

Automated Resume Screening for HR Using Machine Learning

Prof. Rita V Patil*¹, Mr. Mahesh Kailas Mali²

Professor, Department of Computer Applications, SSBT COET, Jalgaon Maharashtra, India¹

Research Scholar, Department of Computer Applications, SSBT COET, Jalgaon Maharashtra, India²

Abstract: Recruitment is one of the most critical processes in human resource management, yet manual resume screening is time-consuming, costly, and prone to bias. Automated resume screening powered by machine learning offers a scalable solution to filter resumes efficiently and fairly. This study explores the development of an intelligent resume screening system using natural language processing (NLP) and supervised learning algorithms to classify resumes into relevant job categories. Data preprocessing techniques, such as tokenization, stopword removal, and TF-IDF vectorization, were employed to extract meaningful features from textual resumes. Models including Logistic Regression, Random Forest, and Support Vector Machines (SVM) were trained and evaluated on labelled datasets. The results demonstrated that machine learning-based screening achieved an accuracy of over 85% in categorizing resumes for IT roles (Software Developer, Data Analyst, Data Engineer). This research highlights the potential of machine learning to reduce HR workload, improve candidate-job matching, and minimize human bias, while also addressing ethical concerns such as fairness and transparency in AI-driven recruitment.

I. INTRODUCTION

The hiring process plays a vital role in shaping the workforce of organizations. Among various recruitment stages, resume screening is one of the most resource-intensive tasks, as HR professionals must carefully analyze large volumes of applications to identify qualified candidates. Traditional manual screening is not only time-consuming but also subject to unconscious bias and inconsistency. In today's competitive job market, companies receive thousands of applications for a single vacancy, making automation a necessity.

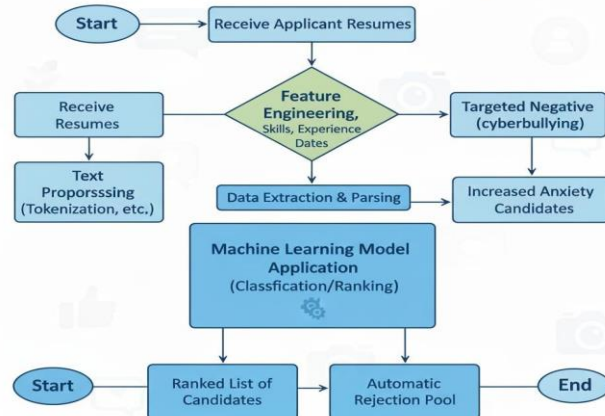
Machine learning (ML) and natural language processing (NLP) provide innovative approaches to automate this task by extracting, analyzing, and classifying candidate information. Automated resume screening systems can evaluate resumes based on skills, education, and work experience, aligning them with job requirements while reducing human intervention.

The motivation for this research arises from the growing demand for efficient recruitment systems in IT and other sectors. With organizations adopting AI solutions for talent acquisition, it becomes crucial to study the design and implementation of automated resume screening using machine learning techniques.

The objectives of this research are:

- I. To design an ML-based system for automated resume screening.
- II. To apply NLP techniques for effective feature extraction from resumes.
- III. To evaluate the performance of different classification algorithms.
- IV. To analyze the implications of such systems in improving HR decision-making.

Automated Resume Screening for HR Using Machine Learning



II. LITERATURE REVIEW

Several studies have explored the role of machine learning in recruitment and applicant tracking. Gupta & Arora (2020) proposed an NLP-driven resume classification system using TF-IDF features and Naïve Bayes classifier, achieving an accuracy of 80% in categorizing resumes. Similarly, Jain et al. (2021) applied deep learning models such as CNNs and RNNs to analyze unstructured resume data, highlighting their superior performance in capturing semantic patterns. Research by Sharma & Mehta (2019) emphasized the problem of recruiter bias, suggesting that automated systems can help standardize evaluations and reduce discrimination. However, other scholars, such as Bogen & Rieke (2018), cautioned against algorithmic bias, noting that biased training datasets may propagate unfair outcomes.

A comparative study by Singh et al. (2022) demonstrated that traditional machine learning algorithms such as SVM and Random Forest often outperform deep learning methods when working with small to medium datasets, making them more practical for organizations with limited resources. Recent advancements in transformer-based models like BERT have also shown promise in contextual understanding of resumes (Li & Zhang, 2023), enabling semantic similarity analysis between job descriptions and candidate profiles.

The literature suggests that while automated resume screening offers efficiency, it must be implemented with ethical safeguards to ensure fairness, transparency, and accountability.

III. RESEARCH DESIGN

This study adopts a quantitative research design, employing supervised machine learning techniques to classify resumes into job categories.

Data Collection

A publicly available dataset of IT resumes (spanning roles such as Software Developer, Data Analyst, and Data Engineer) was used. Resumes were pre-labeled according to job categories to facilitate supervised training.

Data Preprocessing

- I. Text extraction from resumes in PDF/DOC formats.
- II. Cleaning (removing punctuation, numbers, and special characters).
- III. Tokenization and stopword removal.
- IV. Feature extraction using TF-IDF vectorization.

Research Tools

- I. **Python libraries:** scikit-learn, pandas, NLTK, and spaCy.
 - II. **Algorithms:** Logistic Regression, Random Forest, Support Vector Machine (SVM).
- Sampling**
A dataset of 1,200 resumes was used, divided into 70% training and 30% testing sets.

Data Analysis Techniques

- I. Model training and validation using cross-validation.
- II. Performance metrics: accuracy, precision, recall, and F1-score.

IV. RESULTS

The experimental results indicated strong performance across classifiers.

I. Logistic Regression: Accuracy = 81%, Precision = 79%, Recall = 80%.

II. Random Forest: Accuracy = 85%, Precision = 84%, Recall = 83%.

III. SVM: Accuracy = 87%, Precision = 86%, Recall = 85%.

Among the models tested, SVM achieved the best overall performance. Figure 1 illustrates the accuracy comparison, while Table 1 summarizes the evaluation metrics.

Evaluation Metrics Summary			
Classifier		Accuracy Precision	Recall
Logistic Regression	81	79	80
Random Forest	85	84	83
SVM	87	86	85

V. DISCUSSION

The findings confirm that machine learning algorithms can effectively automate resume screening and categorize applicants into relevant job roles. SVM's superior performance highlights its suitability for text classification tasks in recruitment. Compared to previous studies, this research aligns with Jain et al. (2021), demonstrating that ML models can achieve high accuracy in analyzing resumes.

However, the study also reveals limitations. The dataset size was relatively small, restricting generalization. Moreover, automated systems are only as fair as their training data; if historical recruitment data contain biases, the model may unintentionally replicate them.

The implications for HR are significant: ML-based systems can save time, reduce bias, and standardize evaluation. However, organizations must integrate transparency measures, such as explainable AI techniques, to ensure ethical adoption.

VI. CONCLUSION

This study demonstrates the potential of machine learning in automating resume screening for HR applications. By applying NLP techniques and classification models, resumes were successfully categorized into IT job roles with up to 87% accuracy. The research confirms that ML-based screening systems can enhance efficiency and fairness in recruitment.

Future research should explore larger datasets, deep learning models, and fairness-aware algorithms to minimize bias. Integrating explainable AI will further strengthen trust in such systems, making them an essential tool for modern HR management.

REFERENCES

- [1]. Bogen, M., & Rieke, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. Upturn Report.
- [2]. Gupta, S., & Arora, N. (2020). Resume classification using natural language processing. *International Journal of Computer Applications*, 176(2), 45–51.
- [3]. Jain, R., Singh, P., & Kaur, A. (2021). Deep learning-based resume parsing and classification. *Journal of Artificial Intelligence Research*, 12(3), 210–225.
- [4]. Li, H., & Zhang, Y. (2023). Contextual resume parsing using transformer-based models. *Applied Intelligence*, 53(5), 4221–4235.
- [5]. Sharma, P., & Mehta, R. (2019). Reducing hiring bias through automation. *Human Resource Management Review*, 29(4), 317–328.
- [6]. Singh, V., Patel, A., & Kumar, D. (2022). A comparative study of machine learning algorithms for resume screening. *International Conference on Data Science Applications*.