iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42295

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

CV Analysis Using Machine Learning

Avisha Anand[1], Mr. Sandeep Dubey[2]
[1]*Department of Computer Science and Engineering*
[2]*Guide*

Abstract: ***Recruitment is a tiresome process wherein the very first task of a recruiter is to screen resumes (Curriculum Vitae). Nowadays, many companies prefer online job application in comparison to paper resumes. The proposed system is designed in such a way that applying for job openings & screening could be made easy for job applicants as well as the recruiters. The recruiters from the various companies can post their requirements for particular job openings available in their respective companies and on the other hand will allow the job applicants to submit their resumes and apply for job postings they are interest in. The resumes submitted by the candidates are then compared with the job profile requirements as stated by the company's recruiter by using technologies like Machine Learning, Natural Language Processing which will not only help the recruiter to select the best candidates from the large pool of candidates but also help them in saving time that is invested in manual analysis of the resumes.***

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

I. INTRODUCTION

Recruitment is a 200-billion-dollar business that deals in hiring best fit candidates from a large pool of candidates having relevant skills for a given job profile. Scores of candidates mail their resumes to apply for the company's job openings for a position.

Any company having job opening for a position will have their mail inboxes flooded with thousands of emails from the aspiring job applicants every single day. Hence, in any recruiting process, the very first task for any recruiter is to scan all the submitted resumes of the job applicants. Research shows that 90% of all CVs/Resumes are checked for much less than 2 minutes via the employers. This indicates that in maximum cases employer study only the points of interest within the CVs/Resumes and ignores the rest. Thus, to make it simpler to research and recognize the necessary data, it should contain the precise segmentation scheme of general CV/Resumes.

However, it is very difficult task for a company's recruiter to manually go through thousands of resumes and select the most deserving candidate for the job. Since, out of those 1000 of resumes submitted, above 75% of them do not showcase the relevant skills required for the job profile as stated by the company's requirement due to which recruiters more often find it very difficult to pen down the most appropriate candidates from a large applicant pool.

In recent times, more than 50000 e-recruitment websites have been developed, with the usage of various approaches to identify the deserving candidates for a given job profiles of a company. The basic aim of these sites is to showcase the results to the candidates to which category they are best fit into by summarizing their CVs/Resumes on the basis of the keywords used in it. But these techniques have one of the major disadvantages of time complexity for acquiring the results.

The approach discussed in this paper is by using Machine Learning to train the dataset for a particular type of job position and section-based segmentation for data-extraction using Natural Language Processing. In order to improve the time efficiency of the proposed system, the candidate's resume will only be matched to those job openings where their skills matches with the job requirement descriptions of the company to which will in turn reduce the time complexity. Besides, resume matching results of all the candidates will be visible only to the particular company's recruiter.

This paper is structed as follows: Section 2 briefly describes the overview of the existing system available; Section 3 gives detailed design of the system which has been proposed to solve this problem by analysis of the applied algorithm, Section 4 includes a bird's eye view of implementation of this ongoing project and Section 5 presents the conclusion and illustrates the future scope.

II. LITERATURE REVIEW

In today's world, the recruitment process has witnessed a major change with the evolution of the technologies like Internet. Various researchers have contributed to the task of resume screening. The following section summarizes some of the literary work performed in this domain of e-recruitment systems.

©IJRASET: All Rights are Reserved

There have been more than 50000 online recruitment sites which ask the aspiring job applicants to submit the resume on their website, out of which some websites don't even have classification techniques for resume screening. So it is the job of the company recruiter to go through all of them manually, which is a tiresome process for them in order to select the most capable candidates for the subsequent rounds of hiring process. Some of these websites are Indeed, Monster.com, Adecco.com, Top Resume, Ideal, etc.

Let's consider a case study of such a website called Top Resume which uses the concept of Natural Language Processing to analyze aspiring job seeker's resume. Here, first the candidate will upload their resume on the portal and with the help of techniques like Natural Language Processing, only the text data from the submitted resumes are extracted and strength of candidate's profile is displayed in terms of percentage and additional attributes like percentage of candidate's skill is according to the keywords of words like education, certification course and candidate's work experience, which are even made visible to the candidate as the result of their resume screening but the websites does not contain any provision, according to which job applicants can apply only for a particular job opening nor does have any option of providing the recruiter of a company with a rank list of deserving candidates according to the relevant skills for the particular job position. Hence, there are many other such web applications available, providing mostly similar features.

III. PROPOSED SYSTEM

As an attempt to overcome some of the major disadvantages of the existing system, the proposed system will eventually help in reducing their workload of the company recruiters. Thus, the proposed solutions use various approaches with the aim of achieving automated screening of candidate's resume that mainly focuses on the content of the resumes where we perform the extraction of skills and related parameters to match candidates with the job description of the company.

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

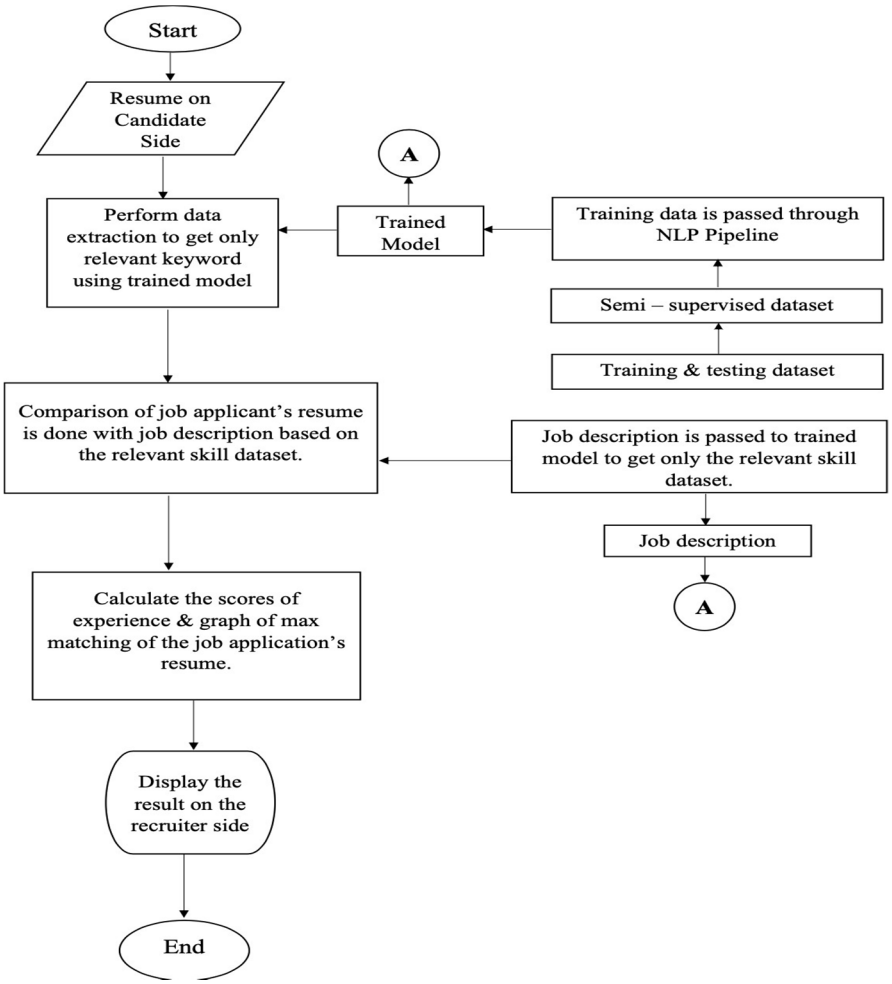


Fig: Activity Diagram showing how the overall system works

Since it is not possible to perform operations on a PDF file, so it is converted into text file. After conversion, on parsing relevant keywords is extracted from the LinkedIn and GitHub profile mentioned in the applicants resume which is also taken into consideration while scoring the candidates profile and these texts are fed to pertained model, which takes only those data which is relevant to the system. Then, these data are sent for comparison with the job description requirements which is already passed through pre-trained model. Now, the jobs requirement data is compared with the resume and data, after which it can calculate the score of the relevant experience and then the plot on the graphs of maximum matching of the job applicant's resume with the company's recruitment requirements. Finally, the screened resume of the aspiring candidates will be ranked based on the score the candidate's profile will be displayed on to the recruiter for further recruiting process. The proposed system for resume screening and rating according to the job requirement posted by a company recruiter has various modules mainly comprising of three parts which are as follows:

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

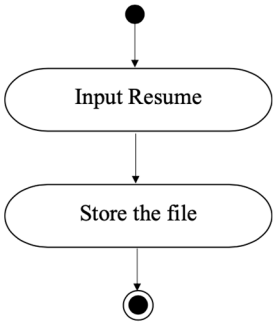


Fig: Activity diagram for accepting resume as input

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.

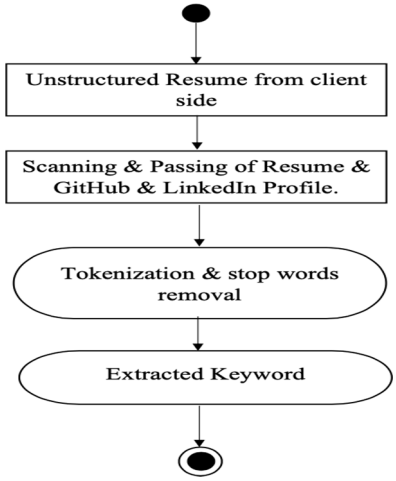


Fig: Activity Diagram showing how data extraction happens for resume input

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', 'Converting required skillsets into required format module'.

1) *Training Data Module:* This module deals for training of data for a particular type of job posting e.g.: Associate Software Developer. For this, first the resumes of similar type of job postings has collected, and then the relevant skills it takes from all the resumes are converted to JSON file (by manually uploading the ZIP files of all the resumes on website called dataturks.com), where only selected text could be easily converted into JSON format, then after that JSON file is passed through NPL Pipeline, where it is trained by using spaCy (an NLP framework) which is used for training general data instead of specific dataset. Thus, in order to make spaCy framework work according to the proposed system's needs, change NER (Named Entity Recognition) for the model so that the entities could be correctly identified for the raw data set. Next, these datasets on being converted to JSON format on dataturks website is passed through NLP spaCy pipeline to get the required trained model. Here, while using this approach any semi-supervised learning is used to label important data in a ZIP file of pdf resumes rather than manually typing each and every word for creating dataset, which is then split into two parts: training and testing dataset, which is then passed through spaCy pipeline.

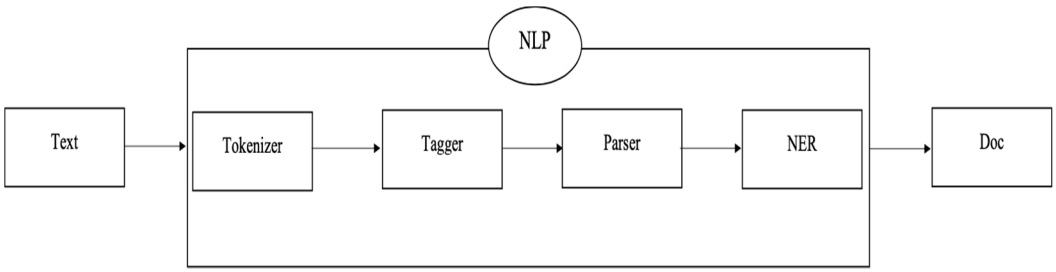


Fig: NLP Pipeline for Training Dataset

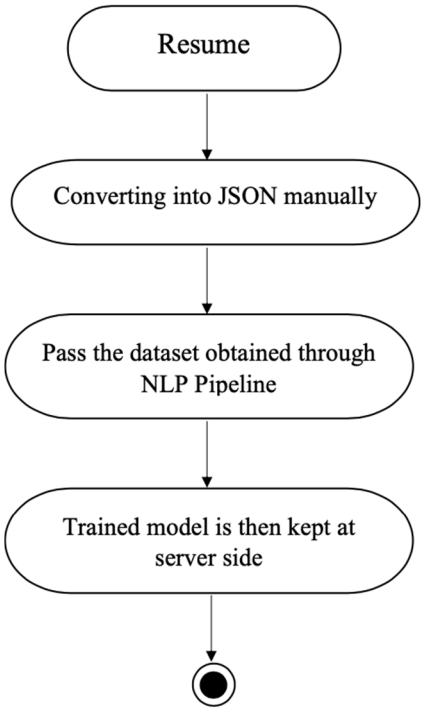


Fig: Activity Diagram for training dataset

2) *Converting Skillsets into Required Format Module:* This module deals with getting text file (made of the same name and stored in the database) of the relevant skills obtained from professional accounts like (LinkedIn and GitHub) mentioned in the candidate's resume that will be compared with the entities as per the trained model from job description uploaded by the recruiter.

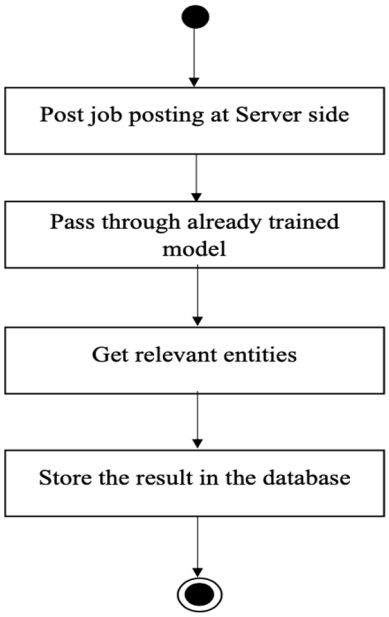


Fig: Activity Diagram showing job posting is converted into require

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 the Resumes, Inputted by the Candidates:* It is the last module but, that mainly deals with the calculation of score for a candidate's resume based on the job posting they have applied. According to this, rank list will be prepared with the candidate's profile receiving higher score placed at the top in comparison with those with least scores and accordingly graphs will also be prepared. This rank list and graphs who can make decision for the selection of next round of hiring process.

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

$$S = \frac{|\text{Sr} \cap \text{RSj}|}{|\text{RSj}|} * 50\% + \frac{|\text{Er} \cap \text{REj}|}{|\text{REj}|} * 20\% + \frac{|\text{Xr} \cap \text{RXj}|}{|\text{RXj}|} * 20\% + \frac{|\Sigma \text{Yw}|}{|\Sigma \text{Cw}|} * 10\%$$

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 post
Xr – 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%

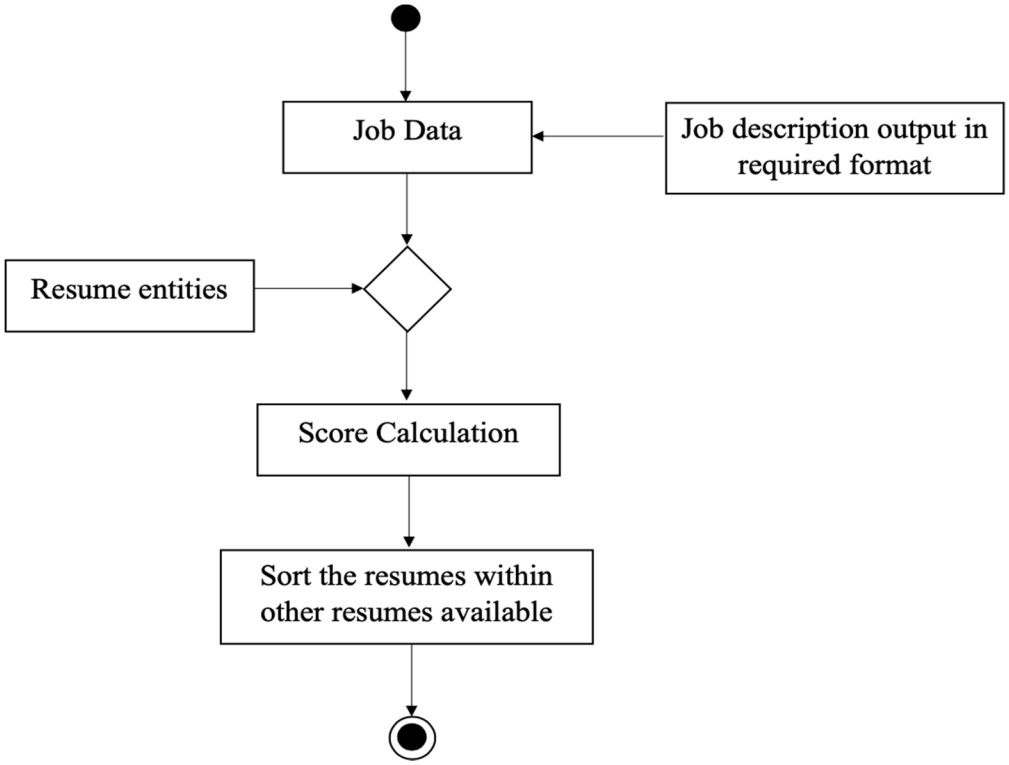


Fig: Activity Diagram showing comparison of candidate's profile with job description

IV. EXPERIMENTATION AND RESULTS

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

A. Screenshot of Training Model for NLP Pipeline

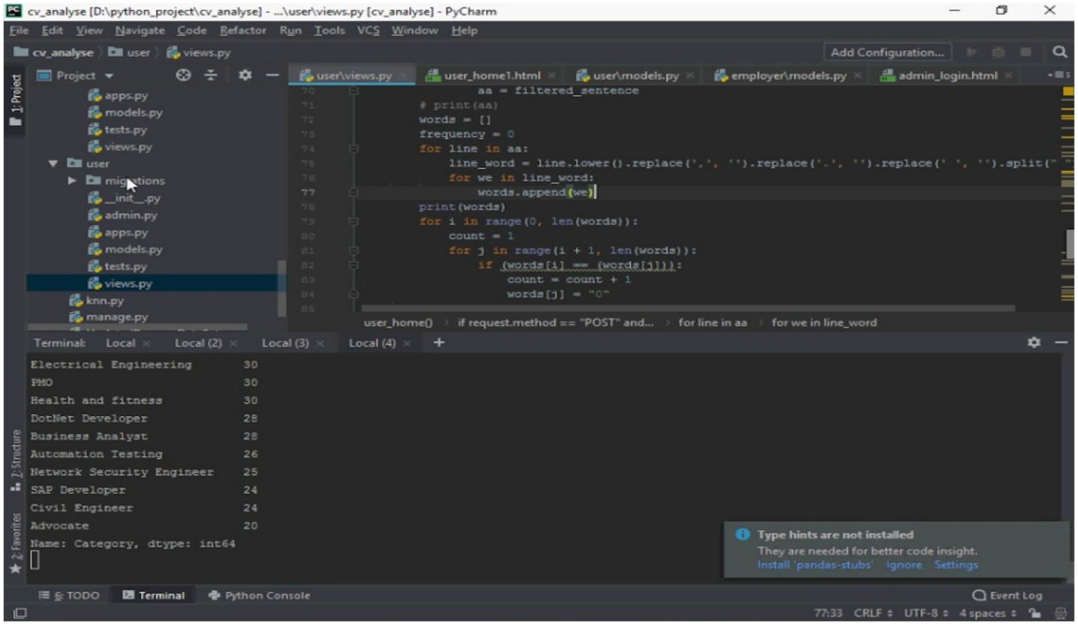


Fig: Training model for NLP Pipeline

B. Accuracy of Trained Model

In order for correct implementation so that system gives the expected accuracy, it is calculated using a model consisting of 220 resumes of which 20 are designed for testing and 200 for training the model.

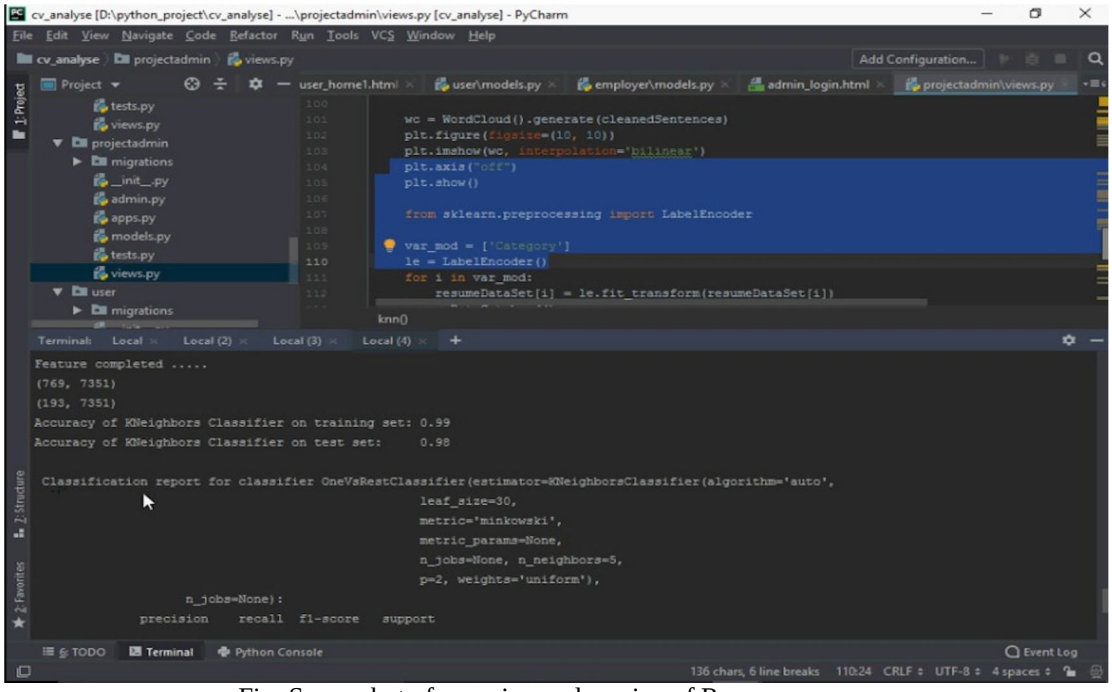


Fig: Screenshot of scanning and parsing of Resume

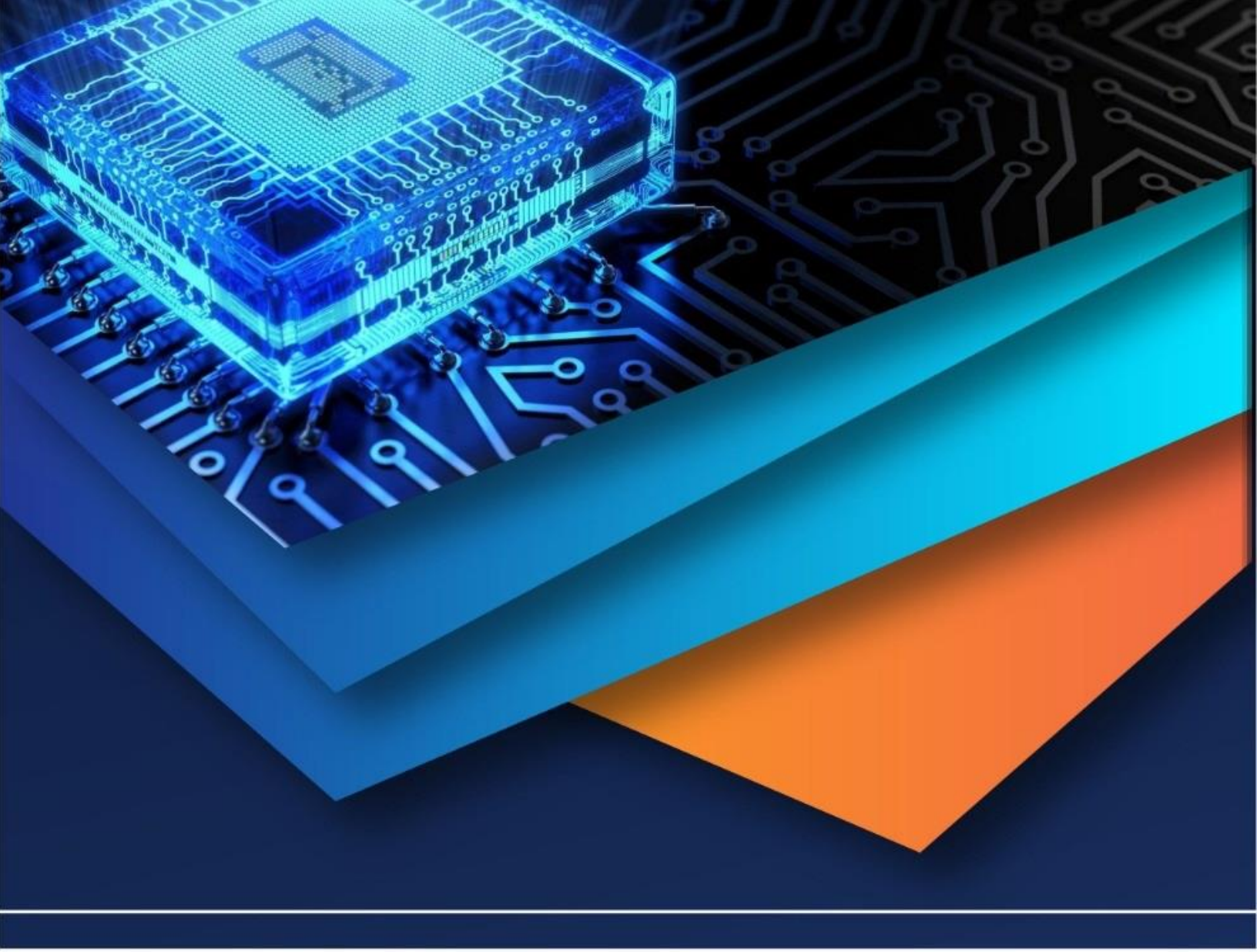
Here, the proposed system takes resume as input and passes it on through Natural language Toolkit from where scanned and parsed text data is produced, that is fed into next module and basic front end is developed.

V. CONCLUSION

The proposed system is under implementation and uses a semi-supervised learning (mainly K-nearest Neighbour) for achieving high accuracy. This system automates the process of requirements specifications and applicants ranking that are relatively highly consistent with those of the human experts. It will enable a more effective way to shortlist submitted candidates CVs from a large number of applicants providing a consistent.

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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)