Resume Analyzer Using Natural Language Processing

Snigdha RP[1] Sibil Arcokian G[2] Rokhinth PB Sarron[3] Ms. M S Bennet Prabha[4]
[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4]SRM Institute of Science and Technology, Ramapuram, Chennai, India

Abstract — Traditional hiring techniques are losing effectiveness as internet hiring grows more and more. It is not easy to manually filter out the resumes cause it would take a lot of time and resources, which the employing organizations cannot bear. Individuals with various specializations and sectors of experience submit a sizable number of unstructured resumes to job portals in a variety of styles and forms. Thus, to effectively channel candidates to their relevant occupational groups as well as facilitate the automatic screening of candidates, structured information must be extracted from application resumes. It's also unfair that many qualified candidates don't get the attention they deserve during the resume screening process. This might result in the hiring of incompetent people or the rejection of competent applicants. We present a method to address these problems by automatically recommending the most qualified job candidates in accordance with the provided job description. Our solution employs NLP to pull relevant information from the unstructured resumes, such as skills, education, and experience, and then summarizes each application. Using the spaCy library in python and the modules such as Entity Recognition, Dependency parsing & LDA (Latent Dirichlet Allocation) for topic modelling.

Keywords: NLP (Natural Language Processing), spaCy, Entity Recognition, Dependency parsing, LDA (Latent Dirichlet Allocation

I. INTRODUCTION

The swift growth of internet connectivity has caused all major firms to modify their employment practices. By advertising job openings online on various job portals and websites, recruiters can connect with a wide pool of potential candidates. Even if e-recruitment has made hiring easier and more affordable for both employers and candidates, there are now a number of new challenges. This efficiency. During the first phase, all relevant candidate information, including certificates, years of education, work experience, and talents, is extracted from the resumes using unstructured language. Using the modules, we rate the resumes in the second stage of our algorithm based on how closely their content matches the job description that has been provided. Utilizing the power of this cutting-edge NLP tool, resume analysis with spaCy automates and expedites the resume screening and assessment process. Because of SpaCy's well-known precision and effectiveness in processing human language, it is a great option for parsing and comprehending the content of resumes and CVs. By incorporating spaCy into the resume analysis procedure, companies can thoroughly and swiftly retrieve essential data from applicants' resumes. This includes having a deep understanding of the language being used, including sentiment, tone, and context, in addition to having the necessary training, experience, and credentials. This project consists of multiple steps to increase its efficiency. During the first phase, all relevant candidate information, including certifications, years of education, work experience, and skills, is extracted from the resumes using unstructured text.

II. LITERATURE SURVEY

In Kondapalli Sai Pranay's work[3] The study methodology screens resumes and compares them to job descriptions using NLP and machine learning.

Anindya Sarkar and Debajyoti Mukhopadhyay's paper, "Automatic Resume Filtering Using Machine Learning,"[4]. The algorithm described in this paper screens resumes using machine learning techniques and assigns a ranking based on how closely each resume matches the job description.

Shweta Agrawal and Sumit Gupta's paper 2019.The idea is that a machine learning and natural language processing system scans resumes and assigns a score based on how well they match the job description is described in the study.

The goal of the career centre is to provide organizational support for students' and graduates' job placements. To support all major activities, the information system was developed by the authors in paper [10]. By acting as a repository for resumes and job openings, the system improves ties between universities and businesses. Conversely, the system ought to function as an electronic recruiter, considering students' individual skills and preferences, open positions, company profiles, job specifications, and available workforce to facilitate informed hiring decisions. explains the text mining- based intelligent management system that supports recruitment services.

A candidate needs to have a strong resume in this cutthroat industry that highlights the necessary details in a way that makes it stand out from other resumes. Numerous semi-structured and unstructured resumes can be found in the company's database. Parsed resumes are required by these companies for the hiring process. Employing a text analytics approach, or keyword search, allows recruiters to find the best candidate for a position by qualitatively evaluating resumes [11].

III. METHODOLOGY

In this paper, it is suggested to develop a resume analyzer using NLP. The majority of current systems use a single methodology to generate the output, but in this paper, we employ three methodologies to improve the efficiency and accuracy of the resume analyzer.

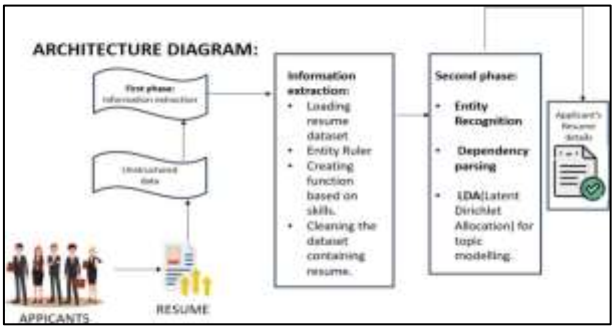


Fig. 1: Proposed Methodology Architecture

Figure 1 represents the architecture diagram of the project. First, candidates send their resumes to the relevant company. After that, the resume goes through a number of processes with the resume analyzer before being shortlisted. The candidate's resume will go through multiple processes in the initial step of information extraction, including loading datasets, entity ruler, creating function based on skills, cleaning the dataset containing resume. Secondly, it passes through three methodology such as Entity recognition, dependency parsing and final LDA to get the final results. The applicants resume details will be displayed at the end.

A. Information Extraction

Relevant data from unstructured resumes is automatically extracted and categorized as part of information extraction for resume screening using NLP. Natural language processing algorithms are trained to identify and extract important entities, including names, contact details, educational backgrounds, professional experiences, certifications, and more. After that, these entities are divided into pre-established categories such as "Education," "Work Experience," "Skills," and "Contact Information," making it possible to extract structured data from the resume documents. This procedure expedites the first screening stage and makes it easier to match a candidate's qualifications with the demands of the position, resulting in a more thorough and impartial evaluation of candidates. Furthermore, NLP lessens the possibility of human bias in the screening process by automating information extraction, which helps to create a hiring process that is more equitable and inclusive.

B. Entity Recognition

When NLP is used for resume screening, the term "entity recognition" describes the technology's capacity to recognize and extract particular entities or bits of data, like names, contact information, skills, work experience, and more, from resumes. By automatically identifying and classifying these entities from resume text, natural language processing (NLP) algorithms facilitate the organization and evaluation of candidate data. This procedure helps recruiters quickly identify the most qualified candidates by streamlining the initial resume screening and assisting in the matching of candidates with job requirements. NLP enhances the efficiency, consistency, and accuracy of the resume screening process by automating this information extraction, which ultimately helps hiring decisions.

C. Dependency Parsing

When employing natural language processing (NLP) for resume screening, dependency parsing entails examining the text's grammatical structure to ascertain the connections between words and phrases on a resume. This method assists in determining the connections between various sections of the document, such as the relationships between job titles and particular companies or the relationships between skills and job descriptions. Dependency parsing facilitates a more accurate extraction of pertinent information by helping to comprehend the context and semantics of the content. NLP systems can increase resume screening accuracy by identifying these linguistic dependencies, enabling recruiters to more quickly and efficiently evaluate a candidate's qualifications and match them to job requirements.

D. LDA (Latent Dirichlet Allocation)

Latent Dirichlet Allocation (LDA) is a topic modelling technique used in resume screening with natural language processing (NLP) that helps identify hidden themes or topics in a set of resumes. LDA groups words in resumes into topics by assigning them probabilities. This is useful for classifying resumes according to their content and pinpointing essential areas of competence or credentials. Recruiters may more easily filter and evaluate candidates based on their qualifications and experiences by using LDA's ability to automatically recognize and tag resumes with pertinent topics. Even with a high volume of resumes, it improves the effectiveness of resume screening by offering a methodical approach to find applicants who closely match job requirements.

E. Mathematical Analysis

The Mathematical analysis of the LDA methodology is expressed below in respect to the various terms and topics calculations:

1) Dataset Split:

N = Total number of resumes (200) Training Data=0.7N resumes Testing Data=0.3N resumes

2) Topic Modelling and LDA:

k = Number of topics (e.g., =5)

3) Model Training:

α = Dirichlet parameter for document-topic density

β = Dirichlet parameter for topic-word density

Max Steps=4000 (maximum number of iterations during training)

Evaluation Frequency=500 (evaluate the model every 500 steps)

4) Evaluation:

Testing Data=0.3N resumes

Coherence Score: CS

Perplexity: P

5) Coherence Score:

CS is a function of the set of high-scoring words in the topics.

6) Perplexity:

P is a function that measures how well the model predicts the test set.

7) Topic Distribution Analysis:

Topic Distribution=[1,2,...,]Topic Distribution=[p1 ,p2,...,pk] where pi is the proportion of resumes assigned to topic i.

8) Optimization:

Adjust parameters α, β, k, Max Steps and Evaluation Frequency to optimize CS and P.

IV. Results

The goal of the project is to make resume analyzing an easier process both for the companies as well as the applicants. On comparison with other existing models are using lesser amount of testing data and are able to only identify individual entities. In that sense, our model is using 70% training data & 30% testing data from the dataset which contained 200 resumes. It is also able to identify multiple entities & relationships among them.

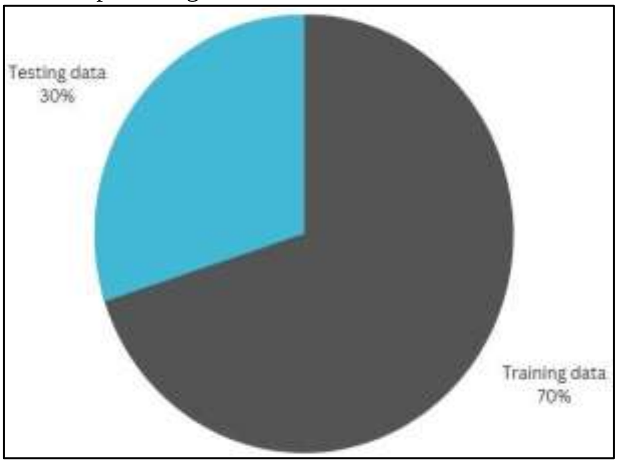
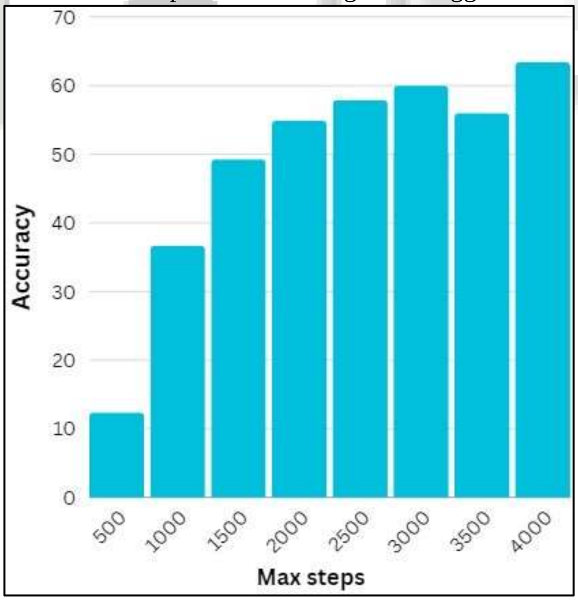


Fig. 1: Training and testing data ratio

A. *Performance Metrics*

According to the used dataset and the trained model an accuracy of 63.40% was achieved on various max steps ranging from 0-4000. With an Evaluation Frequency of 500. It can be further improved on having much bigger datasets.



Graph 1: Performance graph

V. Conclusion

In conclusion by using Natural language processing techniques, the NLP resume Analyzer enhances efficiency and reduces human bias, ensuring a fairer and more objective selection of candidates. By automating the initial stages of candidate evaluation, NLP saves valuable time and resources. Therefore for further improvements of the model much bigger datasets are required, moving on a front-end, back-end etc. can be developed with the trained model being in the background so that the use case of the model can also be increased.

References

[1] Automated Resume Screening Using Natural Language Processing Dr. D. Lakshmi Padmaja1 , Ch. Vishnuvardhan2 , G. Rajeev3 , K. Nitish Sanjeev Kumar4, Journal of emerging technologies and innovative research(JETIR) (March 2023)

[2] Resume Screening Using Machine Learning and NLP : A Proposed System Bhushan Kinge*1, Shrinivas Mandhare2, Pranali Chavan3, S. M. Chaware4, International journal of scientific research in computer science, ngineering and information technology, (April 2022)

[3] Kondapalli Sai Pranay, Resume Screening using Natural Language Processing and Machine Learning, the International Journal of Current Technology and Engineering in 2020

[4] Anindya Sarkar and Debajyoti Mukhopadhyay's, Automatic Resume Filtering Using Machine Learning, the International Journal of Engineering and Technology in 2016.

[5] Shweta Agrawal and Sumit Gupta's, Automated Resume Screening System Using Machine Learning and Natural Language Processing, International Journal of Innovative Technology and Exploratory Engineering in 2019.

[6] Vansh Nawander1, Shrenita Elma2, Andhoju Karthikeya3, Manuka Koushik Yadav4, Sai Karthik Kotala5, M. D. N. Akash6, Modern Resume Analyser For Students And Organisations, International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 10 | Oct 2022

[7] Shreelakshmi CS, Ameena Kausar, Ananya M, Anisha George, Darshan KR, Resume screening rating and recommending skills using NLP, Volume:05/Issue:06/June-2023.

[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] Bhaliya, Nirali, Jay Gandhi, and Yong Luo. "NLP based Extraction of Relevant Resume using Machine Learning."

[10] Yang Y. and Pedersen J. P. "A comparative study on feature selection in text categorization". ICML-97.

[11] Vinaya Ramesh Kudtarkar, Manjula Ramannavar, Dr. Nanadini S.Sidnal, "A Survey on Unstructured Text Analytics Approaches for Qualitative Evaluation of Resumes" International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 14 Issue 2 – APRIL 2015.

