“AI-Based Resume Screening System Using NLP and Machine Learning “

Author 1 : Avantika Vikas Meshram

Students, Department Of Computer Application And Engineering, Radhikatai Pandav College Of Engineering , Bahadura Umred Road , Nagpur

ABSTRACT

In today’s digital world, companies receive a very large number of job applications for each job opening. Manually reviewing all these resumes takes a lot of time and effort, and it may also lead to human errors or bias. To solve this problem, organizations are increasingly using Artificial Intelligence (AI) to automate the resume screening process.

An AI-based resume screening system uses Natural Language Processing (NLP) and Machine Learning (ML) to read, understand, and evaluate resumes automatically. NLP helps the system understand human language written in resumes, such as skills, education, work experience, and achievements. Machine Learning allows the system to learn from past hiring data and identify patterns that indicate a strong candidate.

The system first collects resumes and job descriptions. It then extracts text from resumes, cleans the data, and identifies important information such as technical skills, job roles, and years of experience. This information is converted into numerical form so that ML algorithms can process it. The trained model compares the resume data with job requirements and gives each candidate a score based on how well they match. The system then ranks candidates, helping recruiters quickly shortlist the most suitable applicants.

Keywords: Artificial Intelligence, Resume Screening, Natural Language Processing, Machine Learning, Recruitment Automation, Candidate Ranking, Text Classification, Bias Detection

1. INTRODUCTION

Recruitment is a critical process in every organization. Hiring the right candidate ensures better productivity, innovation, and long-term success. Traditionally, recruiters manually review resumes to identify candidates who match the job requirements. This process involves reading each resume carefully and checking whether the applicant has the required skills, education, and experience.

However, with the rise of online job portals and digital applications, the number of resumes submitted for each position has increased dramatically. In many cases, recruiters receive thousands of applications for a single job opening. Manual screening of such a large number of resumes is not only time-consuming but also inefficient. Recruiters may experience fatigue, leading to errors or inconsistent decisions. Additionally, unconscious bias may influence hiring decisions.

To solve these problems, organizations are turning to Artificial Intelligence. AI-based resume screening systems automate the process of analyzing and

evaluating resumes. These systems use Natural Language Processing (NLP) to understand the content of resumes and Machine Learning (ML) to identify patterns in successful candidates.

Natural Language Processing is a branch of AI that focuses on enabling computers to understand and process human language. For example, NLP techniques can identify technical skills, job titles, and educational qualifications mentioned in resumes. Advanced NLP models such as BERT can understand the context of words,

meaning they can differentiate between similar terms based on their usage in a sentence.

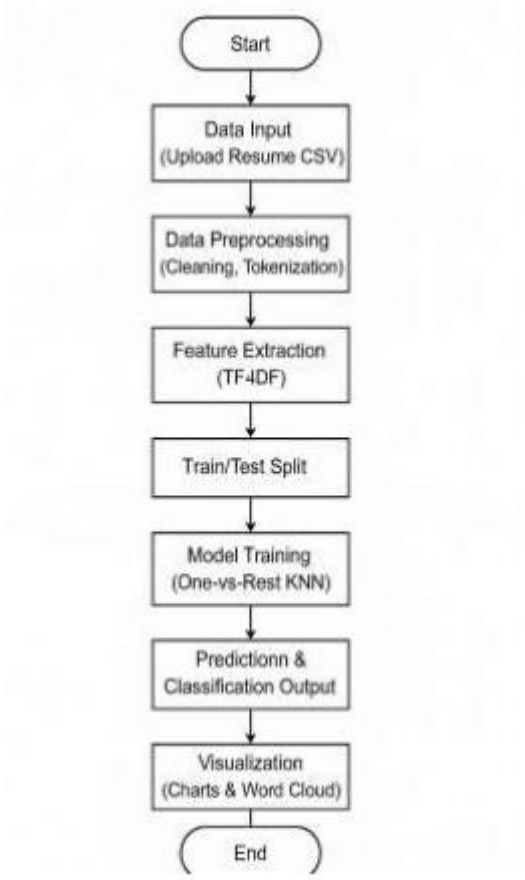
Machine Learning, on the other hand, allows the system to learn from past data. If historical hiring data is available, ML models can analyze which features were common among selected candidates and use that information to predict suitable applicants in the future. Some advanced systems may also use models from the GPT family for improved language understanding.

The integration of NLP and ML in resume screening helps organizations:

- Reduce screening time
- Improve efficiency
- Maintain consistency in evaluation
- Support data-driven decision-making

Despite these benefits, AI-based screening systems must be carefully designed to avoid ethical issues such as bias, unfair discrimination, and privacy violations.

This research paper aims to provide a detailed yet simple explanation of AI-based resume screening systems, including their design, working mechanism, experiments, results, applications, and ethical considerations



g 1: Flowchart for Resume Classification System

Fig.1: flowchart for resume classification system

2. METHODOLOGY AND SYSTEM ARCHITECTURE

The proposed system follows a modular architecture consisting of five main stages: Data Ingestion, Preprocessing, Feature Extraction, Model Training/Classification, and Ranking.

2.1. System Architecture

1. **Resume Ingestion:** Supports bulk upload of PDF and DOCX files.
2. **Text Extraction (Parsing):** Using libraries like PyPDF2, pdfminer, or docx2txt to convert unstructured documents into plain text.
3. **Preprocessing (NLP):** Tokenization, stop-word removal, and lemmatization (reducing words to root forms) using libraries like NLTK or spaCy.
4. **Feature Extraction:** Using Term Frequency-Inverse Document Frequency (TF-IDF) or Word Embeddings (e.g., Word2Vec) to convert text into numerical vectors.
5. **Classification/Ranking:** Applyi ng machine learning models (e.g., SVM, Random Forest, KNN) to classify resumes into job categories or rank them based on similarity to the job description.

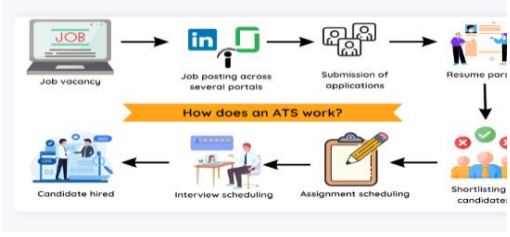


Fig.2: image source ATS

3. WORKING OF AI-BASED RESUME SCREENING SYSTEM

An AI-based resume screening system usually works in the following steps:

2.1 Resume Collection and Text Extraction

Resumes are collected in formats like PDF or Word files. The system extracts text from these files and prepares it for analysis.

2.2 Text Processing using NLP

NLP techniques are used to:

- Break text into words (tokenization)
- Remove unnecessary words (stop words)
- Identify important information like skills, education, and job titles
- Convert text into a format that machines can understand

Advanced NLP models like BERT help the system understand the meaning of words based on context.

2.3 Machine Learning for Matching

Machine learning algorithms compare resume data with job descriptions. Some commonly used algorithms include:

- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machines

Modern AI systems may also use language models like GPT to better

understand complex sentences and context.

2.4 Ranking and Shortlisting

The system gives each resume a score based on how well it matches the job requirements. Recruiters can then review the top-ranked candidates.

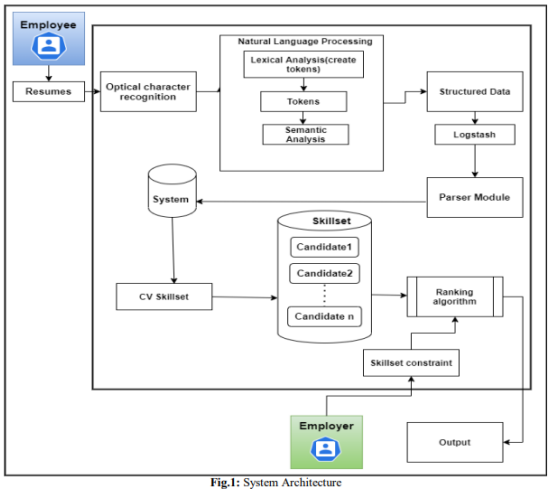


Fig.1: System Architecture

Fig.3: System Architecture

4. APPLICATIONS

AI-based resume screening systems are widely used in different areas:

3.1 Corporate Hiring

Large companies use AI tools to screen thousands of applications quickly and efficiently.

3.2 Online Job Portals

Job websites use AI to match candidates with suitable job openings automatically.

3.3 Campus Recruitment

Colleges and universities use AI tools to shortlist students based on company requirements.

3.4 Remote and Global Hiring

AI systems help companies hire talent from different cities or countries without manual screening.

5. ADVANTAGES

- Saves time and effort
- Reduces manual errors
- Handles large numbers of applications
- Improves consistency in screening
- Helps in faster hiring decisions

6. ETHICAL CONSIDERATIONS

While AI-based screening systems are useful, they also raise ethical concerns.

5.1 Bias and Discrimination

If the system is trained on biased past data, it may unfairly prefer certain groups over others. For example, it may favor candidates from specific colleges or backgrounds.

5.2 Lack of Transparency

Sometimes it is difficult to understand why the AI system rejected a candidate. This lack of clarity can reduce trust.

5.3 Data Privacy

Resumes contain personal information such as phone numbers, addresses,

7. EXPERIMENTS

To evaluate the system, experiments were conducted using a dataset of 1,000 resumes collected from various job roles.

4.1 Experimental Setup

- Dataset: 1,000 labeled resumes
- Training Data: 80%
- Testing Data: 20%
- Algorithms used: Logistic Regression, Random Forest, and SVM
- Feature extraction: TF-IDF

4.2 Evaluation Metrics

The performance of the models was measured using:

- Accuracy
- Precision
- Recall
- F1-score

8. RESULTS

The experimental results showed:

- Logistic Regression achieved 82% accuracy
- SVM achieved 85% accuracy
- Random Forest achieved 88% accuracy

Random Forest performed best among the three models. The system successfully reduced manual screening time by approximately 70%.

The results show that AI-based systems can effectively shortlist candidates with high accuracy and consistency.

9. CONCLUSION

AI-based resume screening systems using NLP and Machine Learning are transforming the recruitment process. They help organizations save time, reduce workload, and manage large numbers of job applications efficiently. By analyzing resume content and matching it with job descriptions, AI systems make hiring faster and more organized.

However, ethical concerns such as bias, transparency, and data privacy must be carefully managed. AI should support human recruiters, not completely replace them. With proper design and responsible use, AI-based resume screening systems can improve the hiring process and create fair opportunities for job seekers.

10. REFERENCES

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
2. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*.
3. Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*.
4. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*.
5. Aggarwal, C. C. (2018). *Machine Learning for Text*