RESUME PARSER USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

J. ADITYA

SK. CHAND BASHA

G. SRI RAMYA

Under the Guidance of

Dr. MARIA CELIN TA

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

AUGUST 2023

BONAFIDE CERTIFICATE

Certified that this minor project report entitled **"RESUME PARSER USING MACHINE LEARNING"** is the bonafide work of **"J. ADITYA (20UEEC0128), SK. CHAND BASHA (20UEEC0319) and G. SRI RAMYA (20UEEC0097)"** who carried out the project work under my supervision.

SUPERVISOR	**HEAD OF THE DEPARTMENT**
Dr.MARIYA CELIN TA	**Dr.G.VAIRAVEL**
Assistant Professor	Associate Professor
Department of ECE	Department of ECE

Submitted for major/minor project work viva-voce examination held on:_____

INTERNAL EXAMINER **EXTERNAL EXAMINER**

TABLE OF CONTENTS

ABSTRACT

Agencies and various IT companies must deal with a large number of new jobs seeking people with various resumes. Resume parsing is the process of extracting information from resumes or CVs. It is a time-consuming procedure that can be automated using Natural Language Processing (NLP) and Machine Learning (ML) techniques. This project provides an overview of an ongoing Information Extraction System project that helps recruiters in identifying the best candidate by extracting relevant information from the resume matching the job description.

The rise of online job application processes, submitting a resume has become easier than ever before. As a result, a larger number of individuals are impacted by this change. Many organizations still accept resumes by mail, which can cause challenges for their human resources departments. Sorting through a large number of applications to find the most suitable candidates can be a time-consuming process for these agencies. Job applicants submit resumes in a wide range of formats, including various fonts, font sizes, colors, and other design elements. Human resources departments are responsible for reviewing each application and selecting the most qualified candidate for the job. My suggestion is to incorporate natural language processing techniques into the projectś parser to assist the human resources department or recruiting manager in analyzing the information provided in resumes. This approach can involve using keyword matching and other natural language processing methods to identify the most suitable candidates and obtain the most effective resumes. By doing so, the organization can improve its recruitment process and identify the best candidates for the job.

LIST OF ABBREVIATIONS

NLP	-	Natural Language Processing
AI	-	Artificial Intelligence
$NLTK$	-	Natural Language Toolkit
ML	-	Machine Learning
EDA	-	Exploratory Data Analysis
$TF-IDF$	-	Term Frequency-Inverse Document Frequency

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Through different accessible channels, including the firm site, outside sites, work adverts, work references, and so on, competitors can go after a job.The employing system for an enterprise starts when competitors submit applications for open positions.

Enlisting organizations and business undertakings survey various continues day to day. That isn't an undertaking for individuals. It would be great if a modernized keen machine would extricate all the significant data from unstructured continues and overhaul it in a routinely organized design that could in this way be positioned for a specific capability. The name, email address, online entertainment profiles, individual sites, long periods of work insight, fine art models, long stretches of training, instructing studies, distributions, certificates, volunteer encounters, key expressions, and at some point or another the group of the resume are completely remembered for the parsed measurements.

The ongoing test is realizing which resumes ought to be arranged and shortlisted in view of the cutoff points. This resume scanner saves you time and diminishes how much exploration you should do. Each set incorporates data about an individual's experience, proficient experience, or instructive foundation.The project will also focus on evaluating the performance of the resume parser by comparing its results with manually parsed resumes.The outcomes of this project have the potential to enhance the hiring process, improve candidate selection, and increase overall productivity in the recruitment industry.

1.1.1 Implementation of module

Natural Language Processing is a branch of AI that integrates languages and computer science to study the patterns and architecture of language. It helps to develop intelligent systems that are based on machine learning and NLP algorithms which can read, analyze, and extract meaning from text and voice.

By collecting the relevant information from the resume, a natural language processing technique is employed to create a Hire ability system. Different NLP libraries, such as NLTK and Spacy, are utilized for extraction. Natural language networks are usually taught as unsupervised techniques, which implies that no previous tagging or labeling is done before the model is trained.

1.1.2 Reason behind implementing NLP

The key reason I chose NLP for Hireability is that it can handle massive volumes of data in seconds or minutes that would take days or weeks to analyze manually. NLP technologies can instantly scale up or down to match demand, allowing us to have as much or as little processing capacity as per requirement. Aside from that, humans are prone to errors or may have personal biases that might distort the findings while conducting repeated jobs like examining resumes one by one and other textual data. In this case, NLP-powered solutions can be taught to understand company's need and requirements in only a few steps. So, once they're up and going, they perform better in terms of accuracy.

1.1.3 Work Flow of NLP

Natural language processing comprises a wide variety of methods for analyzing human language, based on machine learning techniques as well as rules-based and computational approaches. Tokenization, lemmatization and stemming, parsing, part-of-speech tagging, language identification are some basic NLP tasks. NLP tasks, in general, break downs the language into smaller, essential components, attempt to comprehend links between the pieces.The initial stage is generally text wrangling and pre-processing on the collection of documents. Then there's parsing and some basic exploratory data analysis. The next stage is to represent text using word embeddings and then do feature engineering. Following that, we must select a model based on whether we are dealing with a supervised or unsupervised learning scenario. The final step of any ML workflow is model testing and deployment. The early processes of text pre-processing and EDA are covered.

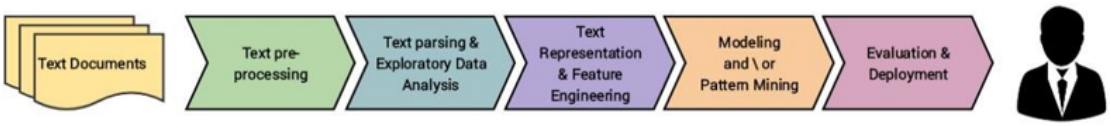

Figure 1.1: Generic Workflow of NLP

1.1.4 Scope and Limitations of the project

A purpose system can help in resolving the challenge of obtaining useful information from a resume in a structured format. By resolving this issue, recruiters will be able to save hours each day by eliminating manual resume screening. Bias in hiring is still prevalent, thus this method may also address the bias hiring process and strengthen a non-bias policy.

Purpose system is not able to solve the complex network issue such as:

• Excessive web traffic can significantly slow down or restrict access to a website entirely. This occurs when the server receives more file requests than it can handle.

• Latency issues.

CHAPTER 2

LITERTAURE SURVEY

2.1 OVERVIEW

Agencies and different high-level companies have to deal with an extreme number of new jobs seeking employees with different resumes. However, looking after those large numbers of text data and filtering out the needed candidates is a burden on the brain and more time consuming. Therefore, the essence of this literature review is on studying resumes in different formats such as single-column resumes, double-column resumes with extension.pdf,.docx, and how the suggested Information Extraction System converts that unstructured information into structured layout through Parsing.

2.1.1 Resume Parser Analysis Using Machine Learning and Natural Language Processing

With the rise of online job application processes, submitting a resume has become easier than ever before. As a result, a larger number of individuals are impacted by this change. Many organizations still accept resumes by mail, which can cause challenges for their human resources departments. Sorting through a large number of applications to find the most suitable candidates can be a time-consuming process for these agencies. Job applicants submit resumes in a wide range of formats, including various fonts, font sizes, colors, and other design elements. Human resources departments are responsible for reviewing each application and selecting the most qualified candidate for the job. My suggestion is to incorporate natural language processing techniques into the project's parser to assist the human resources department or recruiting manager in analyzing the information provided in resumes. This approach can involve using keyword matching and other natural language processing methods to identify the most suitable candidates and obtain the most effective resumes. By doing so, the organization can improve its recruitment process and identify the best candidates for the job.

2.1.2 Resume Parser with Natural Language Processing

Multiple NLP approaches such as bigram trigram and n-gram and text classification were utilized to choose the best candidates from a pool to accomplish these sorts of jobs; this model employs Machine Learning to accomplish the categorization using the methodology.

2.1.3 Resume screening and recommendation system using machine learning approaches

We can use Natural Language Processing (NLP) techniques to extract the relevant information from the resume to save time and effort. Also, a Machine Learning (ML) model is trained to check whether a candidate's skills, experiences, and other aspects are suitable for that particular role. In addition to that, our system will also recommend the other available job roles based on the candidate's skillset.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 OVERVIEW

The methodology for building a resume parser using machine learning involves several steps. The first step is data preparation, which involves collecting a large number of resumes in various formats such as PDF, Word, and HTML. The next step is to preprocess the data by converting it into a structured format that can be easily analyzed and searched. The next step is to use NLP (natural language processing) and machine learning techniques to train the model to identify and extract relevant information from the resumes. Statistical parsers are the most advanced technique of resume parsing using machine learning. They apply statistical models to text to identify and cull out structures in CVs.

The development of a machine learning-based resume parser involves a systematic approach. It begins with assembling a diverse dataset of resumes, encompassing various formats and industries. The data is then preprocessed, removing noise and inconsistencies. Annotation follows, where key information like contact details, work history, education, and skills are labeled. These labeled examples facilitate feature extraction, converting the text data into numerical representations using techniques such as TF-IDF or word embeddings. A suitable machine learning model, such as Conditional Random Fields or Recurrent Neural Networks, is selected and trained on the annotated dataset. Performance evaluation using metrics like precision, recall, and F1-score guides model refinement. Post-processing steps correct errors and inconsistencies, enhancing accuracy. A user-friendly interface is developed for easy resume upload and parsed information retrieval. Continuous improvement is achieved by incorporating user feedback and updating the model to handle new resume formats. Finally, the system is deployed and maintained, ensuring adaptability to evolving resume structures and market trends.

CHAPTER 4

CONCLUSION

he proposed approach is a comprehensive plan that aims to revolutionize the way organizations hire their employees. The goal is to improve the efficiency and effectiveness of the recruitment process, making it easier for both employers and candidates. One of the primary objectives of the approach is to streamline the hiring process by eliminating unnecessary steps and automating certain tasks. This will ensure that the process is completed in a timely manner, saving both the employer and candidate valuable time Moreover, the approach focuses on delivering highly qualified candidates to the firms. To achieve this, the approach will use a ranking system that is based on the technical skills of the candidates. This will ensure that the firms receive resumes of candidates who have the necessary qualifications and expertise to perform the job duties. The result generated shows precision, accuracy, F1-Score and other statistics. Ultimately, the approach aims to make the work of companies and candidates easier and more effective. By providing a streamlined recruitment process, employers can quickly and easily identify qualified candidates, while candidates can apply for jobs with confidence, knowing that they will be evaluated fairly and objectively. The outcomes of this project have the potential to enhance the hiring process, improve candidate selection, and increase overall productivity in the recruitment industry.

REFERENCES

[1] Amit Pimpalkar; Aastha Lalwani; Roshan Chaudhari; Mohd. Inshall; Mahak Dalwani; Tarandeep Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning", *Electrical, Electronics and Computer Science (SCEECS), 2023 IEEE International Students',* vol. 09, no. 03, 2022.

[2] Abdul Wahab and M Nachappa, "Resume Parser with Natural Language Processing", *International Research Journal of Engineering and Technology,* vol. 09, no. 03, 2022.

[3] S Lokesh, Balaje. S Mano, E Prathish and B. Bharathi, "Resume screening and recommendation system using machine learning approaches", *Computer Science Engineering: An International Journal ,* vol. 12, no. 1, 2022.