



REAL TIME RESUME SCREENING USING NLP AND TOKEN BASED INDEXING

¹Bussa Sai Sweshika, ²Golla DivyaSri, ³Madapati Aishwarya, ⁴K. PhaniSree

¹Student, ²Student, ³Student, ⁴Student

¹Information Technology,

¹Gokaraju Rangaraju institute of engineering and technology, Hyderabad, India

Abstract: The goal of resume screening is to find the best candidates for a position. Our system is a resume ranking software, Input would be resumes and job descriptions, output is highly ranked candidate's resume and acquired instantly in real-time. We will be using Mong for string matching, Cosine Similarity, TF-IDF. The existing systems are simple and effective but are not robust in terms of accuracy, efficiency, and processing and could lead to inaccurate assumptions and loss of human potential. We propose a web application that aims to order the resumes, by intelligently reading job descriptions as input and comparing the resumes which fall into the category of given Job Descriptions. In order to match and rate candidates in real-time, the software provides a ranking after filtering and recommends the better resume for a given textual job description. The Advantages of the proposed system are Secured, Interpretability, High accuracy, Lightweight model & fast processing. It could be used in MNC's where multiple resumes must be screened every single day for multiple jobs.

Index Terms - Resume Screening, TF-IDF, Cosine Similarity

I. INTRODUCTION

The evolving technology is creating many chances of employment for many. Nowadays to apply for any job the most essential document is a resume. A resume tells a lot about the person's achievements and the skill sets in all walks of life. The person applying for the job highlights the strong points and skillsets required for the company. Multinational organizations receive thousands of emails from such people who send their resumes for them to apply for a certain post. Now the real challenge is to know which resume is to be sorted and shortlisted according to the constraints. One method is to manually check and sort the resume. Now this method is the most time-consuming and also can lead to a lot of errors because of human interventions. Also humans cannot keep on working continuously. Hence there is a problem of less efficiency as well. Thus we have proposed a system that will easily find the required skill set by scanning the document or the resume and sort according to the skill sets which is a specified constraint of the organization. We are going to use the concept of Machine learning. Machine learning for recruiting is an emerging category of HR technology designed to reduce or even remove time-consuming activities like manually screening resumes.

Hence a system which is intelligent enough to sort all these documents without any error and within time is needed which is exactly our aim. machine learning for recruiting is the application of machine learning, such as the learning or problem-solving that a computer can do, to the recruitment function. This new technology is designed to streamline or automate some part of the recruiting workflow, especially repetitive, high-volume tasks. For example, software that applies machine learning to resumes to auto-screen candidates or software that conducts sentiment analysis on job descriptions to identify potentially biased language. This is exceedingly rare and is also known as monochromacy. You could also have problems seeing clearly and be more sensitive to light, depending on the type.

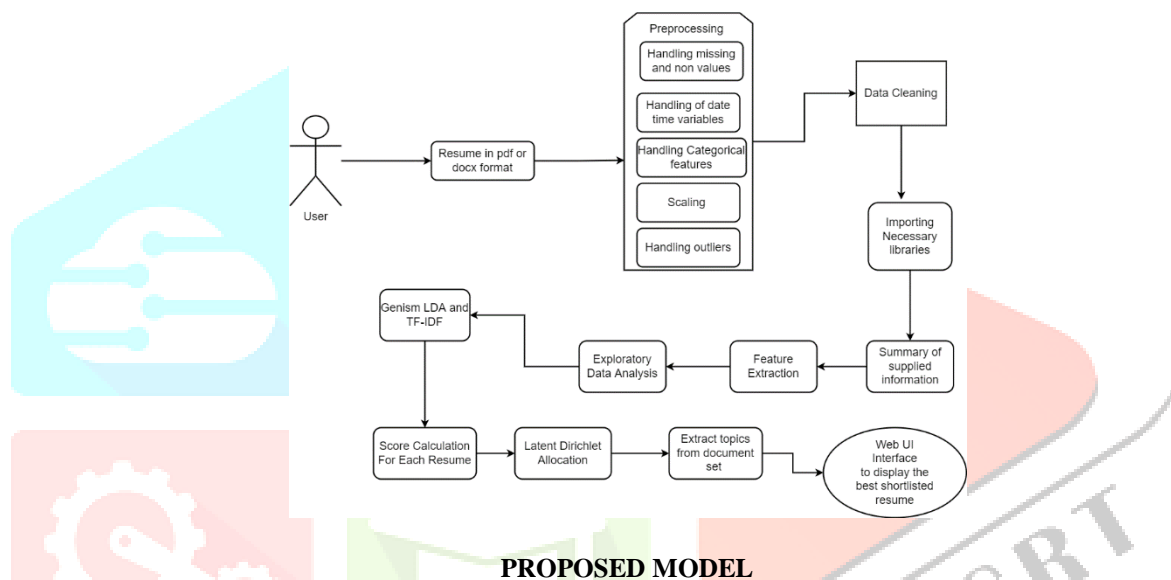
Due to this, the recruiters quite often find it really arduous to narrow down the most appropriate candidates from a large applicant pool. In recent years, there have been more than 50,000 e-recruitment sites have been developed. The developers of these online recruitment sites have used various approaches to identify the prospective candidates for a given job profile of a company. Some of these, have managed to employ classification techniques that will classify the candidate resumes into various categories for every job posting given by every company. In these approaches, every candidate's resume is tried to match with every given job posting on the recruitment site. The aim of these recruitment sites is to throw up the results to the candidate to which they are the best fit. The techniques used by these sites have resulted in high accuracy and precision, but one of the major disadvantages is the factor of time complexity. If every candidate's resume is matched with every other job posting given on the online recruitment site, the time complexity for acquiring the results is very high. The world of Artificial Intelligence [AI] and Machine Learning [ML] has grown significantly in recent times. The availability of large amounts of data brought about by advancements in technology which has made the internet cheap and accessible to previously inaccessible regions of the world has contributed to a great increase in the performance of the ML models in recent times.

1.1 EXISTING MODEL

The existing system is a traditional Machine Learning based system. It gives lower rates of accuracy. It has lower efficiency. It gives lower inaccurate results. This system may lead to loss of human potential. The methods used are NER (Named Entity Recognition), section based segmentation. Existing working methods are time consuming and may review the same resume multiple times if not recognized. The amount of manual work involved in recruiting processes has been reduced, and the initial screening of candidates has been completed. By removing the redundant candidates, you can keep only the ones who are relevant.

1.2 PROPOSED MODEL

Our system is a resume ranking software that uses natural language processing (NLP) and machine learning. This AI-powered resume screening program goes beyond keywords to contextually screen resumes. Following resume screening, the software rates prospects in real-time depending on the recruiter's job needs. The web application aims to order the resumes, by intelligently reading job descriptions as input and comparing the resumes which fall into the category of given Job Descriptions. In order to match and rate candidates in real-time, the software employs natural language processing. Unlike generic processes, this app utilizes Mong for string matching, Cosine Similarity, Overlapping coefficient Natural Language. Our work takes a different approach as it focuses mainly on the content of the resumes where we perform the extraction of skills and related parameters to match candidates with the job descriptions. The interactive web application will allow the job applicants to submit their resumes and apply for job postings they may still be interested in. The resumes submitted by the candidates are then compared with the job profile requirement posted by the company recruiter by using techniques like machine learning.



II . RELATED WOEK

[1] stated that the interactive web application will allow job seekers to submit their resumes and apply for any job postings in which they are still interested. The candidates' resumes are then compared to the job profile requirements posted by the company recruiter using techniques such as machine learning and Natural Language Processing (NLP). The resumes can then be given scores and ranked from best match to worst match. This ranking is only visible to the company recruiter who is looking to hire the best candidates from a large pool of applicants.

[2] stated many new methods were introduced in this process to reduce human involvement and errors. This paper discusses one such process that is very efficient in resume screening. Natural Language Processing (NLP), an automated Machine Learning Algorithm for screening resumes, is included. This paper describes the end-to-end operation of a Python application that efficiently screens candidate resumes based on the organization's requirements.

[3] stated this web application assists us in screening and ranking candidate resumes. The submitted resumes are compared to the job description, and the best-profiled resumes are identified using a technique known as Natural Language Processing (NLP). Finally, the resumes are scored and ranked in order of best to worst fit. This ranking is only available to corporate recruiters looking to select the best-profiled candidates from a large pool of applications.

[4] Candidate job profiles can be screened by analyzing their CV/Resume for selection using the most intellectual and advanced process based on HR requirements. Each profile is read one by one and marked as selected or not for the preliminary stage. The researchers hope to revolutionize the entire recruitment process by making employer tasks easier than ever in this work. They propose to achieve this by ranking the CVs of individuals by matching the total fields in the CVs with the required fields.

[5] Candidate job profiles can be screened by analyzing their CV/Resume for selection using the most intellectual and advanced process based on HR requirements. Each profile is read one by one and marked as selected or not for the preliminary stage. The researchers hope to revolutionize the entire recruitment process by making employer tasks easier than ever in this work.

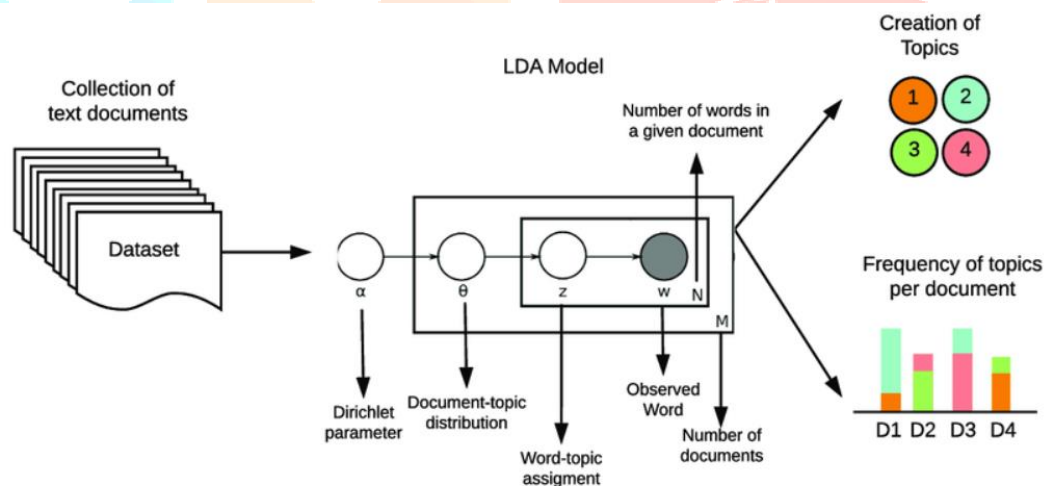
III . RESEARCH METHODOLOGY

NATURAL LANGUAGE PROCESSING : This application makes use of Natural Language Processing (NLP) which helps in data training and feature extraction of the text data. NLP is an analysis of natural languages so that computers can understand them. Natural language, whether spoken, written, or typed, is the most natural means of communication between humans, and the mode of expression of choice for most of the documents they produce. Using NLP methods, semi-structured text data is converted to a structured format with required extracted features. Resume collection is being performed and folder Structure creation is being done. Data Cleaning such as removing clutter and unnecessary punctuation would be taken off, Feature Engineering would be performed for enhancement. This includes removing stop words, punctuation, and stemming.

COSINE SIMILARITY: We've chosen the Cosine Similarity Algorithm, in which the employer's Job Description is matched against the content of resumes in the space, and the topmost similar resumes are suggested to the recruiter. Cosine Similarity is a measurement that quantifies the similarity between two or more vectors. The cosine similarity is the cosine of the angle between vectors. The vectors are typically non-zero and are within an inner product space. The cosine similarity is described mathematically as the division between the dot product of vectors and the product of the euclidean norms or magnitude of each vector. Cosine similarity is a commonly used similarity measurement technique that can be found in widely used libraries and tools such as Matlab, SciKit-Learn, TensorFlow etc.

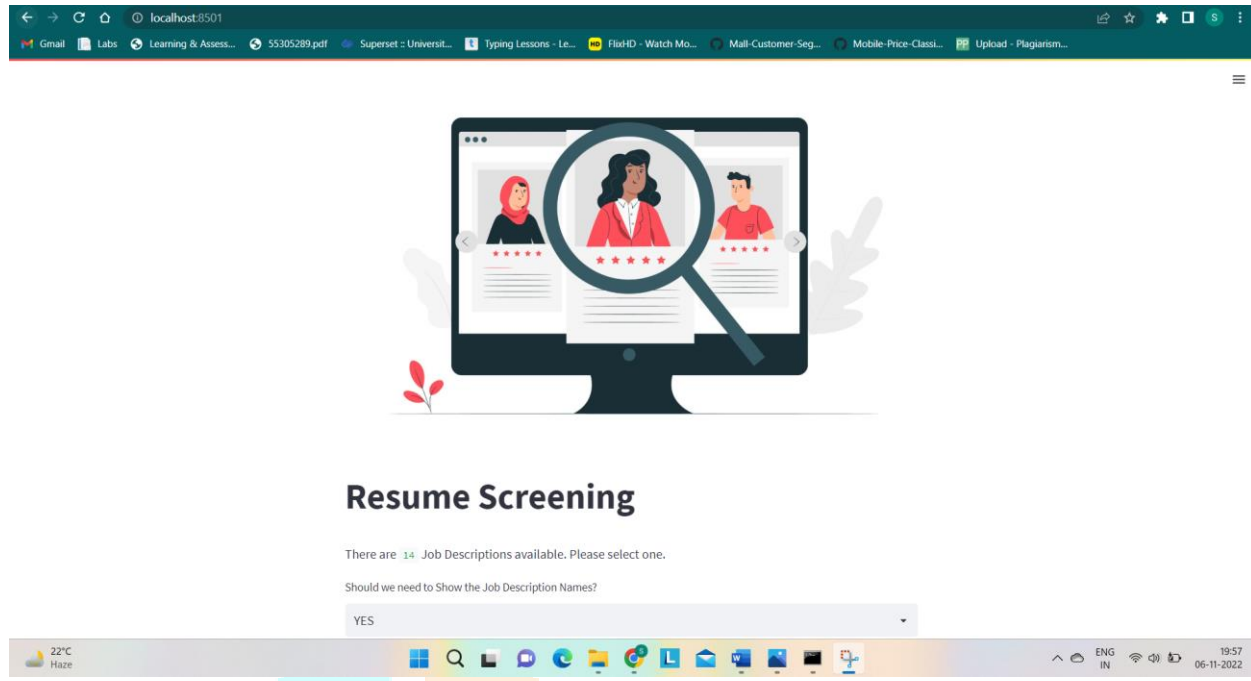
TF-IDF: At this stage, a dynamic Script for the Tf-Idf approach is written. Term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection TF-IDF is word frequency scores that aim to emphasize phrases that are more interesting, e.g., common in a text but not across texts, without delving into the arithmetic. The TF-IDF Vectorizer tokenizes texts, learns vocabulary, inverts frequency weightings, and allows encoding new ones. It provides information on a word frequency in the documents. Higher the TF- IDF score of a term which is computed using the above equations represents more relevance in a document.

Latent Dirichlet Allocation: A tool and technique for Topic Modeling, Latent Dirichlet Allocation (LDA) classifies or categorizes the text into a document and the words per topic, these are modeled based on the Dirichlet distributions and processes. Latent Dirichlet Allocation has been used in the application for the following functions-Discovering the hidden themes in the data. Classifying the data into the discovered themes. Using the classification to organize/summarize/search the documents. The application, then deals with the calculation of the score for a candidate's resume according to the job posting they have applied for.



IV. RESULTS AND DISCUSSION

4.1 WEB-APP



Job Description :

Billing cum Logistics Manager

Job Summary Urgent requirement of Billing cum Logistics Manager for Ecommerce Company Locatio - Gurgaon Company profile : detel-india.com Responsibilities and Duties Total Billing Monitor & Controlling Invoice validation before sending to Customer with Accuracy E-Commerce Billing, Monthly Fixed and Variable Billing. Follow up & Clearance from Operation for specific Remarks in PODs (Shortages/Damages Etc) Deductions booking off Proper Documentation (Invoices/PODs Etc) Reconciliation: POD Reconciliation (Damages/Shortages/Other Remarks) & Resolution Deductions Reconciliation & Recovery Preparation Pending POD Controlling & Monitor

4.2 INPUT AND OUTPUT

Slide to select te JD :



Do you want Job Description to be displayed ?

YES ▾

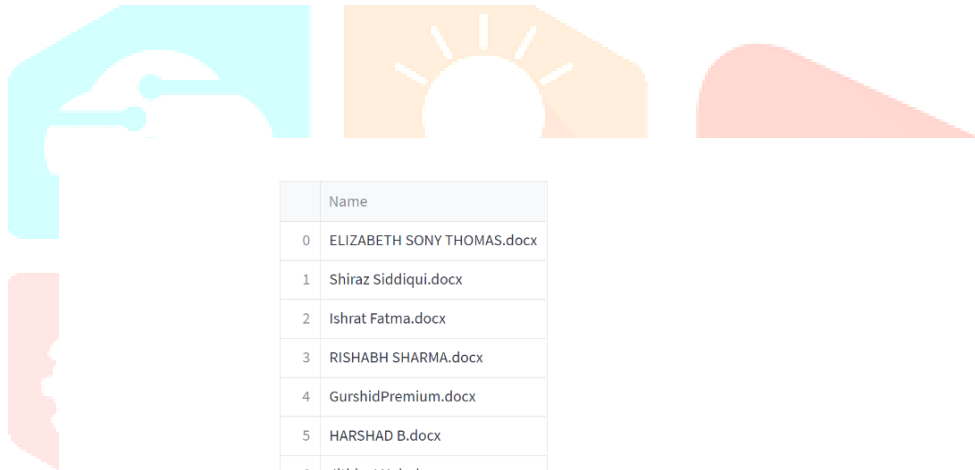
Job Description :

HTML Developer

Slidesigma India LLC is a web design and development company based in South Delhi. Due to the active COVID-19 crisis, we are working remotely. We are looking for an HTML developer with up to 3 years of experience. You are expected to work full-time and will be collaborating with our team and supervisors virtually.

Years of Experience:

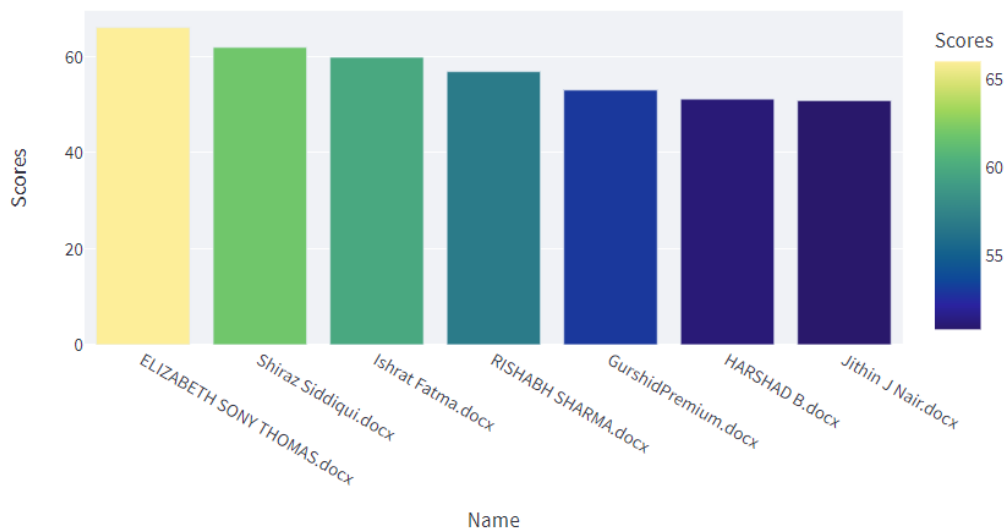
3 years experience



	Name
0	ELIZABETH SONY THOMAS.docx
1	Shiraz Siddiqui.docx
2	Ishrat Fatma.docx
3	RISHABH SHARMA.docx
4	GurshidPremium.docx
5	HARSHAD B.docx
6	Jithin J Nair.docx

	Scores
0	66.0027
1	61.8878
2	59.8262
3	56.8368
4	53.0054
5	51.1262
6	50.8152

Score and Rank Distribution



V. CONCLUSION AND FUTURE SCOPE

Our algorithm was successfully able to screen and shortlist the best candidates with the help of NLP. Highly accurate results were obtained by using Latent Dirichlet Allocation to display the best-shortlisted resume on Web Ui. The web application was successfully able to order the resumes, by intelligently reading job descriptions as input and comparing the resumes which fall into the category of given Job Descriptions. The results from the model are encouraging. The resume classifier application is successful in automating the manual task of project allocation to the new recruits of the organization based on the interests, work experience, and expertise mentioned in the profile.

The application can be extended further to other domains like Telecom, Healthcare, E-commerce, and public sector jobs. We also wish to put into effect and present a smart evaluation in the consistent database to survey with the present models. Efforts can be made to explore whether it is possible to identify in advance what lists might be ranked worse than the current baseline and to investigate whether there is another way to transform word embeddings into document embeddings.

VI. REFERENCES

[1] Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu, M. Kiruthika, Ambarish Gurjar, "Web Application for Screening Resume", International Conference on Nascent Technologies in Engineering (ICNTE), 2020

[2] Tumula Mani Harsha, Gangaraju Sai Moukthika, Dudipalli Siva Sai, Mannuru Naga Rajeswari Pravallika, Satish Anamalamudi, MuraliKrishna Enduri, "Automated Resume Screener using Natural Language Processing(NLP)", 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022

[3] M. Alamelu, D. Sathish Kumar, R. Sanjana, J. Subha Sree, A. Sangeerani Devi, D. Kavitha, "Resume Validation and Filtration using Natural Language Processing", 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 2022

[4] Rasika Ransing, Akshaya Mohan, Nikita Bhrugumarharshi Emberi, Kailas Mahavarkar, "Screening and Ranking Resumes using Stacked Model", 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2022

[5] M.F. Mridha, Rabeya Basri, Muhammad Mostafa Monowar, Md. Abdul Hamid, "A Machine Learning Approach for Screening Individual's Job Profile Using Convolutional Neural Network", International Conference on Science & Contemporary Technologies (ICSCT), 2021