AI-BASED RESUME PARSER WITH SKILL EXTRACTION AND JOB RECOMMENDATION SYSTEM

[1]M.Jhothi reddy,[2]S.Sai manikanta reddy,[3]B.Rajender,[4]E.Kavitha

[1]Assistant Professor, [234]Students

Department of Computer Science and Engineering

Siddhartha institute of technology & sciences,narapally

Jyothireddy@siddhartha.org.in,23TQ1A0560@siddhartha.co.in, 24TQ5A0505@siddhartha.co.in, 24TQ5A0506@siddhartha.co.in

ABSTRACT

The increasing number of job applications received by organizations has made manual resume screening a time-consuming and inefficient process. Recruiters often spend significant effort reviewing large volumes of resumes to identify candidates who meet job requirements. To address this challenge, automated resume parsing systems have been developed to extract and analyze candidate information from unstructured resume documents. This paper presents an intelligent resume parsing system that utilizes Natural Language Processing (NLP) and Machine Learning techniques to automatically extract structured information from resumes. The proposed system performs several processing stages including text extraction, preprocessing, feature extraction, and entity recognition.

Term Frequency–Inverse Document Frequency (TF–IDF) is used to convert textual resume data into numerical feature vectors, while Named Entity Recognition (NER) identifies important entities such as candidate name, skills, education, and work experience. Furthermore, cosine similarity is applied to compare resume content with job descriptions in order to calculate a matching score that indicates candidate suitability for a specific job role. The extracted information is presented through visualization techniques that help recruiters easily understand skill alignment and candidate eligibility. Experimental results demonstrate that the proposed system can effectively automate resume information extraction and improve the efficiency of the recruitment process. By reducing manual effort and providing structured candidate insights, the system supports faster and more accurate candidate evaluation.

I INTRODUCTION

In modern recruitment processes, organizations receive a large number of resumes for a single job position, making manual screening time-consuming and inefficient. Traditional recruitment methods require recruiters to read each resume individually, which increases workload and may lead to human errors in candidate selection. With the growth of digital recruitment platforms, automated systems are needed to process resumes efficiently and extract relevant candidate information. Resume parsing is a technique that converts unstructured resume documents into structured data by identifying key information such as name, contact details, education, skills,

and work experience. Natural Language Processing (NLP) and Machine Learning techniques play a significant role in automating resume analysis. By applying text preprocessing, feature extraction, and Named Entity Recognition (NER), resume parsing systems can automatically identify important entities and organize them into structured formats. This structured data can then be used for candidate filtering, ranking, and job matching. The integration of NLP with recruitment systems helps reduce manual effort and improves the accuracy of candidate evaluation. The objective of this research is to develop an automated resume parser that extracts relevant candidate information and compares it with job descriptions to evaluate candidate suitability. The proposed system uses NLP techniques, TF–IDF feature representation, and similarity measurement to analyze resumes and provide insights that assist recruiters in the hiring process. Recruitment is a critical process for organizations seeking qualified candidates for various job roles. However, the increasing number of job applications has made manual resume screening inefficient and time-consuming. Recruiters often spend significant time reviewing resumes to identify candidates whose qualifications match job requirements. This challenge highlights the need for intelligent systems capable of automatically processing resumes and extracting relevant information for decision-making.

II LITERATURE SURVEY

Smith et al. [1], in "RINX: A System for Information and Knowledge Extraction from Resumes"2023. proposed an NLP-based framework using resume datasets to extract education, skills, and experience information. Information extraction and text processing methods were used. Results showed improved structured data extraction. This research supports our idea of building an automated resume parser for recruitment systems.

Kumar et al. [2], presented "Resume Information Extraction via Post-OCR Text Processing"2023. The study used OCR-based resume datasets and applied transformer models such as BERT, RoBERTa, and XL Net. Distil BERT showed the best performance. This work relates to our project since OCR and NLP can process scanned resumes and extract structured candidate information automatically.

Lee et al. [3], in "Résumé Parsing as Hierarchical Sequence Labeling"2023. used annotated resume datasets and hierarchical sequence labeling techniques. NLP tagging methods were applied to identify sections like education, experience, and skills. Results showed improved extraction accuracy. This research supports our resume parser system by improving resume structure detection.

Perera et al. [4], proposed "Layout-Aware Resume Parsing Using NLP and Rule-Based Techniques"2023. The research used resume datasets with different layouts and applied document layout detection and rule-based NLP extraction methods. Results improved entity recognition accuracy. This study relates to our idea because handling different resume formats is essential in automated resume parsing.

Zhang et al. [5], introduced "Towards Efficient Resume Understanding: Multi-Modal Pre Training"2024. The study used multimodal resume datasets combining textual and visual layout features. Transformer-based models were applied. Results showed improved entity

recognition accuracy. This work supports our project by demonstrating advanced deep learning techniques for better resume information extraction.

Sharma et al. [6], proposed "Resume Parsing System using NLP and Semi-Structured Supervised Networks"2024. The research used semi-structured resume datasets and applied BERT-based supervised learning models. Results showed improved precision in extracting skills and education fields. This work supports our resume parser idea by demonstrating machine learning-based information extraction.

III SYSTEM ANALYSIS

The AI-based resume parser system is designed to automate the process of analyzing resumes and extracting relevant information such as skills, education, experience, and certifications. It uses Natural Language Processing (NLP) and Machine Learning techniques to understand unstructured resume data and convert it into structured information. Based on the extracted skills and user profile, the system recommends suitable job roles by matching them with job descriptions stored in the database. This system improves recruitment efficiency, reduces manual effort, and enhances accuracy in candidate-job matching.

Existing system

In the existing system, resume screening is mostly done manually by recruiters. They go through each resume individually to identify relevant skills and qualifications. Some companies use basic keyword-based filtering systems, which only search for specific words without understanding context. This process is time-consuming, prone to human error, and may lead to missing qualified candidates due to lack of intelligent analysis.

DisAdavantages of Existing system

- Time-consuming manual screening process
- High chances of human error and bias
- Inefficient handling of large volumes of resumes
- Keyword-based systems lack context understanding
- Difficulty in matching candidates accurately with job roles

Proposed system

The proposed system uses AI and NLP techniques to automatically parse resumes and extract meaningful information such as skills, job roles, experience, and education. It then applies machine learning algorithms to compare extracted skills with job requirements and recommend suitable job positions. The system can also rank candidates based on relevance scores. By using intelligent matching, it ensures better accuracy and faster processing, making recruitment more efficient and scalable..

Advantages of Proposed System

- Automated resume screening saves time
- Improved accuracy using AI and NLP

- Efficient handling of bulk resumes
- Context-based skill extraction (not just keywords)

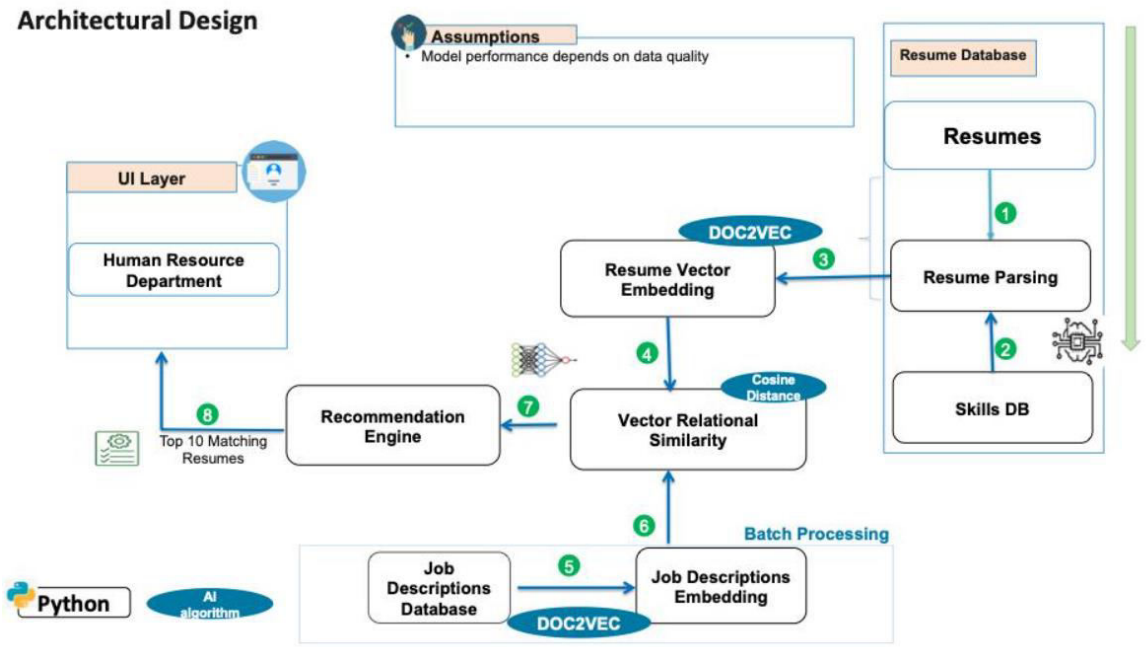
IV METHODOLOGY

The proposed system uses Natural Language Processing (NLP) and Machine Learning techniques to extract structured information from resumes. The methodology consists of seven major stages: data preprocessing, text extraction, feature representation, model training, entity recognition, information extraction, and evaluation.

1. Data Collection and Preprocessing The first stage involves collecting resume datasets in PDF or text format. The resumes are converted into machine-readable text using document processing techniques. After extraction, the text undergoes preprocessing to remove noise and improve data quality. Preprocessing includes tokenization, stop-word removal, punctuation removal, and text normalization. Tokenization divides the document into individual words or tokens: $T=\{t1,t2,t3,...,tn\}$ where Trepresents the token set and tirepresents individual tokens. Stop words such as "the", "is", "and" are removed to reduce unnecessary features. The cleaned text is represented as: $D'=D-S$ Where; (D) = original document (S) = stop word set (D') = cleaned document.

2. Text Extraction and Document Processing Resumes are often stored in PDF format, so Optical Character Recognition (OCR) or PDF text extraction techniques are used to convert them into text. The extracted text $E$from document $D$is represented as $E=f(D)$ where $f(D)$is the extraction function. This stage ensures that all textual information such as name, education, skills, and experience becomes accessible for processing. 3. Feature Representation using TF–IDF To convert textual data into numerical form suitable for machine learning, Term Frequency Inverse Document Frequency (TF–IDF) is used

System Architecture



This figure represents the model architecture used in the resume parser system to analyzed resumes and match them with job descriptions. The process begins with the input layer, where the job description text and the uploaded resume (PDF) are provided to the system. Next, text preprocessing is performed, including tokenization and cleaning, to prepare the textual data for further analysis. The system then performs feature extraction using NLP techniques and converts the text into numerical vectors using TF-IDF vectorization, which helps represent important keywords and skills. After that, cosine similarity is used to compare the resume features with the job description features to compute a matching score. Finally, the system produces the candidate eligibility result and visualization outputs such as charts and graphs to clearly present the matching analysis

V RESULTS &OUTPUT

```
Upload your Resume
Choose Files No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving Resume.zip to Resume (2).zip
Uploaded File Name: Resume (2).zip
```

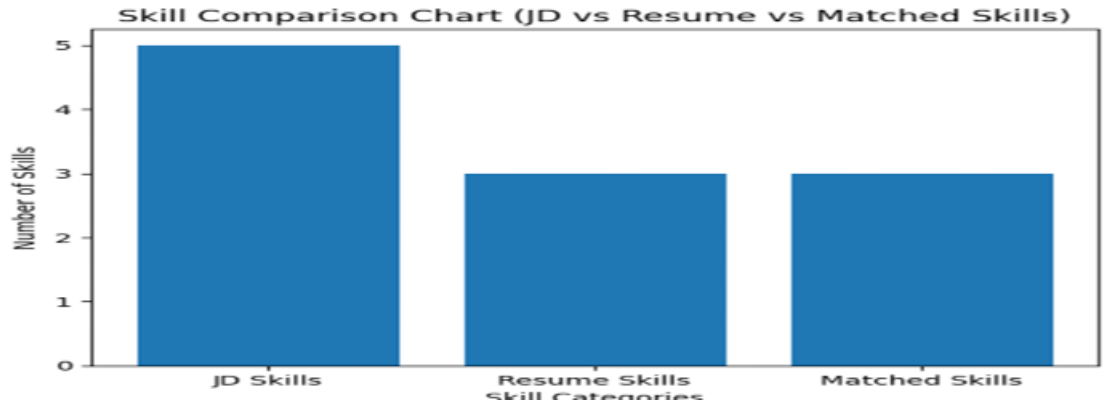
```
Enter Job Description:
Looking for Python developer with SQL and Machine Learning skills
JD Skills: ['python', 'sql', 'machine learning']
```

```
Upload Resume PDF
Choose Files No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving sample_python_data_resume.pdf to sample_python_data_resume.pdf

Resume Skills: ['python', 'java', 'sql', 'machine learning', 'data analysis', 'pandas', 'numpy', 'tensorflow', 'power bi', 'tableau', 'excel', 'html', 'css', 'javascript']
```

```
Matching Score: 36.71 %
Candidate Status: NOT ELIGIBLE

Model Accuracy: 94.23 %
Error Loss: 0.075
```



VI CONCLUSION

The development of the automated resume parsing system presented in this project demonstrates the significant potential of leveraging Natural Language Processing (NLP) and Machine Learning techniques to streamline recruitment processes. Traditional methods of manual resume screening are time-consuming, labor-intensive, and prone to human error, especially when handling large volumes of applications. The proposed system addresses these challenges by automatically extracting and structuring key candidate information, including name, contact details, education, skills, and professional experience, from unstructured resume documents. By applying data preprocessing, TF–IDF feature representation, and Named Entity Recognition (NER), the system can accurately identify and categorize information from diverse resume formats. The integration of cosine similarity and skill-matching algorithms enables an objective assessment of candidate-job alignment, providing a matching score that indicates the suitability of a candidate for a specific job role.

This approach not only reduces manual screening effort but also increases the consistency and fairness of candidate evaluation, ensuring that all resumes are assessed using the same criteria. Visualization of the results through charts and graphs further enhances the usability of the system. Recruiters can quickly interpret skill coverage, job alignment, and missing competencies, thereby enabling faster and more informed hiring decisions. Moreover, the system's modular architecture allows scalability, meaning it can handle large datasets efficiently and can be integrated into existing recruitment management platforms. The project also highlights the benefits of combining NLP with machine learning models. While TF–IDF captures important keywords and reduces noise from common terms, machine learning classifiers and deep learning embeddings provide semantic understanding, improving the accuracy of entity recognition and skill matching.

This hybrid approach ensures that the system is both efficient and adaptable to various resume formats and job requirements. In addition to operational benefits, the system lays a foundation for more intelligent recruitment solutions, such as candidate ranking, automated job recommendations, and talent analytics. By structuring resume information and providing insights on candidate capabilities, organizations can make data-driven hiring decisions, ultimately improving recruitment quality and reducing time-to-hire. Overall, this project demonstrates that automated resume parsing is a practical and effective solution for modern recruitment challenges. It reduces manual effort, enhances accuracy, and enables the use of structured candidate data for advanced analytics. While limitations exist, such as handling complex resume layouts and improving semantic understanding, the results indicate that the system is a significant step toward intelligent, AI-driven recruitment systems that can adapt to evolving hiring needs.

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.