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 - Today, there is a vast amount of data being generated. Big data refers to large and complex data sets characterized by high Velocity, Volume, and Variety. Big data encompasses various types of data such as social data, machine data, and transaction-based data. Social data is gathered from platforms like Facebook and Twitter, while machine data includes information from RFID chip readings and GPRS. Transaction-based data consists of data from retail websites. Text data, a significant component of big data, is structured data. Text analytics involves extracting high-quality structured data from unstructured text. One text analytics technique is CV parsing, which involves parsing resumes or CVs. CV parsing integrates a candidate's resume with the recruitment workflow and automatically processes incoming CVs. This paper introduces a CV parser model that utilizes text analytics to extract entities necessary for the recruitment process in companies.

Key Words**:** _CV, Parser, Big-Data, Text Analytics._

1.INTRODUCTION

Big data refers to intricate data sets that go beyond traditional data processing methods. Big data incorporates predictive analysis, user behavior analysis, and various value extraction techniques. The volume of data sets is expanding rapidly and is being collected through numerous sensors. Big Data involves addressing challenges such as capture, storage, and analysis. Computational analysis of data sets can unveil patterns, trends, and associations, notably among these being text data. Text data is already structured in a predefined format and requires analysis using text analytics methods to derive high-quality structured data from unstructured text. This process enhances customer master data, generates new consumer insights, and identifies product and service locations. Text databases consist of extensive collections of documents sourced from books, articles, and more, with information retrieval conducted from these text-based documents. The focus of this paper is on CV parsing, a crucial aspect of the text analytics process that relies on entity extraction. Entity extraction involves parsing and extracting entities from raw text, serving as a search process for various integrated software systems. CV parsing typically involves converting free-form CV documents into structured XML format to facilitate data storage and evaluation, commonly utilized by recruitment companies. In certain instances, Hadoop can be leveraged as a big data tool for various applications, such as using Hadoop MapReduce for word processing tasks aimed at reducing the frequency of consecutive words within a system. The paper's structure is outlined as follows: Section 2 provides an overview of previous works in text analytics.

2. LITERATURE SURVEY

2.1 A Prototype to Enhance Recruitment Process with NLP based Resume Parsing

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 papers similarly to a human. For assessing and analyzing the unstructured data, various methods including named entity recognition, tokenization, text classification, and approaches of other natural language processing techniques have been explored.

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

Today, the amount of data generating are very large. Big data are large and complex data sets with an alarming Velocity, Volume and Variety. Depending upon the variations of data, big data constitutes social Data, machine data and transaction based Data. Social data collected from Facebook, Twitter etc. Machine data are RFID chip reading, GPRS etc. Transaction based data includes retail website's data. Around the variations of different types of data major part is text data. Text data is structured data. Deriving of high quality structured data from unstructured text is text analytics. Converting unstructured data into meaningful data is text analytics process. CV parsing is one of the text analytics technique. It is resume parsing or extraction of CV. CV parser integrate candidate's resume with recruitment work flow and automatically processes incoming CV's. This paper proposes CV parser model using text analytics. The proposed CV parser model extracts entities required in recruitment process in the companies.

## 2.3 NLP based Extraction of Relevant Resume using Machine Learning

The estimations set of immense records are enormous and jumbled in nature. Thusly, various item programs have been added to deal with such enormous databases.CV parsing is such a strategy for social occasion CV's.CV parser reinforces more than one language, Semantic mapping for limits, development sheets, determination agents, effortlessness of customization. Parsing with lease limit bears us accu-cost results.

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 presents a smart evaluation in the consistent database to survey with the present models

2.4 Resume Screening Using Machine Learning and NLP

Resume screening is the process of analyzing the resumes where the candidates apply for the different types of jobs where the company feel the tedious job to find the appropriate candidate due to the complexity in resumes formats since it has different styles. As a result, selecting applicants for the appropriate job within a company is a difficult task for recruiters. We can extract the key information from the CV using NLTK, Natural Language Processing (NLP) techniques to save time and effort. This system could work with a large number of resumes for classifying the right categories using different classifiers like KNN, SVM, MLP, LR. Furthermore, this system attempts to find the accuracy and performance of the proposed methodology and incorporate it in the IT firms and other regulations for the prevention of manual screening and establish a safe allocation of resources for the companies.

3. MODULE DESCRIPTION
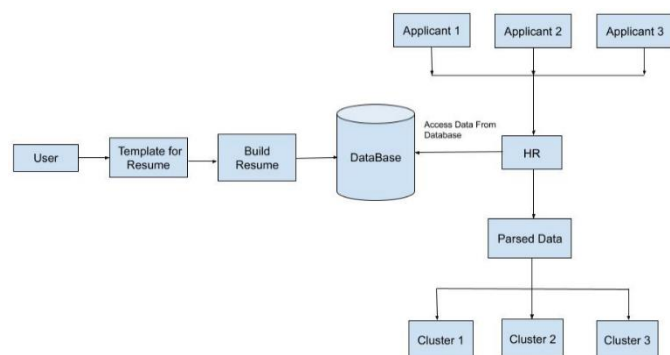
Fig3.1 System Architecture

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.
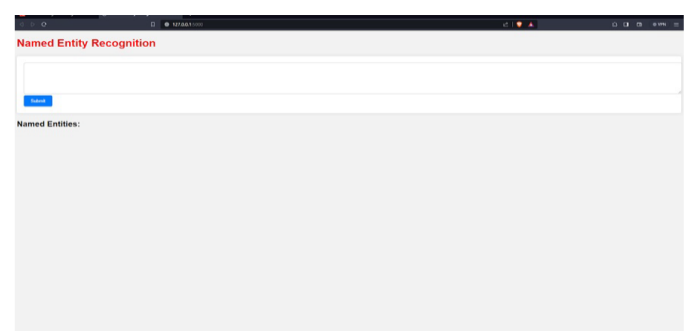
5. Resume Parsing Algorithm:

Design an algorithm that processes resumes 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. GUI/WORKING MODULES
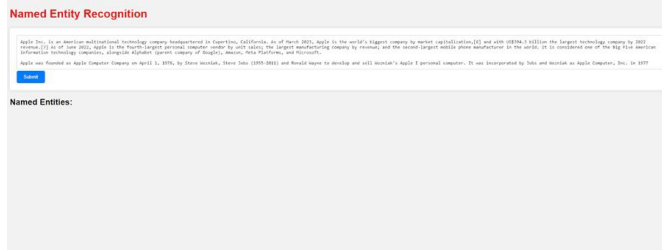

Fig4.1 : Name Entity Recognition(a)

Fig4.2 : Name Entity Recognition(b)

Fig4.3 : Name Entity Recognition(c)

5. ALGORITHM

5.1 Support Vector Machine Algorithm (SVM):

Support Vector Machine (SVM) is a supervised machine learning algorithm that is commonly used for classification tasks. While SVM itself is not typically used directly for resume parsing, it can be a component of a larger resume parsing system. Resume parsing is the process of extracting structured information from resumes, such as candidate names, contact information, work experience, skills, and education. Here^as how you can use SVM as part of a resume parsing system.

5.2 Natural language processing (NLP):

NLP algorithms are typically based on machine learning algorithms. Instead of hand-coding large sets of rules, NLP can rely on machine learning to automatically learn these rules by analyzing a set of examples (i.e. a large corpus, like a book, down to a collection of sentences), and making a statistical inference.

6. 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 powerful tools to efficiently identify and engage with the most qualified candidates, thereby streamlining the entire hiring process and improving the hiring environment. I promise to start a revolution.

6. 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.