RESUME ANALYZER USING NATURAL LANGUAGE PROCESSER WITH PYTHON

Karne Nikhil
Scholar. Department of MCA
Vaageswari College of Engineering, Karimnagar

Dr.Ravikumar Thallapalli
Professor
Vaageswari College of Engineering, Karimnagar

Dr. P. Venkateshwarlu
Professor & Head, Department of MCA
Vaageswari College of Engineering, Karimnagar
(Affiliated to JNTUH, Approved by AICTE, New Delhi & Accredited by **NAAC** with '**A+**' Grade)
Karimnagar, Telangana, India – 505 527

ABSTRACT

In the modern recruitment process, organizations receive a large number of resumes for every job posting, making manual screening time-consuming and prone to errors. This project presents a **Resume Analyzer** system using **Natural Language Processing (NLP)** techniques implemented in **Python**. The system automatically parses and extracts essential information from resumes, such as personal details, educational qualifications, skills, work experience, and certifications. Additionally, it evaluates the relevance of a candidate's profile against job requirements using keyword matching and semantic similarity measures. By leveraging NLP libraries and algorithms, the system provides a **fast, accurate, and unbiased assessment**, thereby assisting HR professionals in efficient candidate shortlisting. The proposed solution also supports ranking of candidates and visualization of extracted skills, making the recruitment process more streamlined and data-driven.

Keywords:

Resume Analysis, Natural Language Processing (NLP), Python, Information Extraction, Candidate Screening, Keyword Matching, Semantic Similarity, Recruitment Automation, HR Analytics

1.INTRODUCTION

Recruitment is a critical process in any organization, and the quality of hiring directly impacts overall productivity and growth. Traditionally, human resources (HR) professionals manually review resumes to shortlist candidates, which is **time-consuming, labor-intensive, and prone to bias**. With the increasing number of applicants for every job opening, organizations face significant challenges in efficiently identifying suitable candidates.

To address these challenges, **Resume Analysis using Natural Language Processing (NLP)** has emerged as an effective solution. NLP is a branch of artificial intelligence (AI) that enables computers to **understand, interpret, and manipulate human language**. By applying NLP techniques, resumes can be **automatically parsed, analyzed, and compared** against job requirements, allowing for a faster and more objective evaluation of candidates.

This project focuses on building a **Python-based Resume Analyzer** that can extract key

information such as personal details, education, work experience, skills, and certifications. Moreover, the system evaluates candidates' suitability using **keyword matching and semantic similarity algorithms**, which consider not only the presence of specific terms but also their context within the resume. The system aims to **assist HR professionals in making data-driven decisions**, improve the efficiency of the recruitment process, and reduce human error and bias in candidate selection.

2.LITERATURE REVIEW

The process of resume screening and candidate evaluation has been the subject of significant research in recent years. Traditional methods relied heavily on manual evaluation, which is **time-consuming, inconsistent, and prone to human bias**. With the advancement of **Artificial Intelligence (AI) and Natural Language Processing (NLP)**, automated systems have been developed to enhance efficiency, accuracy, and objectivity in candidate selection.

Automated Resume Parsing: Many studies focus on extracting structured information from unstructured resume documents. Techniques such as **regular expressions, rule-based parsing, and machine learning classifiers** have been used to identify fields like name, contact details, education, and work experience. For example, a study by Kaur et al. (2020) demonstrated that NLP-based parsing can significantly reduce the manual effort required in the screening process.

Keyword Matching and Relevance Scoring: Several systems employ **keyword matching algorithms** to compare candidate skills with job requirements. However, simple keyword matching often fails to capture semantic meaning or contextual relevance. To overcome this, researchers have incorporated **semantic similarity measures** using word embeddings, such as Word2Vec, GloVe, or BERT, to understand the context and relationships between terms in resumes and job descriptions.

Machine Learning for Candidate Ranking: Some studies explore **supervised learning approaches** to rank candidates based on features extracted from resumes. Features include educational qualifications, years of experience, skill sets, and certifications. Machine learning models, including Random Forest, Support Vector Machines (SVM), and neural networks, have been applied to predict candidate suitability and provide a scoring mechanism.

3. EXISTING SYSTEM

In traditional recruitment processes, HR professionals manually screen resumes to shortlist candidates. This process involves **reading through each resume individually**, verifying qualifications, experience, skills, and other relevant details. While effective for small-scale recruitment, this approach has several limitations:

1. **Time-Consuming:** Manually reviewing hundreds or thousands of resumes can take days or even weeks, delaying the hiring process.
2. **Prone to Human Error:** Important qualifications or skills may be overlooked, leading to unsuitable candidates being shortlisted or deserving candidates being rejected.
3. **Inconsistent Evaluation:** Different HR professionals may have varying standards, resulting in inconsistent candidate assessments.
4. **Difficulty Handling Large Volumes:** With the increasing number of job applications, manual screening becomes inefficient and resource-intensive.
5. **Limited Data Analysis:** Traditional systems cannot provide insights such as skill trends, candidate demographics, or ranking based on suitability.

Some modern HR tools use **basic keyword matching or filtering software**, which partially automates resume screening. However, these systems often fail to understand the **context of skills or experiences**. For example, a resume mentioning "Python for data visualization" may not be matched with a job requiring "Python for machine learning" if only exact keyword matching is used.

Overall, the existing systems **lack advanced Natural Language Processing (NLP) capabilities** and rely heavily on human judgment or simplistic filtering techniques. This limitation highlights the need for a **Python-based Resume Analyzer** that can intelligently parse, extract, and evaluate resumes in a fast, accurate, and unbiased manner.

4.PROPOSED SYSTEM

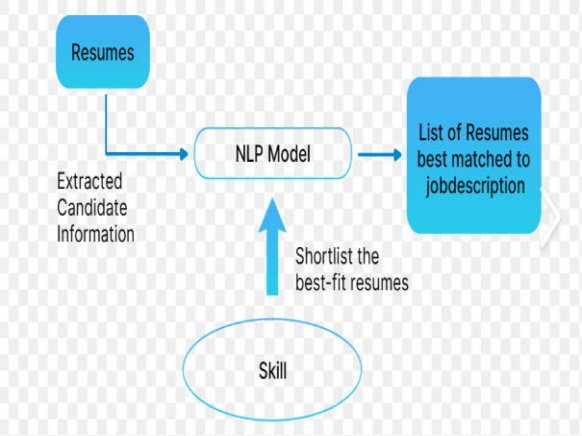
The proposed system is a **Python-based Resume Analyzer using Natural Language Processing (NLP)** designed to automate and improve the recruitment process. It can parse resumes in multiple formats such as PDF, DOCX, and TXT, extracting key information including personal details, education, work experience, skills, certifications, and projects. Using **keyword matching and semantic similarity algorithms**, the system evaluates candidates' suitability by comparing their profiles with job descriptions, considering not only exact matches but also the context of skills and experiences. Candidates are then ranked based on relevance scores, allowing HR professionals to **efficiently shortlist the most qualified applicants**. Additionally, the system can generate visualizations such as skill distributions and experience levels, supporting **data-driven decision-making**. By automating the screening process, it **reduces human bias, saves time, and improves accuracy**, providing a scalable and intelligent solution for modern recruitment challenges. Python libraries like **NLTK, spaCy, PyPDF2, docx, and scikit-learn** are utilized to implement these functionalities effectively.

5.METHODOLOGY

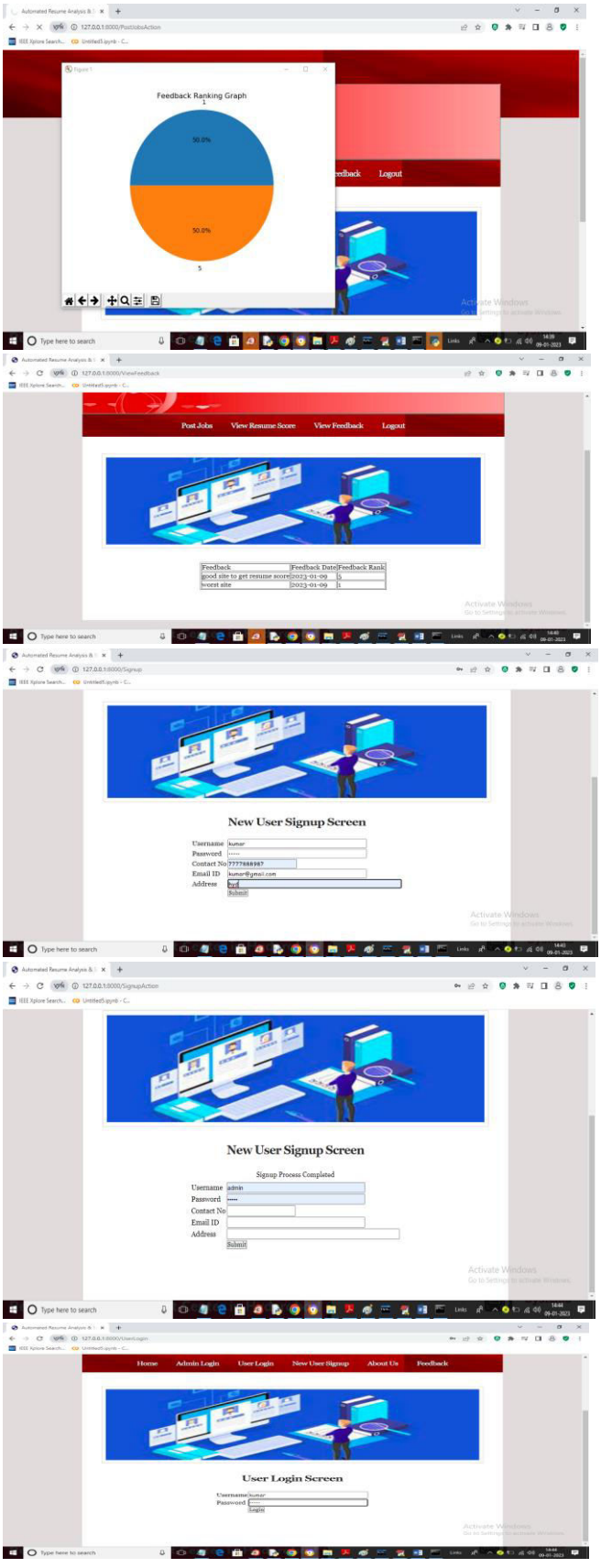
The methodology of the proposed Resume Analyzer system involves several key steps to automatically process and evaluate resumes. First, resumes are collected in different formats such as PDF, DOCX, and TXT. Using Python libraries like **PyPDF2** and **python-docx**, the text content is extracted from these files. The extracted text is then preprocessed using **Natural Language Processing (NLP) techniques** such as tokenization, stopword removal, and lemmatization to clean and standardize the data. Key information including personal details, education, work experience, skills, and certifications is identified and extracted using **Named Entity Recognition (NER)** and pattern-based methods. To match candidates with job requirements, the system employs **keyword matching** along with **semantic similarity models** (like Word2Vec, GloVe, or BERT embeddings) to evaluate the context and relevance of skills. Each candidate is assigned a **suitability score** based on the extracted features and relevance to the job description, enabling efficient ranking. Finally, the system provides **data visualization** for skills distribution, experience levels, and other insights, assisting HR professionals in making informed, unbiased, and quick decisions. This structured methodology ensures that large volumes of resumes can be processed accurately, consistently, and efficiently.

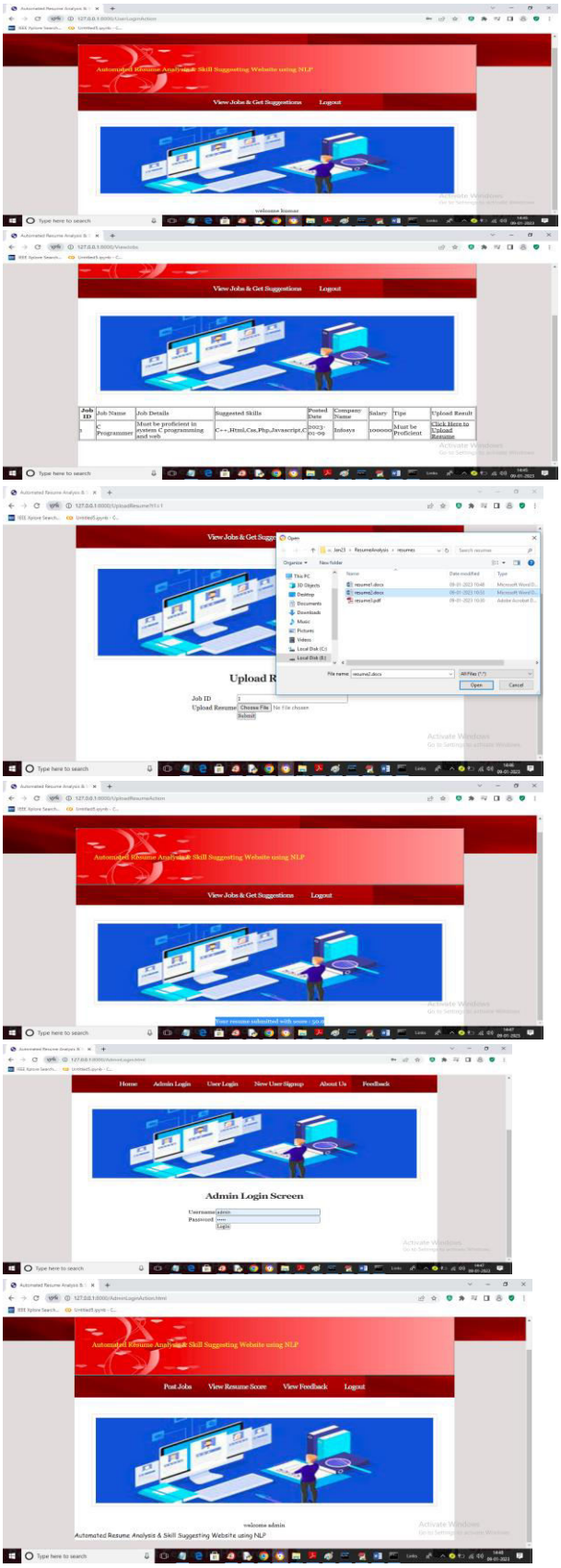
6.System Model

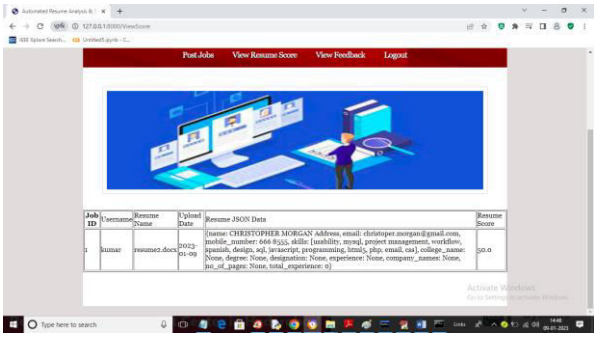
SYSTEM ARCHITECTURE



7..Results and Discussions







8. CONCLUSION

provides an efficient and intelligent solution to the challenges faced in modern recruitment processes. By automating the extraction and analysis of resumes, the system significantly reduces the time, effort, and human bias involved in manual screening. Leveraging NLP techniques, semantic similarity models, and Python-based tools, the system is capable of understanding the context of candidates' skills, experience, and qualifications, allowing for accurate matching with job requirements. The candidate ranking and data visualization features further enable HR professionals to make **data-driven and objective hiring decisions**. Overall, the proposed system demonstrates that integrating NLP into recruitment not only improves efficiency and accuracy but also enhances the fairness and scalability of the hiring process, making it a valuable tool for organizations dealing with large volumes of applications.

9.REFERENCES

- Kaur, P., & Kaur, R. (2020). *Automated Resume Screening Using Natural Language Processing*. International Journal of Computer Applications, 175(22), 1–6.
- Jain, S., & Gupta, R. (2019). *Resume Parser: An Intelligent System for Recruitment*. International Journal of Advanced Research in Computer Science, 10(5), 34–40.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv preprint arXiv:1301.3781.
- Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global Vectors for Word Representation*. EMNLP.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL.
- spaCy Documentation. Available at: https://spacy.io/
- NLTK Documentation. Available at: https://www.nltk.org/
- PyPDF2 Documentation. Available at: https://pythonhosted.org/PyPDF2/
- python-docx Documentation. Available at: https://python-docx.readthedocs.io/en/latest/