International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

AI Powered Resume Screening System using NLP and ML

VIJAYALAKSHMI P[1], SRINANTHITHA L[2], SABISRI S[3], MRS.PRADEEPA K[4]

1, 2, 3 Members - 6th Semester B.E Students, Department of Computer Science and Engineering, E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India
4 Professor, Department of Computer Science and Engineering, E.G.S.Pillay Engineering College, Nagapattinam, Tamilnadu, India

How to Cite this Article:

P, V., L, S. & S, S. (2026). AI Powered Resume Screening System using NLP and ML. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.799

https://doi.org/ 10.55041/ijcope.v2i4.799

Abstract *— The rapid increase in job applications has made manual resume screening inefficient and time-consuming. This paper proposes an AI-powered resume screening system using Natural Language Processing (NLP) and Machine Learning (ML) techniques to automate candidate evaluation. The system extracts relevant information such as skills, education, and experience from resumes and ranks candidates based on job requirements. NLP techniques are used for text processing, while ML models are applied for classification and ranking. The proposed system improves recruitment efficiency, reduces bias, and enhances decision-making accuracy.*

Keypoints — Artificial Intelligence, Resume Screening, Natural Language Processing, Machine Learning, Recruitment Automation

I. INTRODUCTION

Recruitment plays a crucial role in organizational success. Traditional resume screening is manual, time-consuming, and often biased. With the advancement of Artificial Intelligence (AI), automated systems are being developed to enhance recruitment processes.

Natural Language Processing (NLP) helps in understanding and extracting meaningful information from resumes, while Machine Learning (ML) enables classification and prediction of candidate suitability. This paper presents an automated system that reduces human effort and improves efficiency in hiring.

II. PROBLEM STATEMENT

In the current recruitment scenario, organizations receive a large number of applications for each job opening, making manual screening highly inefficient. Recruiters often spend significant time reviewing resumes, which leads to delays in hiring. Additionally, human bias and inconsistency can affect the selection process, resulting in unfair evaluation of candidates. Important information may be overlooked due to fatigue or lack of standardized evaluation criteria. Therefore, there is a need for an automated system that can efficiently process large volumes of resumes, ensure consistent evaluation, and accurately identify suitable candidates based on job requirements.

III. OBJECTIVES

The primary objective of this research is to develop an AI-based system that automates the resume screening process. The system aims to extract relevant information from resumes using NLP techniques and classify candidates using machine learning algorithms. Another objective is to reduce the time and effort required in recruitment while improving the accuracy and consistency of candidate evaluation. Additionally, the system seeks to minimize human bias and provide a scalable solution that can handle large datasets effectively.

1.To design and develop an AI-powered system for automated resume screening.

2.To extract relevant candidate information using NLP techniques such as tokenization and Named Entity Recognition.

3.To apply Machine Learning algorithms for classification and ranking of candidates.

4.To reduce the time and effort involved in manual screening.

5.To ensure fairness and consistency in the recruitment process.

6.To create a scalable system capable of handling large datasets.

IV. LITERATURE REVIEW (SUMMARY)

Various research studies have explored the application of Artificial Intelligence in recruitment systems. Early approaches relied on simple keyword matching techniques, which lacked contextual understanding and often produced inaccurate results. With the introduction of Natural Language Processing, more advanced techniques such as tokenization, part-of-speech tagging, and Named Entity Recognition have been used to extract meaningful information from resumes. Machine Learning models such as Naïve Bayes, Support Vector Machines, and Random Forest have been widely used for classification tasks. Recent advancements in deep learning, including the use of neural networks and transformer-based models, have further improved the accuracy of text analysis. Despite these developments, challenges such as data bias, lack of semantic understanding, and limited adaptability across domains still exist.
However, challenges such as bias, lack of contextual understanding, and data imbalance still remain.

V. SYSTEM ARCHITECTURE

The proposed system architecture consists of multiple interconnected modules that work together to automate resume screening. The process begins with the input module, where resumes and job descriptions are provided to the system.
The system consists of multiple modules that work together in a pipeline.

1. Input Layer

Accepts resumes in formats such as PDF, DOCX.
Accepts job descriptions.

2. Preprocessing Layer

Converts documents into plain text.
Removes noise such as symbols, formatting.
Normalizes text (lowercase conversion, punctuation removal).

3. NLP Processing Layer

Tokenization (splitting text into words).
Stop-word removal (removing common words like “the”, “is”)
Named Entity Recognition (NER) to identify skills, names, organizations
Parsing for sentence structure

4. Feature Extraction Layer

Converts text into numerical vectors
Uses techniques like TF-IDF or Word Embeddings

5. Machine Learning Layer

Trains classification models

Predicts candidate suitability
Ranks candidates based on scores

6. Output Layer

Displays ranked candidates.
Provides selection recommendations.

VI. METHODOLOGY / ALGORITHM

Methodology:

The system follows these steps:

1. Collect resume dataset

2. Convert resumes into text format

3. Perform preprocessing:

Tokenization

Stop-word removal

Stemming/Lemmatization

4. Apply NLP for data extraction

5. Convert text into features using TF-IDF

6. Train ML models (SVM, Logistic Regression)

7. Evaluate model performance

8. Rank candidates based on scores

Algorithm:

Step 1: Input resume dataset
Step 2: Preprocess text
Step 3: Extract features using NLP
Step 4: Apply TF-IDF vectorization
Step 5: Train ML classifier
Step 6: Predict candidate suitability
Step 7: Rank candidates

VII. ADVANTAGES

Automates repetitive tasks.

Reduces hiring time.

Improves candidate matching accuracy.

Ensures consistency in evaluation.

Scalable for large recruitment processes.

Reduces human bias.

VIII . LIMITATIONS

Despite its advantages, the system has certain limitations that need to be addressed.
The performance of the system depends heavily on the quality and size of the training dataset.
If the data contains bias, the system may produce biased results.
The system may also struggle to interpret complex or non-standard resume formats.
Additionally, implementing and maintaining such a system requires computational resources and technical expertise.

IX . RESULTS AND DISCUSSION

The system significantly reduces the time required for resume screening.

It improves accuracy in matching candidates with job requirements.

The automated process ensures consistency in evaluation.

The system can handle large volumes of data efficiently.

Performance depends on the quality of training data and feature extraction methods.

Complex resumes and ambiguous data may affect accuracy.

Continuous model training can improve system performance over time.

X. FUTURE SCOPE

Integration with deep learning models such as BERT and transformer-based architectures.

Development of real-time resume screening systems.

Support for multilingual resumes to enable global recruitment.

Integration with job portals and HR management systems.

Implementation of explainable AI to improve transparency in decision-making.

XI. CONCLUSION

The AI-powered resume screening system provides an efficient and scalable solution for modern recruitment challenges.

The integration of NLP and ML techniques enables accurate extraction and classification of resume data.

The system reduces manual effort, improves efficiency, and ensures consistent evaluation.

It serves as a valuable tool for organizations aiming to streamline their hiring process.

The use of Artificial Intelligence in recruitment represents a significant advancement in hiring practices.

The proposed system demonstrates how automation can improve decision-making and reduce bias.

With further improvements, such systems can become essential tools in modern HR management.

REFERENCES

[1] J. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., 2021. https://web.stanford.edu/~jurafsky/slp3/

[2] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," 2013. https://arxiv.org/abs/1301.3781

[3] J. Pennington et al., "GloVe: Global Vectors for Word Representation," 2014 https://nlp.stanford.edu/pubs/glove.pdf

[4] I. Goodfellow et al., Deep Learning, MIT Press, 2016. https://www.deeplearningbook.org/

[5] C. Cortes and V. Vapnik, "Support-Vector Networks," 1995. https://link.springer.com/article/10.1007/BF00994018

[6] L. Breiman, "Random Forests," 2001. https://link.springer.com/article/10.1023/A:1010933404324

[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," 2011. https://jmlr.org/papers/v12/pedregosa11a.html

[8] S. Bird et al., Natural Language Processing with Python, 2009. https://www.nltk.org/book/

[9] A. Vaswani et al., "Attention Is All You Need," 2017. https://arxiv.org/abs/1706.03762

[10] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers," 2019. https://arxiv.org/abs/1810.04805

[11] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes," 1998. https://people.cs.umass.edu/~mccallum/papers/naivebayes-aaaiws98.pdf

[12] T. Joachims, "Text Categorization with Support Vector Machines," 1998. https://dl.acm.org/doi/10.5555/645326.649721