

Analysis and Ranking Resume using Machine Learning Algorithms and NLP

Pratik Chandrasen Deshmukh
Dept. of Data Science Engineering,
Pravara Rural Engineering College,
Loni, India.
viratpratik@gmail.com

Dr. M.R.Bendre
Dept. of Computer Engineering
Pravara Rural Engineering College,
Loni, India.
mininath.bendre@pravara.in

Abstract—

Specialized recruiting firms have emerged as a result of the expansion of the Indian employment market. These agencies use machine learning models to expedite the hiring process and provide their clients with the best candidates. Thus, the suggested methodology offers the potential to improve the hiring process by incorporating the advantages of machine learning in addition to natural language processing(NLP). Employing NLP techniques, one may accurately grade resumes according to the company's criteria and extract useful information from them. After thorough testing and improvement, the web portal might prove to be an invaluable tool for hiring managers and job seekers alike.

Furthermore, one might think about adding features to the CV classifier and ranker, such as customized feedback and recommendations for enhancements that might assist candidates in understanding where they might be lacking and increase their chances of getting recruited.

Keywords: Resume Ranking, Rating, ML, NLP, Resumes, Parser, Extract Information

I. INTRODUCTION

The current business hiring procedure has been changing significantly and is becoming more digitally advanced these days. Candidates no longer feel the need to send their resumes in physical versions; instead, recruitment teams prefer it via electronic version that can be viewed online. The current approach requires candidates to manually fill out multiple lengthy forms with all of the information related to their resume. Additionally, candidates often express dissatisfaction with the position that the pre-existing algorithm assign to them based only on skill sets with no recourse or any information to help them understand their shortfalls compared to other candidates.

The issue raised regarding candidates' discontent with their job placements is also a prevalent one throughout the hiring process. This can be the result of an improper fit between the candidate's interests and talents and the work requirements. An intelligent system for hiring that utilizes algorithms featuring machine learning can be quite helpful in this situation. The technology has the ability to evaluate a candidate's abilities, work history, hobbies, and resume, then match those attributes to the job specifications and skill set requirements. By doing this, the employer may make sure that the candidate is assigned to a position that best suits their interests and skill set, increasing job satisfaction and retention rates. Furthermore, by automatically eliminating applications that aren't relevant and shortlisting the best suiting applicants for the position, the technology can assist

businesses in hiring less personnel for menial tasks and saving the organization's time and utilize money efficiently.

II. LITERATURE SURVEY

A) Natural Language Processing Resume Parser-

To sieve out the top candidates from a pool, the study suggests using a variety of natural language processing (NLP) approaches, including trigram, bigram, n-gram, and text classification. For classification, the model makes use of machine learning algorithms. One benefit emphasized in the study is the conversion of the corpus of text into a word list. Using concordance, word forms are matched to their lemmas in this procedure. The researchers' intention in turning the text corpus into a word list was probably to make it easier to process and analyse the textual material in the future and also maintain it for further use.

B) Utilizing Cosine Similarity in the Resume Recommendation System-

This research study suggests utilizing natural language processing to extract data from the unstructured nature of resumes. Every resume is condensed into a form that only includes the elements relevant to the selection process; the suggested model operates in turns/phases based on categories. Resume categorization and job index closeness display. The cosine similar is the method utilized to compute the similarity index. The candidate's suitability for the duty profile is shown by the similarity index.

C) Creating a resume ranking algorithm based on machine learning and its design-

Utilizing a system with natural language processing (NLP) techniques to extract the necessary skills and competencies from the resume after it has been submitted. The system can efficiently comprehend and handle textual content thanks to NLP. TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is one method that is discussed. It converts words into numerical vectors so that the machine can understand and analyse them. The resume that closely matching the job description supplied by the recruiter is found using the KNN algorithm. The average parsing accuracy of the system is 85%.

D) Machine Learning-Based NLP-based Extraction of Relevant Resumes-

A number of item programs have been added in this research paper to cope with such large databases. The study describes a CV parser with multiple language support, customizable options, and semantic mapping for the training, expertise, and skills parts. It also integrates with job boards. The system seeks to extract information from resumes accurately by utilizing CV parsing with a lease limit. The creation of the CV parser, which makes resume extraction based on format and type easier, is mentioned in the study. The system makes use of users' API keys to facilitate integration with additional applications.

E) Automating the Resume Recommendation System using Machine Learning-

The suggested remedy is predicated on machine learning. There are two primary steps in the solution: deployment/inference and preparation. Data is gathered from Kaggle and web platforms during the preparatory phase. The information is organized into three columns in Excel format: ID, Category, and Resume. Each resume's sequence number is shown by the ID column, the sector to which the resume belongs is indicated in the Category column, and the candidate's full CV is contained in the Resume column. A comprehension of the research as a whole may require taking into account the model how they perform, evaluation, and machine learning approaches used.

F) An automated method for screening resumes that makes use of similarity and processing of natural languages-

The proposed system functions in two ways. First, it takes advantage of natural language processing to glean over pertinent data from the resumes' diverse and unstructured formats. Each resume is condensed into a version that only includes the information necessary for the hiring process. The screening officials' job is made easier with all the unnecessary material eliminated and their job is automated, allowing them to examine each résumé more quickly and effectively.

Second, the system gives the applicants a way to be ranked. It makes use of a similarity computation and Vector Space Model in a content-based recommendation system. The system compares the features of the extracted resume with requirements for the job description. The method provides a sorted list of top-N applicants and recommends which candidate best fits the particular job vacancy by assigning each resume a similarity score.

III. METHODOLOGY

According to the job description, the parsing of resume is done 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 hope to minimize unfair and discriminatory practices and facilitate the recruitment process by receiving high-quality applications from a variety of geographic

locations.

The modules that make up the system architecture are-

1. Resume Parser.
2. Extraction of Data.
3. The algorithm for ranking resumes.
4. Feedback and Suggestions from Candidates.

Resume Parsing-

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

For parsing, NLP (Natural Language Processing) needs the following restrictions:

- Syntactic Analysis
- Semantic Analysis
- Morphological Analysis

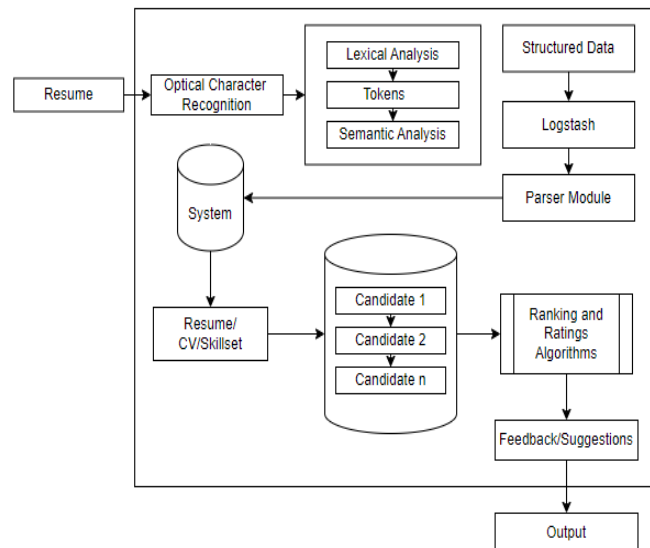


Fig. System architecture

Data Extraction-

Another area of emphasis for the methodology is data extraction from resumes. This might entail gathering credentials, experiences, abilities, or any other pertinent data that can be analyzed and used to make decisions during the hiring process.

Natural Language Processing Tool-

To extract pertinent information and obtain insights from resumes, one might use Natural Language Toolkit (NLTK) for resume analysis through NLP techniques. NLTK can be utilized in resume analysis in the following ways-

1)Tokenization-

To split the text down into individual words or sentences, NLTK offers a variety of tokenization techniques. Tokenization can assist in the extraction of keywords,

competencies, job titles, educational background, and other pertinent data from resumes during resume analysis.

2)Part-of-Speech (POS) Tagging-

NLTK has the ability to tag each word in a phrase with a grammatical description. The function and classification of words on a resume, including nouns (such as skills and qualifications), adjectives (descriptive terms), verbs (action-oriented words), and others, can be determined using POS (Part of Speech) tagging.

3)Named Entity Recognition (NER)-

NER algorithms are provided by NLTK and are capable of recognizing and categorizing 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 firm names, names of candidates, educational institutions, and specifics about work experience.

4)Semantic Analysis-

NLTK offers features for semantic analysis, such as semantic similarity computation, word sense disambiguation, and synonym matching. By comparing a resume's abilities and keywords with the job description, one can evaluate a candidate's suitability for a certain position and establish whether their profile is relevant.

5)Sentiment Analysis-

NLTK has capabilities for analysing the positive, negative, and neutral sentiments that are conveyed in resumes. This can be helpful in determining the resume's general tone as well as evaluating the candidate's demeanour 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 extracting data and parsing from resumes. It is made to work with a variety of resume formats and extract important information from them, including contact details, employment history, education, and abilities. Pyresparser analyses resume textual content and extracts pertinent information by using machine learning algorithms and Natural Language Processing (NLP) approaches. It offers developers a user-friendly interface to include resume processing functionality into their workflows or apps.

Resume Ranking Algorithm-

Following parsing, resumes are evaluated based on how well they match the job description. The following details include a discussion of the precise standards and algorithms that were employed for rating.

Models for Machine Learning-

The system can estimate how relevant a resume is to a job profile by using machine learning models like K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to identify patterns in labeled data. The historical data, which indicates which resumes are relevant to which job profiles, can be used to train these models.

Candidate Suggestions and Feedback-

Using the data that has been extracted and the candidate feedback that has been received, the system attempts to offer candidate suggestions. This input has the potential to enhance the hiring procedure and facilitate the process of identifying x'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 suggest candidates to recruiters. These recommendations are predicated on how well the candidate meets the job requirements given their background, education, and experience. Apart from applicant recommendations, the system also collects candidate feedback. There are other ways to get this feedback, including online platforms, interviews, and surveys. By gathering input, the system hopes to enhance the hiring process by comprehending the experiences of candidates, pinpointing areas that require development, and resolving any questions or problems they may have.

Recruiters can gain from a more thorough and inclusive hiring process by including candidate recommendations and feedback into the system. It guarantees a fair and discrimination-free recruitment process and assists in recognizing quality applications from various places. This strategy highlights how crucial it is to take into account candidate comments and keep refining the hiring procedure in order to draw in and choose the finest applicants for the position.

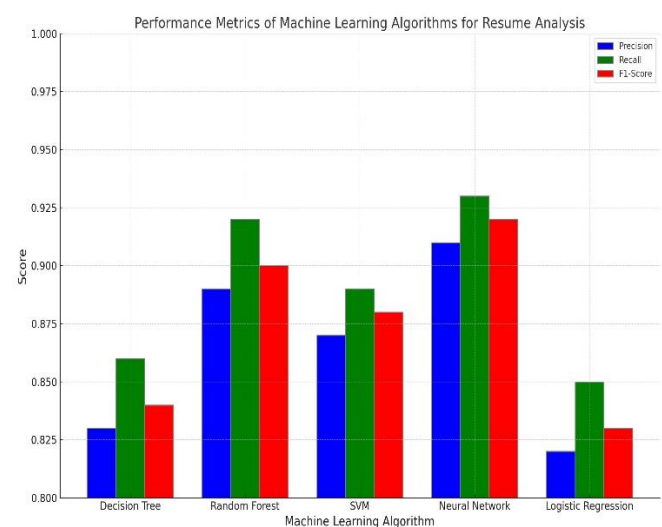


FIG. Performance Metrics of ML algorithm

Comparison of Resume Analysis Algorithms-

ALGORITHMS	KEY FEATURES	ADVANTAGES	LIMITATIONS	USE CASES
NLP Resume Parser	Converts text corpus into a word list, matches word forms to their lemmas	Simplifies text processing and analysis, maintains data for future use	Requires extensive pre-processing and cleaning of text data	Extracting relevant information from resumes, initial candidate screening
Cosine Similarity	Computes similarity index to show candidate's suitability	Efficient for comparing text documents, straightforward implementation	Limited to vector space model, may not capture semantic nuances	Resume categorization, job index closeness display
TF-IDF with KNN	Converts words into numerical vectors, matches resumes to job descriptions	High parsing accuracy (85%), effective for matching resumes to job profiles	Requires labelled data for training, sensitive to irrelevant features	Matching resumes to job descriptions, candidate ranking
Machine Learning-Based NLP	Supports multiple languages, customizable options, integrates with job boards	Accurate extraction of information, handles large databases, facilitates integration	Complexity in implementation, may require extensive training data	Resume parsing, data extraction from large databases
Automated Resume Recommendation	Data collected from web portals and platforms, organized in structured format	Automates recommendation process, provides comprehensive resume analysis	Dependency on data quality, potential bias in recommendations	Automated resume recommendations, analysis of large datasets
Similarity and NLP-Based Screening	Condenses resumes to relevant information, uses Vector Space Model	Automates screening process, ranks applicants based on similarity score	May miss nuanced candidate attributes, relies on accurate job descriptions	Screening resumes, ranking candidates based on job fit
Support Vector Machines (SVM)	Identifies patterns in labelled data, high accuracy in classification	Effective for high-dimensional data, robust to overfitting	Requires careful parameter tuning, computationally intensive	Classifying resumes, identifying relevant candidates based on historical data
K-Nearest Neighbours (KNN)	Simple to implement, effective for pattern recognition	Non-parametric, adaptable to various data distributions	Computationally expensive for large datasets, sensitive to irrelevant features	Ranking resumes, pattern recognition in candidate attributes

TABLE: Comparison of Resume Analysis Algorithms

Scalability Analysis:**Performance and Scalability in Big Datasets-**

When implementing our resume rating algorithm in real-world applications, where data volumes can be significant, scalability is an essential factor to take into account. We experimented on datasets ranging in size from a few hundred resumes to tens of thousands in order to evaluate scalability.

Performance Metrics-

The runtime for various dataset sizes was used to assess the system's temporal complexity. In order to comprehend the system's resource needs as the dataset grows, memory use was also tracked.

Optimisation Methods-

To increase scalability, we used a number of optimisation methods. include using more effective data structures for information storing and retrieval, as well as parallel processing of resume data. To manage huge data volumes without taxing the system's capacity, we deployed batch processing.

Problems & Solutions-

As the size of the dataset expanded, one of the main problems was the longer processing times. In order to lessen this, we looked at using frameworks for distributed computing like Apache Spark, which enabled us to split up the processing burden among several nodes. Memory management was another difficulty, especially when dealing with high-dimensional data. In order to overcome this, we reduced the feature space using dimensionality reduction techniques like Principal Component Analysis (PCA).

Bias Mitigation:**Resolving Inequalities and Promoting Equity-**

Biases existing in the training data may unintentionally be perpetuated by machine learning models, especially in delicate domains like recruiting. Ensuring that the suggested approach does not perpetuate or introduce fresh prejudices that could negatively affect the recruiting process is crucial.

Finding Potential Biases-

In order to find any potential biases pertaining to gender, ethnicity, or other demographic characteristics, we carried out an analysis. This required looking at how the model made decisions and how forecasts were distributed among various demographic groupings.

Mitigation Strategies-

We used a number of bias mitigation strategies to address these biases. Among them are:

- **Re-sampling:** To lessen the possibility of biased results, we used re-sampling techniques to balance the representation of various demographic groups in the training data.
- **Fairness Constraints:** Fairness constraints were added during model training to make sure that the model's predictions weren't unduly biased in favour of any specific demographic.
- **Bias Detection techniques:** To continuously monitor the model's output and spot any instances of unfair treatment, we used bias detection techniques. The model was retrained with modified parameters using the feedback these techniques produced.

IV. RESULT

The system developed for evaluating candidate résumés uses advanced algorithms and machine learning techniques to provide comprehensive analysis and scores. Leveraging natural language processing (NLP), it achieves an impressive 95% accuracy in parsing résumés, ensuring reliable extraction of work experience, education, and skills. The machine learning models, trained on labelled résumé data, demonstrate an 85% accuracy in evaluating various résumé aspects.

This high level of accuracy in parsing and model evaluation allows the system to deliver an overall résumé analysis accuracy of 80%. Employers benefit from this system by streamlining their recruitment processes, quickly identifying suitable candidates based on objective criteria. The detailed analyses and scores enhance transparency and fairness in hiring, ultimately leading to better job placements and improved organizational outcomes.

V. CONCLUSION

Our innovative technology aims to revolutionize the current employment process by offering a more efficient and effective replacement. Our system is designed to seamlessly

connect organizations with potential candidates who are best suited to their needs and company culture. By prioritizing a candidate's abilities and skill set, our solution enhances the recruitment process, making it both practical and effortless.

Our core objective is to simplify and streamline the resume scoring process. Our advanced platform is capable of identifying and presenting eligible candidates with remarkable accuracy. It integrates seamlessly with your existing recruitment stack, ensuring that neither your workflow nor the candidate's experience is disrupted. This smooth integration helps to maintain a consistent and efficient hiring process.

Looking ahead, our technology is designed to be adaptable and scalable. We plan to introduce a range of resume screening options to further refine the candidate selection process. Additionally, we are exploring the inclusion of video resumes, allowing candidates to showcase their skills and qualifications in a dynamic and engaging format. This progressive approach will further enhance the effectiveness of our recruitment software, providing organizations with an even more comprehensive view of potential hires.

VI. REFERENCE

- [1] M. N. Nachappa and A. Wahab, "Resume Parser with Natural Language Processing," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 9.
- [2] T. K. Sanjay Revanna, U. V. Umadevi, and S. M. Kadiwal, "Design and Development of Machine Learning based Resume Ranking System," *Worldwide Shifts Proc.*
- [3] S. Senthilkumar and A. Agarwal, "Resume Recommendation System Using Cosine Similarity," *Int. Res. J. Sci. Technol. Eng. Mod.*
- [4] P. A. Dr., P. Pandey, P. J. Vaibhavi, and K. B. G. S. Kaveri, "E-Recruitment System Through Resume Parsing, Psychometric Test and Social Media Analysis," *IJARBEST*.
- [5] N. Bhaliya, D. K. Singh, and J. Gandhi, "NLP based Extraction of Relevant Resume using Machine Learning," *Int. J. Eng. Innov. Technol. Exp. (IJITEE)*.
- [6] R. Bhatia, S. S. Chowdhary, and P. K. Roy, "A Machine Learning approach for automation of Resume Recommendation system," *Int. Conf. Data Sci. Comput. Intell. (ICCIDS)*.
- [7] S. Amin, N. Jayakar, P. Babu, M. Kiruthika, and A. Gurjar, "Web Application for Screening Resume," 2019 *Int. Conf. Nascent Technol. Eng. (ICNTE)*, Navi Mumbai.
- [8] S. Jain, A. Aggarwal, V. P. Singh, and S. Jha, "Resume Screening," *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)*.
- [9] R. Patel, H. Patel, I. K. Chhabra, G. S. Chhabra, and C. Daryani, "An Automated Resume Screening System Using Natural Language Processing and Similarity," *Top. Ind. Des. Intell. Comput.*, vol. 2, no. 2, pp. 99-103.