HR Tech Analyst: Automated Resume Parsing and Ranking System through Natural Language Processing

Nisha B
Department of CSE
National Engineering   College
Kovilpatti,Tamilnadu,India
2012056@nec.edu.in

Manobharathi V
Department of CSE
National Engineering College
Kovilpatti,Tamilnadu,India
2 012055@nec.edu.in

Jeyarajanandhini B
Department of CSE
National Engineering   College
Kovilpatti,Tamilnadu,India
2012059@nec.edu.in

Dr.G.Sivakamasundari, Assistant Professor(SG)
Department of CSE
National Engineering   College
Kovilpatti,Tamilnadu,India
gsscse@nec.edu.in

Abstract—The rapid evolution of technology has prompted organizations to adopt innovative solutions to streamline their recruitment processes. It presents an Automated Resume Parsing and Ranking System (ARRS) powered by Natural Language Processing (NLP) techniques. The primary goal of ARRS is to enhance the efficiency and effectiveness of the talent acquisition process. ARRS leverages NLP algorithms to extract valuable information from candidate resumes, such as skills, qualifications, and work experience. These extracted data points are then used to rank candidates based on predefined criteria, effectively eliminating the manual and time-consuming screening process. Moreover, ARRS can be customized to align with specific job requirements and organizational preferences, making it a versatile tool for various industries and roles. In addition to its parsing capabilities, ARRS offers robust candidate ranking functionality. By considering multiple factors such as skill match, experience level, and education, the system provides recruiters and hiring managers with a prioritized list of candidates, allowing them to focus their attention on the most qualified individuals. ARRS is designed with user-friendliness in mind, featuring an intuitive interface that simplifies the recruitment workflow. It integrates seamlessly with existing Applicant Tracking Systems (ATS) and provides real-time insights into the recruitment pipeline, enabling data-driven decision-making.

Keywords—**Natural Language processing, Resume Parsing and Ranking candidates**

I.INTRODUCTION

In the dynamic landscape of today's job market, the influx of resumes has given rise to a bottleneck in the hiring process, necessitating a transformative solution to navigate through this sea of candidate profiles efficiently. The traditional manual screening methods, fraught with inefficiencies and susceptibilities to biases, have proven inadequate in coping with the overwhelming volume of resumes. This not only consumes valuable time but also introduces the risk of overlooking critical information, leading to potential missed opportunities and perpetuating unfair hiring practices. For HR professionals and recruiters, it's too difficult to involve in the arduous task of manually reviewing countless resumes. The time and effort expended in extracting vital information from these documents often translate into delays in identifying suitable candidates. This bottleneck in the hiring pipeline not only hampers the efficiency of recruitment processes but also hinders the timely acquisition of top-notch talent, impacting both employers and job seekers alike. To revolutionize this antiquated paradigm, our proposed solution advocates for the development and implementation of an advanced Natural Language Processing (NLP)-based resume parsing system. This groundbreaking system is envisioned to possess the precision and automation needed to extract crucial details from resumes, including employment history, skills, qualifications, and other pertinent information. Its prowess lies in its ability to comprehend sequential data within resumes, capturing nuanced relationships between various elements of the text. By ushering in a new era of efficiency, the proposed solution seeks to liberate HR professionals and recruiters from the shackles of manual screening. This, in turn, empowers them to redirect their expertise towards higher-level tasks, such as conducting insightful candidate interviews, assessing cultural fit, and strategically driving talent acquisition initiatives. The overarching objective is to enhance the competitiveness of organizations by enabling them to swiftly identify and secure the most qualified and diverse candidates.

Resume Parsing with NLP has the following benefits.
Precision   and   Accuracy: NLP-driven   parsing significantly reduces the risk of human error, ensuring that relevant data is consistently extracted and categorized correctly.

Time and Resource Savings: By automating the initial screening process, recruiters and hiring managers can allocate their time more efficiently to strategic aspects of recruitment, such as candidate interviews and assessments.

Enhanced Candidate Experience: The quicker response times enabled by automated parsing enhance the overall experience for candidates, improving an organization's reputation.

Data-Driven Insights: The structured data generated by NLP parsing allows for in-depth data analysis and reporting, empowering informed hiring decisions.

Adaptability: Customizable parsing rules make it easy to tailor the system to the unique requirements of different industries and job roles.

Exploring further into the document, the intricacies of NLP algorithms and the practical implementation of a Resume Parsing System are delved into.

II. RELATED WORK

Many researchers have initiated research on automated resume parsing systems. The literature survey aims to explore and analyze the existing body of research on resume parsing systems in the field. AmitPimpalkaret.al (2023) proposed the study delves into the nuances of resume analysis, including aspects like writing style, word choice, and grammar. It seeks to identify gaps in existing research and potential future directions in the field of unstructured written communication analysis.It explores various algorithms and methodologies while also considering the subtleties of resume analysis, offering insights into potential advancements in this vital area of recruitment[1].Hira Sajid et al.(2022) ,presents a novel resume parsing framework. This framework begins by extracting raw text from resumes and categorizing them into distinct blocks using text block classification. It then employs named entity recognition and ontology to extract and enrich entities from the resumes. The proposed approach aims to improve the accuracy of information extraction from resumes, ultimately facilitating the identification of the best-suited candidates for specific positions [2].Anushka Sharma et al.(2021) aims in automatically extracting essential information from resumes and presenting it in a structured format. This automation helps recruiters quickly identify the most relevant and qualified candidates, ultimately saving time and effort in the recruitment process.[3] Nithya Jayakumar et. al.(2023)introduces a solution that combines data preprocessing techniques and machine learning algorithms to predict employee attrition effectively. This solution encompasses visual representations of attrition trends, resume information extraction, candidate suitability assessments, and AI-based candidate recommendations.[4] Joonsung Kim et al.(2021) addresses the growing complexity of large-scale Natural Language Processing (NLP) models and the need for scalable and flexible computer infrastructure to support them effectively[5].

Anuruth Lertpiya et al.(2018) addresses a gap in existing Thai Natural Language Processing (NLP) research, which has predominantly focused on well-structured, formally written texts with accurate spellings and clear sentence and word boundaries. However, real-world NLP tasks often involve user-generated web content (UGWC), which is less formal and characterized by imperfect spellings and unstructured text.[6].Akshay Gugnani et al. (2018) proposed that the system relies on a novel approach centered around skills. Skills are seen as a common thread that can be standardized across organizations and industries. The system extracts user profiles from various sources, such as PDFs or publicly available data, using an Open Information Extraction (Open IE) pipeline. It then transforms the extracted education and experience details into a unified skill graph, which captures both spatial and temporal relationships among skills [7]. Ahmad Alsharef et al.(2023) investigates the potential use of text similarity measures as a substitute for human hiring managers in the resume screening process, particularly on online recruitment platforms. Three specific text similarity measures, namely Cosine, Sqrt-Cosine, and Improved Sqrt-Cosine (ISC), were employed as algorithms to evaluate resumes for two distinct job positions: a business development manager and a software engineer[8].Raj Thali et al.(2023) aims to create a comprehensive database of real-time job opportunities by scraping information from various websites. It also collects user information from submitted CVs and uses this data to provide tailored job recommendations. The goal is to simplify the job search process by eliminating the need to search across multiple websites[9].O Pandithurai et.al(2021) discusses the importance of knowledge, particularly in the context of employment and recruitment. It highlights that knowledge comprises both theoretical and practical information gained through experience and learning[10].

V.V.SatyanarayanaTallapragadaet.al(2023) process has undergone a significant shift towards online platforms, with companies relying on automated systems to sift through a large volume of resumes[11].Spoorthi M et al.(2023) proposes the use of an automated system capable of efficiently processing resumes in various formats, such as Word or PDF. This system's primary function is to categorize resumes based on the skills mentioned within them[12].B Gunaseelan et al. (2020)highlights the growing popularity of online recruitment and automated resume processing systems as efficient tools benefiting both employers and job seekers[13].G. Sudha et al. (2021) recognizes the significant role that an individual's personality plays in both their personal development and an organization's progress. It highlights that traditional methods of assessing personality and qualifications for job candidates involve manual review of

resumes[14].M.Alamelu et al.(2021) To simplify and streamline this process, a web application has been proposed. This application employs Natural Language Processing (NLP) techniques to automate the screening of resumes [15]. The proposed system is different from existing systems in terms of candidate ranking and has an accuracy of 99%.

III. METHODOLOGY

Key Components of a Resume Parsing System using NLP are Semantic Understanding, Contextual Analysis, Data Standardization and Customization. In Semantic Understanding, NLP technology dissects resumes to grasp the meaning behind words and phrases. In Contextual Analysis, NLP systems understand the relationships between different sections of a resume, such as linking a specific skill to a particular job experience. Data Standardization formats the extracted information into a structured database, ensuring uniformity and ease of comparison. Then recruiters can customize the parsing rules to prioritize specific qualifications and skills that are crucial for their job openings.

Fig 1.System Architecture of Resume parsing and ranking system

A. PREPROCESSING:

Performing preprocessing on text involves several steps to clean and enhance the quality of the text data. Here's a breakdown of the preprocessing steps typically involved in this phase:

Text Extraction: After uploading a resume (usually in PDF or Word format), the text content must be extracted from the document. The libraries like PyPDF2 or pdf2txt (for PDFs) and python-docx (for Word documents) are used to extract text, images, and other elements from PDF documents.

Lowercasing: It involves converting all characters in the text to lowercase. This ensures consistency in the analysis, preventing the model from treating words with different cases as distinct entities.

Tokenization: Tokenization involves breaking down the text into individual words or tokens. This step is essential for further analysis as it transforms the text into a format that a machine can understand. **Text Cleaning:** Remove any unnecessary formatting, special characters, and symbols that may not be relevant to the analysis, like Removing extra spaces, line breaks, or non-alphanumeric characters. This step ensures that the text is in a clean, standardized format.

Removing stop words: Stop words are common words which are frequently used but may not carry much meaning in certain contexts. It helps to reduce the dimensionality of the data and focuses on more meaningful words.

Text Normalization: Normalize the text to ensure consistent representation of words and phrases. This may include converting text to normalizing dates to a common format, etc.

B. MORPHOLOGICAL ANALYSIS:

The morphological phase typically involves analyzing the internal structure of words and understanding how they are formed by breaking them down into morphemes (the smallest units of meaning). In the context of resume parsing and ranking, the morphological phase is less common compared to other NLP tasks like sentiment analysis or entity recognition. However, it can still be useful in certain situations. Here's how the morphological phase might be applied:

1. Stemming: Stemming is a common morphological analysis technique where words are reduced to their base or root form, like Converting "running," "ran," and "runs" to "run" using stemming. This helps in treating different variations of a word as the same, which can be useful for skill extraction, keyword matching, and ranking.

2. Lemmatization: Lemmatization is a more sophisticated morphological analysis method that not only reduces words to their base form but also considers the word's grammatical context. It provides more accurate results compared to simple stemming. For example, Converting "better" to "good," "running" to "run," and "mice" to "mouse" using lemmatization.

C. FEATURE EXTRACTION

The feature extraction phase involves identifying and extracting relevant information or features from the parsed resumes. These features are used to assess the suitability of candidates for specific job openings and assign ranking scores. Here are key steps and features to extract during this phase:

Named Entity Recognition: It identifies and categorizes specific entities within the text such as name, location, organization and skills.

Keyword Extraction: Identify and extract relevant keywords and phrases that represent the skills, qualifications, and experiences of the candidate. This can involve techniques like TF-IDF, TextRank, or machine learning models trained for keyword extraction.

Section Identification: Determine the structure of the resume by identifying different sections such as contact information, education, work experience, skills, and additional sections like certifications or projects.

Contextual Information: Capture contextual information by analyzing the relationships between words and phrases. Contextual understanding is particularly important for interpreting the meaning of sentences and paragraphs.

Quantitative Information Extraction: Identify and extract quantitative information, such as years of experience, education levels, and other numerical data relevant to the candidate's profile.

Data Structuring: Organize the extracted information into a structured format, such as a table or database, for easy retrieval and analysis.

TABLE1. Key Steps for Feature Extraction

FEATURES TO EXTRACT	KEY STEPS
Skills Extraction	Identify and extract technical and soft skills mentioned in the resume. Skills are essential for ranking candidates based on their qualifications. Techniques like keyword matching, named entity recognition (NER), or custom skill taxonomies are used to recognize skills.
Education Details	Extract information about the candidate's educational background, including the degree earned, institution, graduation date, and GPA (if available). Data structuring is used to organize the information into a structured format for analysis.
Work Experience	Quantitative Information Extraction extract details about the candidate's work experience, including job titles, company names, employment dates, and job descriptions. Contextual Information analyze the relevance of each job to the job opening's requirements.
Certification and Qualification	Identify certifications, licenses, and other qualifications held by the candidate. This information can be valuable for ranking.

D. FILTERING CANDIDATES

The Candidates filtration phase involves recruiters to filter out the candidates based on their expectations.Recruiters may set specific criteria or preferences, such as desired skills, qualifications, or experience, and the system filters and ranks candidates accordingly. This process streamlines the recruitment process by presenting recruiters with a more tailored pool of candidates who closely match their specified requirements.

E. RANKING

The ranking phase involves assigning scores or ranks to each parsed resume based on its relevance to a specific job opening. This phase leverages the extracted features and data from the preprocessing and feature extraction phases.Assign weights to each feature based on its importance to the job opening. Features like skills and qualifications that closely match the job requirements may have higher weights.

F. VISUALISING RESULTS

In this phase,the output or results from the resume parsing and filtering processes are represented graphically. This could involve creating charts, graphs, or other visualizations to present information in a more accessible and interpretable format. Visualizing results can help recruiters or users quickly grasp patterns, trends, or key insights from the data, making it easier to make informed decisions during the hiring process.

G. CONNECT SELECTED CANDIDATES VIA EMAIL

The system likely facilitates communication between recruiters and selected candidates. It may automatically generate or assist in composing emails, allowing recruiters to reach out to potential candidates directly. This phase aims to streamline the communication process, enabling recruiters to efficiently connect with individuals who have been identified as suitable matches during the earlier phases of the resume parsing and filtering system.

Resumes come in various formats, styles, and structures. NLP techniques can handle this variability more effectively rather than other techniques that are widely used in the market. Also NLP is more efficient by adapting to different linguistic styles and expressions. This adaptability is often challenging for rule-based or template-based methods.NLP models can continuously learn and improve over time. As the system processes more resumes, it can refine its understanding and accuracy, ensuring a more effective and adaptive solution in the long run. Also there is no restriction for the resumes to follow a uniform template. This is one of the major advantages of preferring NLP rather than other techniques.

IV. RESULTS AND DISCUSSION

In the proposed System, HR personnel can upload PDF resumes which have been received from the candidates, select attributes (skills, certificates, etc.) from a dropdown list, and input relevant information in text boxes, you will need a combination of web development technologies and backend processing. Applicants can only able to upload their resume. The output of a resume parser and automatic ranking project is a comprehensive and organized compilation of information extracted from resumes submitted by job applicants. This output typically comprises a structured representation of key details from each resume, such as the candidate's name, contact information, work history, education, and skills. This structured data is invaluable for recruiters as it eliminates the need to manually sift through unformatted resumes, saving time and effort.

In a job recruitment system with distinct login roles for applicants and recruiters, the user experience is tailored to the specific needs and objectives of each

group. For applicants, the login role offers a personalized gateway to explore job opportunities and manage their job-seeking journey. This role empowers job seekers by providing a centralized platform to manage their applications and streamline the job-hunting process. On the other hand , Recruiters can access a database of applicant profiles, search for potential candidates using various filters, and review submitted applications. Moreover, this role enables recruiters to schedule interviews, communicate with applicants, and assess candidates for suitability.

The first screen in the app System is the user role selection as per Fig.2. User as an applicant can able to upload their resume as per Fig.3. Once applicants successfully upload their resumes through the designated interface, the system responds by displaying a confirmation message, reassuring users that their submission has been received and acknowledged.

User as a recruiter can enter their username and password as per Fig 4. The recruiters can efficiently manage candidate profiles by filtering resumes based on specific criteria such as skills, areas of interest, projects, and internships.
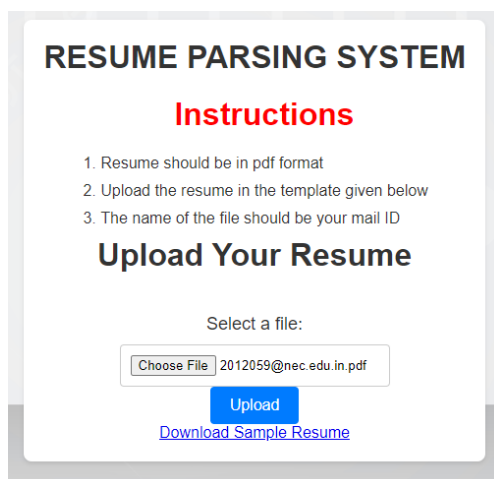
Fig. 2. User Role Selection
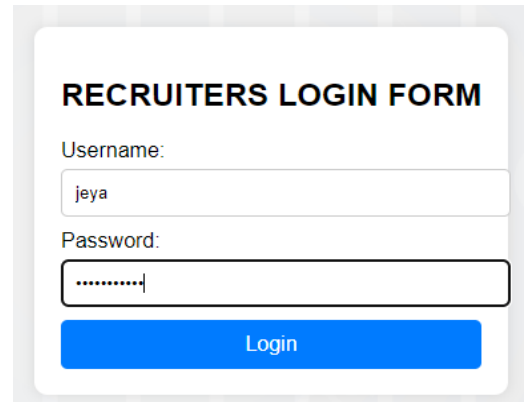

Fig. 3. Applicants uploading resumes

Recruiters can upload a folder of resume and specify desired skills and expertise required for a particular role as per Fig 5, enabling the system to sift

through applicant profiles and highlight individuals whose qualifications align with the job's skill requirements. Additionally, recruiters can filter resumes based on the candidate's declared areas of interest or specialization, ensuring a closer fit between the applicant's career aspirations and the job role.
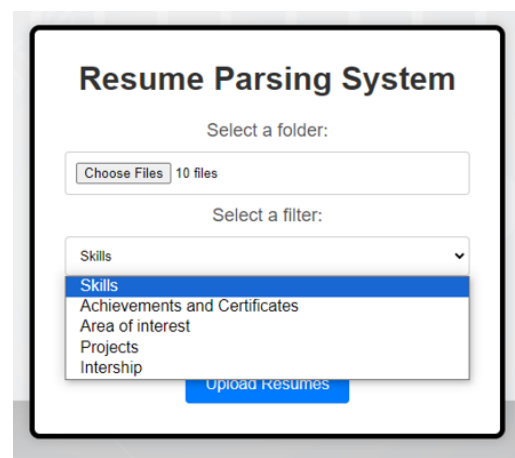

Fig. 4 Recruiters login page

Fig. 5. Recruiters uploading the folder of resumes

Fig.6. Ranking candidates

Candidates are ranked and listed according to a ranking algorithm, which assess various features like qualifications, skills, and experience as per Fig 6. This algorithm assigns scores to candidates, allowing recruiters to easily identify the

most suitable candidates for a specific job position, streamlining the selection process and displays the output as shown in the Fig 7.

Fig.7. Graphical format

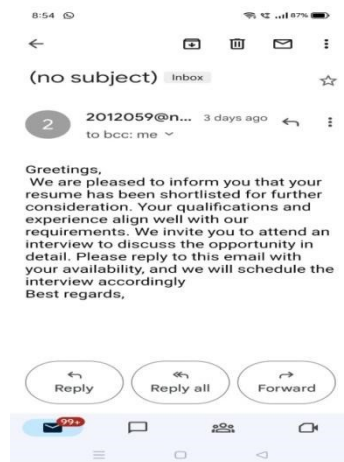Email for the shortlisted candidates will be sent automatically as per Fig .8.

Fig 8 Sending email

V. CONCLUSION

The Resume Parsing and Automated Ranking project represents a significant step forward in streamlining the talent acquisition process. By harnessing the power of natural language processing and machine learning, a robust system has developed that can efficiently analyze and categorize resumes, enabling recruiters to focus their time and attention on the most qualified candidates. The automated ranking feature further enhances the efficiency by identifying the best-fit candidates based on predefined criteria. As a result, this project not only saves valuable time and resources but also ensures that organizations can make well-informed hiring decisions quickly and effectively. With the potential to revolutionize the recruitment industry, our Resume Parsing and Automated Ranking system is poised to bring about substantial improvements in talent acquisition, enhancing both the candidate and employer experience.

Expanding the project to support multiple languages can make it more versatile and inclusive,

catering to a wider range of candidates from diverse backgrounds. Integrating AI chatbots for initial candidate interaction and screening can further automate the early stages of the recruitment process. Providing recruiters with insightful analytics and data visualizations can help them make data-driven decisions, identify trends in their hiring process, and optimize their recruitment strategies.

REFERENCES

[1] Amit Pimpalkar et.al, 2023, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning", IEEE.
[2] Hira Sajid et.al ,2022, "Resume Parsing Framework for E-recruitment", IEEE.
[3] Aushkasharmaet.al, 2021, "Intelligent Recruitment System Using NLP",IEEE.
[4] Nithya Jayakumar et.al, 2023, "On-Demand Job-Based Recruitment For Organisations Using Artificial Intelligence", IEEE.
[5] Anuruth Lertpiya et.al, 2018, "A Preliminary Study on Fundamental Thai NLP Tasks for User-generated Web Content", IEEE.
[6] Akshay Gugnani et.al, 2018," Generating Unified Candidate Skill Graph for Career Path Recommendation",IEEE.
[7]AhmadAlsharef et.al,2020,"FacialEmotionRecognition Using Deep Convolutional Neural Network",IEEE.
[8] Raj Thali et.al, 2021," A Deep Learning Model for Face Expression Detection" IEEE.
[9] V.V. Satyanarayana Tallapragada et.al, 2023, "Improved Resume Parsing based on Contextual Meaning Extraction using BERT", IEEE.
[10]Spoorthi M, 2023,"Automated Resume Classification System Using Ensemble Learning ",IEEE .
[11] B.Gunaseelan, 2021,"Automatic Extraction of Segments from Resumes using Machine Learning", IEEE.
[12] G.Sudha et.al,2021,"Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process", IEEE.
[13] M.Alamelu, 2021, "Resume Validation and Filtration using Natural Language Processing", IEEE.
[14] Vivian lai et.al,2017,"CareerMapper: An automated resume evaluation tool", IEEE.
[15] Bhaliya et.al, 2020,"NLP based Extraction of Relevant Resume using Machine Learning ", IEEE.
[16] Björn Schuller et al. ,2020,"NLP methods for automatic candidate's CV segmentation" ,IEEE.
[17] Saswat Mohanty, 2023,"Resumate: A Prototype to Enhance Recruitment Process with NLP based ResumeParsing",IEEE.