RESUME PARSER USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Nagiredla Laxmi [1], Jampana Vamsi [2], Padhi Srilatha [3], Koda Diwakar Rao [4], Mr.N.V.S.S.Prabhakar[5]

[1,2,3,4]*CSE Department, Raghu Institute of Technology, Visakhapatnam, India*
[5] *Asst.Prof.CSE Department, Raghu Engineering College, Visakhapatnam, India*

Abstract— The primary objective of this project is to develop an intelligent system capable of automatically extracting and analyzing relevant information from resumes using advanced Natural Language Processing (NLP) techniques. The project addresses the challenges associated with manual resume screening by employing machine learning algorithms and linguistic analysis to enhance the accuracy and efficiency of information extraction. The resume parser aims to comprehend diverse resume formats, language variations, and expressions commonly used by job applicants, ensuring robust performance across a wide range of documents.

Key features of the project include the utilization of NLP algorithms to identify and categorize essential details such as personal information, education history, work experience, and skills.

1.INTRODUCTION

In today's competitive job market, the task of talent acquisition is often burdened by the sheer volume of resumes received for a single job posting. Manual screening and analysis of these resumes are not only time-consuming but also prone to errors and biases the extraction and interpretation of critical information from resumes. The primary goal of this project is to develop a sophisticated system capable of parsing and understanding the content of resumes through the lens of Natural Language Processing (NLP). NLP is a branch of artificial intelligence that focuses on the interaction between computers and human language, enabling machines to comprehend, interpret, and generate human-like text. By integrating NLP techniques, the resume parser aims to bridge the gap between the overwhelming influx of resumes and the need for an efficient, accurate, and unbiased candidate evaluation process. The project's significance lies in its potential to revolutionize the initial stages of recruitment. Traditional methods often rely on manual scrutiny, which is not only time-intensive but also susceptible to oversight. With the implementation of NLP, the system can autonomously analyze resumes, extracting pertinent details such as personal information, educational background, work experience, and skills. This automated approach not only expedites the screening process but also reduces the likelihood of overlooking qualified candidates. Furthermore, the project aims to address the challenges posed by the variability in resume formats, language nuances, and expressions used by job applicants

Through the integration of machine learning algorithms, the system will continuously adapt and improve its ability to understand diverse language patterns, ensuring versatility in handling resumes of varying styles and origins. As the project progresses, it is expected to culminate in a powerful tool that not only enhances the efficiency of the recruitment process but also contributes to a more objective and data-driven candidate evaluation. The "Resume Parser Using Natural Language Processing Techniques" project holds immense promise in revolutionizing the way organizations sift through resumes, ultimately leading to more informed hiring decisions and a streamlined recruitment workflow.

2.LITERATURE REVIEW:

Ghosh et al. (2018) explored the application of NLP techniques in resume parsing. The study emphasized the importance of semantic analysis in extracting meaningful information from resumes, highlighting the need for

advanced algorithms to handle diverse language structures.[1]

Choi and Song (2017) investigated the challenges of traditional resume parsing methods and proposed a framework that leverages NLP for improved accuracy. Their work emphasized the need for robust syntactic and semantic analysis to handle variations in resume content.[2]

Duan et al. (2019) discussed the integration of machine learning models in the recruitment process. The study emphasized the potential of machine learning algorithms to enhance the efficiency of candidate screening by learning from historical data and adapting to evolving patterns.[3]

Kuhlmann and Bergmann (2020) provided insights into the role of machine learning in mitigating biases in recruitment. The authors discussed the importance of unbiased algorithms in resume parsing to ensure fair evaluation and selection of candidates.[4]

Kaya and Salah (2017) investigated the impact of automated recruitment systems on HR processes. Their work highlighted the potential benefits of automation in reducing workload and improving the overall effectiveness of talent acquisition.[5]

Van den Heuvel and Bondarouk (2017) discussed the ethical considerations of automated recruitment. The study emphasized the need for transparency and fairness in algorithmic decision-making to maintain trust in automated systems.[6]

Borrajo et al. (2016) identified challenges in resume parsing, including the handling of unstructured data, multilingual content, and diverse formatting. The study emphasized the importance of robust NLP techniques to address these challenges.[7]

Saranya et al. (2020) discussed the limitations of rule-based resume parsing systems and advocated for the integration of machine learning to improve adaptability to varying resume styles.[8]

3.EXISTING SYSTEM

Talent Acquisition Software: Commercial talent acquisition platforms, such as Taleo, Greenhouse, and Workday, often incorporate resume parsing capabilities. These systems use rule-based approaches and machine learning algorithms to extract and categorize information from resumes, streamlining the initial stages of the recruitment process.

Open-source Libraries: Open-source libraries like SpaCy, NLTK (Natural Language Toolkit), and Apache OpenNLP provide tools for various NLP tasks, including tokenization, named entity recognition, and part-of-speech tagging. Developers can use these libraries to build custom resume parsers tailored to their specific requirements.

Resume Parsing APIs: Companies like Sovren, TextKernel, and RChilli offer resume parsing APIs that developers can integrate into their applications. These APIs typically support multilingual parsing, handle diverse resume formats, and use machine learning techniques for improved accuracy in information extraction.

AI-powered Recruitment Platforms: Advanced recruitment platforms, such as Smart Recruiters and Lever, incorporate AI-driven features for resume parsing. These systems go beyond basic keyword matching and employ NLP algorithms to understand context, identify skills, and assess the relevance of experiences mentioned in resumes.

Job Portals and Applicant Tracking Systems (ATS):
Popular job portals like Indeed, LinkedIn, and Glassdoor often include built-in resume parsing functionalities. Additionally, many Applicant Tracking Systems (ATS) used by organizations employ resume parsing to automate the screening and sorting of candidate profiles.

Rule-based Systems: Some legacy systems rely on rule-based approaches to extract information from resumes. These systems define specific rules to identify and categorize information, but they may lack the adaptability and flexibility provided by more advanced NLP and machine learning approaches.

4.PROPOSED SYSTEM

Machine Learning for Adaptive Parsing: The system will incorporate machine learning

models to adapt and improve its parsing accuracy over time. By continuously learning from new resumes and user feedback, the system aims to enhance its ability to handle evolving language patterns, industry-specific terminology, and changing job market trends.

Multilingual Support: The proposed system will be designed to support resumes in multiple languages, addressing the global nature of the job market. It will utilize language-agnostic NLP techniques to ensure accurate parsing and extraction of information from resumes written in different languages.

Context-aware Information Extraction: To overcome the challenges of ambiguous terms and context-dependent information, the system will focus on context-aware information extraction. By understanding the context in which skills, experiences, and qualifications are mentioned, the parser aims to provide more accurate and relevant insights for candidate evaluation.

Formatting and Structure Recognition: The system will be equipped with advanced formatting and structure recognition capabilities to handle diverse resume layouts. This includes the ability to accurately identify and extract information from resumes with varied formatting styles, ensuring robust parsing across different document structures.

User-friendly Interface: The proposed system will feature a user-friendly interface that allows recruiters and HR professionals to easily interact with the parsed data. The interface will provide a comprehensive overview of extracted information, enabling quick and efficient candidate evaluation.

5.METHODOLOGIES

1. Data Collection Module: Gather a diverse dataset of resumes to train and validate the NLP and machine learning models.

2. Preprocessing Module: Clean and prepare the raw resume data for effective processing by the NLP and machine learning models.

3. NLP Module: Apply Natural Language Processing techniques to understand and interpret the content of resumes.

4. Machine Learning Module: Train machine learning models to enhance the adaptive parsing capabilities of the system.

5.Multilingual Support Module: Enable the system to parse resumes written in multiple languages.

6. Context-aware Information Extraction Module: Enhance the system's understanding of the context in which information is presented in resumes.

7. Bias Mitigation Module: Implement measures to identify and mitigate biases in the parsing process.

8. User Interface Module: Create an intuitive and user-friendly interface for recruiters and HR professionals.

9. Customization and Integration Module: Allow customization based on organizational requirements and seamless integration with existing Applicant Tracking Systems (ATS).

10. Testing and Evaluation Module: Validate the performance of the system and ensure its accuracy, efficiency, and adaptability.

11. Documentation and Deployment Module: Document the project details and deploy the resume parser for real-world use.
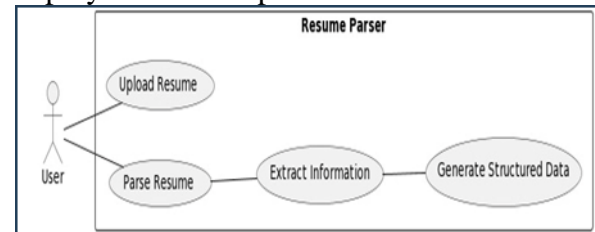
7.RESULTS

Basic information including education details, job experience, skills, category (whether the resume is technology related or other category), current position and company are identified and retrieved by the parser. Overall, the resume parser is a reliable and efficient tool for automatically extracting and providing essential data from resumes.
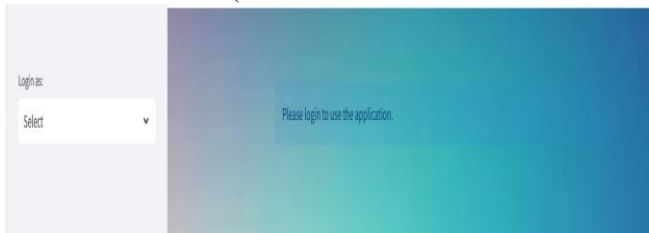
1.Algorithms used:

i)K-Nearest Neighbour
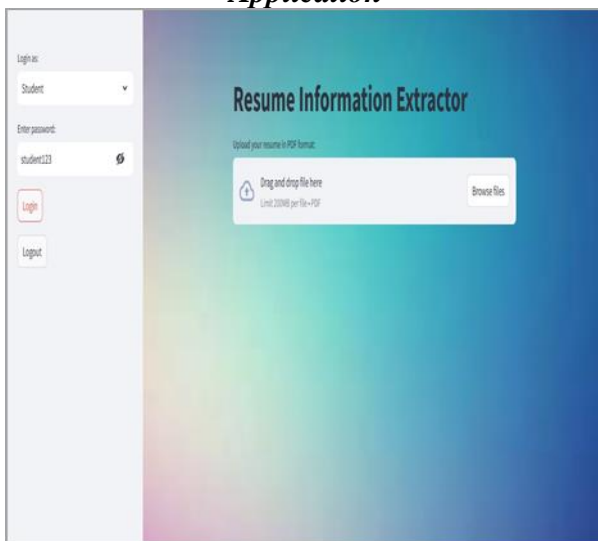
ii)Multinomial Naive Bayes

2.Accuracy and precision:

K-Nearest Neighbour : 1.0(Accuracy)

1.0(Precision)

Multinomial Naive Bayes 0.9827(Accuracy)

0.9838(Precision

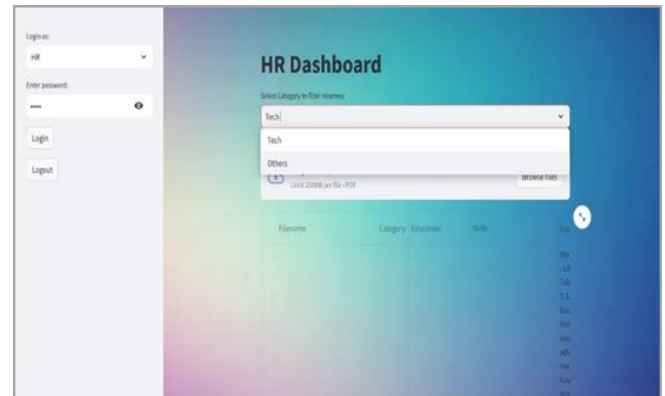**Figure 7.1: User Should Login To The Application**

Figure 7.2: Student Will Login To The Application And Uploads The Resume

Figure 7.3: The Resume Parser Will Extract The Details From The Uploaded Resume

3. HR will login to the application, the resumes which were uploaded will be displayed in the HR portal in the form of table. And HR can filter resumes based on their categories like whether it is related to technology or other related resume.

Figure 7.4: HR will login to the application,the resumes which were uploaded will be displayed in the HR portal in the form of table. And HR can filter resumes based on their categories like whether it is related to technology or other related resume

8.CONCLUSION

The "Resume Parser Using Natural Language Processing Techniques" project represents a significant advancement in the realm of human resources and recruitment by leveraging cutting-edge technologies. Through the integration of natural language processing (NLP) and machine learning (ML) techniques, the system demonstrates an ability to extract valuable information from resumes, providing HR professionals with a powerful tool for candidate evaluation.

9.REFERENCES

Han, Q., et al. (2018). A Resume Parsing Method Based on Keyword Extraction. 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C).

Khan, F. N., & Chaudhary, K. P. (2020). An efficient resume parser using machine learning techniques. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

Nagar, R., et al. (2020). Resume Parsing Using NLP and Machine Learning. 2020 IEEE International Conference on Computing, Communication and Automation (ICCCA).

Bansal, V., et al. (2019). Automatic Resume Parser Using Natural Language Processing. 2019 4th International Conference on

Communication and Electronics Systems (ICCES).

Kumar, P., et al. (2021). Intelligent Resume Parsing Using Natural Language Processing. 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA).

Patil, A., et al. (2019). Resume Parsing Using Machine Learning Techniques. 2019 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

Nanda, S., & Jain, A. (2020). Resume Parsing Using NLP and Machine Learning Techniques. 2020 International Conference on Smart Electronics and Communication (ICOSEC).

Gupta, A., et al. (2018). Resume Parser Using Machine Learning and NLP Techniques. 2018 IEEE 9th Annual Information Technology, Electronics IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

Singh, S., & Joshi, M. (2021). Resume Parsing System Using NLP Techniques. 2021 International Conference on Intelligent Computing and Control Systems (ICICCS).

Kumar, V., et al. (2019). Resume Parsing Using Natural Language Processing Techniques. 2019 International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT).