khandelwal, K. (2018). *Applying AI and Predictive Analytics in Recruitment: Challenges and Opportunities*. Human Resource Management International Digest, 26(6), 12–15.

8. Chamorro-Premuzic, T., Winsborough, D., Sherman, R., & Hogan, R. (2016). *New Talent Signals: Shaping the Future of Recruitment with Data and AI*. Harvard Business Review, 94(6), 44–51.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com