Research Result

Using Machine Learning to Evaluate Resumes

Akash Sharma[1], Saurabh Sharma[2]

[1,2]*Global Nature Care Sangathans Group of Institutions*

ABSTRACT

A recruiter's first duty during the laborious hiring process is to examine resumes (Curriculum Vitae). Nowadays, many businesses prefer online applications to paper ones when receiving job applications. The suggested approach is designed to make applying for openings and going through screening easier for both recruiters and job applicants. Recruiters from different companies might advertise their requirements online in order to fill specific job openings. On the other hand, candidates for employment can upload their resumes and apply for jobs they are interested in. The resumes supplied by the candidates are then compared to the job profile requirements as specified by the company's recruiter using tools like machine learning and natural language processing. This not only helps the recruiter narrow down the field of prospects to the best candidates, but it also saves them time from manually reviewing each resume.

KEYWORDS

Machine Learning, Natural Language Processing, Curriculum Vitae, Parser, Natural Language Toolkit, Beautiful Soup, SpaCy, Named Entity Recognition

1. INTRODUCTION

Recruitment, a 200-billion-dollar industry, entails choosing the top candidates from a large pool of applicants who have the skills required for a specific job profile. In response to job openings, many people mail the organization their resumes.

Any company with a job opening receives tens of thousands of emails from potential employees every day. Consequently, the first step for any recruiter in any hiring procedure is to browse through each and every resume that a job seeker has submitted. 90% of all CVs and resumes are only read by employers for two minutes or less, according to study. This demonstrates that the majority of the time, recruiters only scan the sections of resumes or CVs that immediately grab their attention and ignore everything else. For this reason, it should contain a precise segmentation mechanism for generic CVs and resumes to make it simpler to search for and locate the pertinent data.

However, it is a very difficult task for a company's recruiter to manually filter through hundreds of resumes and select the most qualified candidate for the position. Recruiters frequently struggle to select the top candidates from a large applicant pool since more than 75% of the 1,000 applications received do not display the requisite talents needed for the job profile as defined by the company's needs.

The creation of more than 50000 e-recruitment websites, which employ a variety of methods to locate the best candidates for a company's stated job descriptions, is a new development. The primary objective of these websites is to help job seekers choose which category best fits them by summarizing their CVs or resumes based on the keywords they contain. The time needed to receive the results, nevertheless, is one of these technologies' key downsides.

In order to extract data from a dataset using the method described in this paper, section-based segmentation is used, and machine learning is used to prepare the dataset for a certain type of work position. In order to reduce time complexity and boost time efficiency, the suggested method will only match a candidate's resume to job opportunities when their skills match the job requirement descriptions of the organization. Additionally, the results of all the CV matches will only be visible to the recruiter for that particular organization.

This document is organized as follows: Section 2 provides a brief overview of the current system; Section 3 provides a detailed design of the system that has been proposed to address this problem by analyzing the employed algorithm; Section 4 provides a bird's eye view of the project's implementation; and Section 5 presents the conclusion and future scope, respectively.

2. LITERATURE REVIEW

Today's society has seen a substantial change in the recruitment process as a result of the growth of technologies like the Internet. Numerous researchers have aided in the process of assessing resumes. The section that follows provides a summary of the literary work that has been done in the field of e-recruitment systems.

Potential employees can submit their resumes to more than 50000 online job boards. Some of these websites don't even have resume screening classification techniques. The firm recruiter must go through each application manually, which is a hard process for them, in order to select the top candidates for the following stages of the hiring process. These websites include Top Resume, Ideal, Adecco, Indeed, Monster, and others.

Let's examine a case study of Top Resume, one such website that uses the concept of natural language processing to assess resumes submitted by potential employers. Here, the applicant will start by submitting their résumé to the portal. Only the text from the supplied resumes is retrieved using methods like Natural Language

Processing, and the percentage representing the candidate's profile strength is shown. The candidate is also made aware of additional characteristics, such as the proportion of their expertise that corresponds to words like "education," "certification course," and "work experience." As a result, there are many other web programs like this one that provide nearly identical features.

3. PROPOSED SYSTEM

The proposed system will eventually assist in lessening the workload of the firm recruiters in an effort to address some of the primary drawbacks of the current approach. As a result, the suggested solutions make use of a variety of techniques in order to achieve automated screening of applicant resumes. This screening primarily focuses on the content of the resumes, from which we extract skills and other relevant information in order to match applicants with the job description of the organization.

1) Overall Working Model: In the first step, the system accepts the resume from the aspiring job applicants and performs keyword extraction on it.
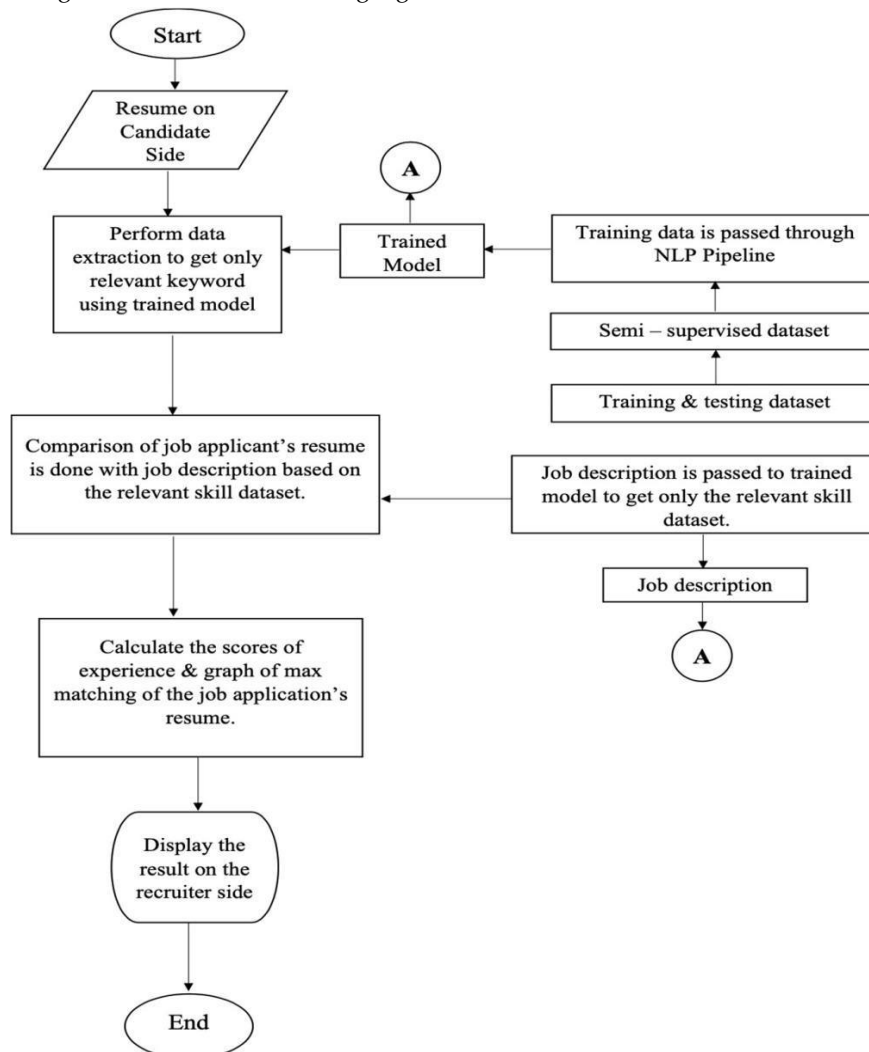

Fig: Activity Diagram showing how the overall system works

Since operations cannot be performed on a PDF file, it is converted into a text file. Following conversion, pertinent keywords are taken from the applicant's resume's mention of their GitHub and LinkedIn pages. These texts are provided to the pertained model, which only uses information that is necessary to the operation of the system and is taken into account when rating the applicant's profile. Then, these data are made available for comparison with the pre-established job description requirements of the pre-trained model. It is now feasible to calculate the score for relevant experience and plot on graphs the proportion of applicants whose resumes most closely match the demands of the hiring company after the resume and data have been compared to the job specifications. After the prospective candidates' screened resumes are graded in accordance with the score, the candidate's profile will then be supplied to the recruiter for use in the following phase of the hiring process. A firm recruiter's suggested system for screening resumes and assessing them in accordance with job requirements includes a variety of modules, most of which are composed of the following three components:

A.   Client Side

In this part the job aspiring candidate will upload their resumes for screening which mainly consist of two important modules i.e.; 'Accepting Resumes as Input' and 'Keyword extraction module'.
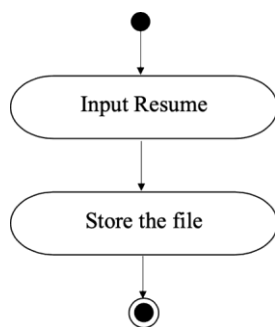
1)   Accepting Resume as Input

Fig: Activity diagram for accepting resume as input

The proposed system will have two types of users namely: Job applicant side and recruiter site, of which both will have a login on the system where recruiter would be able to post about particular job openings available in the company and the job applicants will upload their resume as input for screening. The input resume submitted in .pdf file will be then stored in the database. base64 encoding will be used to store this pdf resume file since MySQL or other primary SQL cannot store pdf files directly in the database.

2)   Keyword Extraction Module: This module deals with scrapping keywords from the resume in order to compare those with the job profile description so as to decide whether the resumes are shortlisted for further job

recruitment process or not based on their education, experience and other information captured on their resume. This keyword extraction is done using section-based segmentation with Natural language Processing.
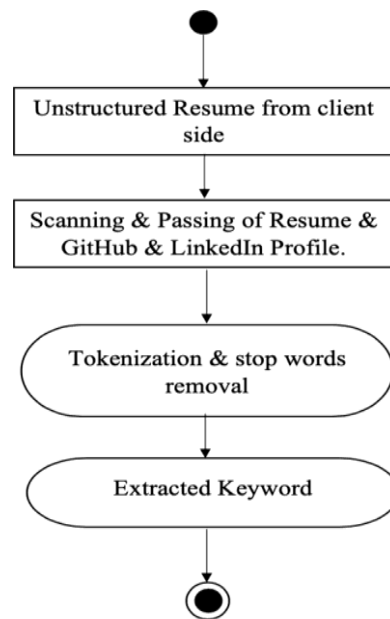

Fig: Activity Diagram showing how data extraction happens for resume input

This module will only start those keywords from resume which are relevant to the job posting. For this, unstructured resume is taken and parsed along with the candidate's GitHub and LinkedIn profile as mentioned in the resume where only tokenized (important) keywords are taken into consideration and stored in text file and otherstop words like 'this', 'and', 'for', etc. are removed.

B.   Server Side

This part discusses the design of the server side, which mainly consists of modules like, 'Training Data Modules', 'Convertingrequired skillsets into required format module'.

1)   Training Data Module: This module focuses on data training for a certain job posting type, such as Associate Software Developer. To do this, first gather resumes from similar job postings. Next, manually upload the ZIP files of all the resumes to dataturks.com, where only selected text could be easily converted to JSON format. Finally, pass the JSON file through NPL Pipeline, where it is trained using spaCy (an NLP framework used for training general data ins). Change NER (Named Entity Recognition) for the model in order for the entities to be appropriately identified for the raw data set in order to make the spaCy framework function in accordance with the needs of the proposed system. These datasets are then put through the NLP SpaCy pipeline after being converted to JSON on the Dataturks website in order to obtain the necessary trained model. Instead of manually entering every word to create the dataset, this approach uses semi-supervised learning to label the important data in a ZIP file of PDF resumes. The

dataset is then divided into training and testing datasets, which are then passed through the spaCypipeline.
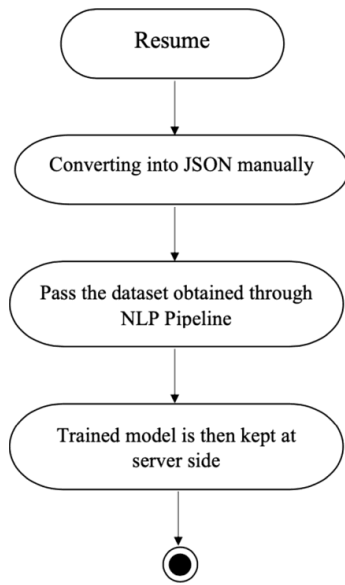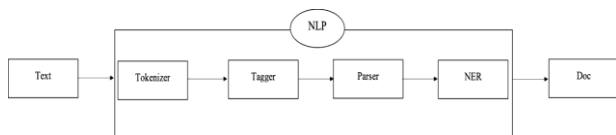

Fig: NLP Pipeline for Training Dataset Fig: Activity Diagram for training dataset

2)   Converting Skillsets into Required Format Module: 1) This module is concerned with obtaining a text file (made of the same name and stored in the database) of the pertinent skills from professional accounts like (LinkedIn and GitHub) mentioned in thecandidate's resume that will be compared with the entities as per the trained model from the job description uploaded by the recruiter.
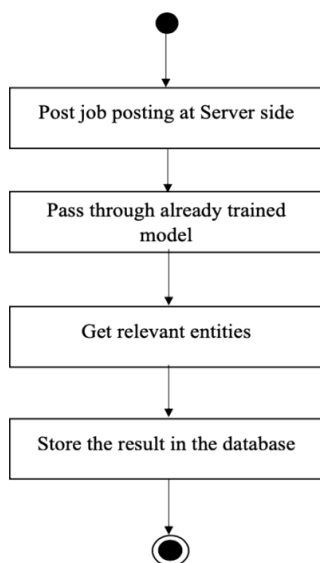

Fig: Activity Diagram showing job posting is converted into require

Further, that candidate's resume text file could be used for scoring and plotting the graphs accordingly. For this to be

the result, job description is passed through trained model relevant to job position.

C.   Recruiter Side

This is the recruiter side, which contains the module – 'Calculation of scores for the resumes, inputted by the candidate'.

1) Calculation of Scores for Candidate-Submitted Resumes: *Although it is the final module, it mostly focuses on calculating a candidate's resume score based on the job ad to which they have applied. In accordance with this, a rank list will be created, with candidate profiles with better scores appearing at the top in comparison to those with lower scores, and graphs will also be created in the appropriate manner. Who can decide who moves on to the next round of the hiring process is shown in this rank list and graph.*
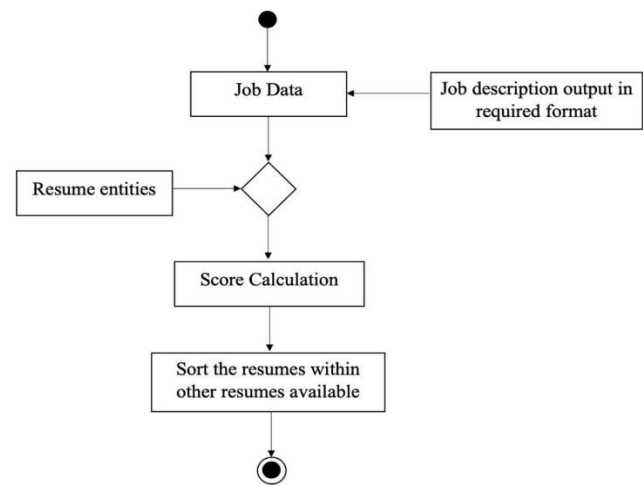

Fig: Activity Diagram showing comparison of candidate's profile with job description

The job description is then compared with the relevant skillsets of resume text file and a score is assigned based on the formula:

$$S = \frac{|\{S_r\}|}{|\{RS_j\}|} * 50\% + \frac{|\{E_r\}|}{|\{RE_j\}|} * 20\% + \frac{|\{X_r\}|}{|\{RX_j\}|} * 20\% + \frac{|\sum Y_w|}{|\sum C_w|} * 10\%$$

Where,

S – Score calculate

Sr – Skill-set of a candidate

RSj – Required job skills by job post

Er – Educational information of a candidate REj – Required education by job postXr – Experience of a candidate

RXj – Required work experience by job post Yw – Years of experience Cw – Total companies the applicant has completed service in.

Thus, from the formula, we have set the following weighting values: Skills weight = 50%, Job experience

weight = 20%, Education level weight = 20%, Loyalty level weight = 10%

This is an ongoing project, where some part of the proposed system has been implemented till now.

4. EXPERIMENTATION AND RESULTS

A. Screenshot of Training Model for NLP Pipeline

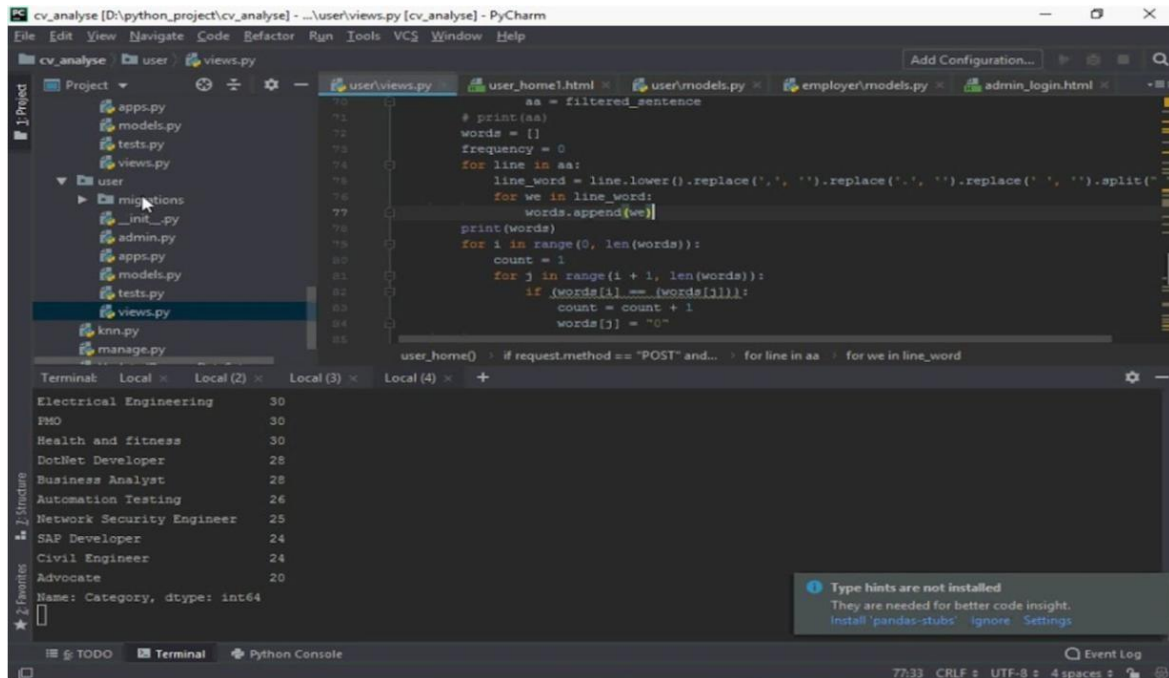Fig: Training model for NLP Pipeline

B. Accuracy of Trained Model

It is estimated using a model made up of 220 resumes, of which 20 are designed for testing and 200 are used to train the model in order to ensure that the system performs as predicted.
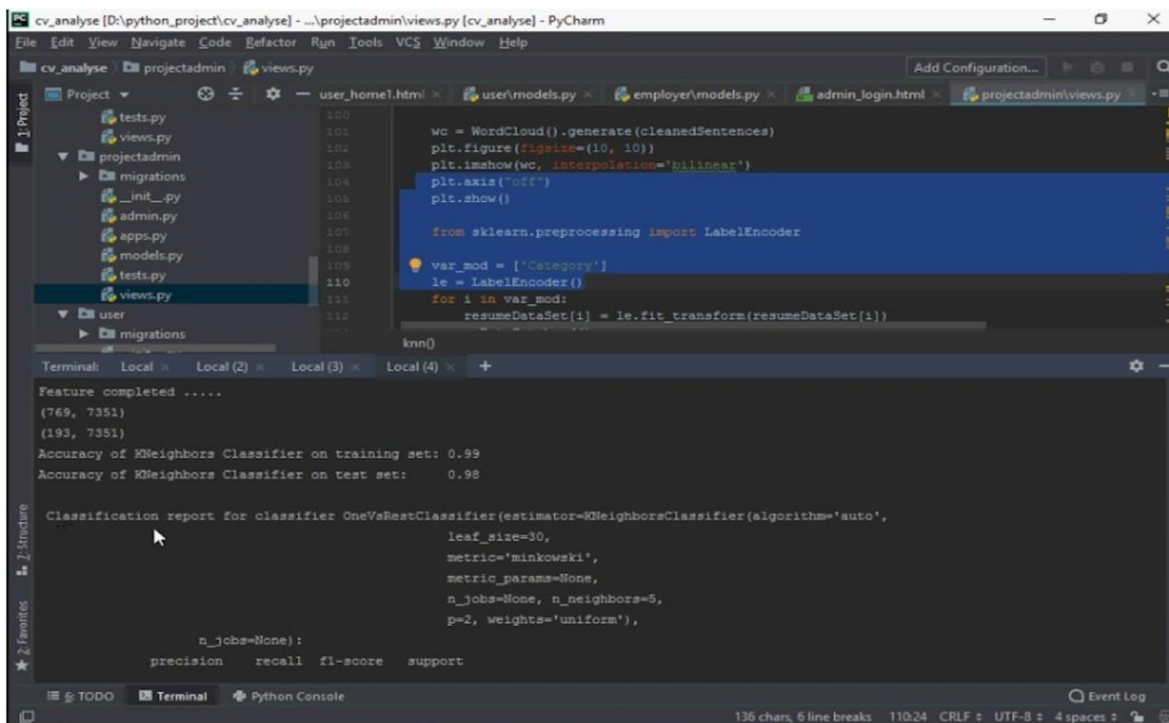

Fig: Screenshot of scanning and parsing of Resume

The suggested system in this instance receives a résumé as input and sends it through Natural Language Toolkit, which produces scanned and parsed text data. This data is passed into the next module, where a simple front end is constructed.

5. CONCLUSION

The proposed system is currently being put into use, and it primarily relies on K-nearest Neighbour semi-supervised learning to achieve high accuracy. This approach automates the process of defining requirements and ranking applications in a way that is generally quite consistent with the judgements of human experts. It will make it possible to more quickly narrow down the CVs of prospects from a broad pool of applications while maintaining consistency.

REFERENCES

[1] Ahmad, Noraziah, and Ahmed N. Abd Alla. "Smart Evaluation for Job Vacancy Application System." 2009 Second International Conference on the Applications of Digital Information and Web Technologies. (pp – 452,455), IEEE, 2009.

[2] Bhaliya, Nirali, Jay Gandhi, and Dheeraj Kumar Singh. "NLP based Extraction of Relevant Resume using Machine Learning." , (pp – 13) ,(2020).

[3] Amin, S., Jayakar, N., Sunny, S., Babu, P., Kiruthika, M., & Gurjar, A. (2019, January). Web Application for Screening Resume. In 2019 International Conference on Nascent Technologies in Engineering (ICNTE) (pp. 1-7). IEEE.

[4] Purohit, J., Bagwe, A., Mehta, R., Mangaonkar, O. and George, E., 2019, March. Natural language processing based jaro-the interviewing chatbot. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 134-136). IEEE.

[5] Kopparapu, S.K., 2010, December. Automatic extraction of usable information from unstructured resumes to aid search. In 2010 IEEE International Conference on Progress in Informatics and Computing (Vol. 1, pp. 99-103). IEEE.

[6] Halde, R.R., 2016, September. Application of Machine Learning algorithms for betterment in Education system. In 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) (pp. 1110-1114). IEEE.

[7] Nimbekar, R., Patil, Y., Prabhu, R. and Mulla, S., 2019, December. Automated Resume Evaluation System using NLP. In 2019 International Conference on Advances in Computing, Communication and Control (ICAC3) (pp. 1-4). IEEE.