RESUME SCANNER AND ANALYZER
RESEARCH PAPER

Ritu Singh, Priyesh Soni, Rishabh Mishra, Madhav Muskan,
Asst. Prof. Subhash Chandra Mauyra

Computer Science Engineering Department

Maharana Pratap Group of Institution, Kanpur

Abstract This paper focuses majorly on the design of the web application which will be used to screen resumes (Curriculum Vitae) for a particular job posting. In the proposed system, a web application will encourage the job applicant candidates as well as the recruiters to use it for job applications and screening of resumes. Recruitment is a tedious process wherein the first task for any recruiter is to screen the resumes. The proposed web application is designed in such a way that job applicants as well as recruiters can use it with ease for applying for job openings and screening respectively. The recruiters from various companies can post the details of the job openings available in their respective companies. The interactive web application will allow the job applicants to submit their resume and apply for their 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 and Natural Language Processing (NLP). Scores can then be given to the resumes and they can be ranked from highest match to lowest
match. This ranking is made visible only to the company recruiter who is interested in selecting the best candidates from a large pool of candidates.

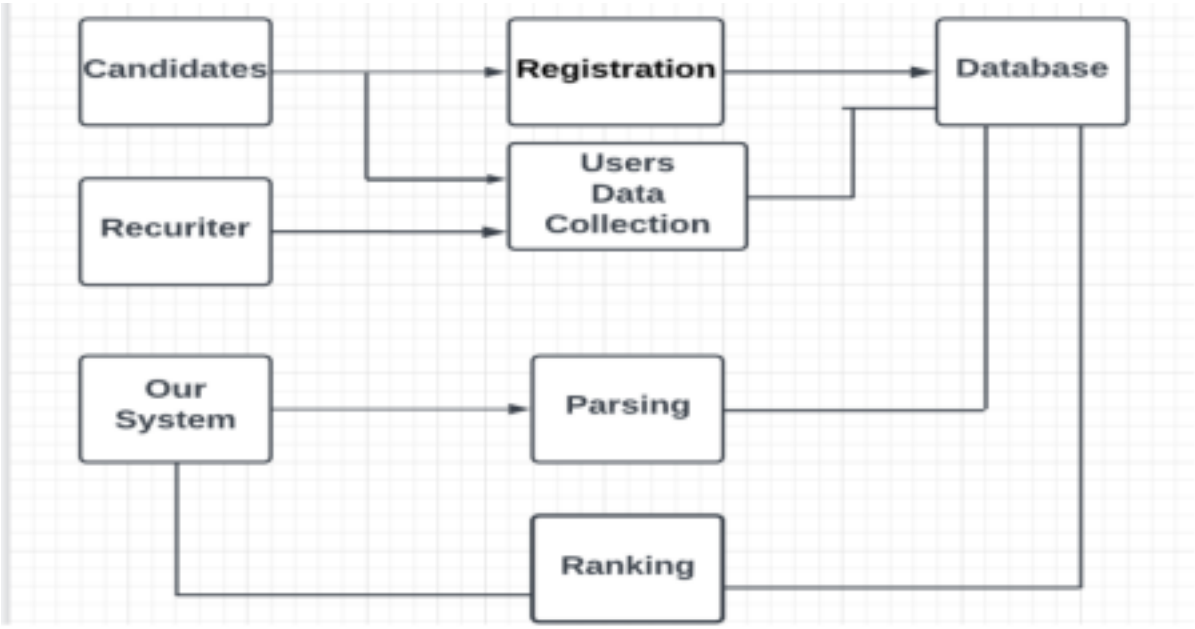
Keywords: Natural Language Processing (NLP); NER (Named Entity Recognition);

Introduction

Recruitment is a 200-billion-dollar business. It deals with hiring the best fit candidates having the relevant skills for a given job profile from an immensely large pool of candidates. If a company has any job opening for a position, scores of candidates mail their resumes to the company to apply for that opening. In the hiring process, the first task for any recruiter is to screen the resumes of all the job applicants. Any company having a job opening for a particular position will have their mail inboxes bombarded with thousands of emails from the aspiring job applicants every single day. Selecting the prospective candidates for that job position from a large pool of candidates for any recruiter is very tedious. It is an extremely daunting task for the recruiters of a company to manually go through a system where they can minimize the cost of thousands of resumes and select the most appropriate candidates for the job. Out of those thousands of resumes submitted to the company for the given job posting, about 75% of them do not showcase the relevant skills that are required for the job profile.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.

Literature review - Machine Learning approach for automation Resume Recommendation System : Pradeep Kumar Roy in their research [1], created a system where they can minimize the cost of hiring new candidates for the job positions in the company. They focused on 3 major problems in this process ● Picking the right candidates from the applicants ● Making sense of their CV's ● Finding out if the candidate is fit for the job role They performed NER, NLP, and text classification using n-grams and used Machine Learning to perform the classification using the algorithms of Random Forest with 38.9% accuracy, Multinomial Naïve Bayes with 44.39%, Logistic Regression with 62.4%, and the highest accuracy was obtained by Linear Support Vector Machine Classifier with an accuracy of 78.53% . B. Skill Finder: Automated Job-Resume Matching System : In research conducted by Thimma Reddy Kalva [2], they have developed a custom dataset of 3000 jobs and 80 resumes from the website indeed using the web service API [9]. This data is then used to rank the student's resumes comparing their skills required in the job, this is done using the Named Entity Recognition(NER) like Apache OpenNLP [10] and Stanford Name Entity Recognizer. The Skill finder efficiently matches the Resumes according to the job role posted and successfully sends emails to the desired candidates . C. ResumeNET : A Learning-based Framework for Automatic Resume Quality Assessment : In research conducted by Yong Luo [3], they have developed a custom dataset of 10,343 resumes which was acquired by a private resume management company. 98.82% (i.e 10,221 resumes) data is unlabeled and the remaining 1.18% (i.e. 122 resumes) data is labeled in 2 categories positive and negative, 33 and 89 of them are labeled as positive and negative. D. Web Application for Screening Resume : The goal for Sujit Amin [4] was to develop a web application for resume screening, with the help of 220 resumes out of which 200 were used for training and 20 used for testing purposes, further, the web application is divided into 3 divisions A) Job Applicant side B) Server-Side C) Recruiter Side The applicant side is where the applicant will provide his/her resume, the server-side will process the resume and then be trained using the NLP Pipeline which uses SpaCy which is an NLP framework. On the recruiter's side, the rank list of the resumes will be shown which was decided from a score calculator [13] so the recruiter can select the best fit candidate for the job. E. Design and Development of Machine Learning based Resume Ranking System : The proposed system hereby Tejaswini K [5] is where the resume is submitted by the candidate after an MCQ test which has a face detection system to detect malpractice. Once the resume is submitted it's run through NLP techniques to get the relevant skills from the resume and use TF- IDF vectorization to convert the words into vectors so the machine can understand it. The classifier used here is the KNN algorithm to identify the resume that closely matches the JD provided by the recruiter. The system has an average parsing accuracy of 85%. F. Resume Classification and Ranking using KNN and Cosine Similarity : Riza Tanaz Fareed, Rajath V, Sharadadevi Kaganumath came up with a method to implement the Resume classification with the addition of cosine similarity [4]. The process is the candidate provides his/her resume to the system. The resume is then passed through an NLP pipeline where the words are extracted out of the resume. Techniques like stop words, lemmatization are used to get the correct set of words. TF-IDF vectorizer is used to vectorize the words for the KNN model to

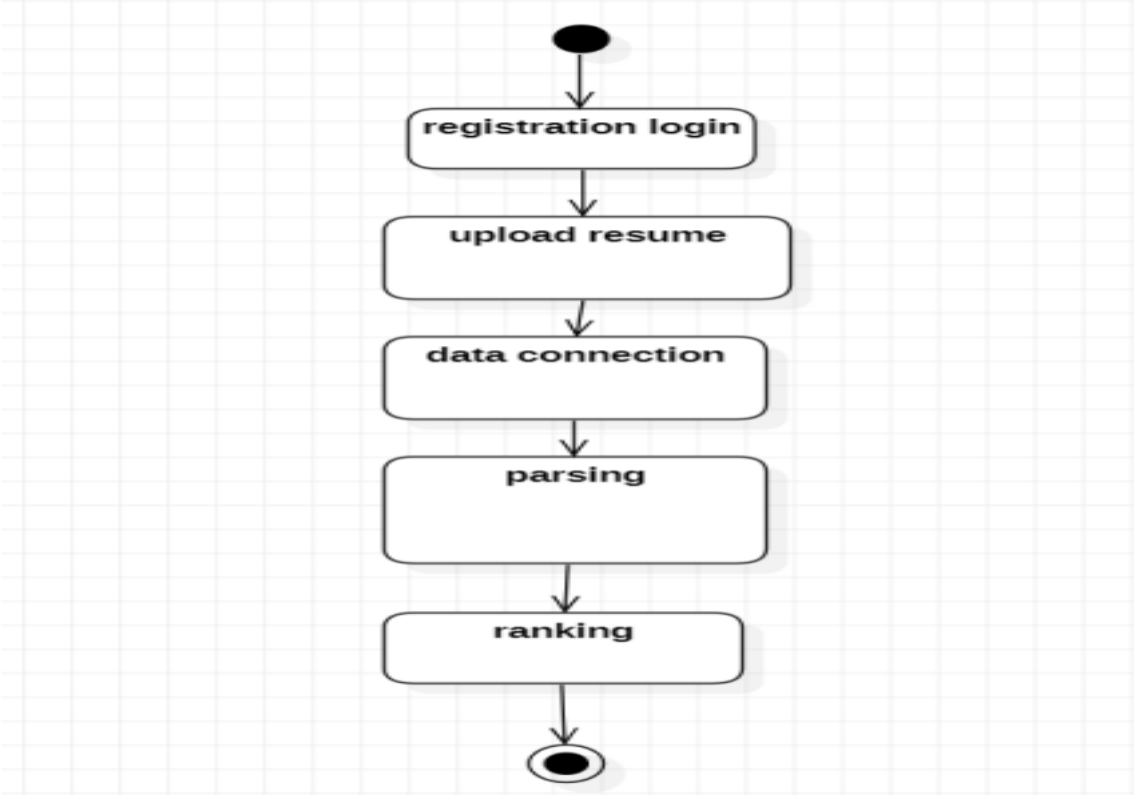
classify the resume into various categories. Now to evaluate the resume on the given JD document similarity detection is necessary so the Cosine Similarity Algorithm is used in which the JD content is matched with the candidate's resume. The accuracy for this trained model is 98.96% . G. Automated Tool for Resume Classification using Semantic analysis : This study conducted by Suhas Tangadle Gopalakrishna and Vijayaraghavan Varadharajan [16] provides a descriptive view of how they use semantic analysis for resume classification. The received resumes by the HR team are parsed through the Natural Language Processing Pipeline (NLPP) where the Stop deletion is used to delete the words like "and, or, the, etc.". Other techniques like Parts of Speech tagging and NER are also used. Total 6 types of classification models are used: Naïve bayes, Multinomial Naïve Bayes, Linear SVM, Bernoulli Naïve Bayes, Logistic Regression and KNN. These classification models are run on a dataset of 30,000 employees' resumes which was split in a 9:1 ratio where 27,000 were used for training and 3000 used for testing across various domains including AI, Computer Architecture, CG, Databases, Distributed Computing, CN, Web Technologies and Cloud Computing. Out of all the 6 Classifiers Multinomial Naïve Bayes came out with a top accuracy of 91%

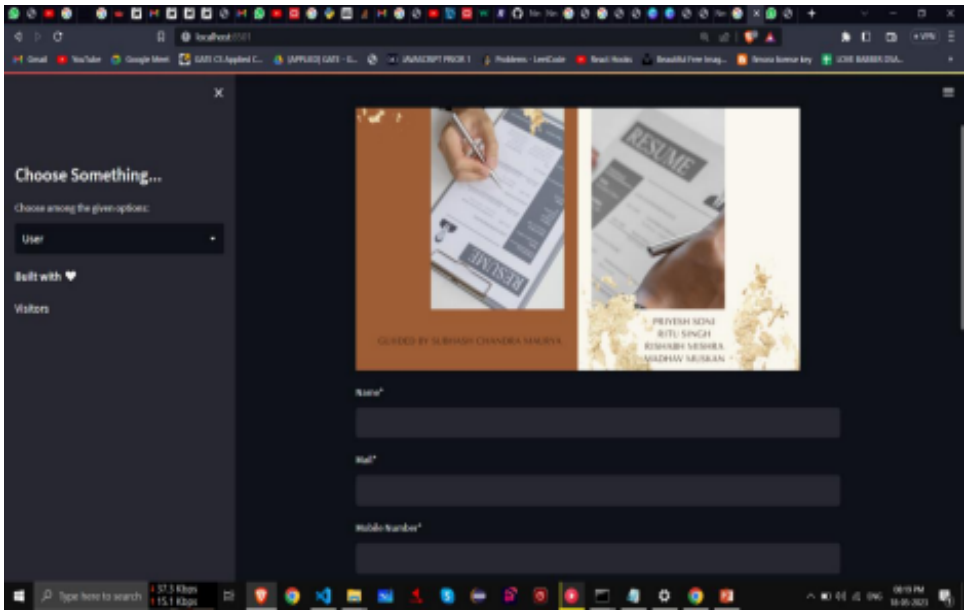


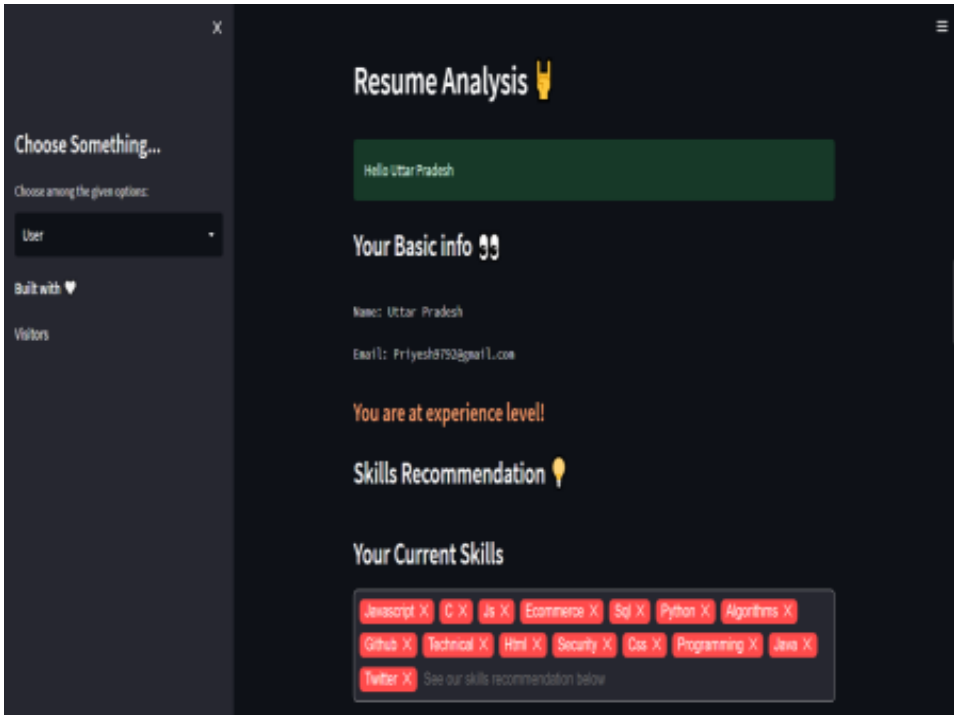
Proposed Methodology

In the first step, the system accepts the resume from the job applicant and performs text extraction on it . Since it is not possible to perform operations on a file in pdf format, it is first converted to text. Text is then fed to the pre-trained model that takes only that data which is relevant to the system. After this step, the data extracted will be displayed to the job applicant in editable format. The system allows the job applicant to edit information in this text before performing any operation on it. Once the job applicant is satisfied and submits the data, it can be sent for comparison with the job description requirements which are already passed through a pre-trained model and the requirements like number of years of experience,qualifications, etc. is already known. Now, the job data(after passing through a pre-trained model) is compared with the resume data(after passing through a pre pre-trained model) after which it can calculate the score based on the formula. Finally, the web application will display the candidate profile to the recruiter with other applicants for the same profile in a rank list based on their score . The proposed web application for screening and rating the resumes according to the job requirement posted by a company recruiter has various modules. These modules come under 3 parts, which are as follows:-

1. Job Applicant Side
2. Server Side
3. Recruiter Side.







CONCLUSION - The proposed system is currently under implementation and uses semi-supervised learning to achieve high accuracy. This system will definitely aid the recruiters to filter out the most prospective candidates based on their resumes for further rounds in the hiring process. It will ease the burden of the recruiters and they will not have to manually view each resume.

REFERENCES

[1]. Pradeep Kumar Roy, Vellore Institute of Technology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019. 10.1016/j.procs.2020.03.284.

[2]. Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job Resume Matching system. 3]Yong Luo, Nanyang Technological University, 2018. A LearningBased Framework for automatic resume quality assessment, arXiv:1810.02832v1 cs.IR.

[3]. Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume, IEEE DOI: 10.1109/ICNTE44896.2019.8945869.

[4]. Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and Development of Machine Learning based Resume Ranking System (2021), DOI: https://doi.org/10.1016/j.gltp.2021.10.002.

[5]. Riza tana Fareed, rajah V, and Sharadadevi kaganumath, "Resume Classification and Ranking using KNN and Cosine Similarity" In Research & Technology (IJERT), ISSN: 2278- 0181, Vol.10.

[6]. Suhas Tangadle Gopalakrishna, Vijayaraghavan Varadharajan, "Automated Tool for Resume Classification Using Semantic Analysis", International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 10, No.1, January 2019 .

[7]. Suhas H E, Manjunath AE, "Differential Hiring using a Combination of NER and Word Embedding", In 2020 International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Vol.9

[8]. Centre for Monitoring Indian Economy Pvt Ltd. (CMIE),2022. The unemployment rate in India.

[9]. Howard, J.L., Ferris, G.R., 1996. The employment interview context: Social and situational influences on interviewer decisions 1. Journal of applied social psychology 26, 112- 136.

[10]. Mudit Kapoor, Business Today, 2021. India's formal job creation numbers beat pandemic blues.