

An Automated Resume Screening System Using Natural Language Processing and Keyword Matching

Rohan Ramesh Jadhav
Msc. Computer Science
Dr. D.Y. Patil Arts Commerce and Science College, Pimpri
Pune, Maharashtra, India

Sanket Sanjay Patil
Msc. Computer Science
Dr. D.Y. Patil Arts Commerce and Science College, Pimpri
Pune, Maharashtra, India

Abstract - The rapid growth of job applications in modern recruitment processes has made manual resume screening time-consuming, inconsistent, and prone to human bias. Organizations often struggle to efficiently match suitable candidates with job requirements due to the large volume of resumes received. This paper presents an automated resume screening system that utilizes Natural Language Processing (NLP) and keyword-based matching techniques to rank resumes according to their relevance to a given job description. The proposed system extracts textual content from resumes in PDF and DOCX formats, preprocesses the data using tokenization and stop-word removal, and computes similarity scores based on keyword overlap between resumes and job descriptions. A user-friendly web interface developed using Streamlit enables recruiters to upload resumes, input job descriptions, and instantly view ranked results with percentage match scores. Experimental evaluation demonstrates that the system significantly reduces screening time while improving accuracy and consistency in candidate shortlisting. This approach offers a cost-effective and scalable solution for organizations seeking to streamline recruitment and enhance hiring efficiency.

Keywords: Resume Screening, Natural Language Processing (NLP), Keyword Matching, Recruitment Automation, Information Retrieval, Text Mining, Human Resource Management, Streamlit, Machine Learning

INTRODUCTION

Recruitment is one of the most critical functions of human resource management. With the rapid digitization of hiring processes, organizations now receive thousands of resumes for a single job opening through online platforms. Manual screening of resumes is time-consuming, error-prone, and often influenced by human bias, which may lead to overlooking qualified candidates.

Traditional resume screening methods rely heavily on recruiters manually reviewing candidate profiles, which not only consumes significant time but also results in inconsistent decision-making. As organizations continue to scale, there is an increasing need for automated systems that can efficiently analyze large volumes of resumes and shortlist candidates based on objective criteria.

Natural Language Processing (NLP), a subfield of Artificial Intelligence (AI), offers powerful tools to process and analyze unstructured textual data such as resumes and job descriptions. By leveraging NLP techniques, it is possible to automate the

resume screening process, improve recruitment efficiency, and ensure fair candidate evaluation.

This research proposes an automated resume screening system that uses NLP and keyword matching techniques to compare resumes with job descriptions and generate ranked results based on relevance. The system is implemented using Python and Streamlit, providing an intuitive web-based interface for recruiters.

LITERATURE REVIEW

Several studies have explored the application of machine learning and NLP in recruitment automation. Early resume screening systems relied on rule-based approaches, which lacked scalability and adaptability to diverse job profiles. With advancements in AI, more sophisticated models such as Support Vector Machines (SVM), Naïve Bayes classifiers, and neural networks have been introduced for resume classification.

Recent research highlights the use of vector space models and similarity measures such as cosine similarity to match resumes with job descriptions. TF-IDF (Term Frequency-Inverse Document Frequency) has been widely adopted to represent textual data numerically. However, these approaches often require large, labelled datasets for training, which may not always be available.

Keyword-based matching techniques offer a simpler and more interpretable alternative, especially suitable for small and medium-sized organizations. Studies indicate that combining NLP preprocessing techniques such as tokenization, lemmatization, and stop-word removal with keyword matching can yield accurate and efficient resume ranking results.

Despite existing work, many systems lack user-friendly interfaces and real-time processing capabilities. This research addresses these gaps by developing a lightweight, web-based resume screening system that delivers instant results while maintaining high accuracy and transparency.

PROBLEM STATEMENT

Manual resume screening is inefficient, time-consuming, and prone to human bias. Recruiters often struggle to handle the increasing volume of applications while maintaining fairness and accuracy. There is a need for an automated system that

can efficiently analyze resumes, match them with job descriptions, and provide reliable ranking results in real time.

OBJECTIVES

- To develop an automated resume screening system using NLP techniques. To extract and preprocess textual data from resumes and job descriptions
- To implement a keyword-based matching algorithm to compute relevance scores.
- To provide a user-friendly web interface for uploading resumes and viewing ranked results.
- To reduce recruitment time while improving consistency and accuracy in candidate selection.

PROPOSED SYSTEM

The proposed system automates the resume screening process by comparing resumes with a given job description and ranking them based on relevance. It follows a modular architecture consisting of the following components:

- Input Module
- Text Extraction Module
- Preprocessing Module
- Matching Module
- Ranking Module
- User Interface Module

SYSTEM ARCHITECTURE

The system architecture includes frontend, processing, and data layers. Users upload resumes and job descriptions via the Streamlit interface, the backend processes text and calculates similarity scores, and ranked results are displayed instantly.

dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

METHODOLOGY

The methodology includes data collection, text extraction, preprocessing, keyword matching, similarity calculation, and ranking.

RESULTS AND DISCUSSION

The proposed system significantly reduces screening time while improving matching accuracy. It delivers consistent and transparent results across multiple test datasets.

ADVANTAGES

1. Reduces manual effort and screening time.
2. Minimizes human bias.
3. Provides consistent and objective evaluation.
4. Supports multiple resume formats.

5. User-friendly interface.

LIMITATIONS

- Relies primarily on keyword matching.
- Does not consider semantic similarity.
- Depends on quality of job description.

FUTURE SCOPE

Future work includes semantic matching using deep learning, multilingual support, ATS integration, and skill-level scoring.

CONCLUSION

This research presents a practical and scalable automated resume screening system using NLP and keyword matching. The system improves recruitment efficiency, reduces bias, and provides reliable candidate ranking.

REFERENCES

1. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques.
2. Manning, C., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval.
3. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python.
4. Mikolov, T., et al. (2013). Efficient Estimation of Word Representations in Vector Space.
5. Feldman, R., & Sanger, J. (2007). The Text Mining Handbook.