

Smart Resume Analyser: A Case Study using RNN-based Keyword Extraction

Patlolla Sruthi^{1}, P. N. V. K. G. Adithya¹, MD. Suleman¹, Pallerla Kunal¹, Surya Prakash Gairola²*

¹Department of AIMLE, GRIET, Hyderabad, Telangana State, India.

²Uttaranchal Institute of Technology, Uttarakhand University, Dehradun, 248007, India

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 to it. 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. With the additional help from Python and its packages like streamlit, pymysql, etc. 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. We also suggest a few sources through which the user can enhance their skills.

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

* Corresponding author: sruthinag.reddy@gmail.com

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 Literature survey

Satyaki Sanyal and team [1] represents the approach of resume analysis software that automates the extraction and analysis of relevant information from uploaded resumes. The software employs NLP, ML, and data mining techniques to identify keywords, patterns, and trends. It evaluates resumes based on job requirements, filtering and ranking them accordingly. This approach simplifies recruitment, reduces recruiter workload, and identifies the most suitable candidates for a job. Vaidya and team [2] the approach described here is a method used by resume analysers to extract relevant information. The process involves eliminating stop words and applying the Soundex algorithm to group similar-sounding words. Keywords are then identified and stored in a database.

Daryania and team [3] represents an automated resume screening system using NLP to analyse resumes and rank them based on similarity to job requirements. The system identifies keywords, skills, and qualifications, comparing them to the job description to generate a relevance score. Resumes can be ranked accordingly, saving time for recruiters. Nawander and team [4] represents an approach using NLP and Streamlit modules to extract data from PDF resumes, store it in a database, and analyse it for ranking. The system suggests improvements for resume representation, including formatting, layout, and language.

Pokhare [5] represents, the approach using NLP and machine learning to parse, extract, and summarize data from PDF resumes. The system identifies sections such as contact details, education, and work experience. NLP and machine learning techniques extract keywords, skills, and qualifications, and analyse language used in describing the candidate. Data is stored in a database for comparison and analysis. Pravesh and team [6] represents the approach which involves the usage of machine learning and the K-Nearest Neighbors (KNN) algorithm to match job requirements with candidate resumes.

Kelkar and team [7] represents a proposed Company Recommender System for selecting the best-fit candidate. The system utilizes text mining and machine learning to rank candidate resumes based on company requirements. Relevant information is extracted

from resumes using text mining techniques and compared to job requirements, resulting in a score. A machine learning model is trained using historical data to identify patterns between requirements and candidate qualifications. This model ranks resumes, and a recommendation engine provides recruiters with a list of recommended candidates. Sinha and team [8] represent a proposed method for Resume Screening using Natural Language Processing (NLP) and Machine Learning (ML) algorithms. The aim is to automate the time-consuming process of CV screening by training the machine to analyse and extract relevant information from unstructured written language. Shubham Bhor and team [9] represents a proposed solution for resume parsing using NLP techniques. The aim is to simplify the process of finding suitable candidates for job openings. Resumes are uploaded to a job portal and parsed to extract important details.

Sanjana and team [10] represents resume validation and filtration using some inbuilt NLP Techniques and filtering jobs. It also checks resume validation accordingly. Recruiting has become a laborious process nowadays, with an overwhelming number of resumes being submitted for job openings. It's impractical for recruiters to review each one manually. Gupta and team [11] represents an automated resume screening system that uses NLP and machine learning techniques to analyse and rank resumes. The system uses NLP techniques such as part-of-speech tagging and named entity recognition, along with machine learning algorithms. Thakur and Goyal [12] represents Resume Analysing using ML Based techniques Resume Classification System (RCS) using the NLP and ML techniques could automate this tedious process. Moreover, the automation of this process can significantly expedite and transparent the applicants' screening process with mere human involvement. This model only focuses on reducing the recruiter's work, but it does not recommend any courses which can add weight to the applicants resume. This paper suggests us that using NLP & RCS and some ML techniques it does some automation work and gives us desired results.

Authors [12] highlighted the significance of ML in prediction, pattern recognition and error reduction across diverse fields, emphasizing the impact of AI in broad domain. Author [13] presented text classification algorithms for various applications and explores the use of machine learning in detecting phishing attacks. Authors [14] discussed the use of machine learning and neural networks, especially CNN, for recognizing handwriting patterns, with a focus on Telugu film industry names, achieving high accuracy (98.3%). Authors from [15] evaluated the sentiment analysis feelings of mankind based on reviews or comments, followed by categorizing them as negative, neutral, or positive. The present work explores sentiment analysis architecture and tools for user-friendly opinion mining. The approach [16] utilized Advanced Deep Learning with global threshold to improve E-commerce product classification, achieving high accuracy and challenging existing technology.

2.1 Drawbacks of existing approaches

The drawbacks in all the existing papers used ML Based Techniques & some algorithms after screening or parsing the data that is unstructured or structured data it accepts or rejects the applications it only matches to the conditions set by the recruiters it doesn't give any type of recommendation. These are the few drawbacks of the existing papers.

3 Proposed work

The initial step of the interface of proposed method involves obtaining resumes from the participants in the form of either a PDF or Word document format. The participants were informed that the maximum allowable file size for submission is 200MB. Upon receipt of the submitted resumes, a prompt display of the uploaded file is provided, allowing participants to review their submission and edit it if necessary. The uploaded resumes are then securely stored in our system for subsequent phases of the interface of proposed method.

The file that has been uploaded by the user is accepted by us. Now we move into the second phase where we extract the data from the uploaded resume. To do the extraction of the data efficiently we take the help of the module pyresparser. We call the pyresparser functions and store their return values in a variable. The pyresparser when called extracts the user's name, the user's email id, the user's phone number, the user's skills, The user's total experience, and more from the uploaded file and returns it to the place where the function was called. The module pyresparser internally uses spacy and NLTK modules which can provide the pyresparser to do NLP operations. pyresparser takes the uploaded resume as the input and returns a list of dictionary objects, each key of this dictionary represents a medium to identify a person to elaborate this statement with an example the medium can be: name, email id, and phone number these are unique characteristics used to identify a person. The value for these keys can be: the name of the person, the phone number of a person, etc, and this information is extracted from the user's resume by this module. Finally, this model defines a dictionary in a list. The keys in the dictionary are a set of names that are the unique identifiers of any person like name, phone numbers and the values for these keys are the corresponding information in the user's resume.

3.1 Objectives of the proposed work

- To identify most qualified candidates for a particular position for any job opening. Useful data from the resume, e.g., education, skills, achievements, experience and so on, should be automatically generated by the tool.
- To shorten recruitment time and reduce bias during the selection process, with a view to ranking candidates based on various aspects of their resume.

3.2 Architecture of the proposed work

The architectural diagram can be considered as a flow chart, and every component of an architecture diagram has some meaning for us to understand. The circle indicates the start/stop, diamond/rhombus represents a decision or branching point. The lines from the diamond are indicative of several possible scenarios which lead to different sub processes. Some specific operation is shown in the rectangle. The flows of process sequence and direction are shown on the lines.

Now with the help of the variable having the users' data we display the user's data. Now we analyse the skills the user has and predict the type of job that best fits the user. After knowing the role which suits the user, we consider the skills the user has and predict the skills the user does not have. This data is saved, and we move to the next phase. Now we consider the number of pages and the skills the users must generate a rating. Once this rating is generated, we display this rating to the user and proceed to the next stage.

Now we consider the list of skills the user lacks and recommend those skills along with a link to learn those skills from. After this, we analyse the total resume and generate a score. Based on this score we recommend the user to add sentences like add a list of hobbies, your career objective, and more which were initially not present in the uploaded resume. Makes the resume eye-catching. Finally, we recommend additional tips on how to answer the question and the basic question asked in the interviews and more which might help the user to perform better in the interviews, etc.

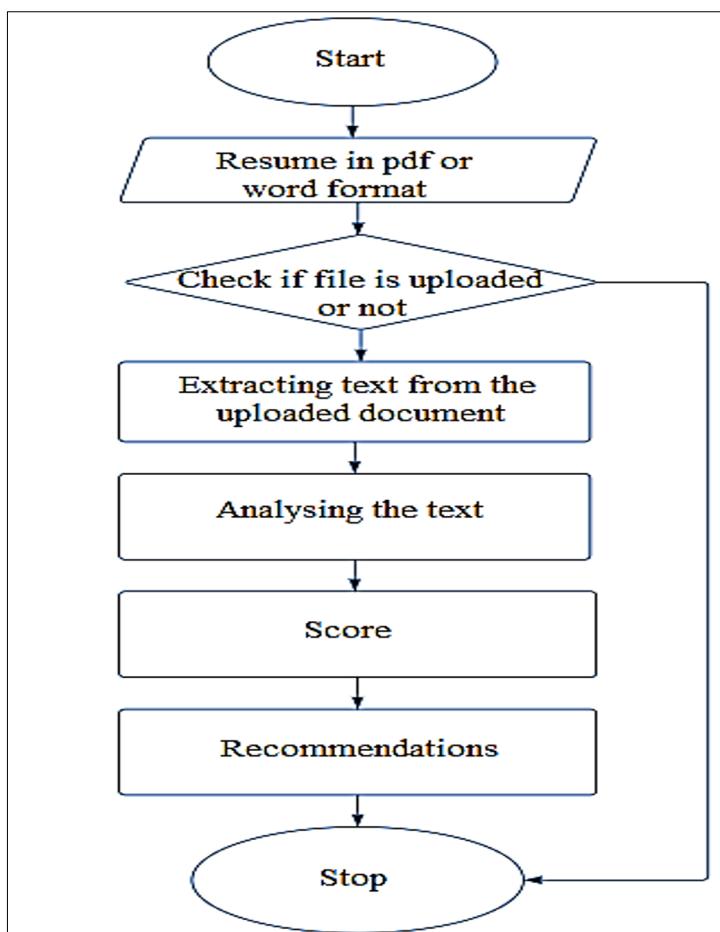


Fig. 1. Architecture diagram.

3.2.1 Automated resume screening

Automated screening of resumes is the primary goal of a smart resume analyser interface of proposed method. To this end, each resume shall be analysed in order to determine whether it is compatible with the applicant's requirements.

3.2.2 Enhanced accuracy

The improved accuracy of resume screening is helped by the use of machine learning algorithms and advanced techniques. This means that there is no inherent bias when it comes to selecting the best suited candidates.

3.2.3 Augmented candidate experience

The objective of smart resume analyser method is also to improve the experience of jobseekers by providing them with feedback and suggestions to improve their CVs and overall applications.

3.3 Modules-connectivity diagram

The module diagram serves as an extended data flow diagram that plays a crucial role in defining the business logic of a system. It consists of the following key elements:

- Components: Components, such as computation or message displays, are functions blocks in the system. These elements have input and output ports which allow data to be exchanged between them, represented as rectangles.
- Data Flow: Moving data in a component is depicted by the flow lines. The sequence in which elements are implemented, and the amount of data flowing through a system, is defined by these lines. Depending on the type of data, they may be defined as lines or colour codes.
- Control Flow Constructs: Control flow constructs, such as those for Schneiderman boxes, allow you to define module logic in line with the principles of Structured Programming; therefore, they are a more satisfying addition to your module diagram. At the same time, those constructs include timed execution, looping, dependent branching and recursive module calls.
- Module Events: One or more modules shall be covered by the module diagram and each of these events is shown individually within it. The details of each module event can be displayed to the user by clicking on it or choosing a set of events. You will see below the module's name that it is a module event.

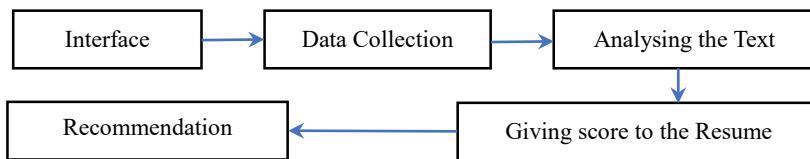


Fig. 2. Module diagram.

4 Results and discussions

4.1 Description about the dataset

In this model, we just use a base resume on which we perform the operations and expect the desired results we took this resume and performed the operations. Detailed Explanation

about the Experimental Result. Figure 3 shows the first Resume to be uploaded. Resumes are uploaded from the local system by browsing the file. After uploading the resume, it is analysed. Figure 5 shows the Resume is being uploaded

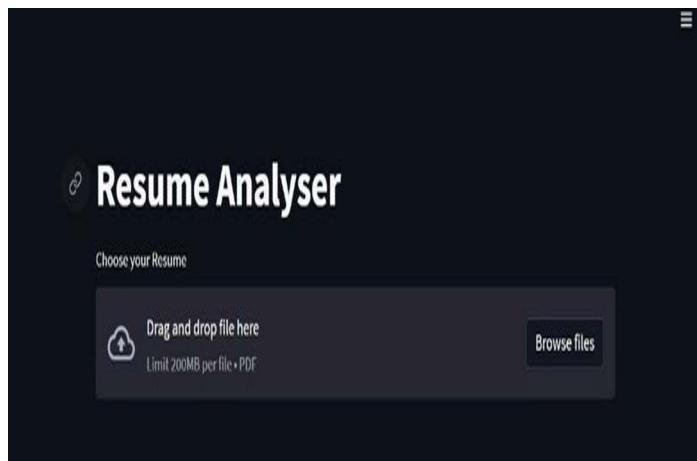


Fig. 3. The interface of the interface of proposed method.

Data Scientist
ROBERT SMITH

Phone: (123) 456 78 99
Email: info@qwikresume.com
Website: www.qwikresume.com
LinkedIn: linkedin.com/qwikresume
Address: 1737 Marshville Road, Alabama.

Objective
Data Scientist with PhD in Physics and 1+ industrial experience. Two years of working experience in Data Analysis team of LIGO Scientific Collaboration (\$3M Special Breakthrough Prize winner of 2016). Over ten years of successful research experience in both theoretical and computational physics. Strong problem-solving and analytical skills. Advanced programming proficiency. Certified in Data Analysis and Machine Learning.

Skills
Data Mining, Data Analysis, Machine Learning, Python, R, MATLAB, Sphinx, LaTeX, Mathematica, Maple, GIT, CVS, HTCondor.

Work Experience

Data Scientist
ABC Corporation - May 1994 – May 2005

- Assisted in determining client needs, deliverable design, estimates and feasibility for analytical projects concerning a custom study for a manufacturer who is using the results to support a litigation claim.
- Served as an internal resource for Jackknife programming and documentation.
- Designed and developed small scale deliverables related to the custom study.
- Participated in the Project Review QIP team.
- Responsible for results reporting in the appropriate media and creation of supporting documentation for the client.
- Monitored products from statistical programs for accuracy, consistency and statistical validity.
- Designed and applied statistical and mathematical methods for corporate analytics that were implemented into client-facing products.

Data Scientist
ABC Corporation - 1993 – 1994

- Maintained automated ETL for reporting.
- Implemented Data mining and machine learning algorithms to describe and predict user behavior on various retailer websites.
- I revamped their "Predictive Marketing" process to be more data driven and profitable.
- The new process was able to hone in on more useful user segments that had a significant increase in conversion.
- Skills Used Data Cleansing and Data Analysis using Python, Scala, R and Spark.
- Cloud computing on AWS.
- Automation of reporting..

Education

© This [Free Resume Template](#) is the copyright of Qwikresume.com. [Usage Guidelines](#)

Fig. 4. The uploaded pdf. (Courtesy: Source [13]).

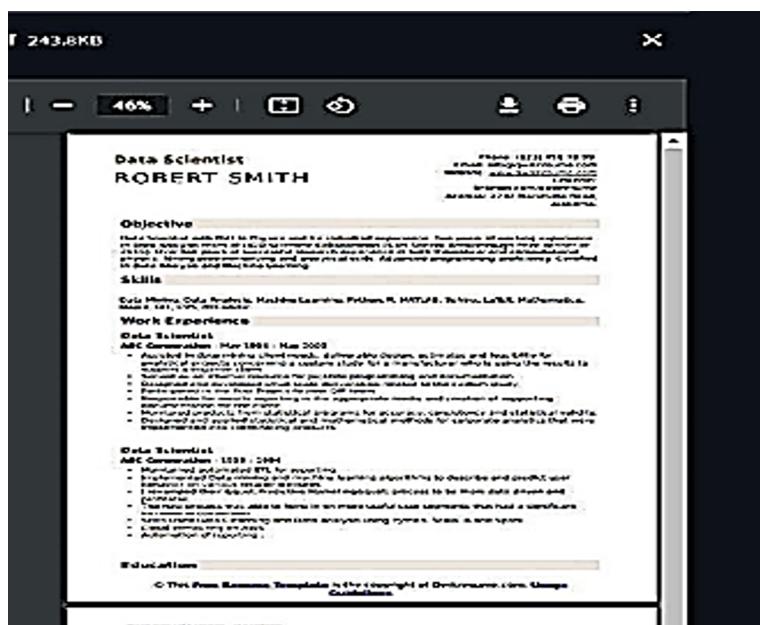


Fig. 5. Uploaded resume.

The screenshot shows the "Resume Analysis" interface. On the left, a sidebar titled "Choose User" allows selecting a user profile (e.g., "for pdf"). The main area is titled "Resume Analysis" and greets "Hello ROBERT SMITH". It displays "Your Basic info" with fields for Name ("ROBERT SMITH") and Email ("info@wikiresume.com"). Below this is a section titled "Skills that you have" containing a grid of skill tags: Programming, Cloud, Analytical, Process, EtL, Documentation, Analytics, Analysis, Python, Sphinx, Aws, Machine learning, Litigation, Website, Conversion, Data analysis, Algorithms, R, Reporting, Automation, Design, Spark, Research, Email, Mining, Scala, Physics, Matlab. A button below the skills grid says "See our skills recommendation". At the bottom, a section titled "Skills Recommendation" suggests a job title: "we recommend a job title based on the skills u wrote in your resume" and "** Our analysis says doing a Data Science Job will be best suited for you. **".

Fig. 6. Basic info.

** Our analysis says you are looking for Data Science Jobs.**

Recommended skills for you.

Data Visualization X Predictive Analysis X Statistical Modeling X Data Mining X
Clustering & Classification X Data Analytics X Quantitative Analysis X
Web Scraping X ML Algorithms X Keras X Pytorch X Probability X
Scikit-learn X Tensorflow X Flask X Streamlit X Recommended skills generated from

Adding this skills to resume will boost 🚀 the chances of getting a Job 💼

Courses & Certificates 🎓 Recommendations

Choose Number of Course Recommendations:

4

1

10

- (1) [Machine Learning A-Z by Udemy](#)
- (2) [Programming for Data Science with Python](#)
- (3) [Machine Learning Crash Course by Google \[Free\]](#)

Fig. 7. Recommendations.

(3) [Machine Learning Crash Course by Google \[Free\]](#)

(4) [Machine Learning by Andrew NG](#)

Resume Tips & Ideas💡

[+] Awesome! You have added Objective

[-] According to our recommendation please add Declaration 🤝. It will give the assurance that everything written on your resume is true and fully acknowledged by you

[+] Awesome! You have added your Hobbies ⚽

[-] According to our recommendation please add Achievements 🏆. It will show that you are capable for the required position.

[-] According to our recommendation please add Projects 🏗️. It will show that you have done work related the required position or not.

Fig. 8. Tips.

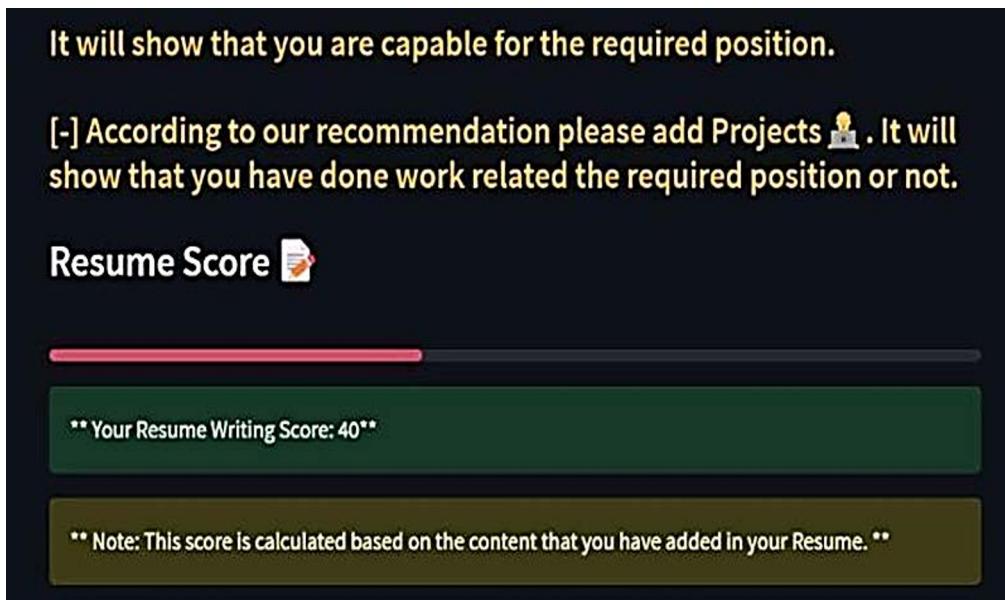


Fig. 9. Score for the resume.

5 Conclusion and future enhancements

Our program accepts resumes in PDF or Word format using the streamlit module and a file uploader. The parsed information is extracted from the uploaded file using pyresparser, which stores it as a list of dictionaries. We display the parsed information to the user and analyze their skills to predict the best-suited position. We identify the skills the user lacks and recommend learning resources. The resume is rated based on skills, content, and other factors, and a rating is provided. We recommend links for interview preparation. NLP and Spacy modules are used for parsing and analysing the resume. Streamlit module enhances the interface design, while paffy.py displays YouTube links for learning missing skills. This program aims to help users improve their resumes and make them more eye-catching. Future enhancements may include analysing social media data for personality insights and work style. Smart resume analyser works if the user's skills are related to the computer science domain or a resume of an IT employee. If a resume has skills beyond the computer science domain, the prosed method cannot recommend courses or give a rating. Rectifying this error is another future enhancement.

References

1. S. Bhor, V. Gupta, V. Nair, H. Shinde, M. S. Kulkarni, *Intl. J. Res. Engg. Sci* **9**, 6 (2021)
2. A. S. Vaidya, P. V. Sawant, *Intl. J. Engg. Tech. Res*. **3**, 8 (2015)
3. C. Daryania, G. S. Chhabrab, H. Patelc, I. K. Chhabrad, R. Patele, *Ethic Infor. Tech* **2**, 2 (2020)

4. V. Nawander, S. Elma, A. Karthikeya, M. K. Yadav, S. K. Kotala, M. D. N. Akash, Intl. Res. J. Engg. Tech **9**, 10 (2022)
5. P. Pokharel, Resume parser using NLP, Tech. Rep. Uni. Wolverhampton, (2018)
6. M. L. Pravesh, R. Samhita, M. Mythili, S. Suprith, Intl. Res. J. Comp. Sci **9**, 8 (2022)
7. B. Kelkar, R. Sheduale, D. Khade, P. Pol, A. Damame, Intl. Res. J. Engg. Tech **11**, 5 (2020)
8. A. K. Sinha, A. K. Akhtar, A. Kumar, Resume Screening Using Natural Language Processing and Machine Learning: A Systematic Review. In: Swain, D., Pattnaik, P.K., Athawale, T. (eds) Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, Springer, Singapore, 1311 (2020)
9. S. Sanyal, S. Hazra, S. M. Adhikary, N. Ghosh, Intl. J. Engg. Sci. Comput **7**, 2 (2017)
10. D. L. Padmaja, Ch. Vishnuvardhan, G. Rajeev, K. N. S. Kumar, J. Emerg. Tech. Innov. Res **10**, 3, (2023)
11. I. Ali, N. Mughal, Z. H. Khan, J. Ahmed, G. Mujtaba, Mehr. Univ. Res. J. Engg. Tech **4**, 1 (2022)
12. R. P. Ram Kumar, P. Sanjeeva, S. F. Lazarus, D. V. Krishna, Intl. J. Inno. Tech. Explor. Engg **8**, 11S2 (2019)
13. 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)
14. B. Sankara Babu, S. Nalajala, K. Sarada, V. Muniraju Naidu, N. Yamsani, K. Saikumar, Machine Learning based online Handwritten Telugu Letters Recognition for Different Domains, in Proceedings of A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems, P. Kumar, A. J. Obaid, K. Cengiz, A. Khanna, V. E. Balas (eds), Intelligent Systems Reference Library, vol 210. Springer, (2022)
15. R. P. Ram Kumar, P. Sanjeeva, Intl. J. Recent Tech Engg **7**, 5 (2019)
16. M. N. Mohammad, Ch. U. Kumari, A. S. D. Murthy, B. O. L. Jagan, K. Saikumar, Mater. Today Proc. 45 (2021)