

CV INSPECTION USING NLP AND MACHINE LEARNING

Rohan A. Bavaskar, Varad D. Bhandarkawthekar, Dhiraj A. Deore, Dr. Shubhangi Vaikole

^{1,2,3}Student, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

⁴Assistant Professor, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Abstract - In the proposed program, the program will encourage job applicants and employers to use it for job applications and apprenticeships. Hiring is a tedious process where the first job of any employer is to filter out beginners. The proposed web system is designed in such a way that job applicants and employers can easily use it to apply for job openings and appraisal. Employers of various companies can submit job creation information available to their companies. The collaborative program will allow job applicants to submit their CVs and apply for job vacancies that they may still be interested in. CVs submitted by nominees are then compared to the job profile requirement submitted by the company employer using techniques such as machine learning. And Indigenous Language Processing (NLP). Scores can then be awarded to beginners and can be rated from the highest similar to the lowest similar. This situation is made manifest only to the person who hires a company who is interested in selecting the best candidates for a large group of candidates.

Key Words: *Natural Language Processing (NLP); NER (Named Entity Recognition); Section-based segmentation; NLTK (Natural Language Toolkit).*

1. INTRODUCTION

1.1 Motivation

The current recruitment process is very complex and time consuming which forces candidates to complete all their skills and knowledge. The HR team also needs additional staff to process re-nominations. So that inspired us to build a flexible and automated solution.

1.2 Need of CV Inspection

In the current system the candidate must complete all the details of his / her CV in a timely manner and the candidates are not satisfied with the job they are choosing in the current system according to their skills. Let me tell you a 5: 1 ratio means that if 5 people get a job it means that out of five, only one young man will be satisfied with his job. Let me tell you an example: If I am a good python developer and the company hires me and makes me work in Java, my python skills are useless. On the other hand, if there is a vacancy in a company so according to the owner of the

company, you will choose the best person for that position. Therefore, this plan will serve as a handshake for two businesses. The company that selects the best candidate and the person who chooses the best job according to his or her skills and abilities.

2. RELATED WORKS

Case study of talent acquisition:

A. First Generation Hiring Systems

In this process the Recruitment Team will publish its vacancies and invite applicants. The methods of publishing were newspaper, television and oral. Interested parties will apply by submitting their CVs. These CVs were then approved and screened by the recruiting team and nominees were called for further rounds of interviews. The whole process will take a lot of time and effort from people to find the right person for their job.

B. Second Generation Hiring Systems

As industries have grown, their employment needs have grown exponentially. In order to provide for these hiring needs there are already some consultation units. Provide a solution where the candidate should upload its details in a specific format and send it to the center. These organizations will then search for candidates based on specific keywords. These agencies were intermediate organizations between the candidate and the company. These systems were not static as the candidate had to reload his CV in a certain format, and these formats changed from program to program.

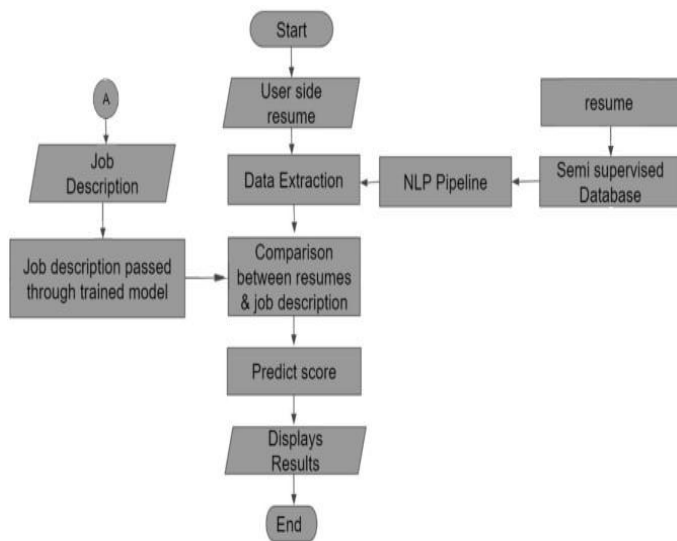
C. Third Generation Hiring Systems

This is a proposed system, which allows candidates to reload their books in a flexible format. This restart is then analyzed by the system, indexed and stored in a specific format. This simplifies the search process. The analytics system operates on an algorithm that uses Natural Language Processing, a sub-domain of Artificial Intelligence. Learns to initiate and understand a natural language / format created by a candidate and translates it

into a specific format. This acquired information is stored in the database. The app gets more information about candidates in its social profiles like LinkedIn and GitHub and updates the database. Listing Qualifications are: 1. Current Compensation 2. Expected Compensation 3. Education 4.

Professional 5. Venue 6. Very Early Date 7. Job Gap 8. Skills Total

3. PROPOSED WORK



- In the first step, the program receives the applicant's CV and makes a copy of it. Since it is not possible to perform tasks on file in pdf format, it is first converted to text.
- The text is then provided with a pre-trained model that only captures the data associated with the system. After this step, the extracted data will be displayed to the job applicant in a formatted format.
- The system allows the job applicant to edit the details of this document before performing any work on it. Once the applicant is satisfied and submitted the information, it can be submitted to compare the job description requirements that have been transferred to the previously trained model and requirements such as years of experience, qualifications, etc. already known.
- Now, work data (after passing a pre-trained model) is compared to resume data (after passing a pre-trained model) and then can calculate points based on a formula.
- Finally, the web application will display the profile of the recruiter and other applicants for the same profile in the rank list based on their score.

4. METHODOLOGY

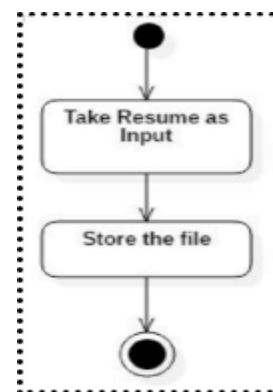
The implementation of the project to evaluate and evaluate the resume according to the job requirement submitted by the company employer has various modules. These modules are subdivided into 2 sections, which are as follows: - 1. Job Applicant Side 2. Server Side

A. Job Applicant Side

In this section, the applicant side of the project or client side project is discussed. It contains the following modules, namely, 'Receiving resume as input' and 'data output module'.

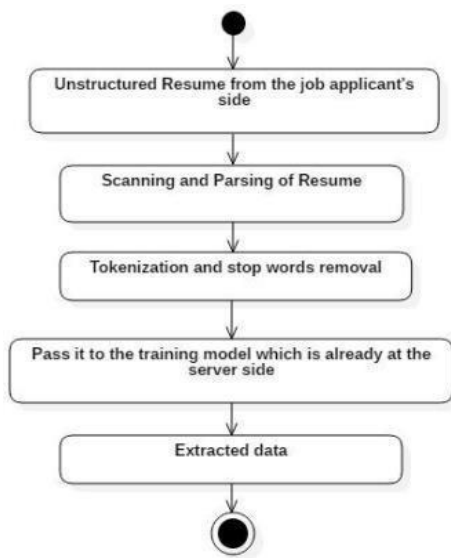
1) Receiving resume as input

The proposed web application will have 2 types of users: - the job applicant side and the employers side. The company's employer will have access to the website. The employer will be able to post vacancies available in his company. The applicant will also have a login that will allow him or her to upload his or her resume as submission. This reboot of the received input in the .pdf file format will then be saved on the website. Since MySQL and any other primary SQL can store pdf files directly on the site, base64 text will be used to store this pdf resume file. The diagram shows the process of retrieving inputs and storage on a website.

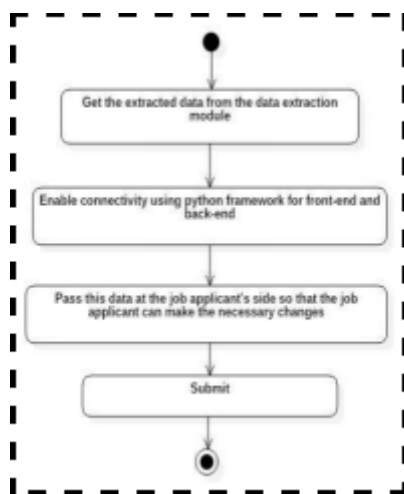


2) Data output module

This module deals with extracting data from a resume given to job applicants as ideas. This data extraction was done using phase-based classification by Natural Language Processing. This module will only extract relevant data from job postings on the applicant's resume. It is considered the informal resume of the applicant. The entire text is then transferred. After this, all the non-essential words like 'this', 'for', 'is' etc. After this, the remaining data becomes token. Only tokens contain relevant data regarding requirements. Transmission businesses are then considered and the text file is produced in the same way. The diagram below gives a diagram of the process diagram above.



The following section of the data extraction module will reflect important data extracted from the applicant. The module below is used to assess whether the applicant is satisfied with the data extracted from his or her resume. Exported businesses will be forwarded using a flash to the job applicant in the form of an edited text file. They are permitted to make any necessary changes to the text file in the event that any information is not extracted by the data extraction module. Once the applicant is satisfied with the changes made to his or her text file, he or she can post it on the website. The figure below shows the detailed mobility of the module mentioned above.

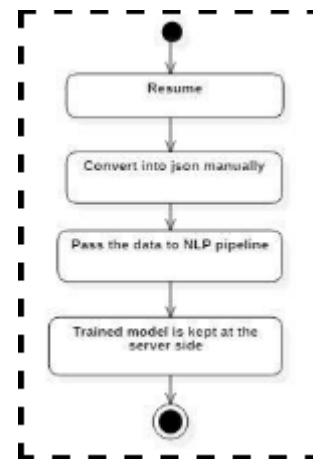


B. Server Side

In this part of the proposed system, the server side configuration is discussed. Contains modules, namely, 'Training data module' and 'Converting job description into the required format module'.

1) Training data module

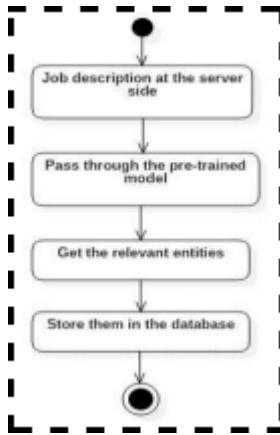
This module in a web application is about training a data set for a specific type of job posting, e.g., a senior web developer. The figure below shows the flow in which data is collected and trained. First of all, resumes related to a particular type of job posting are collected. Then the appropriate text for all restart is converted to a JSON file by manually uploading the ZIP file for all these restarts to a website portal called dataturks.com. It is possible to highlight only selected text that needs to be converted to JSON format in this portal. After converting text annotations to pdf format into JSON format, the JSON file is then transferred to the NLP file. The model was then trained using this NLP pipeline. It can be trained using SpaCy which is an NLP framework. Now, the data set is converted to the JSON format on the dathaturks website and the output is transferred to the NLP SpaCy pipeline for a trained model. In this method, slow reading is used to record important data in a ZIP file for continuous writing instead of typing word for word to create a database. This database is then divided into 2 parts: - Training data and test data and these components are transmitted via SpaCy pipeline.



2) Converting job description into required format

This module is about obtaining a job description or sending it to a text file that is required to be compared with a baptismal candidate and restarting a text file so that the candidate can restart points. For this to happen, job postings are transmitted via a professional model that matches the job position.

Companies that are eligible according to the qualified model will then be excluded from the job submitted by the employer. The text file will be done the same and this will eventually be stored on the website.



5.CONCLUSION

The proposed system will provide a better and more efficient solution to the current hiring system. This will provide potential candidates for the party and the candidate will be successfully placed in an organization that will appreciate his or her skills and abilities.

6.FUTURE WORK

The application can be extended further to the other domains like resume ranking, resume building and public sector jobs.

REFERENCES

- <https://core.ac.uk/download/pdf/55305189.pdf>
- <https://machinelearningmastery.com/natural-language-processing/>
- <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- <https://www.techopedia.com/definition/30343/natural-language-toolkit-nltk>
- <https://towardsdatascience.com/python-libraries-for-natural-language-processing-be0e5a35dd64>