A Major Project Proposal Report on

**Smart Resume Parser and Analyzer**

Submitted in Partial Fulfilment of the Requirements for

the Degree of **Bachelor of Engineering in Information Technology**

under Pokhara University

Submitted by:

**Mr. Binod Dhamala, 191410**

**Mr. Arjun Uchai Thakuri, 191405**

**Mr. Prasiddha Pokhreal, 191425**

**Mr. Swastik Chaudhary, 191447**

Date: 15 May 2024

****Department of IT Engineering

**NEPAL COLLEGE OF INFORMATION TECHNOLOGY**

Balkumari, Lalitpur, Nepal

# **ABSTRACT**

The pursuit of a reputable position in a company is a common goal shared by many people. For making this desire into reality, their resume is the key. Making these resumes is a challenging task for many and modifying it is another time-consuming and burdensome task. Taking these problems into consideration and to make eye-catching resumes we have developed a smart resume analyser that can analyse the user's resume, it intelligently identifies their skills and qualifications, enabling us to suggest the best-suited job titles for their profile. Furthermore, based on this analysis, our system generates recommendations for optimizing and enhancing the user's resume, making it more appealing to potential employers. Through the utilization of NLP and ML technologies, our resume analyser provides personalised and effective solutions to job seekers, helping them stand out in today's competitive job market. The information on the resume can be analysed and understood easily as NLP has the capability to understand and parse the information on the resume and extract the desired information efficiently. Our study explores resume parsing and analysis using SpaCy. By training the model with a dedicated resume dataset, we automate information extraction for personal details, work experience, education, and skills. The resumes are scored based on the job requirement and ranked according to the score obtained. This customized approach streamlines the recruitment process, allowing recruiters to efficiently evaluate and select candidates. Through SpaCy's tailored capabilities, resume analysis becomes faster and more effective.

**Keywords:** Resume Analyzer, NLP, ML, SpaCy’s, Information extraction

**Table of Content:**

[**ABSTRACT..** i](#_Toc166770389)

[**Table of Figure:** iii](#_Toc166770390)

[**1. Introduction** 1](#_Toc166770391)

[**2. Problem Statement** 2](#_Toc166770392)

[**3. Objectives** 3](#_Toc166770393)

[**4. Significance of Study** 4](#_Toc166770394)

[**5. Scope and Limitation** 5](#_Toc166770395)

[**5.1. Scope:** 5](#_Toc166770396)

[**5.2. Limitation:** 5](#_Toc166770397)

[**6. Literature Review** 6](#_Toc166770398)

[**7. Methodology** 7](#_Toc166770399)

[**7.1. Software Development Model** 7](#_Toc166770400)

[**7.2. Requirement Analysis** 8](#_Toc166770401)

[**7.2.1. Functional Requirements:** 8](#_Toc166770402)

[**7.2.2. Non-Functional Requirements:** 8](#_Toc166770403)

[**7.2.3. Software Requirements:** 8](#_Toc166770404)

[**7.3. System Architecture** 9](#_Toc166770405)

[**7.4. Tools and Technology** 12](#_Toc166770406)

[**7.4. Work Division** 13](#_Toc166770407)

[**7.5. Performance Evaluation** 14](#_Toc166770408)

[**8. Deliverables:** 15](#_Toc166770409)

[**9. Project task and time Schedule:** 16](#_Toc166770410)

[**10. Conclusion:** 17](#_Toc166770411)

[**Reference:….** 18](#_Toc166770412)

# **Table of Figure:**

[Figure 1: Step in Agile SDLC model 7](#_Toc166770512)

[Figure 2: System Architecture 10](#_Toc166770513)

[Figure 3: State Chart Diagram 11](#_Toc166770514)

[Figure 4: Compound Diagram 11](#_Toc166770515)

[Figure 5: Confusion Matrix 14](#_Toc166770516)

[Figure 6:Gantt chart Diagram 16](#_Toc166770517)

# **1. Introduction**

A smart resume screening system employs Natural Language Processing (NLP) to extract relevant information from unstructured resumes. The system intelligently identifies the candidate’s skills and qualifications, enabling us to suggest the best-suited job titles for their profile. Furthermore, based on this analysis, our system generates recommendations for optimizing and enhancing the user’s resume, making it more appealing to potential employers. Through the utilization of NLP and ML technologies, our resume analyser provides personalized and effective solutions to job seekers, helping them stand out in today’s competitive job market. With the additional help from Python and its packages, we can store the extracted data and give a rating based on the analysis of the stored data. After the rating is generated, we recommend some modifications like adding sentences like my hobbies are, my goal is, add objectives, and add declaration and more. The traditional process of sifting through resumes can be time-consuming and inefficient, with employers often receiving hundreds or even thousands of resumes for a single job opening. NLP is about enabling computers to understand human language, which is a challenging task. Computers are good at comprehending structured data like spreadsheets and databases, but they struggle with unstructured data such as text and speech. NLP helps computers process and understand this type of data, which can be found in various forms. This interface of proposed method also offers benefits for job seekers, as they can receive personalised feedback on their resumes and recommendations for how to improve their chances of being hired. With Smart Resume Analyser, job seekers can ensure that their resumes are optimised for the positions they are applying for, increasing their chances of landing their dream job. Thus, the Smart Resume Analyser interface of proposed method has the potential to revolutionize the job application process, making it more efficient and effective for job seekers.

# **2. Problem Statement**

Manual resume sorting and ranking present several significant challenges that impact the efficiency and effectiveness of the recruitment process. This method is inherently time-consuming and labor-intensive, requiring recruiters to meticulously review each resume individually. During periods of high applicant volume, this can become overwhelming, making it difficult to manage within tight hiring timelines and leading to delays in identifying and engaging top candidates. Accurate tracking, organizing, and retrieving specific resumes is another major issue with manual sorting. Without a structured system, resumes can easily get lost or misfiled, resulting in missed opportunities and incomplete candidate data, complicating efforts to maintain a comprehensive talent pool for future needs. Objectively comparing candidates with diverse backgrounds and experiences is particularly challenging. Human biases, both conscious and unconscious, can inadvertently influence evaluations, potentially overlooking qualified candidates who do not fit conventional profiles. This lack of objectivity can lead to less equitable hiring practices.

Furthermore, scalability is a critical concern. As organizations grow and the volume of applications increases, the manual process becomes increasingly unsustainable. Recruiters may struggle to keep up with the influx of resumes, creating bottlenecks that slow down the entire hiring process and negatively impact the candidate experience. These challenges underscore the need for more streamlined, automated solutions to enhance the efficiency and fairness of recruitment practices.

## **3. Objectives**

* The primary objective of our proposed system is to develop a custom resume parsing and analysis system that makes job recruitment much easier and faster.
* Explore techniques to extract text from resumes in different formats.

## **4. Significance of Study**

The Smart Resume Analyser project holds substantial significance in both the job-seeking and recruitment landscapes. For job seekers, the tool offers a transformative advantage by providing tailored feedback and recommendations to enhance their resumes. By leveraging NLP and ML technologies, the system ensures that resumes are optimized for specific job roles, significantly improving candidates' chances of securing interviews and landing their desired positions.

For employers, the project addresses critical inefficiencies in the traditional recruitment process. Manual sorting and ranking of resumes are often time-consuming, labor-intensive, and prone to errors. The automated approach of the Smart Resume Analyser streamlines these tasks, enabling faster, more accurate, and objective evaluation of applicants. This not only reduces the workload for recruiters but also ensures a more fair and consistent hiring process.

Furthermore, the scalability of the system allows organizations to handle growing applicant pools without bottlenecks, facilitating a smoother and more efficient recruitment process. Overall, the Smart Resume Analyser enhances the quality and efficiency of job applications and recruitment, benefiting both job seekers and employers.

**5. Scope and Limitation**

Our website provides the following scope and limitation:

## **5.1. Scope:**

* The system will target the extraction and analysis of key information from resumes, including personal details, work experience, education, and skills.
* The scope also includes the automation of the resume parsing process, providing an efficient solution for recruiters to streamline candidate evaluation and selection.

## **5.2. Limitation:**

* System’s Accuracy may be affected if resumes do not provide complete or accurate information. Candidates may omit certain details, exaggerate their qualifications, or include irrelevant information.
* Job requirements and company needs to evolve over time. If the system is not updated regularly, it may become less effective in ranking resumes based on the latest criteria
* Dealing with resumes in different languages adds another layer of complexity.

**6. Literature Review**

This is the first resume short lister present on the internet when searching for one through a browser. It seems to want a job description for the job you are looking for and uses that as a basis for shortlisting your cv by finding out relevant terms related to the specific job or position. It gives you a score on your cv, a resume analysis, a guide for optimizing your resume and uses LLMs (large language models) to compare resume with job description. We want our app to work like this one but at a lower level to start before implementing more features.

**CV Analysis Using Machine Learning [7]**

This particular report published in the iJRASET (International Journal for Research in Applied Science and Engineering Technology) makes a similar model using NLP and tests its accuracy using various means. In particular it uses and shows a formula to calculate the score using various measures such as skillset, experience, required work years etc. Each are given different weightage and used in a formula to calculate the score. This inspired us to make a similar formula of our own to calculate a score.

**7. Methodology**

## **7.1. Software Development Model**

Agile is an iterative and incremental software development approach that focuses on flexibility, collaboration, and continuous improvement. It involves dividing the project into short sprints, regularly delivering working software, gathering feedback, and adapting the product based on customer needs. The process emphasizes close collaboration between stakeholders and development teams, allowing for quicker response to changes and delivering value early in the development lifecycle.

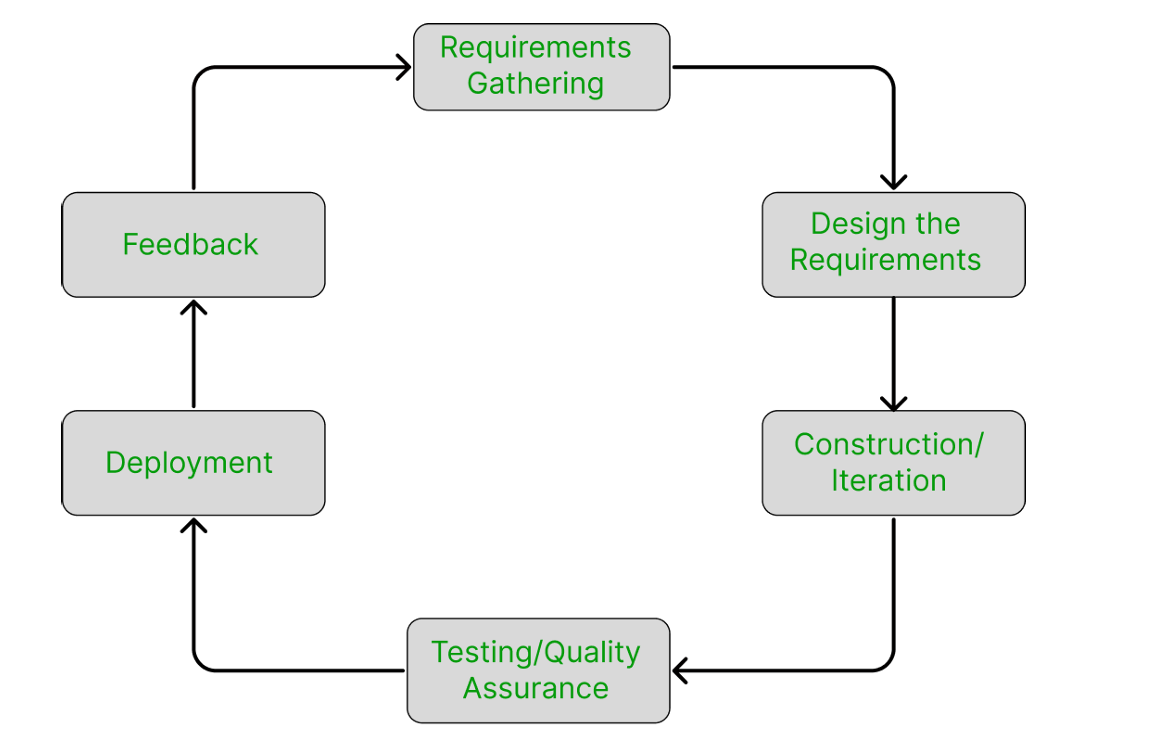


Figure 1: Step in Agile SDLC model

## **7.2. Requirement Analysis**

### **7.2.1. Functional Requirements:**

* Extract essential resume details accurately and comprehensively from various formats and structures.
* Effectively identify and extract specific entities like names, contact details, job titles, companies, institutions, and skills.
* The system should employ a matching algorithm to evaluate how well a candidate's resume aligns with the job description.
* Perform validation checks for data accuracy, consistency and include a ranking mechanism for efficient candidate evaluation and selection.

### **7.2.2. Non-Functional Requirements:**

* Stable and consistent performance with minimal downtime
* The user interface should be visually appealing and easy to navigate, enhancing the overall user experience.
* Flexible to adapt to different resume formats and user requirements. 4.3. Hardware Requirements
* Adequate processing power, memory (RAM), and storage capacity to handle computational demands and store data efficiently.

### **7.2.3. Software Requirements:**

* Natural Language Processing (NLP) Library: SpaCy
* Database Management System: MySQL
* Programming Language: Python, Javascript
* Version Control: Git and Git hub
* Integrated Development Environment (IDE): Visual Studio Code, Jupyter Notebook

## **7.3. System Architecture**

Architecture Description The architecture of the proposed system consists of the following components:

**User Interface:** The system provides a user-friendly web interface where job seekers can upload their resumes, and organizations/recruiters can list job descriptions.

**Backend:** The backend will handle the incoming requests from the frontend and manage the entire processing flow. It will host the NLP models, perform data processing, and implement the ranking algorithm.

**Data Storage:** A database is utilized to store the uploaded resumes, job descriptions, and the extracted resume information. It ensures efficient data retrieval and management.

**Training Data Preparation:** A diverse resume dataset is collected and prepared for training the spaCy NER model. The dataset includes various resume formats, layouts, and terminologies to enhance the model's adaptability.

**Training Model:** The prepared training data is used to train a custom NER model on a labeled dataset to recognize named entities (e.g., skills, education, experience). SpaCy is a good choice for implementing NER models.

**Parsing and Analysis**: When a resume is uploaded, the system applies the trained spaCy NER model to extract relevant information such as personal.details, work experience, education, and skills. The extracted data is stored in the database for further processing.

**Job Description Matching:** The system compares the extracted resume information with the requirements specified in the job descriptions. Factors such as skills, experience, and other relevant criteria are evaluated to assess compatibility.

**Scoring and Ranking:** Based on the evaluation, resumes are assigned scores and ranked to facilitate efficient sorting. This enables recruiters to identify top-matched candidates easily.

**Output and Reporting:** The system generates a summarized output, providing recruiters with a concise overview of the most suitable candidates. This assists in streamlined candidate selection and decision-making.

The system architecture ensures seamless integration and coordination between the different components, enabling efficient resume parsing, analysis, and candidate evaluation. It leverages the custom-trained spaCy NER model and a well-structured data flow to enhance accuracy and performance in resume processing and matching.

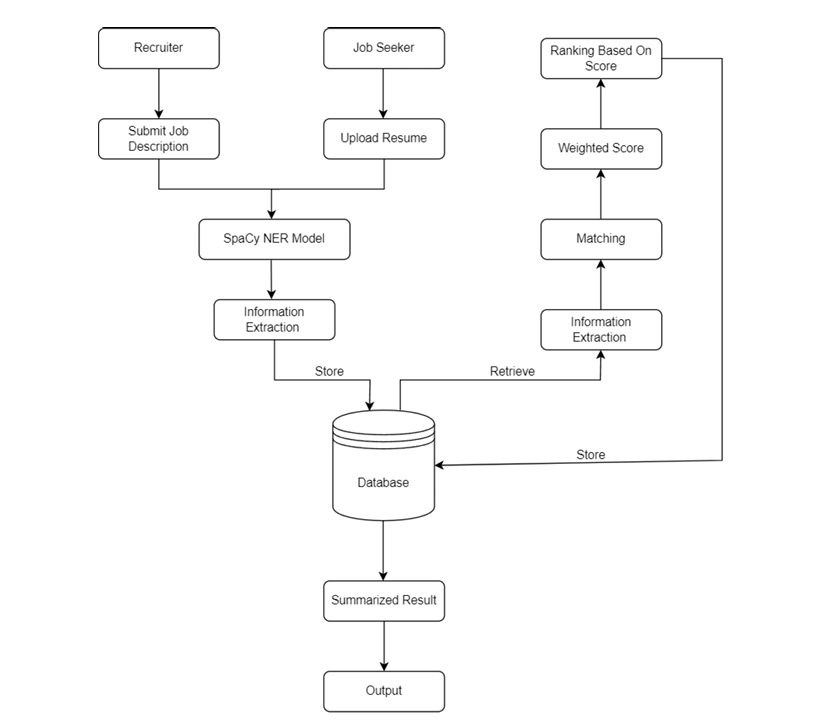
****

Figure 2: System Architecture

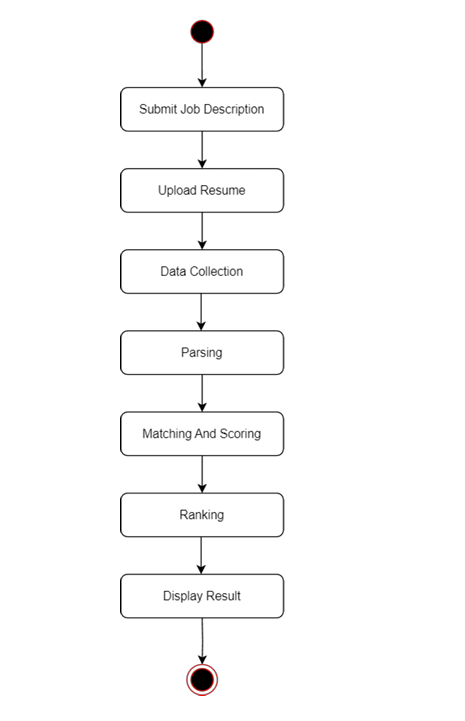


Figure 3: State Chart Diagram

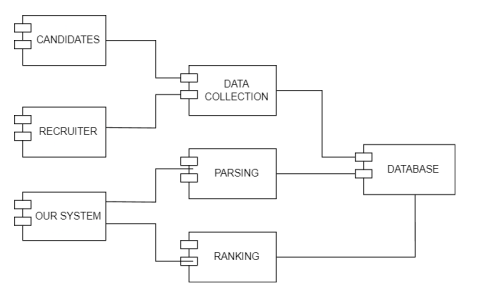
s

Figure 4: Compound Diagram

## **7.4. Tools and Technology**

* Languages:
  1. Python
  2. Javascript
* Framework:

a. Django

* Database:

a. MySQL

* Libraries:
  1. ReactJS
  2. SpaCy
  3. Pathlib
* Text editor:
  1. VS Code
  2. Jupyter Notebook, Google Colab
  3. Visualization tools: Matplotlib

## **7.4. Work Division**

## **7.5. Performance Evaluation**

Confusion matrix will be used for performance evaluation.

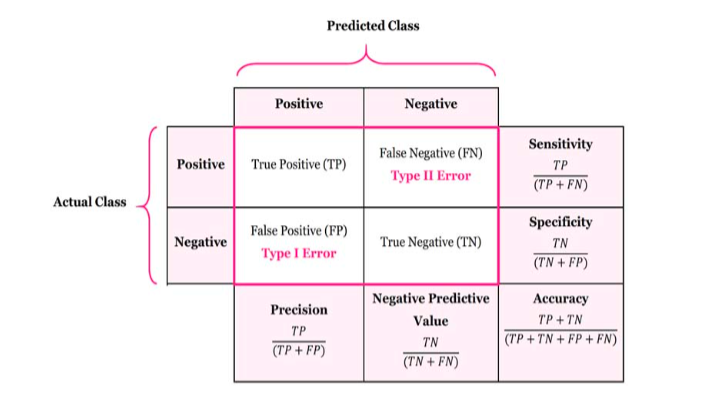


Figure 5: Confusion Matrix

A confusion matrix is a tabular summary of the number of correct and incorrect predictions made by a classifier. It is used to measure the performance of a classification model. It can be used to evaluate the performance of a classification model through the calculation of performance metrics like accuracy, precision, recall, and F1-score.

**8. Deliverables:**

* It should provide user-friendly interface where users can upload their resumes and recruiter can submit job descriptions.
* The complected system should provide the automated output from the resume with organized information, accurately and efficiently extracting details such as name, skills, experiences, and much more.
* The details obtained from the above process should be used by our system to rank the candidates based on which recruiters can focus on the most promising candidate, thereby reducing time and effort.

**9. Project task and time Schedule:**

The following chart presents the tentative schedule for the project.

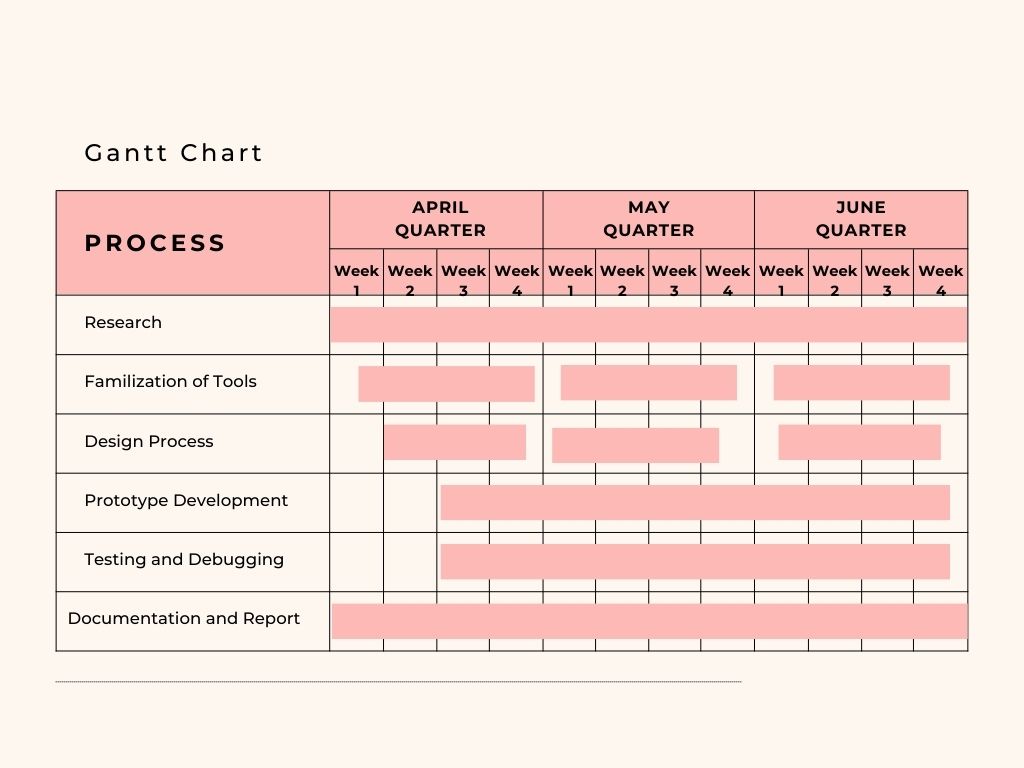


Figure 6:Gantt chart Diagram

**10. Conclusion:**

In conclusion, Smart resume parsing and analysis offers significant benefits for recruiters by saving time, streamlining the hiring process, and enabling more informed decisions. Techniques such as text extraction, contact detail capturing, and leveraging spaCy’s pattern matching with regular expressions and keyword matching ensure accurate retrieval of information, including skills, education, and candidate names. Hands-on experience confirms the practical application and potential of resume parsing, ultimately revolutionizing recruitment practices. By implementing a Smart resume parsing and analysis, recruiters can enhance efficiency and effectiveness, leading to better hiring outcomes. The system will be able to accept or reject a job applicant based on two factors the company's requirements must match the skills listed in the applicant's resume, and the test evaluation will be based on the applicant's skills, ensuring that the resumes uploaded by the applicant are genuine and the applicant is truly knowledgeable about the skills. Building a Smart resume parser requires a combination of technical skills, domain knowledge, and attention to detail. With the right approach and tools, we can develop a powerful resume parser that automates the extraction of crucial information from resumes, saving time and effort in the recruitment process.

**Reference:**

[1] Raj and N. Singh, "A Systematic Literature Review (SLR) On The Beginning of Resume Parsing in HR Recruitment Process & SMART Advancements in Chronological Order," *Journal of Computational Science,* vol. 48, pp. 100-115, 2022.

[2] Gupta and N. Gupta, "The Impact of Smart Resume Parsing on the Recruitment Process," *Human Resource Management Review*, vol. 31, no. 3, p. 100426, 2021.

[3] S. Gupta and R. Mittal, "A Survey on Resume Parsing Techniques," *Journal of Big Data,* vol. 7, no. 1, p. 13, 2020.

[4] Agarwal and M. Agarwal, "A Review of Resume Parsing and Analysis Tools," *International Journal of Information Management,* vol. 49, p. 102070, 2020.

[5] P. Gupta and R. Gupta, "The Future of Resume Parsing and Analysis*," Future Computing and Information Systems,* vol. 5, no. 1, pp. 1-10, 2020.

[6] M. Thejaswee, V. Srilakshmi, K. Anuradha, G. Karuna, *Performance Analysis of Machine Learning Algorithms for Text Classification*, in Proceedings of the Advanced Informatics for Computing Research(ICAICR 2020), A. K. Luhach, D. S. Jat, K. H. Bin Ghazali, Gao, P. Lingras, (eds), Comm. Comp. Inform. Sci. Springer, Singapore 1393 (2021)