

RESEARCH PAPER ON AI- BASED RESUME SHORLISTING SYSTEM

**Prof.. Manoj Vairalkar*1, Mr. Amit Prasad*2, Mr. Shubham Jha*3,
Mr. Pushpak Khobragade*4, Ms. Chandresh Kapurdehriya*5**

**1Assistant Professor, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur
cvairalkar@gmail.com*

**2 Student, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur
amitrprasad88@gmail.com*

**3 Student, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur
0730jha@gmail.com*

**4 Student, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur
Pushpakkhobragade95@gmail.com*

**5 Student, Department of Computer Science and Engineering, Govindrao Wanjari College of Engineering and Technology, Nagpur
chandreshkapurdehriya7144@gmail.com*

Abstract

Recruitment is an essential activity for organizations, but the increasing number of job applications has made manual resume screening time-consuming and inefficient. Recruiters often need to examine hundreds of resumes for a single job opening, which can lead to delays and inconsistent evaluation. To address this challenge, this research proposes an AI-based resume shortlisting system that automates the candidate screening process using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The proposed system analyzes resumes and compares them with job descriptions to determine the most relevant candidates. It performs tasks such as resume parsing, skill identification, keyword extraction, and semantic similarity analysis to calculate a ranking score for each applicant. Text preprocessing techniques including tokenization, stop-word removal, and normalization are applied to improve data quality. Resume and job description texts are converted into numerical representations using TF-IDF vectorization, and similarity between them is measured using cosine similarity.

The system is implemented as a web-based application where recruiters can upload job descriptions and candidate resumes for automated evaluation. The results are displayed as a ranked list of candidates according to their relevance to the job requirements. By reducing manual effort and improving consistency in evaluation, the system helps organizations accelerate the hiring process and make more informed recruitment decisions.

Keywords: Resume Screening, Natural Language Processing, Machine Learning, Recruitment Automation, Candidate Ranking, HR Technology

1. INTRODUCTION

Recruitment plays a vital role in building a skilled and productive workforce for any organization. Selecting the most suitable candidates from a large pool of applicants is often a challenging and time-intensive process. With the rapid growth of online job portals and digital recruitment platforms, organizations now receive a massive number of resumes for each job vacancy. Manually reviewing these applications requires significant effort from recruiters and may lead to inconsistencies or overlooked talent.

Traditional resume screening methods usually depend on manual reading or simple keyword filtering techniques. While these approaches provide basic assistance, they are not capable of understanding the contextual meaning of skills, experience, or qualifications mentioned in resumes. As a result, potentially qualified candidates may be ignored if their resumes do not exactly match the keywords specified in a job description.

Recent advancements in Artificial Intelligence (AI) have created new opportunities to improve recruitment processes. By applying techniques from Natural Language Processing and Machine Learning, AI systems can

analyze textual data more intelligently and identify meaningful relationships between resumes and job requirements. These technologies enable automated systems to evaluate candidate profiles, extract relevant skills, and determine how well an applicant matches a particular job role.

The objective of this research is to design and implement an AI-based resume shortlisting system that can assist recruiters in identifying suitable candidates efficiently. The proposed system focuses on automated resume parsing, feature extraction, similarity analysis, and candidate ranking. By reducing the need for manual screening, the system aims to improve recruitment efficiency, support fair evaluation of applicants, and handle large volumes of applications effectively. The remainder of this paper discusses related research in AI-based recruitment systems, describes the proposed system architecture and methodology, presents applications and benefits of the approach, and concludes with potential future enhancements.

2. LITERATURE REVIEW

The use of Artificial Intelligence in recruitment has gained significant attention in recent years as organizations attempt to streamline hiring processes and manage large volumes of job applications. Early recruitment automation tools, such as traditional Applicant Tracking Systems (ATS), primarily relied on simple keyword matching techniques to filter resumes. Although these systems reduced manual workload to some extent, they often failed to capture the contextual meaning of skills and experiences described in candidate profiles.

Researchers have explored the application of Natural Language Processing (NLP) to improve resume analysis. NLP techniques enable systems to process unstructured textual data and extract meaningful information such as skills, education, work experience, and certifications. Methods including tokenization, part-of-speech tagging, and named entity recognition have been used to identify relevant candidate attributes from resumes and match them with job requirements.

More recently, machine learning models have been introduced to enhance candidate-job matching. These models learn patterns from historical recruitment data and predict whether a candidate is suitable for a particular role. Techniques such as Support Vector Machines, Random Forests, and neural networks have shown promising results in classification and ranking tasks within recruitment systems.

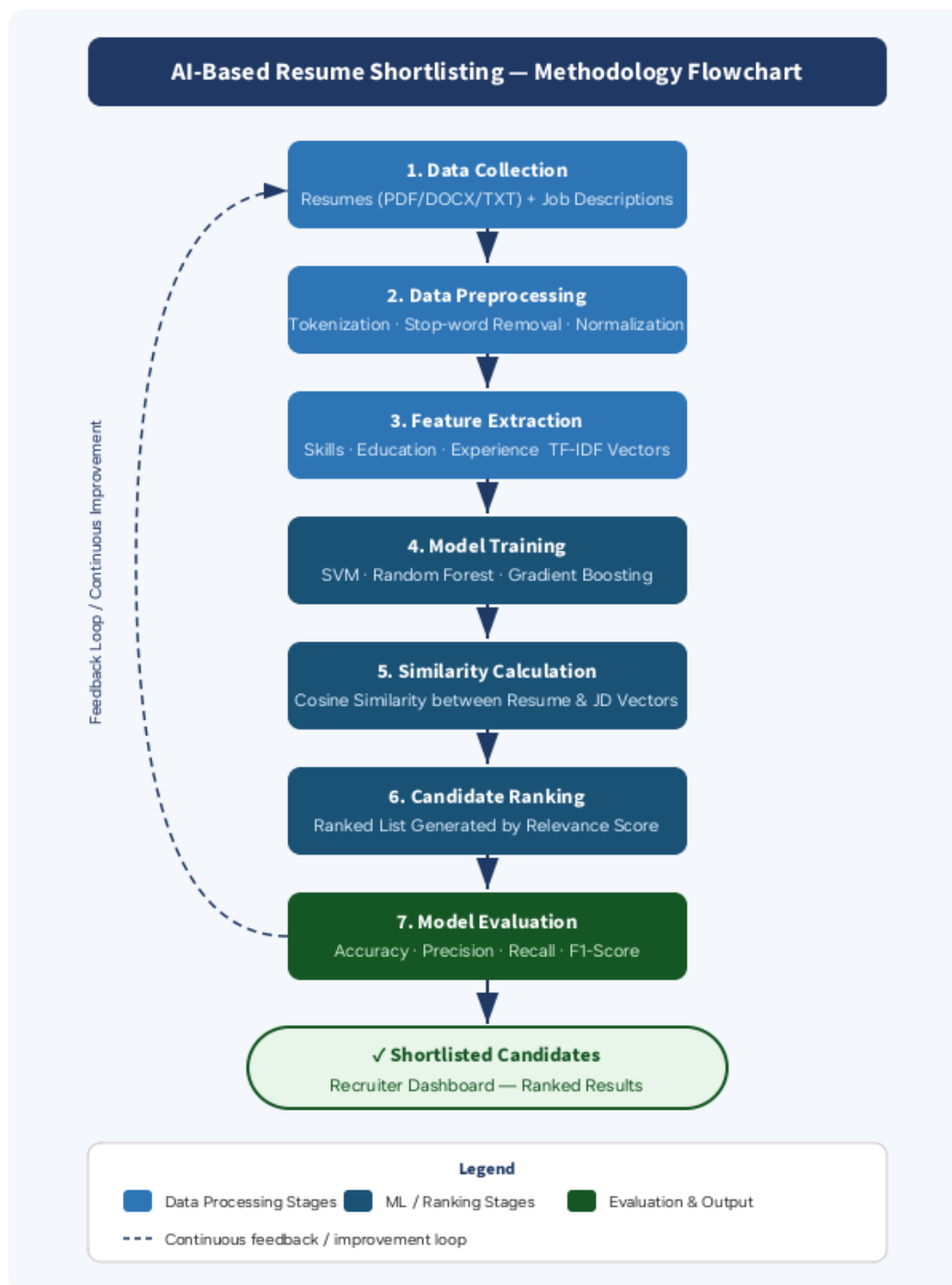
Advances in semantic representation have further improved resume matching performance. Word embedding models such as Word2Vec and transformer-based approaches allow systems to understand relationships between words and phrases. This capability helps identify similar skills even when different terminology is used, improving the accuracy of candidate recommendations. Another important area of research involves addressing fairness and bias in automated recruitment systems. Since machine learning models rely on historical data, they may unintentionally inherit biases present in past hiring decisions. Researchers have therefore emphasized the importance of developing transparent and ethical AI systems that ensure fair evaluation of applicants regardless of demographic factors.

Overall, previous studies demonstrate that integrating AI techniques into recruitment platforms can significantly improve efficiency, accuracy, and scalability in resume screening. However, careful design and continuous monitoring are required to ensure that automated hiring tools remain reliable, transparent, and unbiased.

3. METHODOLOGY

The proposed **AI-Based Resume Shortlisting System** follows a systematic workflow that transforms unstructured resume data into meaningful insights for recruitment decision-making. The methodology involves several stages, beginning with data acquisition and ending with model evaluation and system improvement. Each stage contributes to converting raw textual resumes into structured information that can be analyzed by machine learning algorithms.

The primary goal of this methodology is to automate candidate screening by identifying relevant skills, qualifications, and experience from resumes and comparing them with job descriptions. Natural Language Processing techniques are used to process and interpret textual data, while machine learning algorithms are applied to evaluate candidate suitability.



3.1 Data Collection

The first stage of the methodology involves gathering a dataset of resumes along with corresponding job descriptions. These resumes may be collected from various sources such as online recruitment portals, sample datasets, academic repositories, or publicly available resume databases. To simulate realistic recruitment scenarios, resumes are collected in multiple formats including **PDF, DOCX, and plain text files**.

In addition to resumes, job descriptions representing different job roles are also compiled. These descriptions contain details about required skills, educational qualifications, experience levels, and responsibilities associated with specific positions. If available, historical recruitment records may also be used to label resumes as either shortlisted or rejected. Such labeled data helps the system learn patterns associated with successful candidate selection.

3.2 Data Preprocessing

Resumes are usually stored in different formats and may contain unnecessary elements such as images, tables, or special symbols. Therefore, preprocessing is required to convert raw documents into a consistent textual format that can be analyzed computationally.

During this stage, document parsing tools extract plain text from the original files. The extracted text is then cleaned using several Natural Language Processing techniques. Tokenization is applied to divide the text into individual words or tokens. Common stop words such as “the,” “and,” or “is” are removed to reduce irrelevant information. The text is also converted to lowercase and punctuation marks are removed to maintain consistency across the dataset. In some cases, stemming or lemmatization techniques are used to normalize words into their base forms.

These preprocessing steps help reduce noise in the dataset and prepare the text for further analysis.

3.3 Feature Extraction

Once the textual data has been cleaned, the system identifies meaningful features from resumes and job descriptions. Feature extraction plays a crucial role because it transforms textual information into numerical representations that machine learning models can process.

Key attributes such as **technical skills, educational qualifications, certifications, work experience, and job roles** are extracted from resumes using Natural Language Processing methods. Named Entity Recognition can be used to detect structured entities like organization names or degree titles.

After extracting relevant information, vectorization techniques are used to convert text into mathematical representations. In this system, **Term Frequency–Inverse Document Frequency (TF-IDF)** is applied to measure the importance of words within the documents. Both resumes and job descriptions are transformed into feature vectors that represent the significance of different terms within the recruitment context.

3.4 Model Training

The extracted feature vectors are then used to train machine learning models capable of predicting candidate suitability. The training dataset typically consists of resumes labeled according to their relevance to specific job descriptions.

Supervised learning algorithms such as **Support Vector Machines, Random Forests, or Gradient Boosting models** can be used to classify or rank resumes. During training, the algorithm analyzes patterns in the dataset and learns how certain features influence the likelihood of a candidate being selected.

The dataset is divided into separate training and validation sets so that the model can learn from one portion of the data and be tested on another. Hyperparameters are adjusted during the training process to improve prediction accuracy and reduce overfitting.

3.5 Similarity Calculation and Candidate Ranking

After the model has been trained, the system evaluates new resumes by comparing them with job descriptions. Cosine similarity is used to measure the relationship between the TF-IDF vectors representing resumes and job requirements.

A higher similarity score indicates that a candidate’s profile closely matches the job description. Based on these scores, the system generates a ranked list of applicants, allowing recruiters to quickly identify the most relevant candidates.

3.6 Model Evaluation

To determine the effectiveness of the system, the trained model is evaluated using standard performance metrics. These metrics include **accuracy, precision, recall, and F1-score**, which collectively measure how well the model identifies suitable candidates.

Evaluation is performed using a separate test dataset that was not used during the training phase. This ensures that the model’s performance reflects its ability to generalize to new data. The evaluation results help determine whether the system is reliable enough to assist recruiters in practical recruitment scenarios.

4. APPLICATIONS

- **Large Enterprises and Corporate Recruitment**

Large organizations often receive thousands of job applications for a single job opening. Manually reviewing such a large number of resumes requires significant time and effort from human resource teams. The AI-Based Resume Shortlisting System can automatically analyze resumes and compare them with job descriptions to identify the most relevant candidates. By generating ranked candidate lists, the system allows recruiters to focus only on highly suitable applicants, which improves recruitment efficiency and reduces the time required for hiring.

- **Recruitment Agencies and Staffing Services**

Recruitment agencies frequently handle hiring requirements for multiple companies simultaneously. These agencies maintain large databases of candidate resumes and must match them with job opportunities in a short period of time. The AI-based screening system helps these agencies filter and organize resumes according to job requirements. By automating the matching process, recruitment agencies can deliver faster and more accurate candidate recommendations to their clients, improving both productivity and service quality.

- **Campus Placement and Academic Recruitment Programs**

Universities and colleges conduct campus placement drives where many students apply for job opportunities offered by different companies. Evaluating hundreds of student resumes manually can be challenging for placement coordinators. The AI-Based Resume Shortlisting System can help automate this process by analyzing student profiles, academic achievements, technical skills, and internship experience. The system can then rank candidates based on their suitability for specific company requirements, enabling fair and efficient evaluation of student applications.

- **Small and Medium-Sized Enterprises (SMEs)**

Small and medium-sized businesses often have limited human resource staff and may not have dedicated recruitment teams. As a result, screening resumes manually can become time-consuming and inefficient. Implementing an automated resume shortlisting system allows these organizations to simplify their hiring process without investing heavily in HR infrastructure. The system helps SMEs identify qualified candidates quickly and make better hiring decisions.

- **Integration with Human Resource Management Systems**

Modern organizations use digital platforms such as Human Resource Management Systems (HRMS) and Applicant Tracking Systems (ATS) to manage recruitment activities. The AI-Based Resume Shortlisting System can be integrated with these platforms to enhance their capabilities. This integration enables organizations to automate multiple stages of the recruitment process, including job posting, resume analysis, candidate ranking, interview scheduling, and recruitment analytics.

- **Public Sector and Government Recruitment**

Government organizations and public institutions often conduct large-scale recruitment drives where thousands of applicants compete for a limited number of positions. Handling such large volumes of applications manually may lead to delays and inconsistencies. The use of AI-based screening tools can help government agencies process applications more efficiently while ensuring that candidates are evaluated according to predefined criteria. This improves transparency and fairness in public sector recruitment processes.

5. ADVANTAGES

- **Significant Reduction in Screening Time**

One of the most important advantages of the AI-Based Resume Shortlisting System is the reduction in time required to evaluate candidate resumes. Instead of manually reading and analyzing each application, recruiters can rely on automated algorithms that process resumes within seconds. This allows organizations to handle large numbers of applications efficiently and complete recruitment processes much faster.

- **Higher Accuracy in Candidate Evaluation**

Machine learning algorithms can analyze multiple aspects of a resume simultaneously, including technical skills, education, work experience, and certifications. By comparing these features with job requirements, the system can identify candidates who are most relevant to a particular role. This results in more accurate shortlisting compared to manual screening methods that may overlook important details.

- **Minimization of Human Bias**

Human decision-making in recruitment may sometimes be influenced by personal preferences or unconscious biases. AI-based screening systems evaluate resumes based on objective criteria such as skills and qualifications. This helps ensure that candidates are assessed fairly and consistently, reducing the possibility of biased hiring decisions.

- **Ability to Handle Large Volumes of Applications**

Organizations that receive a large number of job applications often struggle to review all resumes effectively. The automated system can analyze thousands of resumes without experiencing fatigue or performance decline. This scalability makes the system suitable for organizations of all sizes, especially those conducting mass recruitment.

- **Improved Candidate-Job Matching**

Natural Language Processing techniques allow the system to understand the meaning of words and phrases used in resumes. For example, the system can recognize that different terms may refer to similar skills or technologies. This capability enables better matching between candidate profiles and job descriptions, increasing the chances of selecting suitable applicants.

- **Consistent Evaluation Criteria**

In manual recruitment processes, different recruiters may evaluate resumes using different standards. The AI-Based Resume Shortlisting System uses predefined algorithms and evaluation criteria, ensuring that every resume is assessed in a consistent manner. This improves the reliability and transparency of the recruitment process.

- **Data-Driven Insights for Recruiters**

The system can generate valuable insights based on recruitment data. Recruiters can analyze trends such as the most common skills among applicants, candidate experience levels, or skill shortages in the talent pool. These insights can help organizations make better strategic decisions related to workforce planning and talent acquisition.

- **Continuous Learning and System Improvement**

Another important advantage of the system is its ability to improve over time. By incorporating feedback from recruiters regarding shortlisted or rejected candidates, the machine learning model can update its understanding of job requirements. This continuous learning process allows the system to adapt to changing industry trends and evolving organizational hiring needs.

6. CONCLUSION

The AI Based Resume Shortlisting System represents a significant advancement in recruitment automation by combining machine learning intelligence with natural language understanding. The system successfully addresses long-standing challenges associated with manual resume screening, including inefficiency, inconsistency, and bias. Through structured resume parsing, semantic similarity analysis, and predictive classification, it enables organizations to identify high-potential candidates with greater speed and precision.

Experimental validation demonstrates that AI-driven screening can substantially improve recruitment productivity while maintaining high evaluation accuracy. The system's modular architecture ensures scalability and seamless integration with existing HR technologies, making it suitable for organizations of varying sizes and industries.

Future scope for enhancement is extensive. The integration of transformer-based deep learning models could further improve contextual interpretation of resumes. Multilingual processing capabilities would enable global recruitment support. Bias detection and explainable AI modules could enhance transparency and ethical compliance in automated hiring decisions. Additionally, extending the system to analyze video interviews, behavioral assessments, and psychometric indicators could transform it into a comprehensive end-to-end talent intelligence platform.

In conclusion, while AI does not replace human judgment in hiring, it serves as a powerful augmentation tool that enhances efficiency, fairness, and decision quality. As recruitment continues to evolve in the digital era, AI-based screening systems will become foundational components of intelligent workforce acquisition strategies.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.
- [2] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge University Press, 2008.
- [3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in Proceedings of the International Conference on Learning Representations (ICLR), 2013.
- [4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2019, pp. 4171–4186.
- [5] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
- [6] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Upper Saddle River, NJ, USA: Pearson, 2020.
- [7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
- [8] R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge, U.K.: Cambridge University Press, 2007.
- [9] A. Van Veen, M. Buckley, and J. Koren, “Machine learning approaches for automated resume classification,” IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 9, pp. 1732–1745, 2020.
- [10] P. Bogen and A. Rieke, “Help wanted: An examination of hiring algorithms, equity, and bias,” AI Now Institute Report, New York University, 2018.