International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

SMART RESUME ANALYZER

Mr. Ashish Modi[1], Mr. Hassan Raza Chowdhary[2]

[1]Asst. Prof / Department of Information Technology, Nagindas Khandwala College, Mumbai, Maharashtra, India,
[2] Student / Department of Information Technology, Nagindas Khandwala College, Mumbai, Maharashtra, India.

---***---

Abstract - In today's competitive job market, the need for efficient and effective hiring processes is paramount. The Smart Resume Analyzer is an innovative AI-powered tool reimagining the recruitment landscape by seamlessly integrating advanced machine learning (ML) and natural language processing (NLP) techniques. Unlike traditional resume screening methods that are often hampered by manual evaluation, subjectivity, and inefficiencies, this system offers a cutting-edge approach to candidate assessment. It intelligently extracts and analyzes key information from resumes, utilizing sophisticated algorithms such as Cosine Similarity and Term Frequency-Inverse Document Frequency (TF-IDF) to provide real-time, data-driven evaluations.

Moreover, this tool democratizes the job search process by empowering applicants with the knowledge and tools to refine their resumes, thereby leveling the playing field in a competitive job market. The Smart Resume Analyzer is more than just a recruitment aid—it is a transformative solution that bridges the gap between employers and the talent they seek, fostering a more efficient and equitable hiring process.

Key Words: NLP (Natural Language Processing) Resume Parsing, Machine Learning, Recommender Systems, Data Security, Data Privacy, User Engagement, Semantic Analysis, User Data Analysis: & Text Mining.

1. INTRODUCTION

Any Recruiters will often face the challenge of selecting the best candidates from an overwhelming pool of applicants for a given job vacancy. Manually sifting through thousands of resumes to identify the most qualified candidates is not only time-consuming but also prone to errors and inconsistencies. Although job websites have introduced methods to improve accuracy and precision in candidate matching, a significant limitation remains in the time it takes to process and compare every resume against multiple job postings. This high time complexity becomes a bottleneck, especially when dealing with vast datasets.

With over 50,000 e-recruitment platforms emerging in recent years, various strategies have been developed to streamline the process of matching candidates to job profiles. Some of these platforms have successfully implemented techniques to categorize and rank resumes based on their relevance to specific job postings. Among the most effective methods are Content-based Recommendation systems, Cosine Similarity, and K-Nearest Neighbours (KNN) algorithms, which help in identifying the resumes that most closely align with a job description.

However, despite these advances in accuracy and precision, the time required to find suitable candidates remains a significant drawback. The challenge lies in efficiently and accurately matching each resume to the appropriate job posting in a way that reduces processing time without compromising the quality of the results. To address this, innovative approaches are needed that not only improve the speed of candidate identification but also maintain high standards of accuracy and relevance in matching candidates to job opportunities. These advancements will be crucial in optimizing the e-recruitment process, making it more efficient and effective for both recruiters and job seekers.

2. REVIEW OF LITERATURE

Satyaki Sanyal and his team [1] have developed a resume analysis software that automates the extraction and evaluation of pertinent information from submitted resumes. Their approach integrates Natural Language Processing (NLP), Machine Learning (ML), and data mining techniques to detect keywords, patterns, and trends. By assessing resumes in relation to job specifications, the software streamlines the recruitment process, lightens the load on recruiters, and highlights the most qualified candidates.

Vaidya and colleagues [2] describe a technique used by resume analyzers to retrieve relevant information. Their method involves removing stop words and applying the Soundex algorithm to group similar-sounding words.

Identified keywords are then cataloged in a database for further processing.

Daryania and his team [3] present an automated resume screening system that leverages NLP to evaluate resumes and rank them based on their alignment with job requirements. This system extracts keywords, skills, and qualifications from resumes and compares them to job descriptions, generating a relevance score to assist in ranking.

Nawander and associates [4] introduce a method combining NLP with Streamlit modules to extract data from PDF resumes, store it in a database, and analyze it for ranking purposes. Their system also offers suggestions for enhancing resume presentation, including formatting, layout, and language.

Pokhare [5] outlines an approach that employs NLP and machine learning to parse, extract, and summarize data from PDF resumes. This system identifies key sections such as contact details, education, and work experience, and uses NLP and ML techniques to extract and analyze keywords, skills, and qualifications. The extracted data is stored for comparison and further analysis.

Kelkar and team [6] propose a Company Recommender System designed to match candidates to the best-fit company. Their system uses text mining and machine learning to rank resumes based on company-specific requirements. By extracting relevant information from resumes and comparing it to job specifications, the system assigns scores and leverages a machine learning model trained on historical data to identify patterns. This approach ranks resumes and provides a list of recommended candidates to recruiters.

Sinha and colleagues [7] introduce a method for resume screening utilizing NLP and ML algorithms. The goal is to automate the CV screening process, allowing the system to analyze and extract essential information from unstructured text.

Shubham Bhor and his team [8] propose a solution for resume parsing using NLP techniques. Their goal is to streamline the identification of suitable candidates for job openings by extracting key details from resumes uploaded to job portals.

Sanjana and her team [9] focus on resume validation and filtering using built-in NLP techniques. Their approach includes checking the accuracy of resumes and filtering job applications. This addresses the challenge of manually reviewing a large volume of resumes, which has become increasingly impractical for recruiters.

Thakur and Goyal [10] present a Resume Classification System (RCS) that uses NLP and ML techniques to automate resume analysis. While their model improves the efficiency and transparency of the screening process, it does not provide recommendations for enhancing applicant resumes. Their work demonstrates how NLP and RCS can significantly reduce the recruiter's workload and deliver effective results.

3. METHODOLOGY

The development of a smart resume analyzer that focuses on job recommendation, resume classification, and information extraction involves several key steps. This methodology integrates Natural Language Processing (NLP) techniques and machine learning algorithms to automate the resume screening process, enhancing efficiency and accuracy.

i. Data Collection

Acquire a diverse and representative dataset of resumes to ensure the system can handle various job profiles and categories.

- Sources: Gather resumes from job boards, company career pages, and publicly available datasets. Ensuring a diverse set of sources helps capture different resume styles and formats.
- Dataset Composition: The dataset should include key fields such as resume text, job category, skills, education, and work experience. This diversity will support comprehensive analysis and effective classification.

ii. Data Preprocessing

Prepare the collected resume data for analysis by cleaning and standardizing it.

- Text Cleaning: Remove unnecessary elements such as headers, footers, and irrelevant text to focus on the core resume content. Normalize text by converting it to lowercase to maintain consistency.
- Tokenization: Break down the text into individual words or tokens. This step simplifies the text for further analysis.

- Lemmatization/Stemming: Reduce words to their base or root forms (e.g., "running" to "run"). This standardization helps in consolidating similar terms.
- Vectorization: Convert the cleaned and tokenized text into numerical representations using techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) or Bag-of-Words. This transformation makes the text data usable for machine learning models.

iii. Feature Extraction

Extract and identify key features from resumes that are necessary for effective classification and recommendation.

- Skills Extraction: Use NLP methods to identify and extract specific skills mentioned in the resumes. Techniques such as Named Entity Recognition (NER) can help in detecting and categorizing these skills.
- Experience and Education Parsing: Extract structured information about candidates' work experience and educational background. This includes identifying job titles, company names, roles, durations, and academic degrees.

iv. Resume Classification

Develop and train a classification model to categorize resumes into predefined job profiles or categories.

- Model Selection: Implement various machine learning algorithms to classify resumes. Options include Naïve Bayes, Support Vector Machines (SVM), Random Forests, and Gradient Boosting. Consider modern approaches such as transformer-based models like BERT for enhanced accuracy.
- Training and Validation: Split the dataset into training and validation sets. Train the model using the training data and assess its performance on the validation set. Evaluate the model using metrics such as accuracy, precision, recall, and F1-score to ensure it performs effectively.

v. Job Recommendation System

Integrate a recommendation engine to suggest suitable job positions based on the classified resumes.

- Similarity Matching: Implement similarity measures, such as cosine similarity, to compare features extracted from resumes with job descriptions. This comparison helps in identifying relevant job matches.
- Ranking Mechanism: Develop a ranking system to order job recommendations based on their relevance scores. The scores are derived from the similarity measures to prioritize the most suitable job opportunities.

vi. Performance Evaluation

Assess the overall effectiveness of the Smart Resume Analyzer and its components.

- Cross-validation: Conduct k-fold cross-validation to ensure that the model generalizes well across different subsets of the data. This process helps in verifying the robustness of the model.
- Confusion Matrix Analysis: Analyze the confusion matrix to understand the distribution of true positives, false positives, true negatives, and false negatives.

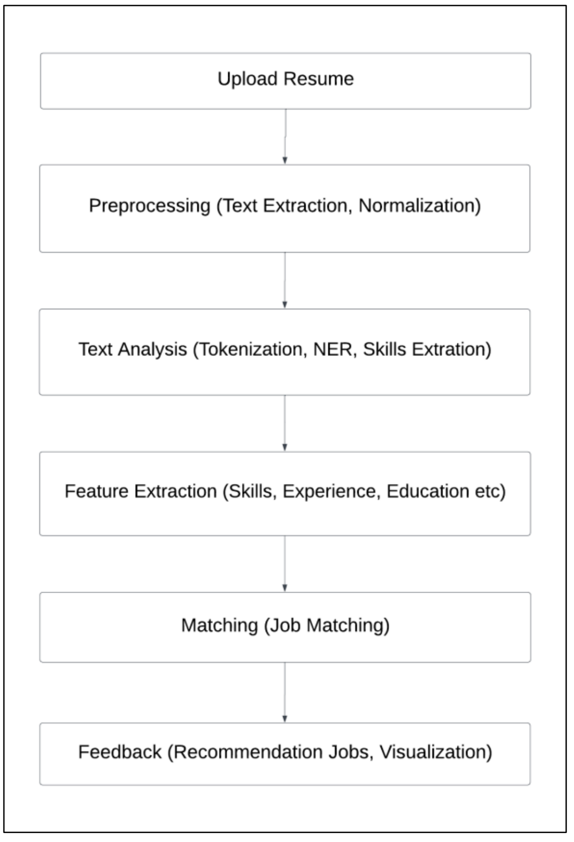


Fig -1: Flow Chart Diagram

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

4. MODEL WITH EXPERIMENTAL RESULT

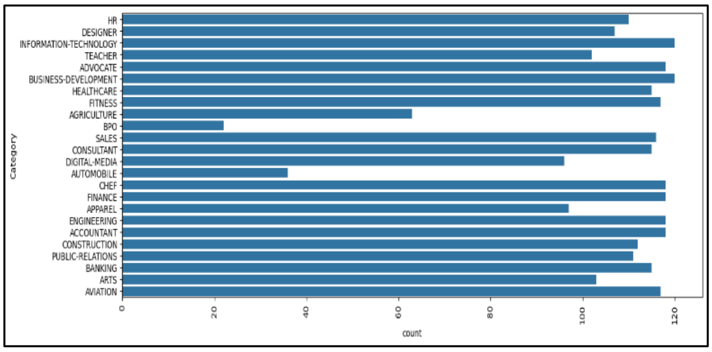
The experimental results of the Smart Resume Analyzer highlight its capability to efficiently process and analyze PDF resumes, subsequently offering relevant job recommendations. The PDF parsing module of the system shows notable proficiency in extracting and standardizing text from resumes. This functionality is crucial as it translates a variety of complex resume formats into a uniform structure that can be evaluated systematically. However, the system does face challenges when dealing with more elaborate resume designs, such as those with intricate layouts, unconventional fonts, and embedded graphics. These elements can complicate text extraction and normalization, indicating a need for further enhancement of the parsing algorithms to handle such variations more effectively.
Nevertheless, the overall performance of the PDF parsing module remains robust. It effectively processes a broad spectrum of resume formats, converting them into a format suitable for the recommendation system. This transformation is essential for the recommendation algorithms to operate efficiently. Consequently, users can generally trust the Smart Resume Analyzer to handle most resumes adeptly, though those with particularly sophisticated designs might encounter some limitations.

In terms of job recommendations, the system's recommendation engine shows significant efficacy in aligning resumes with appropriate job descriptions. This aspect is central to the system's value, as it ensures job seekers receive suggestions that match their qualifications and career goals. The success of the recommendation system is largely due to the Support Vector Machine (SVM) model, which outperforms the Logistic Regression model in terms of accuracy.
The SVM model excels because it effectively manages high-dimensional data and constructs a decision boundary that maximizes the margin between distinct classes. This attribute makes SVM particularly suitable for complex classification tasks, such as matching resumes with job descriptions that may contain nuanced requirements. The SVM model's ability to capture these subtleties results in more accurate and relevant job recommendations for users.

Conversely, the Logistic Regression model, while effective for many classification tasks, falls short in this scenario. Logistic Regression generally relies on a linear decision boundary, which may not sufficiently address the complexities involved in resume and job description matching. Consequently, the superior performance of the SVM model offers a notable advantage for the Smart Resume Analyzer, enhancing its effectiveness in providing tailored job recommendations.

The Smart Resume Analyzer demonstrates significant proficiency in both PDF resume parsing and job recommendation generation. The parsing module manages a wide range of resume formats well, though improvements are needed to better handle more complex designs. Meanwhile, the recommendation system, particularly the SVM model, provides exceptional accuracy in matching resumes with job descriptions. Ongoing refinement of both components will likely further boost the overall effectiveness of the Smart Resume Analyzer, solidifying its role as a valuable tool for job seekers.

Fig -2: Category & Count Plotting for Classification



Accuracy: 0.8402777777777778

	precision	recall	f1 score	support
ACCOUNTANT	0.78	1.00	0.88	21
ADVOCATE	0.96	0.76	0.85	29
AGRICULTURE	0.94	0.70	0.80	23
APPAREL	0.94	0.81	0.87	21
ARTS	0.93	0.64	0.76	22
AUTOMOBILE	1.00	1.00	1.00	19
AVIATION	0.94	0.97	0.95	30
BANKING	0.94	0.74	0.83	23
BPO	1.00	1.00	1.00	15
BUSINESS-DEVELOPMENT	0.46	0.61	0.52	18
CHEF	0.96	0.96	0.96	28
CONSTRUCTION	0.86	1.00	0.93	25
CONSULTANT	1.00	0.58	0.73	31
DESIGNER	0.96	0.79	0.86	28
DIGITAL-MEDIA	0.83	0.90	0.86	21
ENGINEERING	0.76	0.96	0.85	23
FINANCE	0.88	0.67	0.76	21
FITNESS	0.83	0.91	0.87	22
HEALTHCARE	0.85	0.76	0.80	29
HR	0.64	1.00	0.78	21
INFORMATION-TECHNOLOGY	0.73	0.83	0.78	23
PUBLIC-RELATIONS	0.78	0.78	0.78	23
SALES	0.92	0.85	0.88	27
TEACHER	0.72	1.00	0.84	33
accuracy			0.84	576
macro avg	0.86	0.84	0.84	576
weighted avg	0.86	0.84	0.84	576

Fig -3: Model Accuracy Evaluation

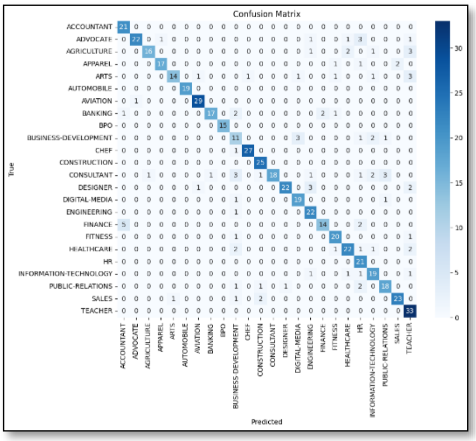


Fig -4: Confusion Matrix

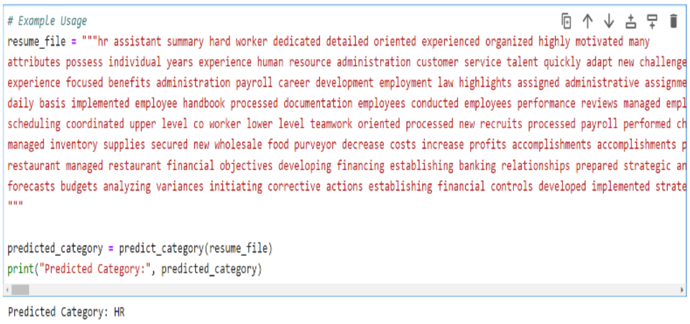


Fig -5: Resume Category Prediction

Fig -6: Job Recommendation

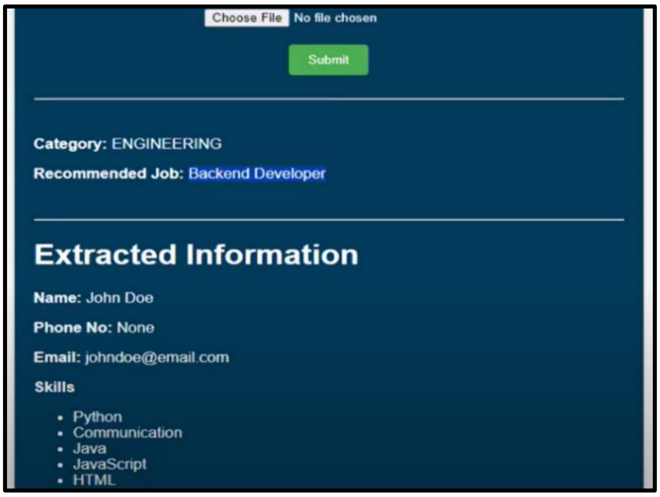


Fig -7: Result

5. CONCLUSION

In this paper, we reviewed the processes of Resume Screening and Shortlisting, focusing on the role of AI-powered resume analyzers. These advanced tools are crafted to support HR professionals by significantly enhancing the efficiency of candidate screening and shortlisting. Through the use of AI, these analyzers streamline the recruitment process by automatically evaluating and ranking resumes, enabling HR teams to quickly identify the most suitable candidates for a given role. This review delves into the methodologies and technologies that underpin AI-driven resume analyzers, highlighting their potential to transform traditional hiring practices.

Smart resume analyzers leverage natural language processing to intelligently extract and align keywords, skills, and qualifications from resumes with the specific criteria outlined in job descriptions. By analyzing language patterns and contextual relevance, these tools ensure a precise match between a candidate's profile and the job requirements, streamlining the hiring process and enhancing the accuracy of candidate selection.

REFERENCES

[1] NLP based Extraction of Relevant Resume using Machine Learning. Author: Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh. Published in International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075 (Online), Volume-9 Issue-7, May 2020.\

[2] A Review Paper on Resume scanning using python. Author: Panil Jain, Nandkishor Kamble, Risha Nadar, Shaikh Luqman. Published in International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 04 — Apr 2022.

[3] Resume Recommendation using Machine Learning. Author: Ramba S Naik, Shrinivas R Dhotre. Published in International Journal of Creative Research Thoughts (IJCRT), Volume 10, Issue 7 July 2022.

[4] Resume Screening Using LSTM. Author: Navale Sakshi, Doke Samiksha, Mule Divya3, Prof. Said S. K. Published in International Journal of Research Publication and Reviews, April 2022.

[5] "The study on the effectiveness of talent acquisition through E-recruitment" by Prof. Sahana G N and Pallav N. Published in the International Journal of Research and Analytical Reviews. May 2023.

[6] "A review of machine learning applications in human resource management" by Swati Garg, Shuchi Sinha, Arpan Kumar Kar, and Mauricio Mani. Published in the International Journal of Productivity and Performance Management. February 2021.