iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52202

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 11 Issue V May 2023- Available at www.ijraset.com

Resume Parser Analysis Using Machine Learning and Natural Language Processing

Pratham Rasal[1], Yashashri Balwaik[2], Mahesh Rayate[3], Rohit Shinde[4], Prof. A. S. Shinde[5]
Information Technology, Savitribai Phule Pune University

Abstract: ***With the rise of online job application processes, submitting a resume has become easier than ever before. As a result, a larger number of individuals are impacted by this change. Many organizations still accept resumes by mail, which can cause challenges for their human resources departments. Sorting through a large number of applications to find the most suitable candidates can be a time-consuming process for these agencies. Job applicants submit resumes in a wide range of formats, including various fonts, font sizes, colors, and other design elements. Human resources departments are responsible for reviewing each application and selecting the most qualified candidate for the job. My suggestion is to incorporate natural language processing techniques into the project's parser to assist the human resources department or recruiting manager in analyzing the information provided in resumes. This approach can involve using keyword matching and other natural language processing methods to identify the most suitable candidates and obtain the most effective resumes. By doing so, the organization can improve its recruitment process and identify the best candidates for the job.***
Keywords: Resume, Natural Language Processing, Machine Learning, Support Vector Machine

I. INTRODUCTION

Through different accessible channels, including the firm site, outside sites, work adverts, work references, and so on, competitors can go after a job.

The employing system for an enterprise starts when competitors submit applications for open positions.

The proposed Worker Suggestion Framework is used to pick people from school grounds and other enlistment processes. Basically, possibility for occupations answer the enlistment staff, who then, at that point, test them to check whether they fit the bill for a specific assignment. This is a difficult calling that requires a lot of relational contact during the recruiting system. It could likewise include different tedious assignments like physically reaching competitors, checking the resumes, etc. The recruiting group should likewise really look at the up-and-comers' experiences to check whether they have taken part in any peculiar ways of behaving, such having a lawbreaker record.

Enlisting organizations and business undertakings survey various continues day to day. That isn't an undertaking for individuals. It would be great if a modernized keen machine would extricate all the significant data from unstructured continues and overhaul it in a routinely organized design that could in this way be positioned for a specific capability. The name, email address, online entertainment profiles, individual sites, long periods of work insight, fine art models, long stretches of training, instructing studies, distributions, certificates, volunteer encounters, key expressions, and at some point or another the group of the resume are completely remembered for the parsed measurements.

An information base is then used to store the parsed records for ensuing utilization. A CV contains an abundance of data on an individual's achievements and abilities in all circles of life. The estate lady who is going after the job stresses the significant components and ranges of abilities required by the business. Many messages from individuals who send their resumes to go after a position are gotten by worldwide partnerships.

The ongoing test is realizing which resumes ought to be arranged and shortlisted in view of the cutoff points. This resume scanner saves you time and diminishes how much exploration you should do. Each set incorporates data about an individual's experience, proficient experience, or instructive foundation.

The project will also focus on evaluating the performance of the resume parser by comparing its results with manually parsed resumes. Metrics such as accuracy, precision, recall, and F1 score will be utilized to assess the effectiveness of the developed system. Additionally, the project will explore techniques to handle various challenges in resume parsing, including noisy data, inconsistent resume formats, and linguistic variations. The outcomes of this project have the potential to enhance the hiring process, improve candidate selection, and increase overall productivity in the recruitment industry.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

[1] This technique communicated parsing of the resumes with least limit and the parser works the utilization of a couple of standards which train the call and address.Scout bunches use the CV parser system for the confirmation of resumes. As resumes are in dumbfounding plans and it has different sorts of certifiable components like set up and unstructured evaluations, meta experiences, etc.The proposed CV parser approach gives the part extraction method from the moved Cv's. [2] It follows a philosophy of 4 stages, the essential stage was to get the data (resume) and convert them into coordinated association and subsequently play out the examination using significant learning systems. Second step consolidates the psychometric test where the text mining is used to create scores for each newcomer. In the third step they perform web scratching on various virtual amusement regions to get the additional information about the contenders and recommend fitting situations to them. In the fourth step, the system will propose the capacities and essentials wherein the students are missing and moreover help them with getting chosen in the best organization. [3] This work employs brain associations and CRF to segment and extract various types of information from resumes. A CNN model is used for segmentation, and a Bi-LSTM model is used for extraction. They extract and process data from personal, educational, and professional blocks. The results are promising, and the resulting JSON file contains 23 data fields.

[4] The objective of this project is to develop an automated resume parser system that can parse resumes based on the job profile and convert unstructured resumes into a structured format. The system also includes a ranking structure that ranks resumes based on extracted information such as skills, education, experience, name, etc. CV parsing is utilized to extract relevant data from resumes. The proposed parser uses rules to extract name and address information, and is capable of parsing resumes in different formats and languages. This method provides an effective way of extracting data from resumes that may contain both structured and unstructured data, as well as metadata. This approach has become increasingly popular among recruitment agencies and employers for resume screening and selection. [5] The work described in presents a method for evaluating the emotional content of resumes based on various quality criteria. The algorithmic system can be used by recruiting managers and human resources departments to evaluate resumes and identify the most suitable candidates based on emotional content, clarity, and perspective.This method involves using a text coherence-based approach to group resumes and assess their emotional inclusion in terms of clarity and perspective. The resulting assessments are then converted into a comprehensive quality rating. All of the limits were in general officially dressed into a solidified 1 to 5 rating scale for accomplice a quality estimation for resumes. The abstract evaluation results gained through the algorithmic system were amicable to and were consequently supported through the knowledge of gatherings.

III.MOTIVATION OF SYSTEM

The current hiring process requires candidates to manually enter all of their qualifications and information, which adds time and complexity. The HR team also requires more workers to review the applications of the applicants.The resumes are also in unstructured format and in different formats like pdf, docs, etc, this makes the job of the company HR team and recruiting team's work difficult and time consuming. Overall, the challenges faced by the hiring team and HR teams in the traditional recruitment process have highlighted the need for a more flexible and automated method for recruitment. The use of resume parsing tools can significantly simplify the recruitment process and improve the efficiency of the HR team, ultimately leading to better hiring decisions.All paragraphs must be indented.

IV.IMPLEMENTATION DETAILS OF MODULE

A. Support Vector Machine

1) *Import the dataset:* This involves obtaining the dataset from a source, such as a file on disk or an online database. Thdata can be in a variety of formats, such as pdf, docs, etc.
2) *Explore the data to figure out what they look Like:* Once the data is imported, you need to take a closer look at it to understand its structure, format, and characteristics. This can involve visualizing the data using charts, graphs, or other visualizations, or using statistical analysis techniques to identify patterns or trends.
3) *Pre-process the data:* Before you can use the data with an SVM algorithm, you may need to clean it up and prepare it for analysis. This can involve removing missing or invalid values, normalizing the data, or transforming it into a format that is easier to work with or a format that you consider as a standard format.
4) *Split the data into attributes and labels:* In order to use an SVM algorithm, you need to separate the input data (attributes) from the output data (labels). This allows you to train the algorithm to recognize patterns in the input data that correspond to the output labels. Once the data is split into training and test sets, we separate the attributes (resume information) and labels (categories/classes) for each set. This allows us to establish a clear distinction between the input features and the target variables for our machine learning algorithm.

5) *Divide the Data into Training and Testing Sets:* To evaluate the performance of the SVM algorithm, you need to split the data into two parts: a training set and a testing set. The training set is used to train the algorithm, while the testing set is used to evaluate how well the algorithm generalizes to new data.
6) *Train the SVM Algorithm:* With the data split into training and testing sets, you can now use the training set to train the SVM algorithm. This involves selecting appropriate hyperparameters for the algorithm.
7) *Make some Predictions:* Once the SVM algorithm has been trained, you can use it to make predictions on new data. This involves feeding the input attributes to the algorithm and obtaining the predicted output labels.
8) *Evaluate the Results of the Algorithm:* Finally, you need to evaluate the performance of the SVM algorithm on the testing set. This can involve calculating metrics such as accuracy, precision, recall, and F1 score, or using visualizations to compare the predicted labels with the actual labels.

B. Random Forest Algorithm :

1) *Select random K data points from the training set :* In this step, a subset of K data points is randomly selected from the training set. This is done by sampling the data points with replacement, which means that the same data point can be selected multiple times.
2) *Build the decision trees associated with the selected data points (Subsets) :* In this step, a decision tree is built using the selected subset of data points. The decision tree is a model that maps input features to output labels by recursively splitting the data based on the values of different features. The decision tree is built by selecting the best feature to split on at each node.
3) *Choose the number N for decision trees that you want to build :* In this step, you need to choose the number N of decision trees that you want to build. This is a hyperparameter that can be tuned to achieve the best performance on the validation set. Typically, a larger number of decision trees will result in a more accurate random forest, but will also increase the computational cost.
4) *Repeat Step1 and Step2 :* In this step, Steps 1 and 2 are repeated N times, resulting in N decision trees that are trained on different subsets of the training data.

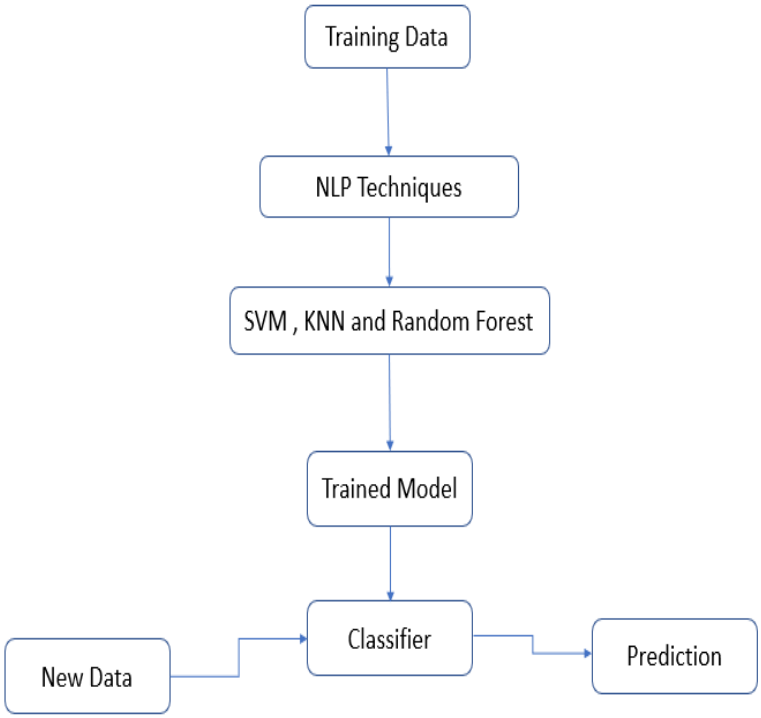


Figure – a. Block Diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

V. RESULT

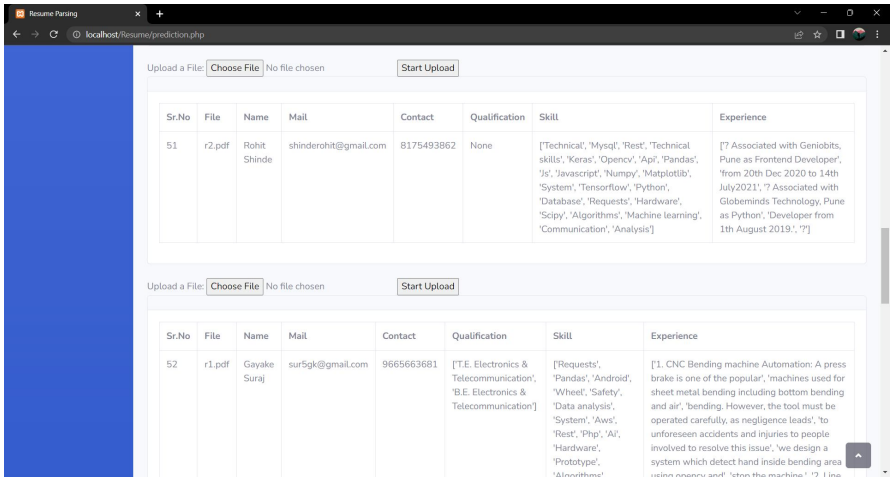


Figure – b. Features Extraction

	precision	recall	f1-score	support
Civil Engineer	1.00	0.80	0.89	5
Electrical Engineer	1.00	1.00	1.00	4
Mechanical Engineer	1.00	1.00	1.00	4
Software Developer	1.00	1.00	1.00	1
Software Engineer	1.00	1.00	1.00	1
Student	1.00	1.00	1.00	1
Web Developer	0.00	0.00	0.00	0
accuracy			0.94	16
macro avg	0.86	0.83	0.84	16
weighted avg	1.00	0.94	0.97	16

Figure – c. Accuracy, Recall, F1-Score, Precision

VI. CONCLUSIONS

The proposed approach is a comprehensive plan that aims to revolutionize the way organizations hire their employees. The goal is to improve the efficiency and effectiveness of the recruitment process, making it easier for both employers and candidates. One of the primary objectives of the approach is to streamline the hiring process by eliminating unnecessary steps and automating certain tasks. This will ensure that the process is completed in a timely manner, saving both the employer and candidate valuable time

Moreover, the approach focuses on delivering highly qualified candidates to the firms. To achieve this, the approach will use a ranking system that is based on the technical skills of the candidates. This will ensure that the firms receive resumes of candidates who have the necessary qualifications and expertise to perform the job duties. The result generated shows precision, accuracy, F1-Score and other statistics. Ultimately, the approach aims to make the work of companies and candidates easier and more effective. By providing a streamlined recruitment process, employers can quickly and easily identify qualified candidates,

while candidates can apply for jobs with confidence, knowing that they will be evaluated fairly and objectively. The outcomes of this project have the potential to enhance the hiring process, improve candidate selection, and increase overall productivity in the recruitment industry.

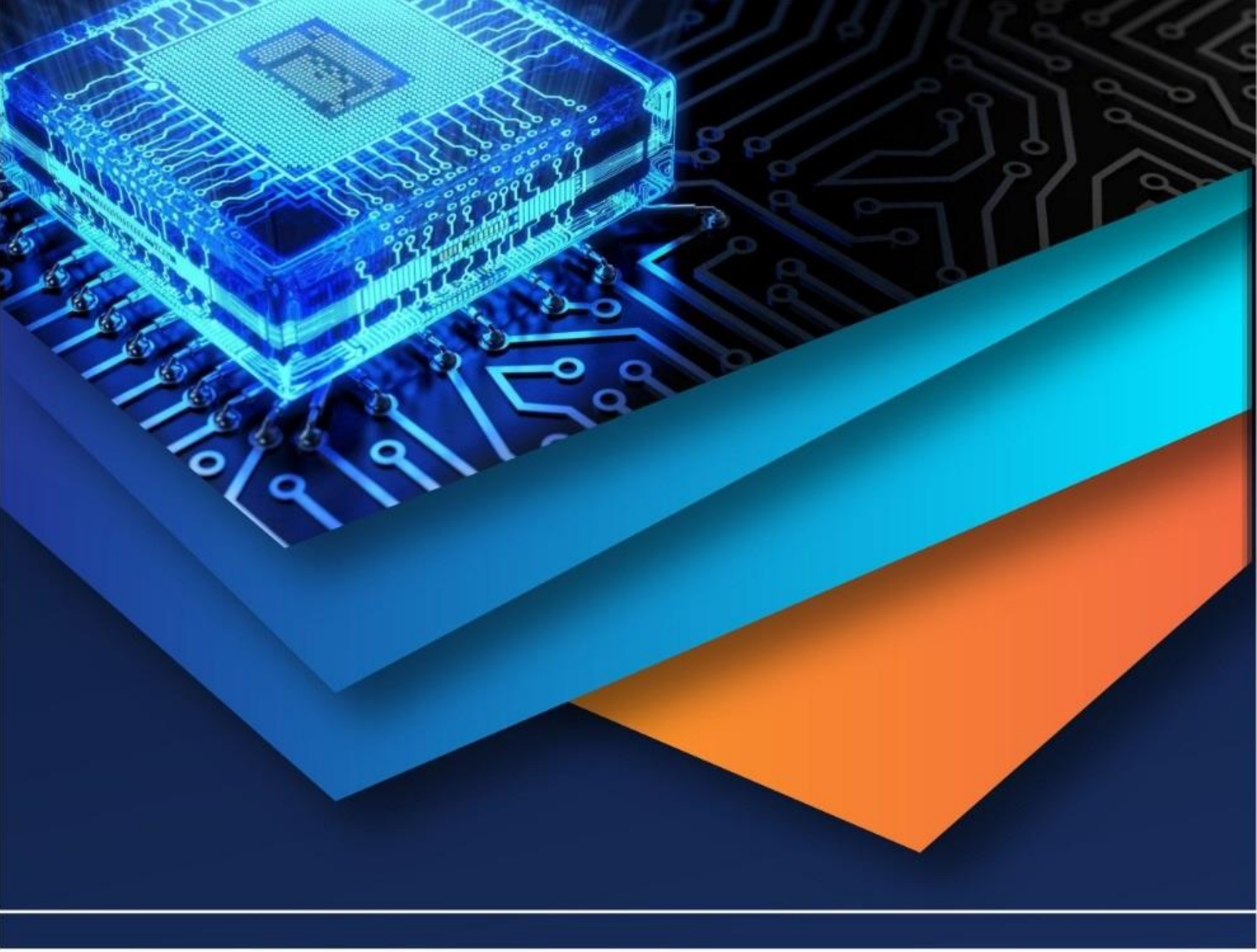
VII. FUTURE SCOPE

The main future scope of our project is to parse resumes from different applications and recruiting websites like LinkedIn, Naukri.com, Indeed, Firstnaukri, Hirect, Internshala, etc. In coming future, this system can be made more flexible in which wide ranges of psychometric tests will be added. As a future work, we can enlarge the resume dataset and improve the performance of the proposed system.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh, "NLP based Extraction of Relevant Resume using Machine Learning ",2020, IJITEE.
[2] Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S,Kaveri BW, "E-Recruitment System Through Resume Parsing, Psychometric Test and Social Media Analysis", 2019,IJARBEST.
[3] Ayishathahira and Sreejith,, "Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing",International CET Conference on Control, Communication and Computing(IC4),2018.
[4] Papiya Das, Manjusha Pandey, Siddharth Swarup Rautaray, "A CV parser Model using Entity Extraction Process and Big Data Tools ", 2018,IJITCS
[5] *Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal "An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes".*



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)