JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Resume Analyser: Analysing and Filtering System

[1]Prof.Deveshree Wankhede, [2]Aniket Ganesh Singh, [3]Vihaan Biju Nair, [4]Manikandan Kanagaraj Nadar

[1]Assistant Professor, [2,3,4]Student [1,2,3,4]Department of Computer Engineering,
[1,2,3,4]Shivajirao S. Jondhale College of Engineering, Dombivli, India

Abstract : The Resume Analyzer and Filtering System presented in this study is an advanced computational tool designed to streamline the recruitment process by automating the analysis and evaluation of job applicants' resumes. The system leverages cutting-edge NLP techniques, data mining algorithms, and machine learning models to extract pertinent information from resumes, enabling efficient candidate selection. Resume Analyzer and Filtering employs a multi-faceted approach to resume analysis, encompassing the extraction of skills, qualifications, experience, and educational background. Additionally, it assesses the relevance of this information to specific job requirements, ensuring a precise match between candidates and positions. The system's graphical prediction capabilities further enhance its utility, allowing for long-term projections based on historical data and market trends.

***IndexTerms* -** Resume Analysis, Recruitment Process, NLP, Data Mining, Machine Learning, Skills Extraction.

An effective resume analyzer system is a fundamental component of any modern HR infrastructure. In today's fast-paced job market, companies receive an overwhelming number of resumes for every open position, making it increasingly challenging for human resources professionals to identify the most qualified candidates efficiently. The "Resume Analyser: Analysing and Filtering System" project is a crucial solution aimed at streamlining this process, ultimately saving time and resources while ensuring that the best candidates are not overlooked.

This project report delves into the design, development, and implementation of a cutting-edge resume analyzer system, which uses state-of-the-art NLP and machine learning techniques. The system has been engineered to autonomously evaluate and categorize resumes, thereby assisting HR teams in the initial candidate screening process. By doing so, it optimizes the recruitment process, minimizes human bias, and enhances the chances of identifying the most suitable talent for a given job opening.

By the end of this report, you will gain a comprehensive understanding of how the Resume Analyzer System can significantly improve the efficiency and effectiveness of the recruitment process, benefiting organizations of all sizes and industries. This project represents a pivotal step towards modernizing HR practices and staying competitive in an ever-evolving job market. In the dynamic landscape of the modern job market, organizations face an ever-increasing influx of job applications and resumes. Managing this volume of data efficiently and effectively is a critical challenge for human resources departments. To address this challenge, the "Resume Analysis and Filtering System" project was conceived. This project report provides an in-depth exploration of the development and implementation of this innovative system, designed to streamline the employee recruitment process, improve decision-making, and enhance the overall hiring experience. It can be defined as the ability to analyze, understand, and generate human speech.

In today's competitive job market, companies contend with an overwhelming influx of job applications for each available job opening. The manual screening of these resumes is a laborious and resource-intensive task that often proves to be impractical, leading to delays in the hiring process. Recognizing this challenge, the implementation of a resume analyzing and filtering system has become not just advantageous but essential for efficient talent acquisition. Such a system employs advanced technologies, including NLP and machine learning, to meticulously evaluate and categorize resumes based on predetermined criteria. The primary objective is to expedite the identification of the most qualified candidates while optimizing the allocation of resources. By automating the initial screening, the system not only saves valuable time and effort but also significantly mitigates the risk of overlooking top-tier talent. Consequently, this approach aligns with the demands of a fast-paced job market where prompt and effective hiring decisions can make all the difference in securing the best candidates for the organization.

I LITERATURE SURVEY

The impact of the a mentioned publications on the field of recruitment and resume analysis has been substantial, contributing significantly to the advancement of technology-driven hiring processes. Each of these studies has introduced novel approaches and methodologies, leveraging Machine Learning (ML), Artificial Intelligence (AI), and Text Mining techniques. Here is a summary of their respective impacts:

The study paper "Recruitment Prediction Using Machine Learning" by Jagan Mohan Reddy, Sirisha Regella, Srinivasa Reddy Seelam :This publication likely focuses on using ML algorithms for predicting candidate suitability for specific roles. Its impact lies in providing organizations with a data-driven approach to streamline their hiring process, potentially leading to more efficient and accurate candidate selection [1].

The second paper "Automatic Extraction of Segments from Resumes using Machine Learning" by Gunaseelan B, Supriya

Mandal, Rajagopalan V. :This study likely focuses on automating the extraction of specific sections from resumes using ML techniques. Its impact may be in reducing manual effort in resume processing and improving the accuracy of information extraction [2].

The third paper "Application of AI Technology in Resume Analysis and Job Recommendation" by Yi-Chi Chou, Han-Yen Yu :This study likely explores the application of AI in resume analysis and job recommendation. Its impact is likely in optimizing the matching of candidates to job openings, potentially leading to better candidate-job fit and reduced recruitment costs [3].

The fourth paper "A Resume Evaluation System Based on Text Mining" by Yi-Chi Chou, Chun-Yen Chao, Han-Yen Yu :This study likely introduces a resume evaluation system based on text mining techniques. Its impact may be in providing organizations with a systematic and data-driven approach to assess candidate resumes, potentially leading to more informed hiring decisions [4]. The fifth paper "Assessment of Graduate Students' Resumes Using Short Text Searching Method" by Sara Nasr, Oleg Vitoldovicz German : This publication likely involves the use of short text searching methods for assessing graduate students' resumes. Its impact may lie in providing a quick and efficient way to evaluate resumes, potentially aiding in the recruitment of highly qualified candidates [5].

The sixth paper "Web Application for Screening Resume" by Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar : This publication likely introduces a web-based application for resume screening. Its impact may lie in providing a user-friendly platform for organizations to efficiently process a large volume of resumes, potentially improving recruitment efficiency [6].

II METHODOLOGY

The development of a resume analyzer and filtering system project entails a structured methodology aimed at enhancing the efficiency of candidate selection processes. The project begins with a thorough analysis of requirements, encompassing the desired features and objectives of the system. This phase involves gathering insights from employers and HR to ensure alignment with market needs.The development phase focuses on implementing core functionalities such as resume parsing, data extraction, and analysis algorithms.using the various algorithm the resume is analysed and the score is given the employee is provided some 0f the courses and videos to enhance its skills.The courses are provided basis on the basis of employes interest using the analysis algorithm.The filtering system is useful for recruiter to filter the best resume according to the requirement and the skills that the recruiter needs in a employee. This systematic approach ensures the successful realization of a robust and effective resume analyzer and filtering system, empowering organizations to streamline their recruitment processes and identify top talent efficiently.

III EXISTING SYSTEM

Introduction

Recruitment is the major aspect in every business organizations because the company growth is mainly depends on employee performance. Every concern facing some major problem for recruiting trained applicants [1].

In this system, it predicts employee will join or not on before resume selection. The prediction based on some key features like employee age, experience, salary hike, job function and distance from home etc. Recruitment is identifying and encouraging prospective employees to apply for job. The effective prediction is helpful for future recruitment process which will help to choose efficient candidates and within a short period of time. The cooperation between employees and managers allows the enterprises to achieve the business goal and vision. Making a good prediction not only increase employee retention but also reduces the cost associated with a bad hire [1].

System Architecture

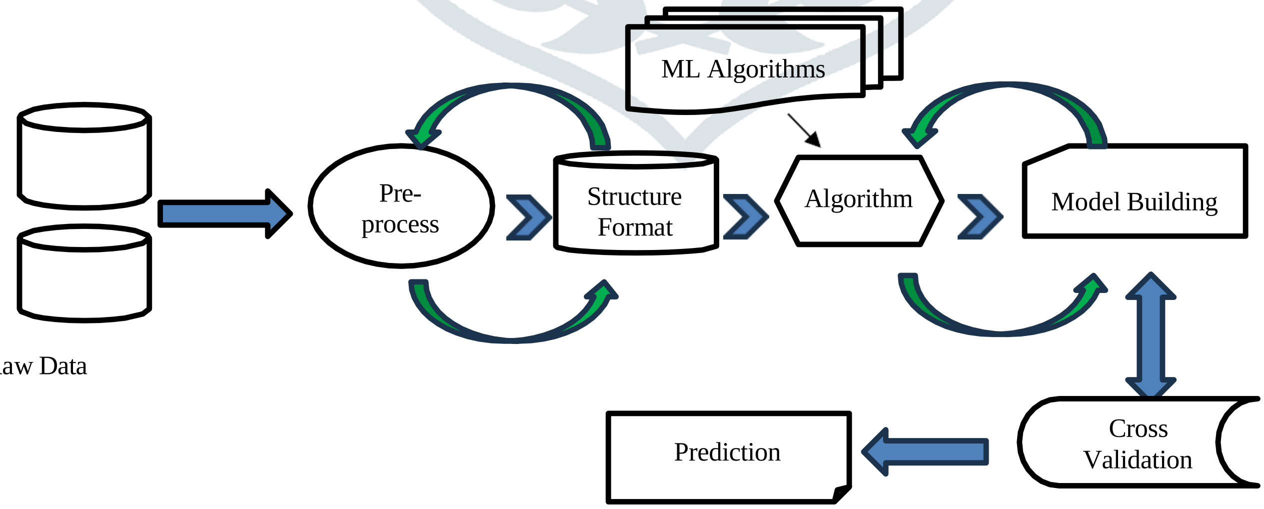

Fig 4.2 Existing System Working

The above Figure demonstrates mainly working of internal process. In this scenario overall system is designed in n tiers separately. The raw data taken as input, then pre-process the raw data by using data pre-processing tools. Therefore, the structured format of raw data is obtained by removing missing values, null values and outliers. After that, a one hot ending technique is applied to enlarge the dataset during the classification process. Our study includes various machine learning algorithms to build models. The results were obtained and predicted the unknown test sample. The measure the perforce of our proposed method, the accuracy and precision is calculated. The given process is continuous until we will get better results from system [1].

Result

The result are achieved by applying various machine learning algorithms and findings of employee recruitment prediction. The experimental are obtained by using decision tree, Random forest, Gaussian Naive Bayes, K-Nearest Neighbor.This System emphasizes on three metric to validate the algorithms viz. accuracy, precision and recall. Initially, the dataset is partitioned into two subsets as training and testing with 70% and 30% respectively.

Table 4.3 Results

Models	Accuracy	Precision	Recall
Decision Tree	0.96	0.75	0.7
Random Forest	0.95	1	0.17
Guassian Naive Bayes	0.99	1	0.88
K-nearest Neighbour	0.93	0.44	0.23

Table 4.3 shows the various algorithm used by existing system and there performance scores. The accuracy of decision tree, Random forest, Gaussian Naive Bayes (GNB) and KNN is 96%, 95%, 99% and 93% respectively. The precision of all the algorithms is 75%, 100%, 100% an d44% only. Similarly, the recall is 70%, 17%, 88% and 23% respectively. Here, GBN is given highest accuracy and precision among all other algorithms [1].

IV LIMITATIONS OF EXISTING SYSTEM

While a resume analysis and filtering system can provide valuable insights, there are several limitations that need to be taken into consideration:

1. Dependency on Data Quality: The accuracy of predictions and recommendations is heavily reliant on the quality and completeness of the data provided in the resume. Inaccurate or incomplete information can lead to less reliable outcomes.
2. Bias and Fairness Concerns: If the system is not designed with care, it may inadvertently perpetuate biases present in the data it's trained on. This could lead to unfair recommendations or predictions, particularly for underrepresented groups.
3. Inability to Assess Soft Skills: While technical skills and qualifications can be extracted from a resume, soft skills such as teamwork, communication, and leadership may not be easily quantifiable from text alone.
4. Lack of Context: Resumes may not provide the full context of a person's experience, skills, or accomplishments. For example, a brief bullet point may not fully convey the scope and impact of a project.
5. Inability to Gauge Personal Growth: A resume typically provides a snapshot of a person's past experiences. It may not reflect their potential for growth, adaptability, or willingness to learn new skills.
6. Overlooking Unconventional Backgrounds: The system may have difficulty evaluating candidates with unconventional career paths or those transitioning between industries.
7. Absence of Qualitative Information: Resumes may not provide qualitative information about a candidate's work ethic, personality, or cultural fit within an organization.
8. Subjectivity in Interpretation: Different people may interpret a resume differently. The system's predictions and recommendations could vary depending on the interpretation of the input data.

Addressing these limitations requires a combination of advanced NLP techniques, rigorous testing, and ongoing refinement based on user feedback. Additionally, it's crucial to maintain transparency and fairness in the system's operation. Regular audits and updates should be conducted to mitigate biases and improve overall performance.

V PROBLEM STATEMENT AND OBJECTIVES

Problem Statement

The current landscape of candidate selection and recruitment processes often lacks the tools and methodologies to provide a comprehensive and forward-thinking assessment of candidates. Conventional systems primarily focus on short-term readiness for immediate job roles, limiting their ability to predict candidates' long-term compatibility and potential for sustained growth and contribution to organizations. This limitation hampers the alignment of recruitment strategies with broader talent acquisition objectives. Additionally, the field lacks cutting-edge predictive modeling techniques and innovative data sources, leaving room for advancement. There is a need for a paradigm shift in recruitment technology that addresses these issues and embraces graphical prediction techniques for more accurate and forward-looking candidate assessments

Objectives

- The primary objective of the Resume Analyzer and Filtering System (RAFS) is to revolutionize the accuracy and efficacy of candidate selection by addressing the aforementioned challenges.
- RAFS aims to provide organizations with a comprehensive, data-driven approach that extends beyond immediate job placement. It seeks to predict candidates' long-term compatibility and their potential for sustained growth, adaptability, and contribution to organizations.
- Through the utilization of intricate graphical representations, including graphs, scatter plots, and line graphs, RAFS aims to provide a high level of granularity and precision in candidate assessments.
- Furthermore, RAFS aims to foster innovation in the field of resume analysis and candidate filtering by serving as a platform for the exploration of cutting-edge predictive modeling techniques and groundbreaking visualization methods. In doing so, it endeavors to stay at the forefront of technological advancements in the realm of talent acquisition.

In summary, the primary objective of RAFS is to provide organizations with a dynamic and adaptable tool for future-oriented candidate selection that aligns with broader talent acquisition goals and leverages the power of graphical prediction techniques.

VI PROPOSED SYSTEM

Introduction

In today's dynamic job market, the recruitment process demands efficiency, precision, and adaptability. Identifying the most suitable candidates from a pool of applicants is a critical challenge faced by organizations across industries. Traditional methods of resume screening are time-consuming, prone to bias, and often fail to leverage the full potential of technological advancements. The proposed Resume Analysis and Filtering System addresses these challenges by introducing a state-of-the-art solution that leverages cutting-edge technologies such as NLP and ML.

Resume Analysis and Filtering System aims to revolutionize the way resumes are evaluated, providing a streamlined and data-driven approach to candidate selection.

- Automate Resume Analysis: Resume Analysis and Filtering System will utilize NLP techniques to extract pertinent information from resumes, including skills, qualifications, experience, and educational background.
- Enhance Prediction Capabilities: The system will implement ML algorithms to predict candidates' suitability for specific positions based on processed text data and market trends.
- Provide Personalized Course Recommendations: Resume Analysis and Filtering System will maintain a database of relevant courses and certifications, offering tailored suggestions based on the extracted resume information.
- Improve User Experience: The system will feature user-friendly interfaces for both administrators and job seekers, ensuring an intuitive and efficient interaction.
- Ensure Data Security and Privacy: Resume Analysis and Filtering System will adhere to strict data protection regulations, safeguarding the confidentiality and security of user information.

Architecture

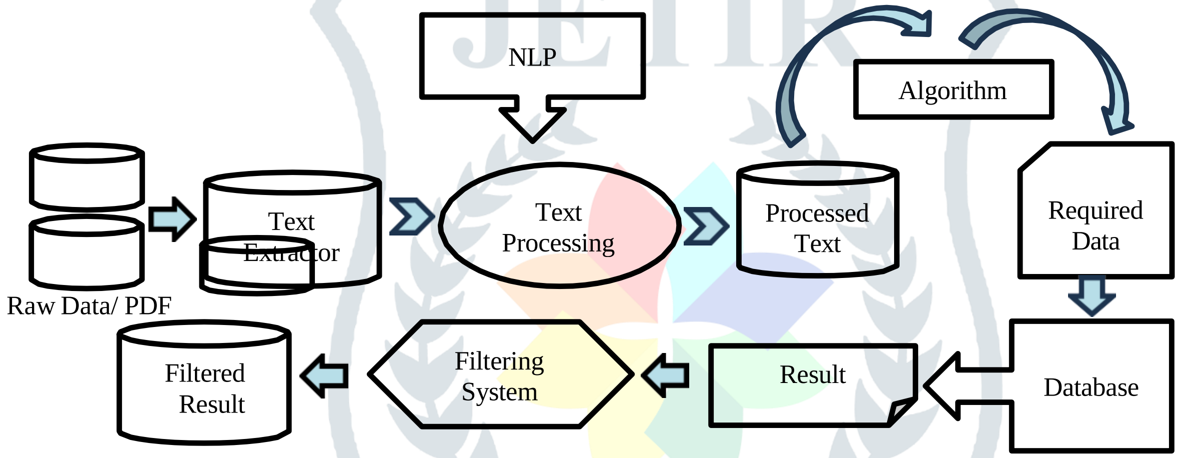

Fig. 7.2 Architecture of Proposed System

From the above figure, we can seek the Architecture of Proposed System. In the initial step the raw data in the form of PDF is been given to the filtering system the text is been extracted from the raw data. The data is been processed and cleaned(filling in the missing values, eliminating the noise, and cleaning of data)by using different NLP techniques we can process the text in the proposed system we have used name entity recognition and py parser the processed text is been passed through a name entity recognition algorithm and we get a required text for filtering purpose

Named Entity Recognition (NER) is a NLP technique that involves identifying and classifying named entities within a text. Named entities are specific objects, people, locations, expressions of times, monetary values, percentages, etc. NER is used to extract structured information from unstructured text and is often used in various NLP applications like information retrieval, question answering, and text summarization. It typically classifies words or phrases into categories such as names of persons (people's names), organizations, locations, dates, and more.

The K-Nearest Neighbors (KNN) algorithm plays a pivotal role in resume analysis, offering a versatile approach to candidate assessment and recommendation. In this context, KNN operates by treating each resume as a data point in a high-dimensional feature space, where attributes such as skills, qualifications, experience, and education serve as dimensions. Upon receiving a new resume, KNN computes the distances between this resume and all others in the dataset, identifying the k nearest neighbors based on predefined similarity metrics. This process enables several key functionalities in resume analysis. Firstly, KNN facilitates candidate classification by assigning labels such as "suitable" or "not suitable" based on the majority class among the k nearest neighbors. Secondly, it aids in similarity assessment, providing insights into the resemblance between resumes and helping identify candidates with comparable profiles. Additionally, KNN powers recommendation systems by suggesting job positions to candidates based on the positions held by their nearest neighbors. By leveraging KNN's ability to capture complex patterns in resume data, recruiters and HR professionals can streamline the recruitment process, identify suitable candidates efficiently, and make informed decisions regarding candidate-job fit.

The required text will be stored in a database for security and for easy retrieval.The result from the data base will pass through a filtering system(filtering system is a mechanism or software designed to sort, categorize, or selectively process data or content based on certain criteria).The filtering system in case of proposed system is used for controlling the flow of information and after filtering we will get the end output or most suitable candidate for a particular job role.

Algorithm/Process Design

Step 1: Input Data Extraction:
Input: Raw data in the form of PDF or text documents. Output: Extracted text from the raw data.

Step 2: Data Preprocessing:
Tasks:Fill missing values.Eliminate noise from the data. Clean the data. Techniques: NLP.
Output: Processed and cleaned text data.

Step 3: Named Entity Recognition (NER):
Task: Identify and classify named entities (e.g., names of persons, organizations, locations) within the processed text. After NER, the extracted named entities along with other relevant features such as experience, education, skills, etc., are used to create a feature vector representation for each resume.
Techniques: NER algorithms like name entity recognition, py parser. Output: Extracted structured information.

Step 4: K-Nearest Neighbors (KNN) Algorithm:
Task: Once the feature vectors are prepared, KNN can be applied for various tasks in resume analysis, such as candidate classification, similarity assessment, or recommendation. KNN calculates the similarity between resumes based on their feature vectors and assigns labels or makes predictions based on the majority class or average of the labels of the k nearest neighbors.
Techniques: Distance Metrics, Efficient Data Structures, Choosing the Value of K Output: Candidate Classification, Similarity assessment, or Recommendation.

Step 5: Store Required Text:
Task: Store the relevant information obtained from NER in a database for security and easy retrieval. Output: Structured data stored in the database.

Step 6: Filtering System:
Task: Sort, categorize, or selectively process data based on specific criteria. Output: Filtered data based on job role requirements.

Step 7: Output Generation:
Task: Select the most suitable candidate for a particular job role based on the filtered data. Output: End result or the most suitable candidate.

Step 8: End of Algorithm.

VII RESULT

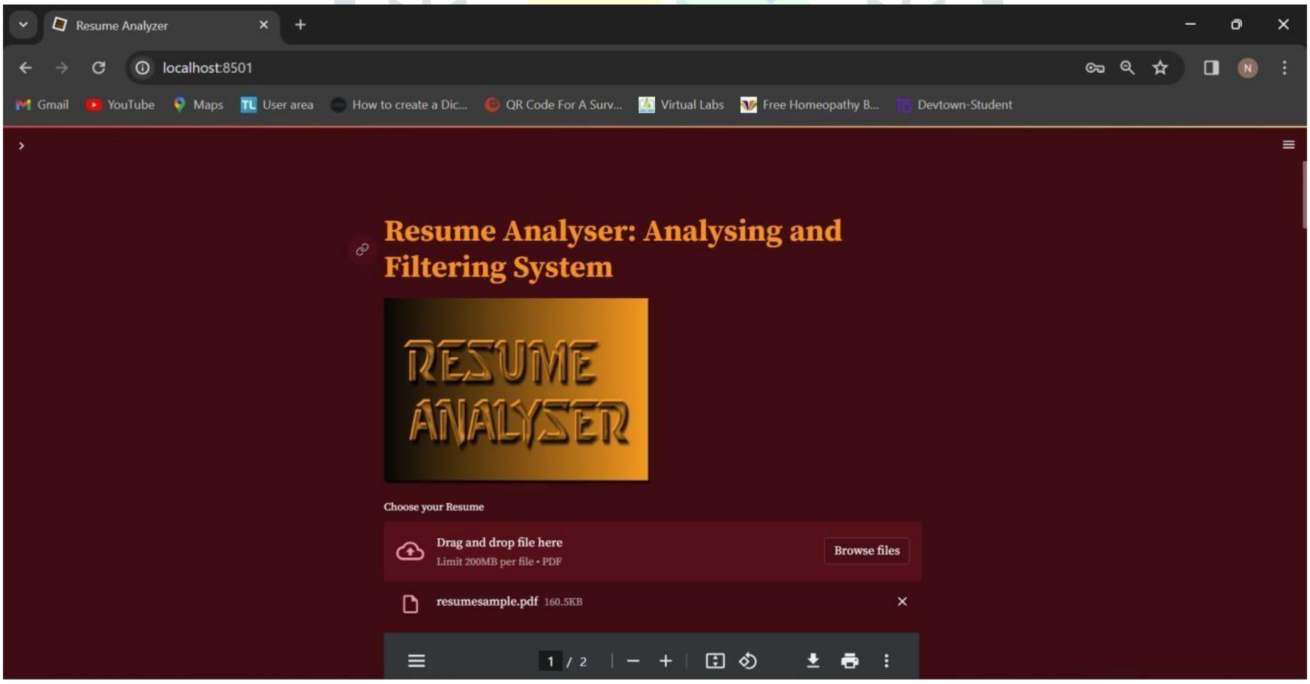

Fig. 8.1 Home Page of RAFS Website.

After running the code, it redirects to local host of 8501 port. It opens the home page of Resume Analyser and Filtering System as seen in Fig. 8.1.

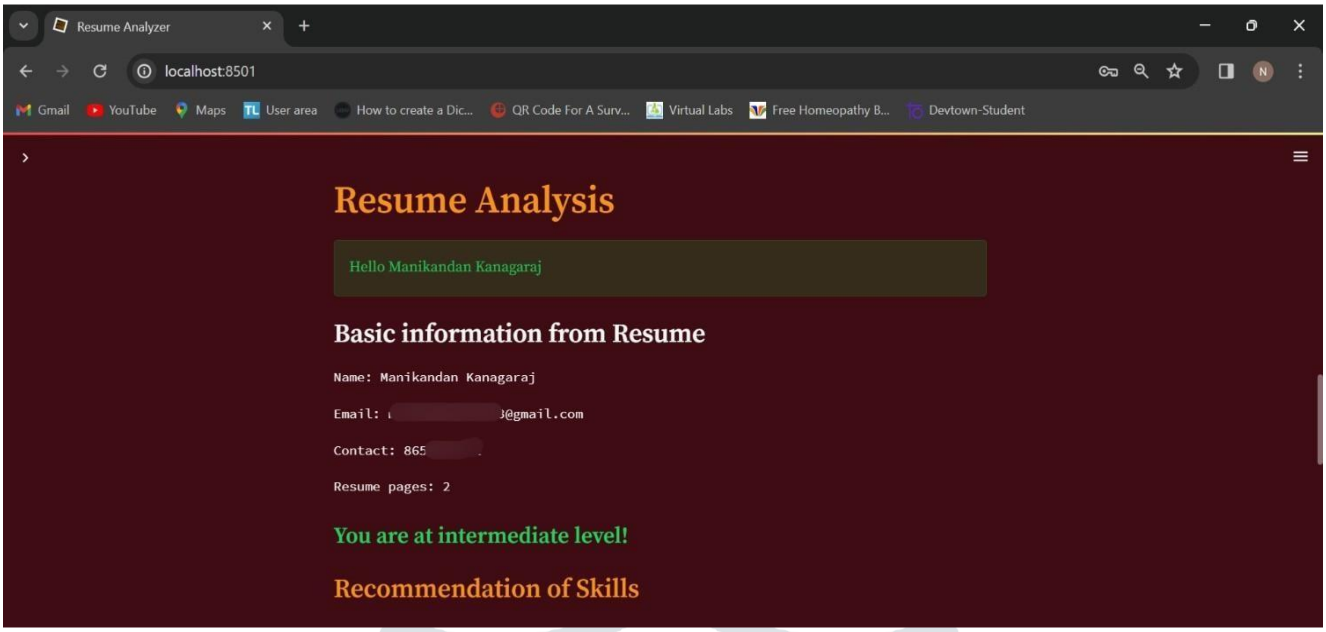

Fig. 8.2 Analysed Result

In Fig. 8.2 the Resume is Analysed and we have got the important data or information of Candidate.

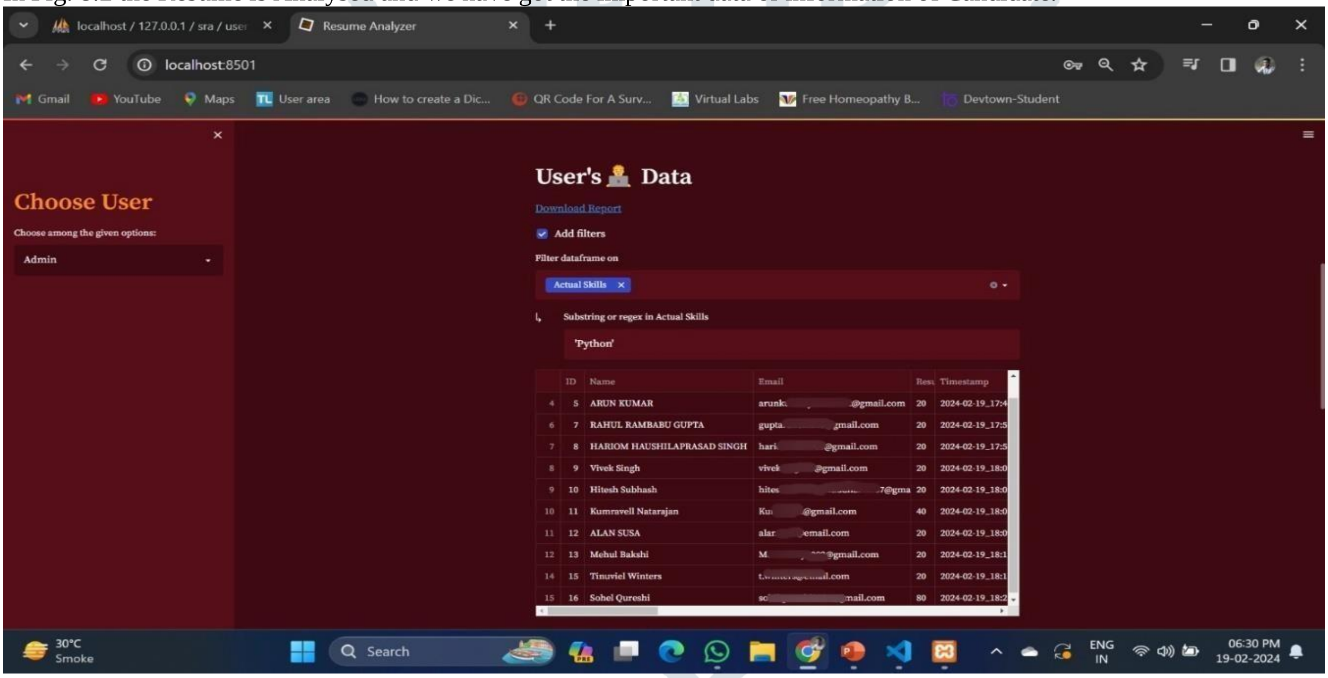

Fig. 8.3. Filtering System

In Fig. 8.3 the Admin can filter the resume to get a desired Resume. One can go through each data and can fetch a Resume easily using the Resume Filtering System.

VIII CONCLUSION

The Resume Analyser: Analysing and Filtering System represents a paradigm shift in modern talent acquisition. It integrates cutting-edge technologies to seamlessly transform raw data from PDFs or text documents into refined, structured information for in-depth analysis. Through a systematic process, missing values are filled, noise is eliminated, and data is meticulously cleaned using NLP techniques.Named Entity Recognition (NER) takes the spotlight, identifying and classifying entities like individuals, organizations, and locations within the text. This yields structured information critical for candidate assessment, securely stored in a database for future access.The Filtering System refines the dataset based on specific criteria, ensuring that candidates align with job role requirements. The system's pinnacle achievement lies in Output Generation, where the most suitable candidate is selected. This outcome is a testament to the system's intelligent interpretation and categorization of data.

In conclusion, the Resume Analyser: Analysing and Filtering System revolutionizes recruitment processes. Its systematic, NLP-driven approach ensures accuracy, efficiency, and objectivity in candidate selection. This empowers organizations to build high-performing teams that precisely match their needs, heralding a new era in talent acquisition.

In future iterations, the Resume Analyser: Analysing and Filtering System aims to elevate talent acquisition processes by integrating a multitude of advancements. Expanding beyond text, the system will embrace multimodal data integration, encompassing audio and video resumes, thereby offering a more holistic view of candidates.Enhanced privacy and security measures will safeguard sensitive candidate information, and user experience optimization will ensure a seamless and intuitive recruitment platform for recruiters, hiring managers, and candidates alike.

References

[1] Jagan Mohan Reddy, Sirisha Regella, Srinivasa Reddy Seelam, "Recruitment Prediction Using Machine Learning" in IEEE, 2020, pp. 1-4.

[2] Gunaseelan B, Supriya Mandal, Rajagopalan V., "Automatic Extraction of Segments from Resumes using Machine Learning", in IEEE, 2020.

[3] Yi-Chi Chou, Han-Yen Yu, " Based on the application of AI technology in resume analysis and job recommendation " in IEEE International Conference on Computational Electromagnetics, 2020, pp. 293-296

[4] Yi-Chi Chou, Chun-Yen Chao, Han-Yen Yu , "A Resume Evaluation System Based on Text Mining", in IEEE, 2019.

[5] Sara Nasr, Oleg Vitoldovicz German, "Assessment of Graduate Students' Resumes Using Short Text Searching Method", in IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering, 2019.

[6] Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar, "Web Application for Screening Resume", in International Conference on Nascent Technologies in Engineering, 2019.