International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

MACHINE LEARNING AND NLP BASED RESUME PARSING FRAMEWORK FOR E-RECRUITMENT

Snehal Nikam, Prathmesh Patil, Supriya Pingale, Yashaswini Rajebhosale, Prof. K. B. Dhomse.

[1] nikamsnehal208@gmail.com, Student of BE. Dept. of Information Technology MET-IOE, Nashik, India
[2] ap698431@gmail.com, Student of BE. Dept. of Information Technology MET-IOE, Nashik, India
[3] supriya.pingale24@gmail.com, Student of BE. Dept. of Information Technology MET-IOE, Nashik, India
[4] yashaswinirajebhosle@gmail.com, Student of BE. Dept. of Information Technology MET-IOE, Nashik, India
[5] Internal Guide Dept. of Information Technology MET's Institute of Engineering, Nashik, India

---

Abstract - The growth of the Indian recruitment market has led to the emergence of specialized recruitment agencies that leverage machine learning models to streamline the recruitment process and deliver the right talent to their clients. So, the proposed model has the potential benefits of the ML model to modify the hiring process and make it more efficient and fairer. By using NLP approaches, you can extract valuable information from resumes and provide accurate ratings based on the requirements of the company. With careful testing and refinement, your web portal could be a valuable resource for both job applicants and hiring managers alike. In addition, you could consider incorporating features such as personalized feedback and suggestions for improvement in the resume. This can help applicants to understand where they may be falling short and how they can improve their chances of being hired.

1. INTRODUCTION

The Machine Learning and NLP based Resume Parsing Framework for E-Recruitment project aims to address the inefficiencies and biases inherent in traditional recruitment processes by introducing a sophisticated Machine Learning (ML) and Natural Language Processing (NLP) based system for resume building and parsing in the realm of e-recruitment. This comprehensive system seeks to revolutionize the way organizations handle job applications, providing a streamlined approach that enhances both efficiency and fairness. At its core, our system will automate various aspects of the recruitment process, starting with resume creation. By employing NLP techniques, it will assist job seekers in crafting high-quality resumes, offering suggestions for content improvement, formatting, and skill highlighting. This not only saves time for applicants but also ensures that their resumes are optimized to stand out to potential employers.

Once resumes are submitted, the system will utilize ML algorithms to extract essential information such as contact details, education history, work experience, and skills. This automated parsing process significantly reduces manual data entry efforts and minimizes errors, ensuring that recruiters have accurate and structured data to work with.Moreover, our system goes beyond simple parsing; it incorporates advanced ML models to assess and rank candidates based on their qualifications and alignment with job requirements. By analyzing the content of resumes in-depth, it can identify the most suitable candidates for specific roles, thereby improving the efficiency of the selection process.

One of the critical aspects of our system is its focus on mitigating unconscious bias in resume evaluation. Through NLP-based sentiment analysis and fairness-aware algorithms, we aim to ensure that all candidates are evaluated fairly and objectively. By removing biases from the equation, our system promotes diversity and inclusivity in the hiring process, ultimately leading to better outcomes for both candidates and organizations. In addition to enhancing efficiency and fairness, our system offers several other benefits. These include cost savings through reduced manual labor, improved accuracy in candidate assessments, and a more positive candidate experience. By streamlining the recruitment process and fostering a more inclusive environment, our ML and NLP-based system represents a significant step forward in the field of e-recruitment.

2. LITERATURE SURVEY

2.1 A Review on Text Analytics Process With a CV Parser Model

Feb 2018, Papiya Das, Bhaswati Sahoo, Manjusha Pandey. Big data makes use of predictive analysis, user behavior analysis and some value extraction methods. Data sets are growing rapidly and are gathered by numerous sensors. Big Data deals with challenges like capture, storage, analysis etc. We can computationally analyze data sets to reveal patterns, trends and associations.

2.2 NLP based Extraction of Relevant Resume using Machine Learning

April 2020, Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh. The estimation set of immense records are enormous and jumbled in nature. Thus, various item programs have been added to deal with such enormous databases. Its age accelerates for mentioning resumes with respect to its sorts and codecs. Its coordination advances customers' API key for blend endeavors. The parser works the utilization of two or three rules which train the call and address. Scout bundles use the CV parser system for the determination of resumes. As resumes are in amazing arrangements and it has different sorts of real factors like set up and unstructured estimations, meta experiences, etc. The proposed CV parser approach gives the component extraction method from the moved CV's. The future degree of work is to put into effect and present a smart evaluation in the consistent database to survey with the present models associated with job applications, optimizing the recruitment workflow for improved outcomes.

2.3 An automated resume screening system using natural language processing and similarity.
June 2020, Chirag Daryania, Gurneet Singh Chhabrab, Harsh Patel. An automated resume screening system that simplifies the e-recruitment process by eliminating the various problems faced by the recruiters as they relied on manual shortlisting of applicants for a given job position.

2.4 Resume Screening Using Machine Learning and NLP
Dec 2022, Bhushan Kinge, Shrinivas Mandhar. To detect, identify and classify various resumes using multiple machine learning and Neural Network models like SVM, KNN, Word2Vec, Cosine similarity, etc.

2.5 A Prototype to Enhance Recruitment Process with NLP based Resume Parsing
April 2023, Anshuman Behra, Deepak Gupta. The process of reviewing resumes takes time. The unstructured written language can be understood and parsed by natural language processing and machine learning, which can then extract the required information. The goal is to teach the computer to analyze written paper similarly to a human.

3. MODULE DESCRIPTION

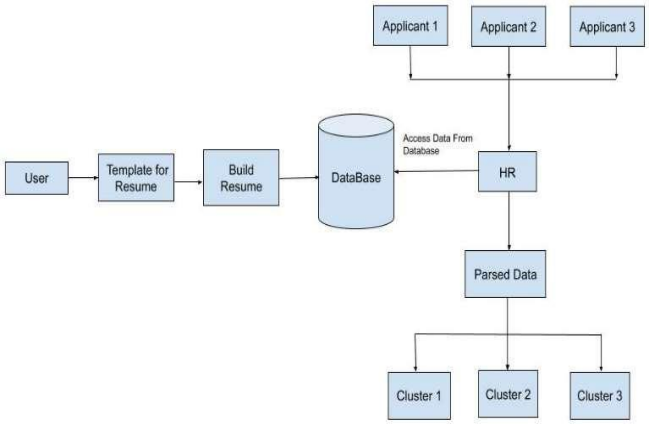

Fig 4.1 System Architecture

The system architecture diagram above represents the flow of the resume analysis system. It is a machine learning-based resume parsing solution designed to automate the extraction of important information from resumes. The system processes resumes submitted by applicants and extracts important details such as contact information, work history, education, skills, and achievements. Its purpose is to simplify and speed up the recruitment process and make it easier for employers to assess a candidate's qualifications and suitability for the position. It has an integrated resume builder for candidates.
The main components are:

1. Data collection and preprocessing:
Collect various CVs for training and testing the model. Sources include job portals, company databases, publicly available datasets, etc.

2. Feature extraction:
Convert text data into numeric vectors using techniques such as TF-IDF (Term Frequency - Inverse Document Frequency) and word embeddings (Word2Vec, GloVe, etc.). Extract meta information such as file format, creation date, file size, etc.

3. NLP Techniques:
Unique Entity Recognition (NER): Identify and classify entities such as names, addresses, dates, etc.
Part of speech (POS) tagging: Assigns words to grammatical categories (noun, verb, etc.) to help understand sentence structure. Dependency Analysis: Analyze grammatical structures to understand relationships between words.

4. Machine learning model:
Train a model to identify and extract relevant entities (skills, education, work history, etc.) from resumes. Develop a classifier to categorize resumes into different jobs or industries.

5. Resume Parsing Algorithm:
Design an algorithm that processes resume one by one and extracts information in a logical order (personal information, education, work history, etc.).

6. Keyword Matching:
Implement a mechanism to match the extracted information with a predefined list of job-related skills and keywords.

4. FLOWCHARTS

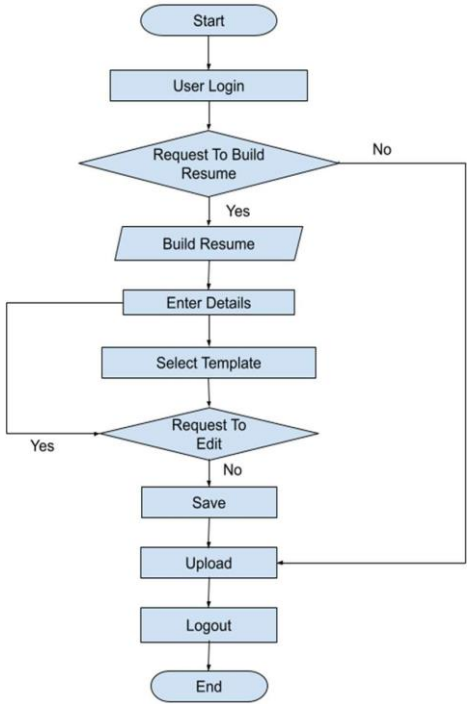

Fig 5.1 Resume Builder

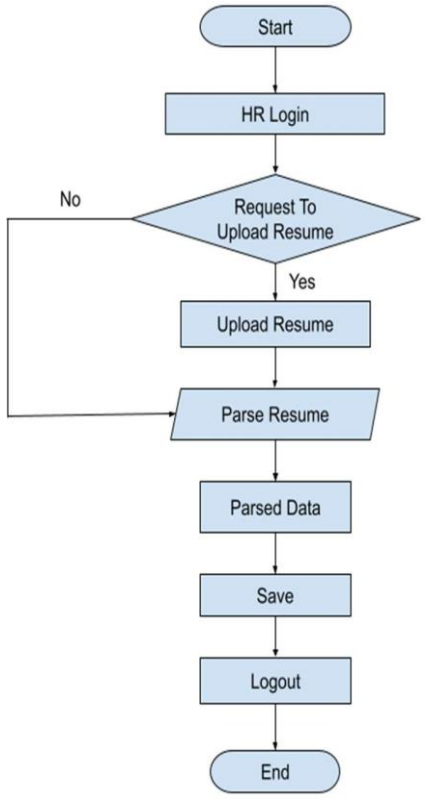

Fig 5.2 Resume Parser

The flowchart for the Machine Learning and NLP Based Resume Parsing Framework For E-Recruitment begins with the start of the process. The user uploads a resume for parsing, which is then preprocessed to clean and standardize the text. Next, relevant features such as education history, skills, and achievements are extracted using machine learning and NLP techniques. These extracted features are then classified into categories (e.g., education, skills) using machine learning algorithms. Finally, the parsed resume data is stored for further processing or analysis, and the process ends. After the resume is parsed and the relevant information is extracted, the framework can perform additional tasks such as matching the candidate's skills and qualifications with job requirements, ranking candidates based on their suitability for a particular job, and generating summaries or insights from the parsed data. The framework can also be integrated with existing recruitment systems to automate the screening and shortlisting process, saving time and effort for recruiters. Additionally, the framework may include features for continuous learning and improvement, where it adapts its parsing and matching algorithms based on feedback and outcomes from previous recruitment cycles.

5. ALGORITHM

1. Named Entity Recognition (NER): NER is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. Libraries like SpaCy, NLTK, and Stanford NER are commonly used for NER tasks.

2. Pattern Matching: Pattern matching involves finding specific sequences of data within a larger dataset. It's commonly used in text processing, searching, and data validation. Python's `re` module provides support for regular expressions, which are powerful tools for pattern matching.

3. Regular Expressions (Regex): Regular expressions are sequences of characters that define a search pattern, mainly for use in pattern matching with strings. They are incredibly versatile and useful for tasks like data validation, text parsing, and text manipulation.

4. SHA256 Encryption: SHA-256 (Secure Hash Algorithm 256-bit) is one of the cryptographic hash functions designed by the National Security Agency (NSA). It generates an almost-unique, fixed-size 256-bit (32-byte) hash. SHA-256 is widely used in various security applications and protocols, including digital signatures, SSL certificates, and cryptocurrencies like Bitcoin.

6. WORKING MODULES

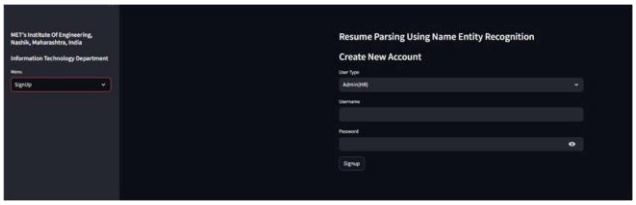

Fig 5.3 Sign up Page

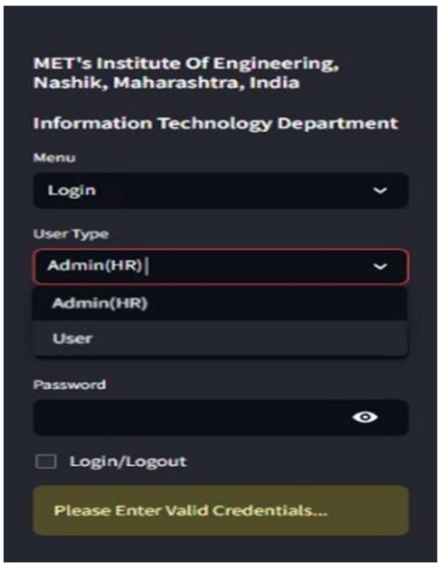

Fig 5.4 Login Page

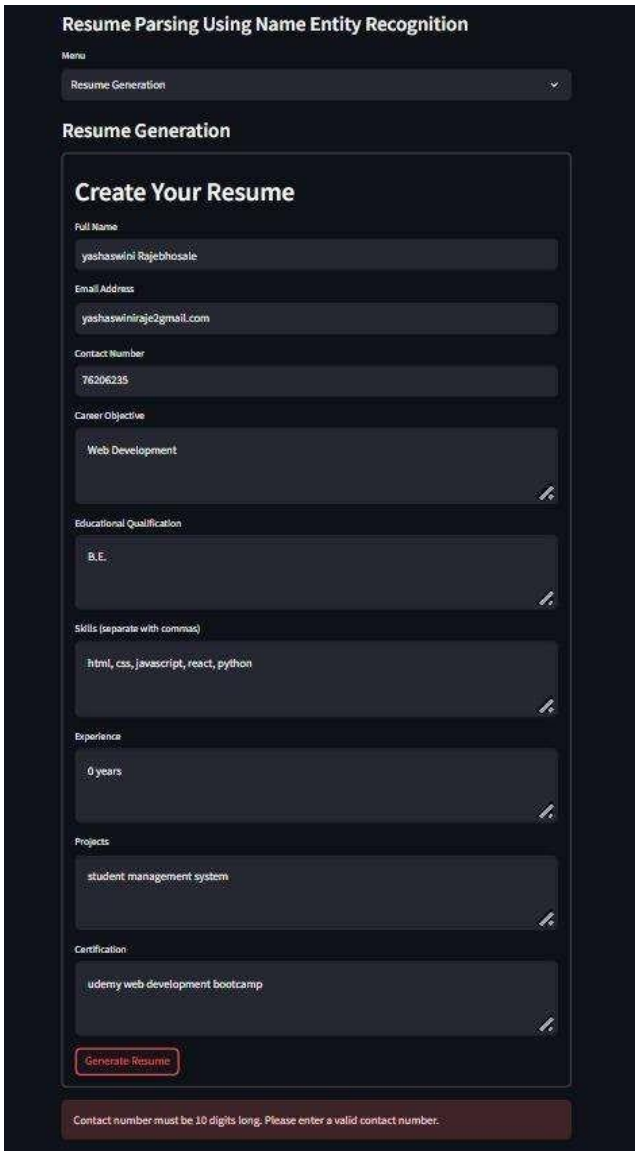

Fig 5.5 Resume Generator

Fig 5.6 Resume Upload

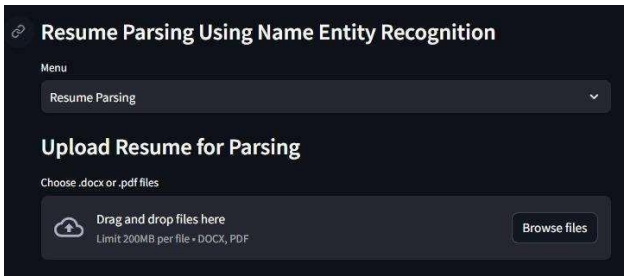

Fig 5.7 Resume Parser(a)

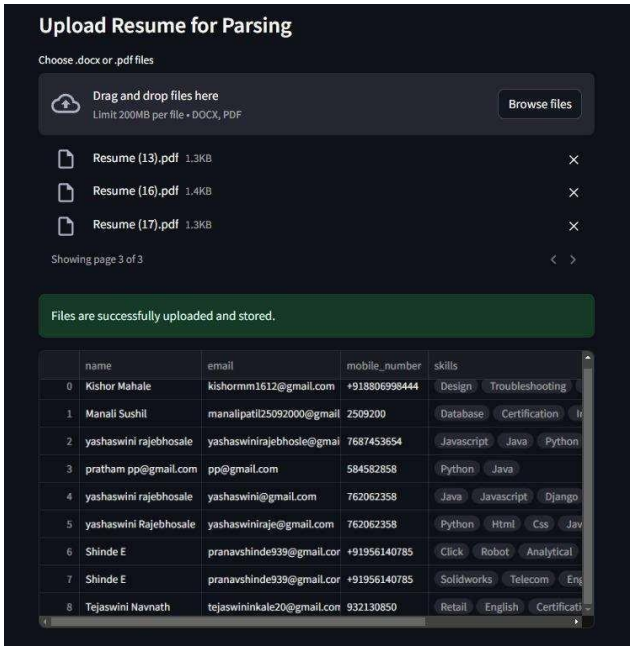

Fig 5.8 Resume Parser(b)

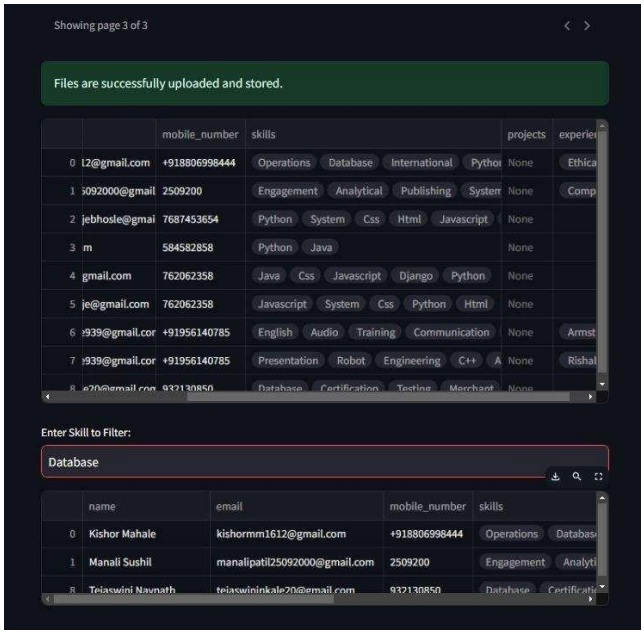

Fig 5.9 Resume Parser(c)

8. CONCLUSIONS

In summary, machine learning and natural language processing (NLP)-based resume parsing frameworks for e-recruiting represent significant progress in automating talent acquisition processes. By leveraging advanced algorithms, the framework not only streamlines the traditionally time-consuming task of resume screening, but also improves the accuracy and efficiency of candidate evaluation. Integrating NLP allows for a more nuanced understanding of contextual language, allowing for more comprehensive candidate profiling. Ultimately, this framework provides recruiters with tools to efficiently identify and engage with the most qualified candidates, thereby streamlining the entire hiring process and improving the hiring environment.

9. REFERENCES

1. Abdul Wahab, Dr. M N. Nachappa. Resume Parser with Natural Language Processing. International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 03, March 2022.

2. Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh. NLP based Extraction of Relevant Resume using Machine Learning. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue- 7, May 2020.

3. S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, A. Gurjar. ”Web Application for Screening Resume.” 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019.

4. Chirag Daryani, Gurneet Singh Chhabra, Harsh Patel, Indrajeet Kaur Chhabra,Ruchi Patel. ”An Automated Resume Screening System Using Natural Language Processing and Similarity.” Topics In Intelligent Computing and Industry Design, 2(2), 99-103, 2020.

5. Suhas Tangadle Gopalakrishna, Vijayaraghavan Varadharajan. ”Automated tool for resume classification using semantic analysis.” International Journal of Artificial Intelligence and Applications (IJAIA), Vol.10, No.1, January 2019.

6. Dr. Parkavi A, Pooja Pandey, Poornima J, Vaibhavi G S Kaveri B. ”E-Recruitment System Through Resume Parsing, Psychometric Test and Social Media Analysis.” IJARBEST, 2019.

7. Astitva Aggarwal, Samyak Jain, Shalini Jha, Ved Prakash Singh. ”Resume Screening.” International Journal for Research in Applied Science Engineering Technology (IJRASET), May 2022.

8. Pradeep Kumar Roya, Sarabjeet Singh Chowdhary, Rocky Bhatia. ”A Machine Learning Approach for Automation of Resume Recommendation system.” International Conference on Computational Intelligence and Data Science (ICCIDS 2019). 27