

---

## RESUME PARSER AND ANALYZER USING NLP

Nimish Patil<sup>\*1</sup>, Shubham Yadav<sup>\*2</sup>, Vikas Biradar<sup>\*3</sup>

<sup>\*1,2,3</sup>Bachelor Of Engineering, Department Of Information Technology, Vidyalkar  
Institute Of Technology, Mumbai, Maharashtra, India.

---

### ABSTRACT

In company's or organization's recruiting process, for a single opening there are multiple applying candidates. First process in recruiting procedure is to go with resumes one by one for the selection of best candidate. But this is time consuming and hectic process for recruiters. The reason behind this is each applicant has its own unique resume with different sections and different formats. So, it is not possible for recruiters to go with resumes one by one. To minimize the efforts of recruiters and investment of timing in this process we have proposed a system where recruiters can easily analyze resumes in simple file formats with the ranking. This project is based on Natural Language Processing (NLP). Two major components of this system are Job Seekers and Recruiters. Applicants will upload their resume in different formats mainly pdf or doc. Then parser will parse these resumes for field extraction, after analysis system will scale the resume and will give suggestions like required skills to be included in resume or courses to be done or fields to be included in resume to increase its rank. This ranking will be saved in databases where recruiter will get idea about most deserving candidates for a particular role.

**Keywords:** Natural Language Processing (NLP), Resume Parser, Resume Analysis, Part-of-speech tagging, Named Entity Recognition (NER).

---

### I. INTRODUCTION

Organization has to deal with thousands or sometimes lakhs of resumes for single job profile. So, it is not possible for recruiters to analyze them one by one. Also, we know that these resumes do not have proper formats. Each resume might be different from others in sections like format, fields, patterns, etc. Though only some information is common or required for eligibility like individual's name, email, education, skills, year of experience, projects, internships, etc. Recruiters only want to look upon this fields. Resume analyzing process may take too much time also it requires more efforts. Hence there is option for recruiters to use Resume parser and Analyzer which will save their efforts and time. Resume parsers structure the unstructured number of resumes into easily readable formats. This process of converting unstructured resumes into structured format is known as resume parsing. Here only required fields which are required are extracted from resumes. So, HR's work become easy. Second user of this system are job seekers. They will apply for role they want. How would they know that where their resume stand. So, our system will analyze the field which are extracted from resumes and according to them system will suggest some areas where resume can be improve like required skills to be added in resume, quality of resume, enhancements required.

### II. LITERATURE SURVEY

**Paper Name:** Natural Language Processing (almost) from Scratch by Ronan Collobert, Jason Weston L'eon Bottou, Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa.

This paper has proposed neural network architecture which can be used for different natural language processing tasks like part-of-speech tagging, name entity recognition, chunking, etc. This paper is focused on benchmarking of proposed neural network using four standards of NLP such as Part-of-Speech tagging (POS), Name Entity Recognition (NER) and Semantic Role Labeling (SRL). POS labels each word with unique tag that represents its syntactic role. Chunking is also called as Shallow Parsing. Chunking labels segments of sentence with syntactic parts such as noun or verb or phrases. Named Entity Recognition atomic elements from sentence into categories such as 'person' or 'location' or 'organization', etc. Semantic Role labelling – It gives semantic role to syntactic constituents of sentence. In this experimental study, authors used F1 scores over chunks for models in three tasks NER, CHUNK and SRL. For POS evaluation, pre-word accuracy is used.

**Paper Name: Resume Information Extraction with A Novel Text Block Segmentation Algorithm by Shicheng Zu, Xiulai Wang and Seth Darren.**

In era of rapid development of deep neural networks, this proposed system fills the gap between lack of systematic investigation in resume parsing using neural networks. In this proposed system, they normalize the resume parsing process by focusing on six important information fields of resume. The fields are personal information, education, work experience, projects, skills and publications. As these are important factors in any resume, authors didn't focus much on other trivial information such as interests and hobbies, leadership and references. Authors found that, BLSTM's performance was better in text block classification and robustness. For named entity recognition, BLSTM-CNNs-CRF algorithm was found effective.

**Paper Name: A Few Shot Approach to Resume Information Extraction via Prompts by Chengguang Gan, Tatsunori Mori.**

Prompt learning shows great performance in most text classification tasks with few training examples, It is useful for NLP methods where samples are dynamic, In this paper, authors proposed a system where they created multiple sets of templates manually and verbalizers based on textual characteristics of resumes. They compared performances of Masked Language Model (MLM), Pre-training Language Models (PLMs), and Seq2SeqPLMs. They further improved the design of prompt templates and verbalizers for knowledgeable Prompt-Tuning. Experiments conducted by authors show that, verbalizers designed based on their rules were more effective and robust than existing manual templates and automatically generated prompt methods.

### **III. METHODOLOGY**

There are several techniques and algorithms are available to solve NLP based problems. As these problems are based on Deep Learning concepts, Python Language is preferred. Libraries available in Python language such as nltk and spacy are used to extract text/information from documents. For text cleaning, Regular Expression (RE) library is used. NLTK and Spacy libraries are used for Natural Language Processing (NLP) related tasks like eliminating stop words, extracting root words, POS, NER. Pre-processing data is a difficult task. As text preprocessing is the initial stage of any NLP project, text pre-processing is done in order to prepare the text data.

**The following are some of the pre-processing steps:**

- Removing Stop words
- Lower casing
- Tokenization
- Lemmatization

**Steps in Parsing:**

- a) Extract the text from pdf
- b) Extract Text from Doc files
- c) Detect the file extension
- d) Extract the entities
- e) Extract the email
- f) Extract Name from Resume
- g) Extract Mobile Number
- h) Extract Skills from the resume
- i) Extract Technical Skills
- j) Extract Education and Year
- k) Extract Experience

After parsing analysis is done on extracted fields and stored in database which can be used by recruiter's visualization and decision process also it will be used for job seeker's suggestions.

#### IV. MODELING

The proposed system is a responsive web application developed using Python and streamlit framework. Natural Language Processing will be done using nltk and spacy libraries available in Python. For analysis of extracted fields from the resume different techniques like regular expressions will be used. Based on analysis of the resume system will calculate score of the resume and according to the score it will give suggestions. Matching of required skills and skills in resume will be done and suggestions will be given. This data of applicants will be saved in database where recruiters will use this data for decision process and visualization. Recruiters will be able to download reports of candidates based on their resume which are minimalistic and effective for selection process, this system is included in Automatic Tracking System (ATS) of an applicant.

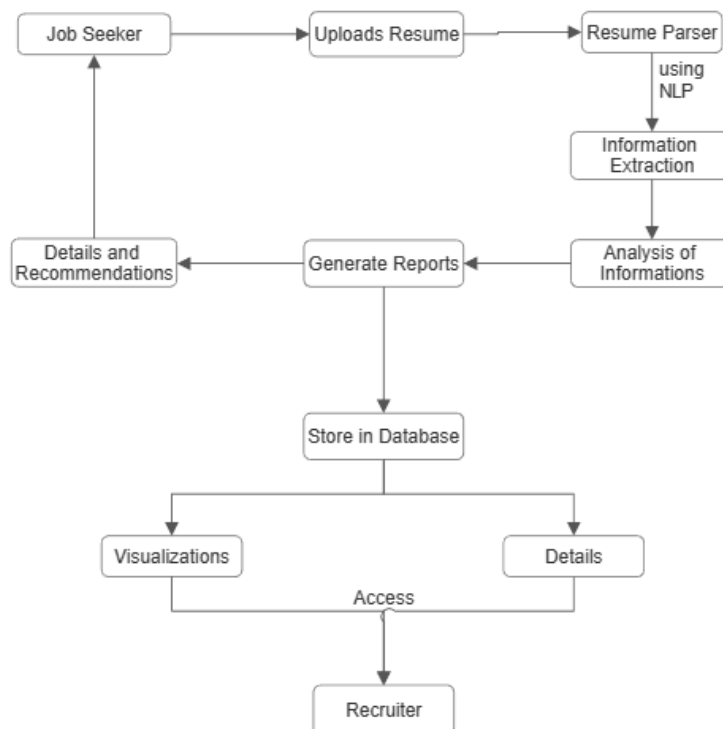


Figure 1: System Architecture

#### V. RESULTS AND DISCUSSION

Step 1: Applicant needs to upload his/her resume to the portal for the analysis. Fig. 2 shows result of the step 1.

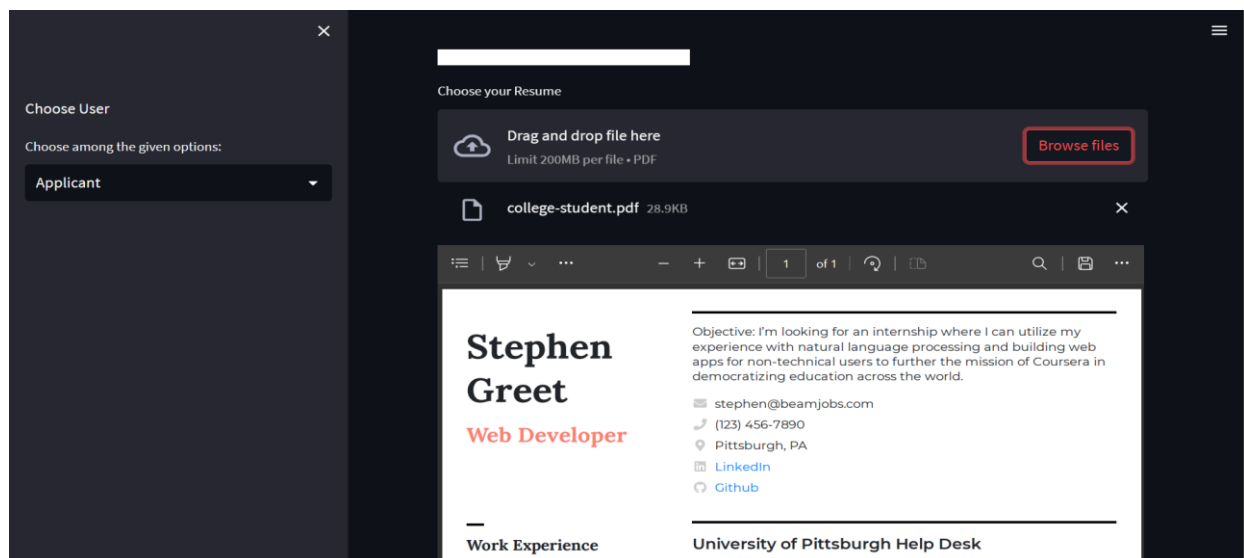
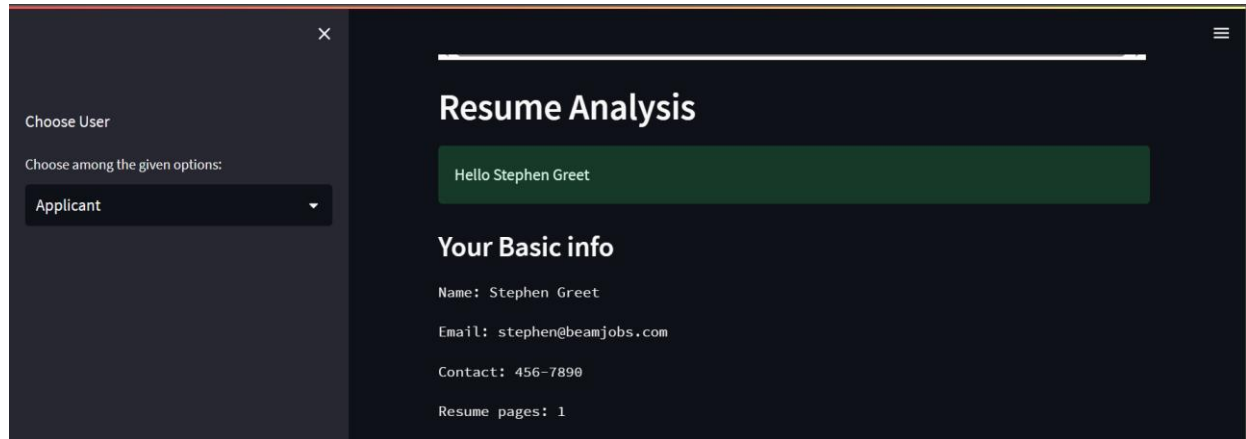


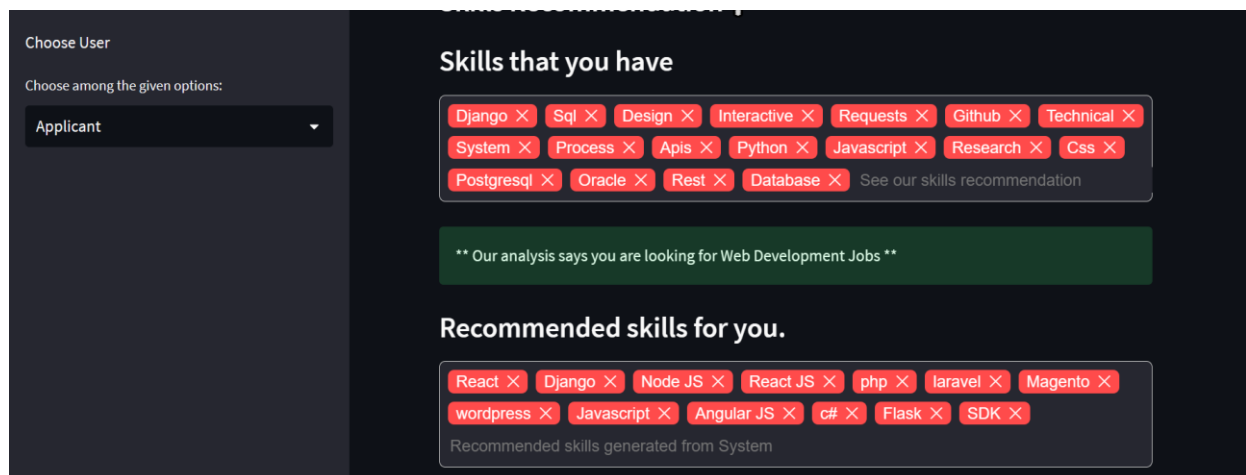
Figure 2: Resume Uploading by Applicant

Step 2: Resume parser will extract the basic information from the resume about applicant. Fig. 3 shows result of current step.



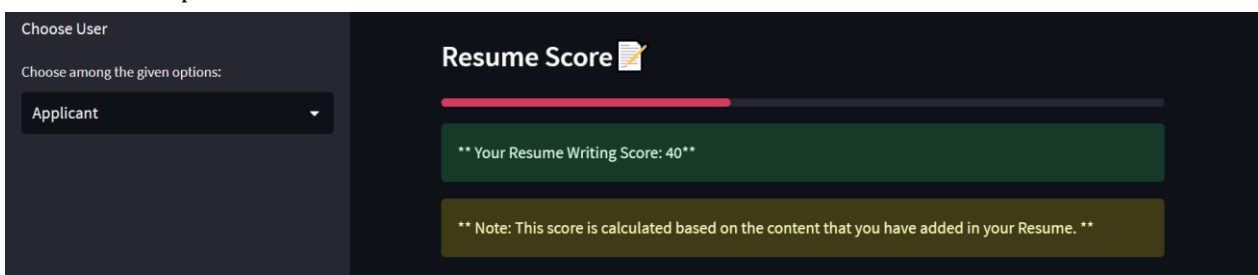
**Figure 3: Basic Information**

Step 3: Resume Parser will extract skills that job seeker have. Analyzer will compare the skills to the pre-defined skills and will suggest required skills to the applicant. Fig. 4 shows results of this step.



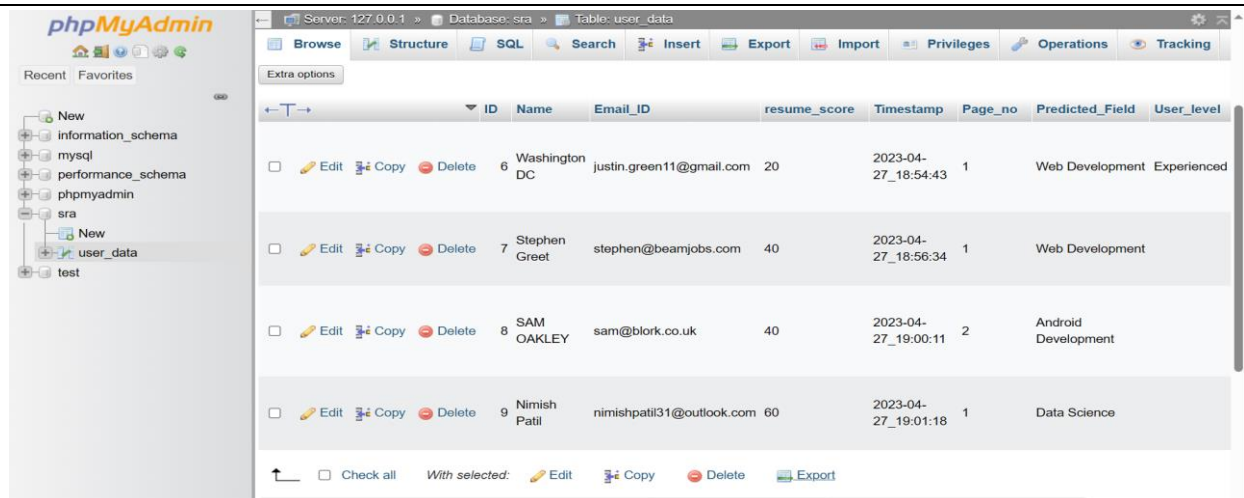
**Figure 4: Recommended Skills**

Step 4: Analyzer will analysis the fields present in the resume and based on that will give score. Fig. 5 shows result of this step.



**Figure 5: Resume Score**

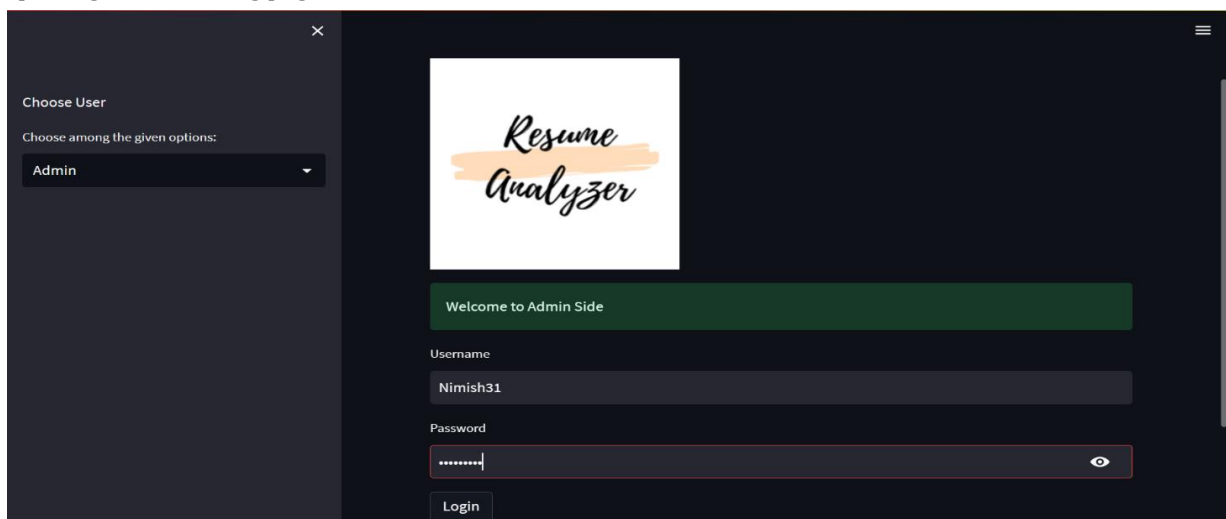
Step 5: These all extracted fields will stored into database. Fig. 6 shows the database where information is stored.



| ID | Name          | Email_ID                  | resume_score | Timestamp           | Page_no | Predicted_Field     | User_level  |
|----|---------------|---------------------------|--------------|---------------------|---------|---------------------|-------------|
| 6  | Washington DC | justin.green11@gmail.com  | 20           | 2023-04-27_18:54:43 | 1       | Web Development     | Experienced |
| 7  | Stephen Greet | stephen@beamjobs.com      | 40           | 2023-04-27_18:56:34 | 1       | Web Development     |             |
| 8  | SAM OAKLEY    | sam@blork.co.uk           | 40           | 2023-04-27_19:00:11 | 2       | Android Development |             |
| 9  | Nimish Patil  | nimishpatil31@outlook.com | 60           | 2023-04-27_19:01:18 | 1       | Data Science        |             |

Figure 6: Database

Step 6: Fig. 7 is a landing page of a recruiter.



Choose User

Choose among the given options:

Admin

Welcome to Admin Side

Username

Nimish31

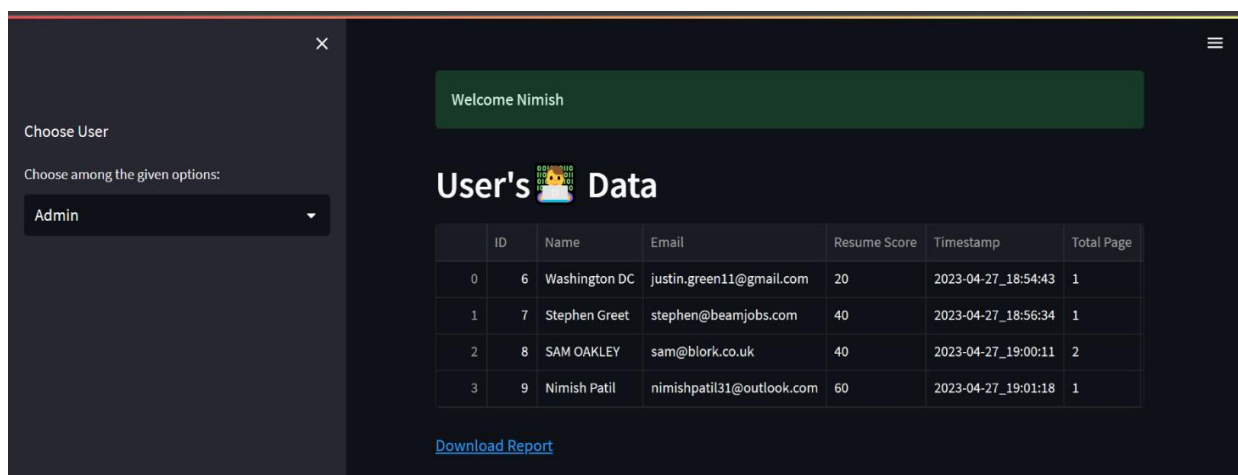
Password

.....

Login

Figure 7: Admin Page

Step 7: Admin as Recruiter can see the applicants information along with the resume score, so that he can take decisions wisely about selecting candidates based on their skills and other factors. Fig. 8 shows reports that Admin can see on portal. Admin also has option to download these reports. Fig. 9 shows the downloaded report in CSV format.



Welcome Nimish

User's Data

|   | ID | Name          | Email                     | Resume Score | Timestamp           | Total Page |
|---|----|---------------|---------------------------|--------------|---------------------|------------|
| 0 | 6  | Washington DC | justin.green11@gmail.com  | 20           | 2023-04-27_18:54:43 | 1          |
| 1 | 7  | Stephen Greet | stephen@beamjobs.com      | 40           | 2023-04-27_18:56:34 | 1          |
| 2 | 8  | SAM OAKLEY    | sam@blork.co.uk           | 40           | 2023-04-27_19:00:11 | 2          |
| 3 | 9  | Nimish Patil  | nimishpatil31@outlook.com | 60           | 2023-04-27_19:01:18 | 1          |

[Download Report](#)

Figure 8: Reports

|   | A  | B           | C           | D         | E         | F          | G                 | H          | I            | J           | K   | L | M | N | O | P | Q | R |
|---|----|-------------|-------------|-----------|-----------|------------|-------------------|------------|--------------|-------------|---|---|---|---|---|---|---|---|
| 1 | ID | Name        | Email       | Resume Sr | Timestamp | Total Page | Predicted         | User Level | Actual Skill | Recommen    | Recommended Course  |   |   |   |   |   |   |   |
| 2 |    | 9 Nimish Pa | nimishpat   | 60        | 2023-04-2 | 1          | Data Science      |            | ['Tkinter',  | ['Data Visi | ['Machine Learning Crash Course by Google [Free]', 'Machine Learning A-Z by Udemy'    |   |   |   |   |   |   |   |
| 3 |    | 8 SAM OAKL  | sam@blor    | 40        | 2023-04-2 | 2          | Android Developme |            | ['Android'   | ['Android'  | ['Android Development for Beginners [Free]', 'Android App Development Masterclass     |   |   |   |   |   |   |   |
| 4 |    | 7 Stephen G | stephen@    | 40        | 2023-04-2 | 1          | Web Development   |            | ['Django',   | ['React',   | ['C['Front End Web Developer by Udacity', 'Python and Django Full Stack Web Developer |   |   |   |   |   |   |   |
| 5 |    | 6 Washingtc | justin.gree | 20        | 2023-04-2 | 1          | Web Deve          | Experienci | ['Testing',  | ['React',   | ['C['Django Crash course [Free]', 'Python and Django Full Stack Web Developer Bootcam |   |   |   |   |   |   |   |
| 6 |    |             |             |           |           |            |                   |            |              |             |   |   |   |   |   |   |   |   |
| 7 |    |             |             |           |           |            |                   |            |              |             |   |   |   |   |   |   |   |   |
| 8 |    |             |             |           |           |            |                   |            |              |             |   |   |   |   |   |   |   |   |

Figure 9: Downloaded reports in CSV format

Step 8: Fig. 10 shows the pictorial representation of applicants based on their skills.

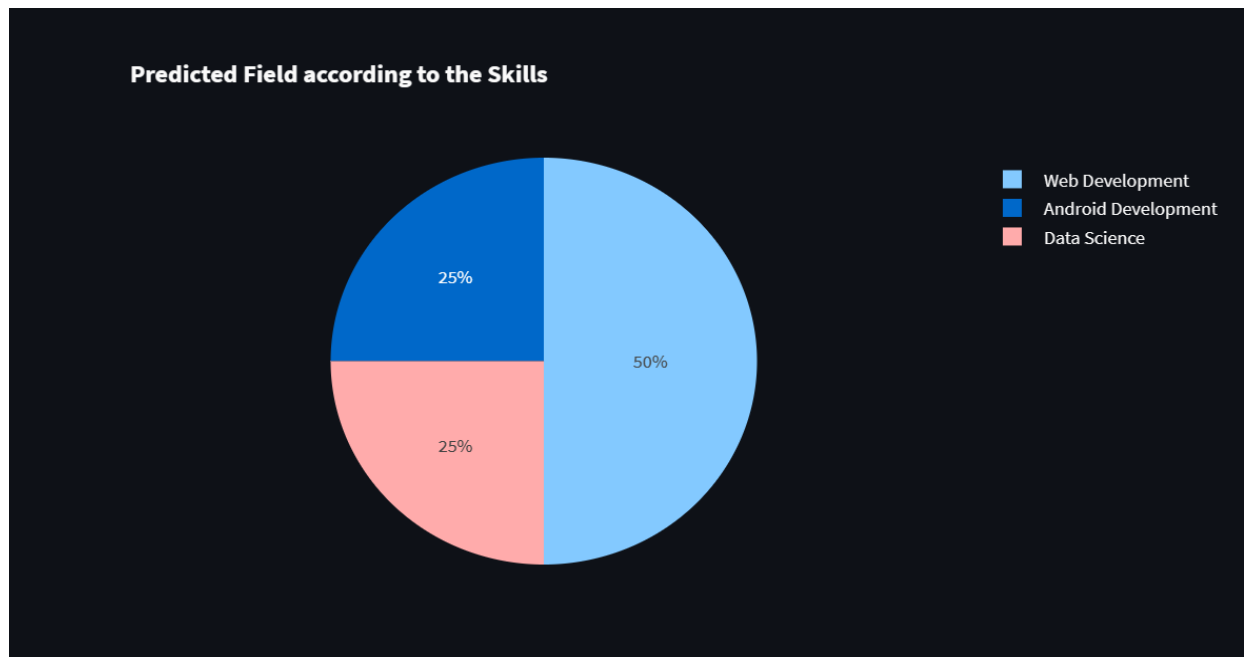


Figure 10: Pie chart of applicants based on their skills

Step 10: Fig. 11 shows pie chart based on the applicants experience level.

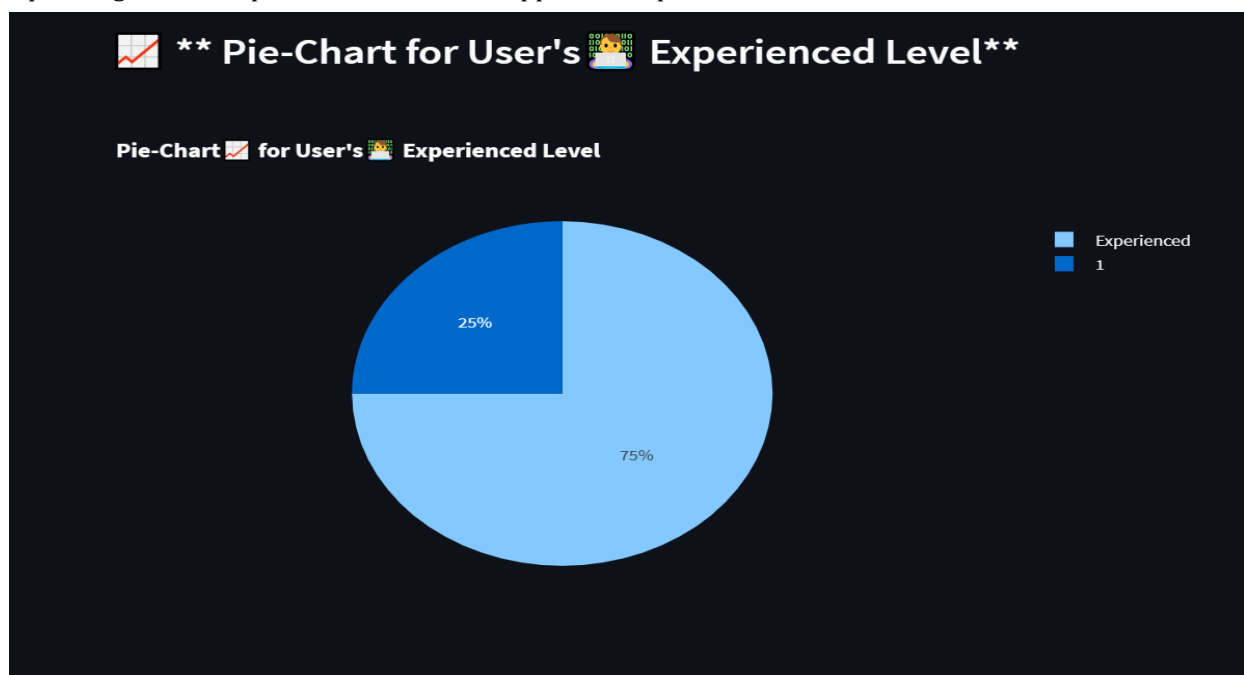


Figure 11: Pie chart of applicants based on their experience level

---

## **VI. CONCLUSION**

This paper represents brief overview of our Resume Parsing and Analyzing software system. Our system is able to parse most of the resumes in accurate way. Using of Natural Language Processing (NLP) for this system was best suitable option. As here we are dealing with raw data and for the processing of this data NLP is suitable. The power of computers is increasing day by day, hence most of the computers are now able to perform machine learning/deep learning tasks. Therefore, there is rapid increase in automation in every field. Our system fulfills one of areas where automation requires. When there are lakhs of applicants applying for same profile in an organization, there will be need of such software systems.

## **VII. REFERENCES**

- [1] Ronan Collobert, Jason Weston L'eon Bottou ,Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa "Natural Language Processing (almost) from Scratch" arxiv.
- [2] Shicheng Zu, Xiulai Wang and Seth Darren "Resume Information Extraction with A Novel Text Block Segmentation Algorithm".
- [3] Chengguang Gan, Tatsunori Mori "A Few Shot Approach to Resume Information Extraction via Prompts"
- [4] Agrawal, R., 2021. analyticsvidhya. Available at: <https://www.analyticsvidhya.com/blog/2021/06/must-knowntechniques-for-textpreprocessinginnlp/#:~:text=Text%20preprocessing%20is%20a%20method,text%20in%20a%20different%20case>. [Accessed 2022]
- [5] Bhatia, V., Rawat, P., Kumar, A. & Shah, R. R., 2019. End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT. arxiv.
- [6] Chavan, J., 2020. medium. Available at: <https://medium.com/@jeevanchavan143/nlptokenization-stemming lemmatization-bag-of-words-tf-idf-pos-7650f83c60be> [Accessed 2022].
- [7] Chen, J., Gao, L. & Tang, Z., 2016. Information Extraction from Resume Documents in PDF. Research Gate. 64
- [8] D., 2021. analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2021/06/text-preprocessing-innlp-with-pythoncodes/> [Accessed 2022].
- [9] Kopparapu, S. K., 2015. Automatic Extraction of Usable Information from Unstructured Resumes to Aid Search. ieeexplore.