International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

RESUME PARSING USING MACHINE LEARNING AND NLP

Mr.R.Vignesh[1], Ms.C. Chithra Devi[2]
[1]Student (M.Sc.Information Technology) & Rathinam College of Arts and Science.
[2]Assistant Professor Department of Information Technology & Rathinam College of Arts and Science.

ABSTRACT - Machine Learning is a branch of computer science that focuses on using data and algorithms to replicate how humans learn and gradually improving the accuracy of its analysis. Agencies and other high-level organisations must deal with a big number of new job openings that require a variety of resumes. The tool that makes the hiring process easier for both candidates and recruiters. We extract crucial data such as education, experience, and abilities from resumes using Natural Language Processing (NLP) techniques, making them easier to digest and analyse. NLP also allows for the construction of visually appealing and remarkable career materials such as resumes, cover letters, and websites. Based on keyword matching, the programme groups terms into sectors and gives suggestions, predictions, and insights. This considerably minimises the load on candidates and recruiters, resulting in a more efficient hiring process.

Keywords: Resume, CV, NLTK, NLP, KNN

1. INTRODUCTION

Every day, corporations and recruitment firms process a large number of resumes. This is not a job for humans. An automated intelligent system is required to extract all relevant information from unstructured resumes and convert it to a common structured format that can subsequently be ranked for a specific employment post.

The parsed information contains (name, email address, phone number, work experiences, education, hobbies, interests, accomplishments, certifications, projects), keywords, and ultimately the resume cluster (ex: Web Development, Data Science, etc.). The parsed data is subsequently saved in a database (in this example, SQLLite) for further usage. Resumes, unlike other unstructured data (email body, web page contents, etc.), are somewhat structured. Discrete sets of data are used to store information. Each set comprises information about the person's contact information, employment experience, and educational background. Despite this, resumes are tough to understand. This is due to differences in information types, order, writing style, and so on. To successfully and efficiently extract data from various types of resumes, the model must not be dependent on the order or type of data. To address this time-consuming process, we use a solution that makes the process quick, simple, and dependable. It collects keywords from resumes using NLP techniques and uses them for forecasts, recommendations, and analytical analysis.

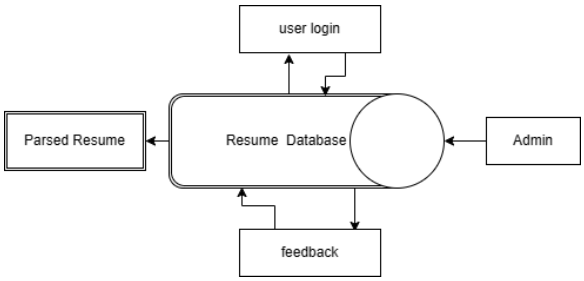


fig: Flow Diagram

2. LITERATURE SURVEY

In recent years, the application of machine learning and natural language processing techniques in the recruitment process has grown in popularity. Many scholars have investigated various methodologies and techniques for automating resume screening and candidate selection. Here are a few research that are relevant:

1. **Ayman Abu Samra, Mohammad Kuhail, and Hossam Faris (2020), "Automated Resume Screening System Using Artificial Intelligence"** The authors present an automated resume screening system that extracts and analyses relevant information from resumes using machine learning algorithms and natural language processing techniques. To match the candidate's abilities and qualifications with the job criteria, the system employs a keyword-based method.
2. **Linying Dong, Zhiyun Ren, and Yanjun Ma (2020), "Resume Parsing: A Survey"** The authors conduct a literature study on resume parsing and present an overview of various strategies and approaches employed in the field. They examine the challenges and limitations of resume parsing and make recommendations for further research.
3. **Shreya Rani and Dr. Sanjeev Sofat's "Resume Screening Using Natural Language Processing Techniques" (2021).** The authors present a resume screening system that extracts relevant information from resumes and matches it with job criteria using natural language processing techniques. They employ the TF-IDF algorithm to determine the keywords and their significance in the resume.

4. **"Automated Resume Screening and Shortlisting System Using Ma- chine Learning Techniques" by Rajendra Patil and Sunil Pise (2021)** The authors present an automated resume screening and shortlisting system that classifies resumes based on their relevance to job requirements using machine learning techniques. To represent the sample data, they employ several feature extraction approaches such as Bag-of-Words, Term Frequency-Inverse Document Frequency (TF-IDF), and Word Embeddings.

According to the literature, using machine learning and natural language processing techniques can considerably increase the efficiency and accuracy of the hiring process. However, several obstacles and limits remain, including the accuracy of keyword matching, the management of unstructured data, and the ethical implications of automated decision-making in recruitment.

3. EXISTING SYSTEM

The manual handling of resumes by hiring managers throughout the current recruitment process makes it time-consuming and error-prone. It might be challenging to compare and accurately evaluate candidates because resumes frequently come in a variety of forms and styles. It gets more difficult to analyse each CV as the volume of job applications rises, which could lead to the recruiting process taking longer or leaving out good applicants. Additionally, the existing method might not offer information beyond what is stated openly in the resume, and it might be ineffective to match qualified candidates with appropriate job ads. Overall, there are gaps in the system's ability to assess applications and pair candidates with appropriate job positions.

3.1 DRAWBACKS OF EXISTING SYSTEM

The present method of managing resumes typically relies on hiring managers to review resumes manually, which can be a laborious and error-prone procedure. It might be challenging to compare and accurately evaluate candidates because resumes frequently come in a variety of forms and styles. It gets harder to manually analyse each CV as the number of job applications rises, which could lead to the hiring process taking longer or missing qualified applicants. Furthermore, beyond what is expressly stated in the CV, the existing method might not offer insights into a candidate's abilities, experiences, or other pertinent information. Due to a lack of effective matching algorithms, there may also be inefficiencies in the matching of competent candidates with appropriate job advertisements.

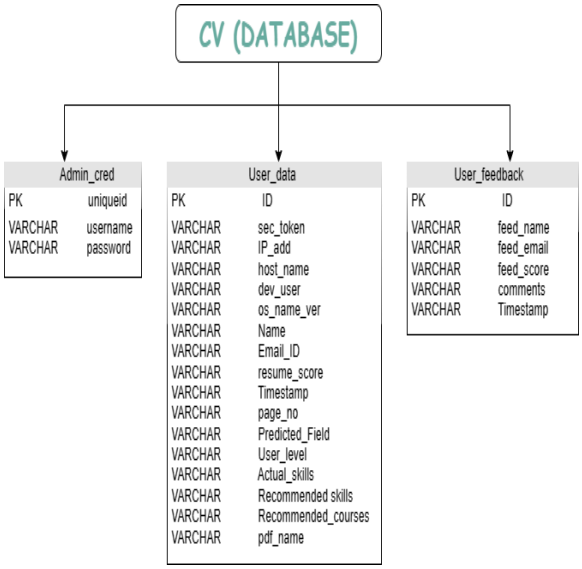


Fig: Database Design

4. PROPOSED SYSTEM

Both job seekers and recruiters appear to find the suggested system's "Resume parser" to be a useful tool for analysing and gleaning information from resumes. Recruiters can save time and effort by using natural language processing to comprehend the resume and then extract information from it. Additionally, the system is quite useful for candidates because it makes forecasts, hints, and recommendations based on the data in their resumes. It's crucial to remember that the system functions best when a resume is uploaded in traditional chronological order. Additionally, the system's robust analytics can offer the administrator or recruiter insightful data. Overall, it appears that the proposed approach could enhance the recruitment process by making it more effective

4.1 ADVANTAGES OF PROPOSED SYSTEM

In comparison to the present manual method, the proposed solution, Resume Parser, has a number of benefits. It tracks and examines resumes based on job roles, making it simpler for recruiters to find the candidate who is the best fit for a certain position. Recruiters may make prompt decisions based on the data offered by the system because the system also offers quick, secure, and in-the-moment predictions. Additionally, the technology can quickly extract information from a resume, which cuts down on the time and effort needed for manual resume screening.

Additionally, the method makes it simpler for recruiters to compare the information of various candidates by collecting minute data from resumes, such as education, experience, and talents. A beneficial and precise tool for

both candidates and recruiters, the system also offers recommendations, tips, and projections depending on the data supplied by the applicants. The recruitment process might be considerably enhanced by the proposed approach by making it more effective, accurate, and efficient overall.

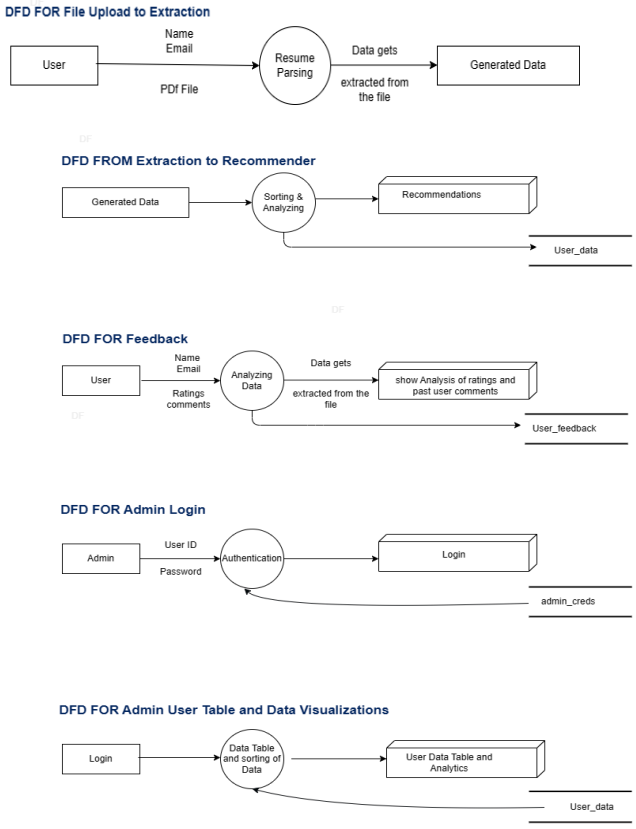


Fig: Dataflow Diagram

5. METHODOLOGY

The methodology for developing the resume parsing tool using natural language processing can be broken down into the following steps:

.

1. **Data collection**: gathering a sizable dataset of resumes in various formats, including PDF, Word, and Text. Resumes from various fields and businesses should be included in the dataset.

2. **Data pre processing:** Cleaning and normalising the gathered dataset to make sure the data is in a dependable and consistent manner. This entails cleaning out any extraneous details like graphics and formatting and transforming the data into a common text format.

3. **NLP (Natural Language Processing:** Extracting pertinent information from the resume, such as the applicant's name, contact information, educational background, talents, and experience, using NLP techniques like Named Entity Recognition (NER), Part of Speech (POS) tagging, and Dependency Parsing.

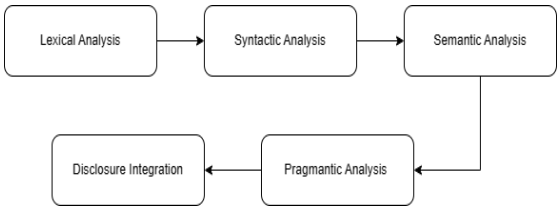


Fig: Phases of NLP

4. **Keyword extraction and clustering:** Using methods like TF-IDF (Term Frequency-Inverse Document Frequency), extracting keywords from the resume, and grouping them into pertinent industries depending on how similar they are.

5. **Recommendation and prediction**: Based on the extracted data and keyword clustering, recommendations and predictions are made using machine learning techniques like Decision Trees, Random Forest, and Naive Bayes.

6. **User interface design:** Creating an easy-to-use interface that enables users to post resumes and read suggestions and parsed information.
7. **Testing and evaluation:** To assess the tool's performance in terms of accuracy and efficiency, testing it on a sample of resumes is necessary. increasing the tool's overall performance by modifying it iteratively in response to user feedback.

8. **Deployment:** Setting up the tool in a live environment and offering users assistance if there are any faults or problems.

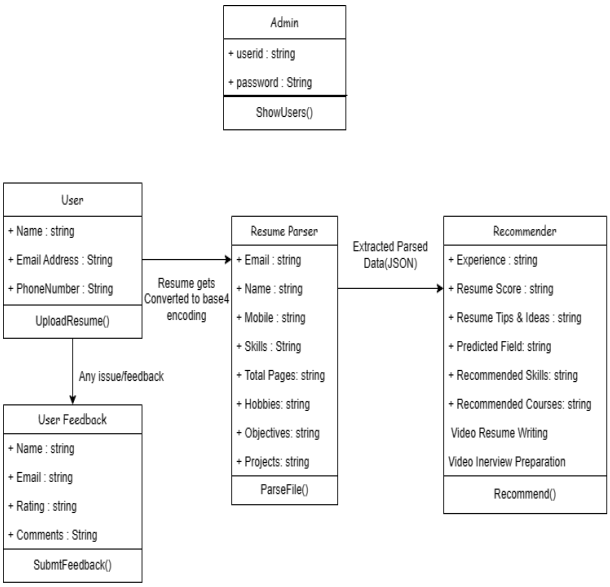


Fig: Class Diagram

6. MODULE DESCRIPTION

• **Resume Parsing**: Using a structured framework, the programme will extract the relevant data from resumes. This will simplify the comparison and analysis of resumes.

```
from pyresparser import ResumeParser
data = ResumeParser('/path/to/resume/file').get_extracted_data()
```

• **Keyword Clustering**: Based on their keywords, the extracted data will be grouped into sectors. This will make it simpler to determine which candidates are most suitable for a given position.

User-Friendly Interface: Both recruiters and candidates will find the tool easy to use. It will have a simple and intuitive user interface that will make it simple to use and navigate.

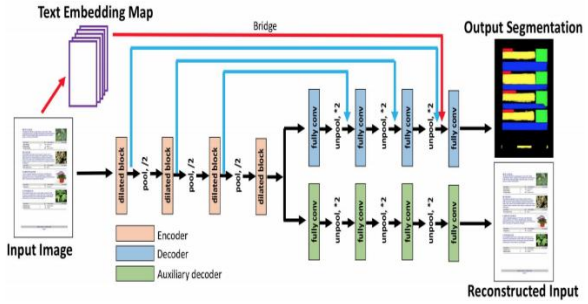
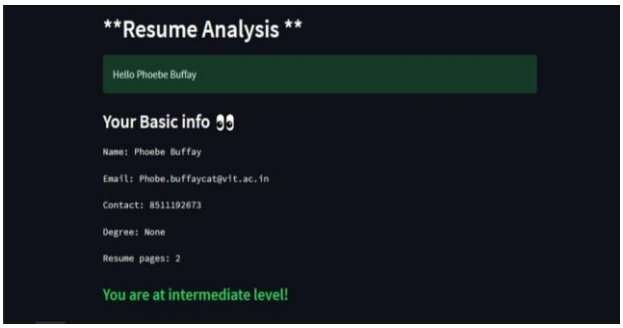


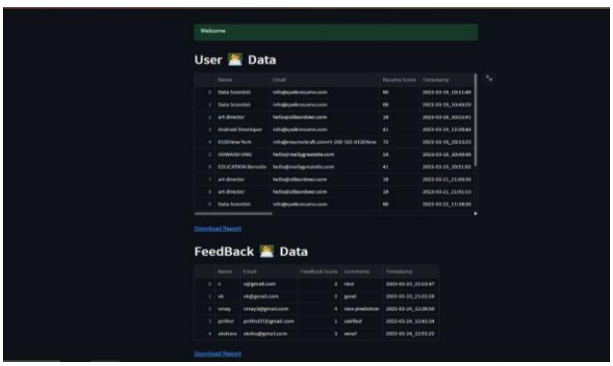
Fig: CNN algorithm extraction

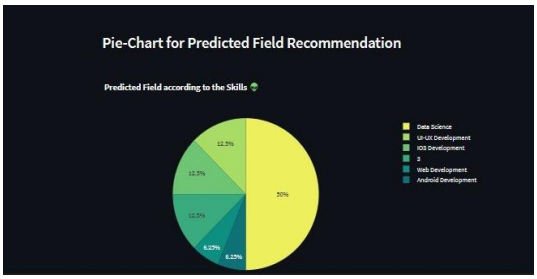
• **Recommendations and Predictions:** Based on keyword matching, the programme will offer recommendations and predictions. This will make it simpler for hiring managers to find the right person for a position and for job seekers to polish their resumes.

• **Analytics and Insights:** The tool will offer analytics and insights on the roles that are in demand, the most sought-after skills and qualifications, and other critical insights that can be leveraged to enhance the hiring process.

7. RESULTS







8. CONCLUSION

This tool is intended to process resumes, extract pertinent data, and produce output that can be read by machines. As a result, candidates can improve their resumes with the help of recommendations and analytics, and recruiters can more effectively organise and rank prospective candidates according to job requirements.

To extract specific information from resumes, such as education, experience, abilities, and qualifications, the application uses Natural Language Processing (NLP) approaches. The retrieved data is then utilised to group resumes according to their pertinent keywords and to determine which applicant is the greatest fit for a given employment role.

Utilising the Resume Parser enables a data-driven hiring method, which is one of the main advantages. It enables recruiters to make better hiring decisions by analysing resumes and offering insights on job seekers' qualifications, abilities, and experience. Additionally, as recruiters must manage a high amount of resumes during the hiring process, the tool's capacity to discover pertinent keywords and group resumes based on them helps to lighten their labour.

8.1 FUTURE WORKS

The resume parser tool's future work will involve integrating it with other recruiting tools, continuously enhancing the algorithms for natural language processing, expanding multilingual support, enhancing the algorithms for matching, and enhancing the data visualisation for better understanding of the recommendations and analytics.

REFERENCES

[1] Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal “An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes”, 2015, IJIRAE.

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

[3] Chen, J., Gao, . L. and Tang, Z., 2016. Information Extraction fromResume Documents in PDF. researchgate.

[4] A Machine Learning approach for automation of Resume Recommendation systems. Pradeep Kumar Roya, Sarabjeet Singh Chowdharyb, Rocky Bhatia, 2020
.

[5] Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S,Kaveri BW, “E-Recruitment System Through Resume Parsing,Psychometric Test and Social Media Analysis”, 2019,IJARBEST

[6] M. J. Cafarella, D. Downey, S. Soderland, and O. Etzioni,“KnowItNow: Fast, scalable information extraction from the web,”in Conference on Human Language Technologies (HLT/EMNLP), 2005
.

[7] S.Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data. Morgan-Kauffman, 2002
.

[8] Ujjal Marjit, Kumar Sharma and Utpal Biswas, “ Discovering Resume Information Using Linked Data” , in International Journal of Web & Semantic Technology, Vol.3, No.2, April 2012.

[9] M. J. Cafarella and O. Etzioni, “A search engine for natural language applications,” in WWW, pp. 442–452, 2005.
.

[10]https://www.analyticsvidhya.com/blog/2021/06/resume-screening-with-natural- language-processing-in-python