SkillSift: Automated Resume Classification and Skill Extraction Platform Using NLP

[1]V Harini and *[2]R Nithya

[1]Student of II M.Sc., Department of Computer Science, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.

*[2]Assistant Professor, Department of Computer Science, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.

Abstract

The rapid growth of job applications has made manual resume screening time-consuming and inefficient for recruiters. To address this challenge, SkillSift is proposed as an automated resume classification and skill extraction platform powered by Natural Language Processing (NLP). The system analyzes resumes submitted in digital formats and automatically identifies key information such as technical skills, educational background, work experience, and certifications. Using NLP techniques such as text preprocessing, tokenization, named entity recognition, and semantic analysis, the platform extracts relevant skill sets and categorizes resumes based on job roles or domains. SkillSift enhances the recruitment workflow by reducing manual effort and enabling faster candidate shortlisting. The system also improves accuracy in identifying candidate competencies by analyzing contextual information within resumes. Recruiters can efficiently filter and rank applicants based on required skills, improving the quality of hiring decisions. Additionally, the platform supports scalable resume processing, making it suitable for organizations handling large volumes of applications. Overall, SkillSift demonstrates how NLP-driven automation can modernize recruitment systems by transforming unstructured resume data into meaningful insights. This approach not only saves time for recruiters but also ensures a fair and consistent evaluation of candidates.

Keywords: Natural Language Processing, Resume Classification, Skill Extraction, Automated Recruitment, Text Mining.

1. Introduction

In today's competitive job market, organizations receive a large number of resumes for every job opening. Manually reviewing and filtering these resumes is a time-consuming and inefficient process for recruiters. Traditional recruitment methods often require human resource professionals to read each resume individually to identify relevant skills, qualifications, and experience. This manual approach not only increases workload but also creates the possibility of human bias and errors during candidate selection. Therefore, there is a growing need for intelligent systems that can automatically analyze resumes and assist recruiters in identifying suitable candidates quickly and accurately. Recent advancements in Natural Language Processing (NLP) and machine learning have made it possible to process large amounts of textual data efficiently. Resumes usually contain unstructured text that includes information about education, technical skills, work experience, certifications, and projects. NLP techniques can be applied to analyze this unstructured information and convert it into structured data that can be easily interpreted by recruitment systems. By extracting meaningful insights from resumes, automated systems can significantly improve the efficiency of the hiring process. To address these challenges, this project proposes SkillSift: an Automated Resume Classification and Skill Extraction Platform using NLP. The proposed system is designed to analyze resumes, extract important skill sets, and classify candidates according to their qualifications and job relevance. The platform uses NLP techniques such as text preprocessing, tokenization, and skill recognition to identify relevant keywords and categorize resumes based on specific job roles or domains. The main objective of SkillSift is to simplify the recruitment process by reducing manual effort and improving the accuracy of candidate screening. By automatically identifying relevant skills and classifying resumes, the system helps recruiters quickly shortlist potential candidates from a large pool of applications. This approach not only saves time but also ensures a more consistent and objective evaluation process. Overall, the integration of NLP-based resume analysis in recruitment systems can significantly enhance hiring efficiency and decision-making. SkillSift demonstrates how intelligent automation can transform traditional recruitment practices by enabling faster, scalable, and more effective candidate evaluation.

*Corresponding Author: R Nithya

< 6 >

2. Review of Literature

Several researchers have explored the use of Natural Language Processing and machine learning techniques to improve recruitment systems. Malinowski *et al.* (2006) [6] studied the concept of automated resume processing and proposed methods for extracting structured information from unstructured resumes. Their research demonstrated that NLP techniques can effectively identify candidate attributes such as education, professional experience, and skill sets. The study emphasized that automated resume analysis can reduce the workload of recruiters and help organizations manage large volumes of applications more efficiently. Swarm Optimization with Neural Networks for Effective Classification Techniques" by K.Kalyani (2021) [2] introduces a hybrid EHBMO-NN model, combining Extended Honey Bee Mating Optimization with Artificial Neural Networks to improve classification accuracy and reduce training time. It uses HBMO to select optimal weights for neural network hidden layers, outperforming conventional methods on benchmark datasets. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray data is used for cancer classification. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray data is used for cancer classification (6). Another study conducted by Balog and de Rijke (2006) [7] investigated the use of machine learning models for identifying experts and classifying professional profiles based on textual data. Their research applied classification algorithms to analyze candidate documents and determine their suitability for specific roles. The results showed that automated classification methods could improve the accuracy of candidate selection and assist recruiters in identifying individuals with relevant expertise. These studies highlight the potential of combining NLP techniques with intelligent algorithms to enhance recruitment systems. More recent research has also focused on improving resume analysis by incorporating semantic understanding and contextual analysis of text. By analyzing the relationships between words and phrases, advanced NLP systems can better understand the meaning of candidate skills and experiences. These developments indicate that intelligent resume analysis platforms can significantly enhance the recruitment process by providing more accurate and efficient candidate screening.

3. Existing System

In the existing recruitment process, many companies still use manual resume screening or basic Applicant Tracking Systems (ATS) to evaluate candidates. Recruiters review resumes individually or filter them using simple keyword searches. This process becomes difficult when organizations receive a large number of applications for a single job position. Most traditional systems rely only on keyword matching, which cannot understand the context of skills and experience mentioned in resumes. Because of this, some qualified candidates may be ignored while irrelevant resumes may be selected. Additionally, different resume formats can cause errors in information extraction. Therefore, the existing system is often time-consuming, less accurate, and inefficient for large-scale recruitment.

4. Proposed System

The proposed system, SkillSift, is designed to automate the process of resume analysis by using Natural Language Processing techniques. The system accepts resumes in digital formats and processes the textual content to extract relevant information about the candidate. Initially, the resume text undergoes preprocessing steps such as removing unnecessary characters, converting text into standardized formats, and dividing sentences into tokens. This preprocessing stage ensures that the text is prepared for further analysis. After preprocessing, the system identifies important information related to candidate skills, education, and work experience using NLP-based extraction techniques. The extracted information is then organized into a structured format that can be easily analyzed by the system. SkillSift compares the extracted skill sets with predefined job requirements and classifies resumes according to their relevance to specific job roles. The system can also rank candidates based on how closely their qualifications match the desired criteria. By automating resume analysis, the proposed platform reduces the time required for candidate screening and improves the overall efficiency of the recruitment process. Recruiters can quickly identify the most suitable candidates without manually reviewing every resume. The system also ensures consistent evaluation by applying the same analysis criteria to all applicants.

5. Experimental Result

The SkillSift system was tested using a dataset of sample resumes representing different job roles and skill sets. The system successfully extracted relevant information such as technical skills, educational qualifications, and work experience from the resumes. After processing the documents, the platform classified candidates into appropriate categories based on their skill profiles. The experimental evaluation demonstrated that the use of NLP techniques significantly improved the accuracy of resume analysis compared to simple keyword-based approaches. The system was able to recognize different variations of similar skills and identify relevant candidate information even when the wording varied. Additionally, the automated classification process reduced the time required for resume screening, allowing recruiters to review shortlisted candidates more efficiently.
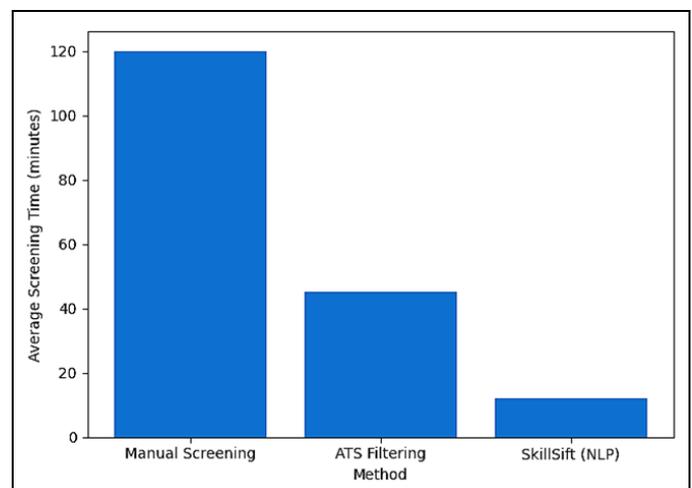
Fig 1: Screening Time Comparison

These results indicate that the SkillSift platform can

< 7 >

effectively support recruitment activities by providing reliable and accurate analysis of candidate resumes. The system demonstrates the potential benefits of applying NLP techniques in human resource management and recruitment processes.
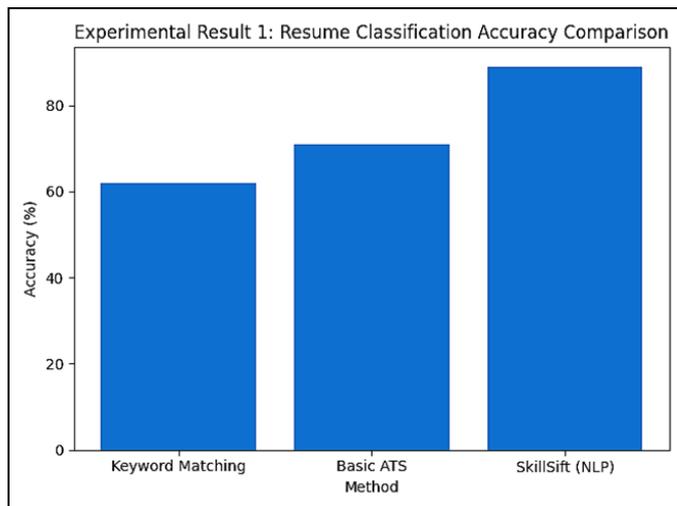
Fig 2: Accuracy Comparison

6. Conclusion

The increasing volume of job applications has made manual resume screening an inefficient and time-consuming process for recruiters. Intelligent technologies such as Natural Language Processing provide effective solutions for automating resume analysis and improving candidate evaluation. The SkillSift platform demonstrates how NLP techniques can be used to extract relevant information from resumes and classify candidates based on their skills and qualifications. By converting unstructured resume content into structured data, the system enables recruiters to quickly identify suitable candidates from large applicant pools. The proposed approach not only reduces manual effort but also ensures consistent and objective evaluation of applicants. The experimental results show that automated resume classification can improve both efficiency and accuracy in the recruitment process. Future improvements may include integrating machine learning models and advanced semantic analysis to further enhance the system's capability in understanding complex resume content.

References

1. Liddy ED. Natural Language Processing. *Encyclopedia of Library and Information Science*. Marcel Dekker Inc.; 2002.
2. Kalyani K. Swarm Optimization with Neural Networks for Effective Classification Techniques. *Annals of the Romanian Society for Cell Biology*. 2021;25(4):7413-7419. "Vasile Goldis" Western University Arad, Romania.
3. Kalyani K. Classification of Microarray Gene Expression using Artificial Neural Network (ANN). *Turkish Journal of Computer and Mathematics Education*. 2021;12(7):1372-1378. Ninety Nine Publication.
4. Kalyani K, Chakravarthy T. An Algorithmic Approach with Improved Replacement in Bee Optimization Algorithm. *ICTACT Journal on Soft Computing*. 2015;5(2):905-910. ICTACT.
5. Kalyani K. Microarray Data Classification using Artificial Neural Network. *International Journal of Engineering and Advanced Technology (IJEAT)*. 2019;9(1S2):54-56. Blue Eyes Intelligence Engineering & Sciences Publication.
6. Malinowski J, Keim T, Wendt O, Weitzel T. Matching People and Jobs: A Bilateral Recommendation Approach. *Proceedings of the 39th Hawaii International Conference on System Sciences*; 2006.
7. Balog K, Azzopardi L, de Rijke M. Formal Models for Expert Finding in Enterprise Corpora. *Proceedings of the 29th Annual International ACM SIGIR Conference*; 2006.
8. Jurafsky D, Martin JH. *Speech and Language Processing*. Pearson Education; 2009.
9. Aggarwal CC, Zhai C. *Mining Text Data*. Springer; 2012.
10. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*; 2019.
11. Zhang Y, Yang Q. A Survey on Multi-Task Learning in Natural Language Processing. *IEEE Transactions on Knowledge and Data Engineering*; 2021.

< 8 >