e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

A SURVEY ON RESUME ANALYSIS USING NLP

Harshitha R*1, Mrs. Veena B*2

*1P.G, Student, Department Of Master Of Computer Applications, University B. D. T College Of Engineering, Davangere, Karnataka, India.

*2Assistant Professor, Department Of Master Of Computer Applications, University B. D. T College Of Engineering, Davangere, Karnataka India.

ABSTRACT

Specialized recruitment companies that use machine learning models to expedite the hiring process and provide their clients with the best personnel have emerged as a result of the expansion of the Indian recruitment market. In order to improve the employment process and make it more equitable and efficient, the suggested model has the potential to benefit from machine learning. You may accurately grade resumes according to the company's criteria and extract useful information from them by applying natural language processing (NLP) techniques. The system presented in this study extracts minute details from a résumé, including education, experience, talents, and experience, using Natural Language Processing (NLP) techniques. Parsing the resume facilitates and expedites the hiring process. Three components comprise the proposed system: an information extraction system, a file upload and parsing system, and an administration management system. The applicant's résumé will be uploaded by the administrator into the system, and pertinent data will be extracted in an organized manner. Based on the demands of the business, HR can choose the ideal applicant for the position using the information that has been parsed from the resume.

Keywords: Resumes, Nlp, Parser, Extract Information, Skillset.

I. INTRODUCTION

In the field of machine learning, a model is trained using a dataset to anticipate the desired result when fresh data is provided. Natural language processing, or NLP, is mostly used to screen resumes. Natural language is the term used to describe how people interact with one another. The goal of natural language processing (NLP) is to enable computers to comprehend spoken and written language in a manner similar to that of humans. NLP blends deep learning, machine learning, statistical, and rule-based modeling of human language with computational linguistics. When these technologies are combined, computers can better comprehend and "understand" human language in the form of text or speech data. According to LinkedIn , millions of new job seekers enter the workforce annually as a result of India's expanding job market. According to the 2021 Employees Provident Fund Organization (EPFO) around 1.3 million additional employment were created. India's unemployment rate as of this year is approximately 7.74%, with 9.06% of unemployed people living in urban areas and 7.13% in rural areas.

Talent acquisition companies therefore arise as solutions to this problem, filling in the void and completing the job with less resources costing the company within an acceptable timeline. As a result, if the companies hire in bulk, there are many applications to find the talent that they need, which will require a considerable amount of resources and time. Since it would take a lot of time to read through the millions of applications received even here, these organizations utilize a variety of machine learning algorithms to identify the best resumes for each job post.

Below is the wireframe of Unstructured Data (Single-Column, Double-Column, Free Format)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

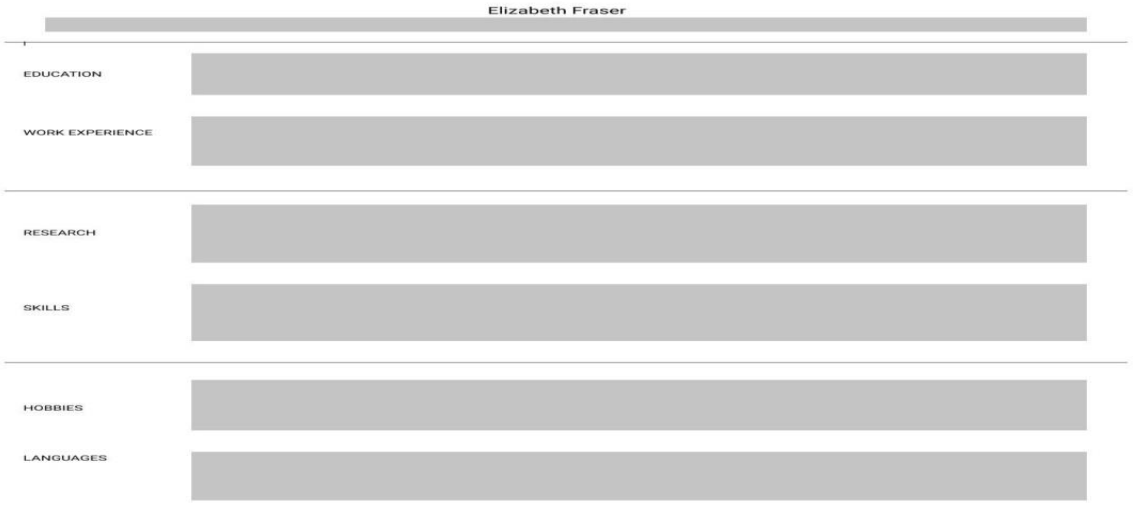

Figure 1: Single Column Resume

Figure 2: Double Column Resume

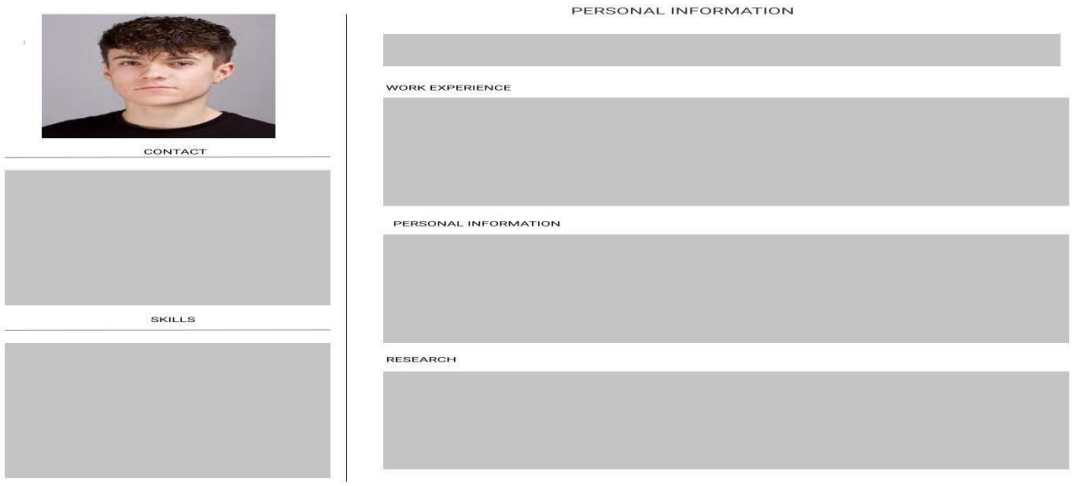

Figure 3: Free Format Resume (Image)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

Structured Data (Name, Address, Email, phone_number, Skills [Hard Skills, Soft Skills], Education, Experience, Languages)

Name	
Address	
Email	
phone_number	
Skills [Hard Skills, Soft Skills]	
Education	
Experience	

The structured data comes in. json format which makes the HR department easy to read the resume.

II. LITERATURE REVIEW

2.1. End-to-End Resume Parsing and Finding Candidates for a Job Description using BRET

The usage of deep learning-based systems has given job seekers an end-to-end solution for evaluating acceptable applicants for various positions based on their compatibility (Bhatia, et al., 2019). They may be completed in two steps: creating a resume parser to extract all relevant data from candidate resumes, and then using BERT phrase pair classification to rank the resumes. Sentence pairs were classified using the BERT algorithm, which predicted the correlation score between the applicant profiles and the job description with 72.77 percent accuracy. They investigated the feasibility of creating a common parser for all resume formats in this study and found that it was not feasible to do so without losing information in every circumstance, leading to the unjust rejection of resumes from certain individuals. The finished task and the data flow are shown in the system diagram below.

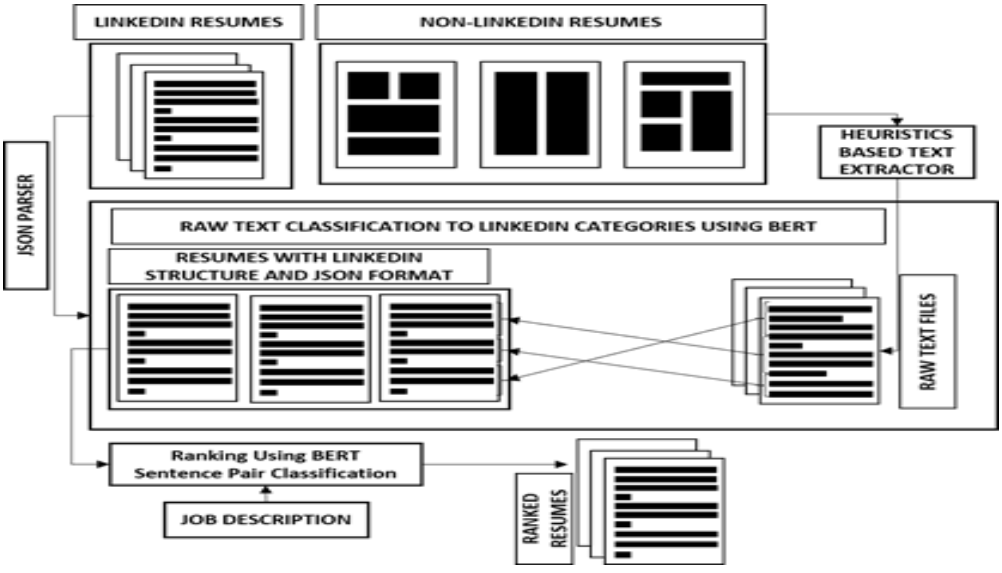

Figure 4: A system diagram showing data transmission and task completion.

2.2. Information Extraction from Free-Form CV Documents in Multiple Languages

A method for choosing pertinent document sections and comparable specific information at the low hierarchy level has been made possible by the application of two natural language processing algorithms to extract significant material from an unstructured multilingual CV (VUKADIN, et al., 2021). In order to achieve a high degree of extraction accuracy, it makes use of an NLP machine learning technique. Authors applied the encoder component of the BERT language model using the transformer architecture in their work. To extract information at the section and item levels from a CV document, two models were created. The extracted Skills portion from the dual model is categorized using a self-assessment model of skill performance.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

The authors assert that by developing an NLP system, they have overcome the difficulty of parsing CVs. It is demonstrated that the freshly introduced tokens [NEW LINE] and [SKILL] have learned to function as predicted.

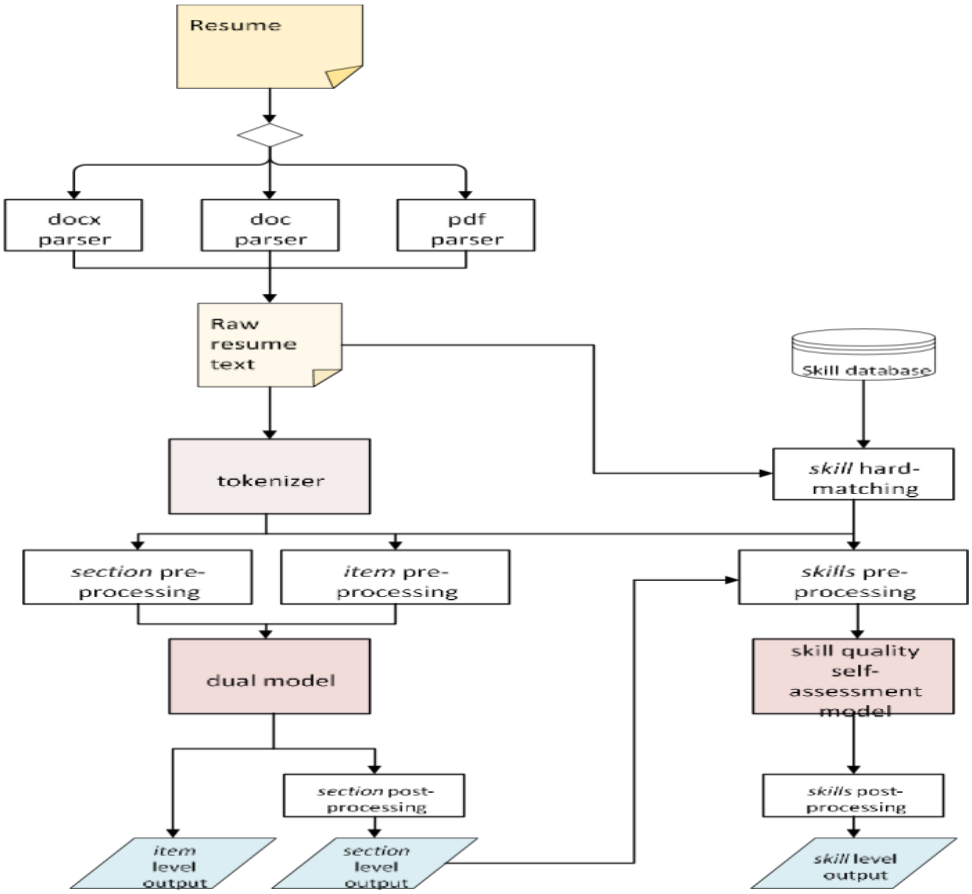

Figure 5: The Information Extraction System from Free-form CVs: A High-Level Overview

2.3. Resume Information Extraction with a Novel Text Block Segmentation Algorithm

Over the past ten years, the recruiting process has changed significantly, moving from traditional job fairs to web-based platforms. Consequently, a resume parsing pipeline utilizing distributed embeddings and neural network-based classifiers is provided by (Wang & Zu, 2019). The tedious task of individually crafting several handmade parts is eliminated by the pipeline. Text block segmentation, integrated word representations, resumption fact detection using position-wise line information, and named entity recognition utilizing multiple sequence labeling classifiers inside labeled text blocks are all combined in the proposed system. By removing the CNN layer from BLSTMCNN's-CRF and contrasting different word embeddings separately, the ablation experiment was carried out.

2.4. Information Extraction from Resume Documents in PDF Formats

(Chen et al., 2016) focuses on the issue of data extraction from resumes in the PDF format and suggests a hierarchical extraction method. Resume articles use heuristic criteria to divide a page into blocks, classify each block using a Conditional Random Field (CRF) model, and treat the task of extracting detailed information as a sequence labeling problem. The layout-based features, which have been found to be particularly helpful for semi-structured information extraction tasks, have increased the average F1score by more than 20% in testing, according to the authors. PDF resumes often provide more extensive information than HTML resumes. The following explains the report's methodology.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

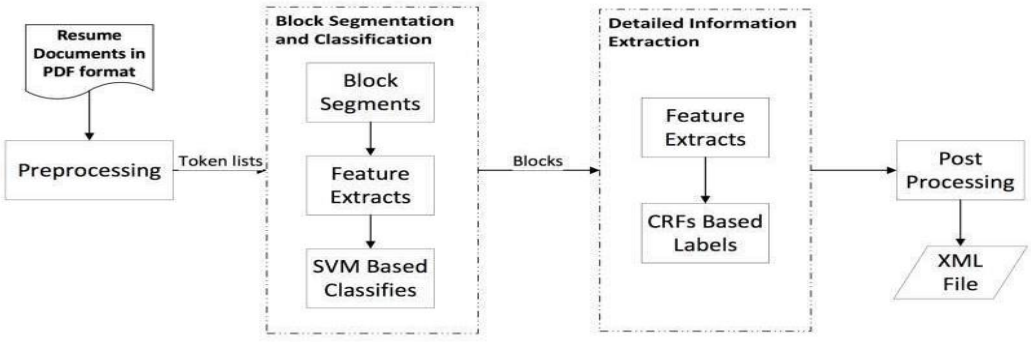

Figure 6: Workflow of Information Extraction from Resume

2.5. Study of Information Extraction in Resume

In order to automatically obtain and analyze various resume forms, (Nguyen et al., 2018) presents Text Normalization, Rule-based Named Entity Recognition, Deep Neural Network Find Name Entities, and Text Segmentation. Convolutional Neural Networks, Bidirectional Long Short-Term Memory, and Conditional Random Field were used to create the Deep Learning model, which the author used to first label the sequence in the resume and then extract Name Entities from the labeled line. When testing on a medium-sized collection of CV data, the developer obtained encouraging results with over 81 percent F1 for NER and compared this model to other systems. Still, there are a few issues with this method that could be fixed. The model cannot be trained with the current amount of data.

2.6. Automatic Extraction of Usable Information from Unstructured Resumes to Aid Search

A natural language processing (NLP) system with an automated information extraction emphasis is proposed by (Kopparapu, 2015) to enable quick resume search and management for both structured and unstructured resumes. They claim that the process consists of two passes. During the first pass, the resume is separated into a series of consecutively named blocks that represent the range of information included in each block. Subsequently, detailed information is acquired in the second pass. For the extraction process, they used a range of heuristics and pattern matching techniques. Experiments on a large number of resumes show that the proposed system can recall 88% of the resumes and achieve 91% accuracy on a variety of resume forms.

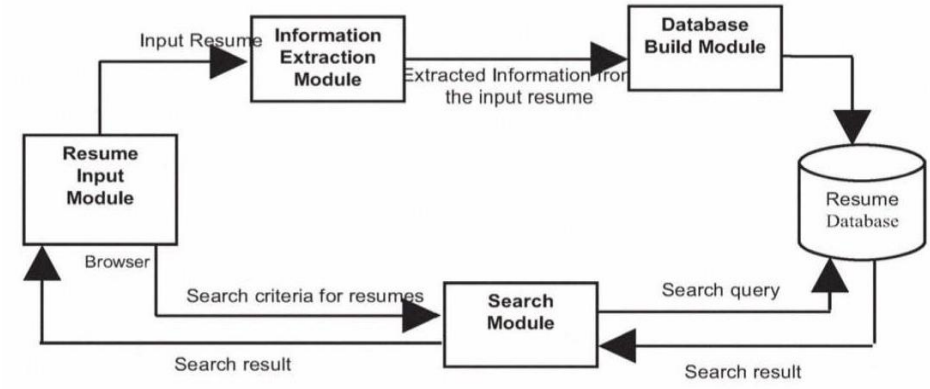

Figure 7: Workflow of the System

2.7. Analysis and Findings

After a thorough examination of the suggested literature study, questions have been raised regarding the development and use of many algorithms to extract the actual number of applicants based on their resumes. Research publications and studies have shown that using several techniques, such as neural networks, NLP, and bidirectional encoder representations from the transformers (bert) model, has been successful in extracting information. Using a large text corpus, like Wikipedia, to train a general-purpose "language understanding" model, BERT is a language representation pre-training technique that applies the model to downstream NLP tasks. The BERT approach outperforms LSTM, RNN, and Bi-LSTM outcomes because it is the first unsupervised system for pre-training NLP and is fundamentally bidirectional.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

The extraction of information from resumes without losing candidate information is perceived as a difficult process requiring sophisticated analysis, despite the fact that the research employs a wide range of methodologies.

III. METHODOLOGY

According to the job description, the resume is processed using a mechanism known as natural language processing in this methodology. The received resumes would be analyzed and sorted based on their content. Our secondary objective is to collect data from resumes, applicant recommendations, and candidate feedback for positions. By doing so, we want to minimize discriminatory and unjust practices and facilitate the recruiting process by receiving high-quality applications from a variety of geographic locations.

The System Architecture consists of Modules:

1. Resume Parser.
2. Data Extraction.
3. Candidate Suggestions and Feedback.

Resume Parsing: NLP approaches are used to parse resumes that are received. Using the job profile as a guide, the parsing process entails collecting pertinent information from the resumes. The goal of this stage is to organize the unstructured resume data so that analysis will be simpler.

NLP (Natural Language Processing) requires the following constraints for parsing:

- ❖ Morphological Analysis
- ❖ Syntactic Analysis
- ❖ Semantic Analysis

Data Extraction: The process also emphasizes data extraction from resumes. This could entail gathering credentials, experiences, abilities, or any other pertinent data that can be analyzed and utilized to make decisions during the hiring process.

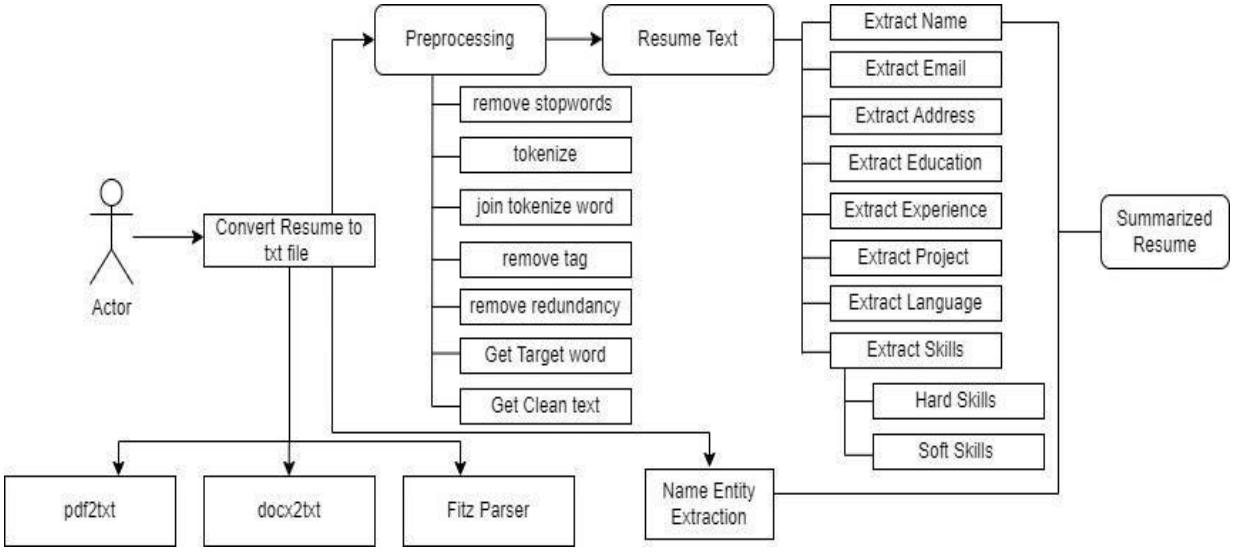

Figure 8: System Work Flow

Natural Language Processing Tool: Utilizing NLP approaches, the Natural Language Toolkit (NLTK) may be used for resume analysis in order to extract pertinent information and get insights from resumes. Here are some examples of how to use NLTK for resume analysis.

Tokenization: Tokenization techniques offered by NLTK allow the text to be divided into individual words or phrases. Tokenization can assist in the extraction of keywords, competencies, job titles, educational background, and other pertinent data from resumes during resume analysis.

Part-of-Speech (POS) Tagging: Every word in a phrase can have a grammatical tag assigned to it by using POS tagging features included in NLTK. The role and category of words on a resume, such as nouns (skills,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

credentials), verbs (action words), adjectives (descriptions), and more, may be identified with the use of POS tagging.

Named Entity Recognition (NER): NER algorithms, like as those provided by NLTK, can recognize and categorize named entities in text, including names of people, places, dates, and organizations. NER can assist in the extraction and classification of crucial data from resumes, including applicant names, firm names, educational institutions, and specifics about job experience.

Semantic Analysis: Semantic analysis features offered by NLTK include semantic similarity computation, word sense disambiguation, and synonym matching. By comparing a resume's abilities and keywords with the job description, one may evaluate a candidate's fitness for a certain position and establish whether their profile is relevant.

Sentiment Analysis: The sentiment conveyed in a resume, whether positive, negative, or neutral, may be examined using the sentiment analysis tools included in NLTK. This can be helpful in determining the resume's general tone as well as evaluating the candidate's demeanor and manner of speaking. Resume analysis systems can automate the process of extracting, classifying, and assessing information from resumes by utilizing the natural language processing (NLP) capabilities offered by NLTK. This makes it possible for recruiters to quickly screen and assess candidates according to their credentials, experience, skills, and other pertinent factors. A Python package called Pyresparser is dedicated to parsing and extracting data from resumes.

Candidate Suggestions and Feedback: Based on the data retrieved and applicant comments received, the system seeks to offer recommendations for candidates. This input has the potential to enhance the hiring procedure and facilitate the process of identifying qualified applications from different geographic areas. It also highlights how crucial it is to steer clear of unfair and discriminatory hiring practices. A section of the suggested system is dedicated to candidate recommendations and comments. The system uses the pertinent data that it has extracted from the resumes to propose candidates to recruiters. These recommendations are predicated on how well the candidate meets the job criteria given their background, education, and experience. Apart from applicant recommendations, the system also collects candidate feedback.

IV. ARTEFACT DESIGN

FUNCTIONAL DECOMPOSITION DIAGRAM

The technique's distinct elements and their various degrees of connection to one another are represented by a chart known as an FDD. An approach is shown in the figure in a way that is top-down. The function name is shown as a rectangle with a linking arrows**.**

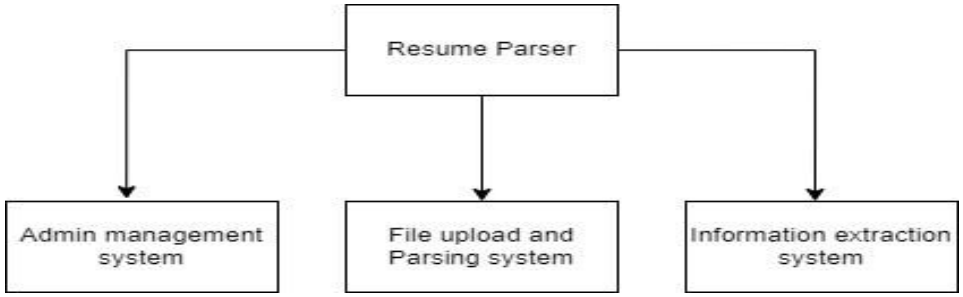

Figure 9: Functional Decomposition Diagram

ADMIN MANAGEMENT SYSTEM (AMS)

Requirement- Code	Requirement Specification	MOSCOW
AMS-F-1	The admin should be able to click the Login button on the system.	Must Have
AMS-NF-1.1	Login form should be opened.	Must Have
AMS-F-2	System should allow admin to login.	Must Have
AMS-F-2.1	Email should be valid.	

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

		Must Have
AMS-F-2.2	Password should be valid.	Must Have
AMS-F-2.3	For a successful login, the admin must have a valid username and password.	Must Have
AMS-F-2.4	The entered username and password must be validated by the system.	Must Have
AMS-UR-2.1	If the login and password are incorrect or empty, the system should produce an error message.	Must Have

V. CONCLUSION

A typical resume provides an overview of a person's educational history, professional experience, credentials, and personal information. These components might be present in several forms or absent at all. Keeping current with the technical terms used in resumes may be challenging. Corporate names, academic institutions, degrees, and other data that may be written in a variety of ways make up a resume. An individual's evaluation of every CV will take time.

Machines perform tasks more accurately and more quickly than humans. As a result, I created a machine learning-based system that can extract crucial information from resumes in one minute or less. This technique may be used by the hiring person to hire any individual.

VI. REFERENCES

[1] Bhatia, V., Rawat, P., Kumar, A. & Shah, R. R., 2019. End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT. arxiv.

[2] Pradeep Kumar Roy, Vellore Institute of Technology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019. 10.1016/j.procs.2020.03.284.

[3] Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job-Resume Matching system. 3]Yong Luo, Nanyang Technological University, 2018. A LearningBased Framework for automatic resume quality assessment, arXiv:1810.02832v1 cs.IR].

[4] Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume, IEEE DOI: 10.1109/ICNTE44896.2019.8945869.

[5] 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.

[6] Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna. "Design and Development of Machine Learning based Resume Ranking System." Global Transitions Proceedings, 2021.

[7] Anushka Agarwal, Dr. Senthilkumar. "Resume Recommendation System Using Cosine Similarity." International Research Journal of Modernization in Engineering Technology and Science, April 2022.

[8] 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.