Proceedings of the International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS-2024)
IEEE Xplore Part Number: CFP242B2-ART: ISBN: 979-8-3315-1809-7

2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) | 979-8-3315-1809-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICICNIS64247.2024.10823187

Deep Learning based Approach to Streamline Resume Categorization and Ranking

RAMYA R
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai, India
ramyarajan1505@gmail.com

GUNDAPANENI NAGARANI
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai, India
nagaranigundapaneni1663@gmail.com

Dr. NATARAJAN S
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai, India
s_natarajan@ch.amrita.edu

***Abstract*— In today's world, a lot people are searching and applying for jobs. When there is a need for company to recruit new employees, it is difficult for them to select the correct resume according to their job description from the lakhs of resume that they received. This project will help the recruiters to select the accurate resume with relevance to their job description. It will produce resumes which matches the best with the recruiter's job description by matching the skills, experience of resume with the job description. Used various advanced natural language processing models and deep learning models to provide more accurate results than any other existing models. With these methods the model achieved accuracy of 98%. Also, system will send mail to those people whose resume has not matched much with the job description stating the reason like what lacks in their resume when compared with job description. This will be helpful for candidates to know what lacks in them. Also, for the HR, interview questions will be generated according to each candidate resume. This will reduce the Reduce HR's work for preparing interview questions for each person.**

Keywords— Resume Categorization, Resume Ranking, Deep Learning, Recruitment Automation, Tailored Interview Questions, Personalized Candidate Feedback, BERT, NER, NLP.

I. INTRODUCTION

In the current competitive job market companies when announced job vacancies they start to get a lot of resumes from which selecting a very few might be a challenging task. Recruitment automation refers to selecting the best top resume that matches the job description without much manual work.

Existing approaches to resume classification often rely on keyword matching or rule-based system. Keyword matching techniques will find the resume that are relevant to job description based on the matching keywords between both resume and job description. Rule based system on the other hand will select resume with relevance to job description based on some predefined rules which may not be applicable to all resumes. While these may fetch relevant resume to given job description the result will not accurate and always correct. They struggle to capture the nuances of language and can miss qualified candidates whose resume don't perfectly align with keywords. But the advancement in Natural language processing (NLP) and machine learning may help to overcome these limitations.

This project investigates the application of advanced NLP models and deep learning techniques for classifying resumes based on their relevance to job descriptions. We aim to develop a system that surpasses the accuracy of existing methods by leveraging the power of deep learning and contextual understanding of language. By achieving a more accurate and efficient classification process, this project seeks to empower organizations to make better hiring decisions and enhance the overall recruitment experience. Also, this system has the ability to send email to those people whose resumes was not matches less with job description so that those will have idea for why they did not get selected for the particular job they applied. Nowadays recruitment process is a tedious task for HR. In order to ease the recruitment, process we have one feature which is tailored interview question generation. According to each person resume interview question will be generated by our system which will reduce the workload of HR.

II. LITERATURE SURVEY

In this second chapter, related works and literature reviews of various other papers and research in the field explored, particularly in the domain of resume categorization and ranking using deep learning techniques. Such applications of NLP models will focus on automating the recruitment process. Further, the literature review will briefly present various approaches and algorithms used to enhance the accuracy, f1 score of job matching and generate tailored interview questions.

Ongoing work aims to understand the models such as BERT, RoBERTa, and GPT-3, and their application in the recruitment field. Also, other type of methods is referenced, such as hybrid Named Entity Recognition (NER), among others that have been developed recently to help improve the extraction and analysis of the relevant information from resumes in other ways. Now, let us have look into literature reviews from a few research articles and papers:

Ahmed Heakl et al. [1] Proposed a revised resume categorization with large number of Datasets and Language Models. With parameters such as transformers, online Recruitment and Large Language Models. This paper tried to overcome the limitations of less datasets, no proper standardized resumes, and some privacy issues of already present models. They overcome those limitations by using the concept of data augmentation where the new datasets are created by doing some changes on existing data and thus, they got large dataset of varied formats for processing. For training they used advanced natural language processing technique like Bert etc.

Dr. Ambareesh S et al. [2] Suggested a method for shortlisting the resumes. The project explored on various natural language processing techniques to automatically evaluate resumes. Some techniques include NER (Name entity recognition) etc. The accuracy that this paper achieved using NLP techniques is more compared to machine learning techniques.

B. Lalitha et al. [3] Introduced research for practical resume screening using NLP techniques. This paper focus on shortlisting the suitable candidates from the pool of the resumes. It was developed in such a way that it can be benefit for both candidates and recruitment Company. The latest techniques like CNN and KNN algorithms which are advanced and time taking. In this they used NLP tools, which automates the process and also decrease the time consumption and provides correct answers.

Thi-Thuy-Quynh Trinh et al. [4] Proposed a domain adaptation approach for resume classification using graph attention networks. with parameters such as pos tagging, graph multi-headed attention networks and domain adaptation. By Comparing the contextual comparability among the resume text and job posts for resume categorization, the proposed model's goal is to categorize resumes successfully without the need of an upgrading model or classified resumes. A very accurate result was obtained when it was tested with a real hiring database.

Dipendra Pant et al. [5] Undertook a pilot study on ranking systems and parsing automated resume through natural language processing. The study focused on automating the resume categorization using tools like NLP. They mainly focused on enhancing the accuracy and more productive in order to select the best resumes. It takes into account only the important information to rank resumes which improves the quality of resumes being selected. They developed a user- friendly interface for HRs to interact easily with the application in order to get the best resumes. Their work is intergraded with applicant tracking system which improves the application's effectiveness.

Tumula Mani Harsha et al. [6] Developed an automated resume Screener. Parameters like skill set, recruitment, etc. This study includes NLP and Machine learning algorithms for resume classification. The outcome of this study was its application to screen resumes.

Tejaswini k et al. [7] The aim of this paper is to propose an innovative technique called content-based recommendation that utilizes the concept of cosine similarity to calculate the similarity between job description and resumes. It also uses KNN for classification and large datasets which is one advantage to increase accuracy. The system got a text parsing accuracy 85% and ranking accuracy nearly 92%.

Elias Abdollahnejad et al. [8] Proposed an integrated BERT based model to reduce the advanced shortlisting process of job applications. The aim of this paper is to use NLP and machine learning approaches to give a best solution for identifying suitable candidates for specific job role. In this paper the modern BERT-algorithm model simulates the recruiter's decision-making process. Compare to other models BERT performs better outcomes.

Vedant Bhatia et al. [9] In this paper, it proposed a new method known as resume parsing where it extracts entire data from candidate resumes. Then one type of BERT techniques was used which is known as sentence pair classification to measure its relativeness with job description. They combined resume datasets that includes both linked in format and non-linked in format.

Pradeep Kumar Roy et al. [10] In this paper, the main goal is to improve selecting candidates based on large number of resumes. It also developed an automated method for matching and resume classification which helps in quit evaluating and selecting the best candidates. Another two classifications on content- based recommendation and KNN algorithm.

III. PROPOSED METHOD

Our research focuses on automating the recruitment process using advanced deep learning techniques. We employ transformer models like BERT, RoBERTa, and GPT-3 to accurately categorize and rank resumes, capturing deep semantic context. A hybrid Named Entity Recognition (NER) approach combines rule-based approach and machine learning methods in order to extract key information. Additionally, our system generates tailored interview questions specific to each candidate's profile using NLP techniques, ensuring alignment with the job description. It also provides personalized feedback to non-selected candidates, guiding them on areas for improvement. This comprehensive solution integrates seamlessly into existing HR workflows, enhancing recruitment efficiency and decision-making.

IV. METHODOLOGY & WORKING

Our system tried to automate the resume categorizing and ranking with at most possible accuracy. In our system they can upload resume of various file formats like .jpg, .png, .pdf,.doc. All these file formats can be processed by our system. Similarly, job description should be in word. This job description will also be given by HR. once the resume and job description are submitted by HR and if he clicks submit button then the resumes will be categorized and ranking will be done internally. Once he inputs the number of top ranked resumes he wants and clicks submit then all those number of top ranked resumes will be provided to hr. if they want interview questions for each resume, then if he clicks the interview question generate button then those interview questions will be generated. once if he clicks the send feedback button then the feedback will be sent to non-selected candidates. The methodology of this project involves various steps.

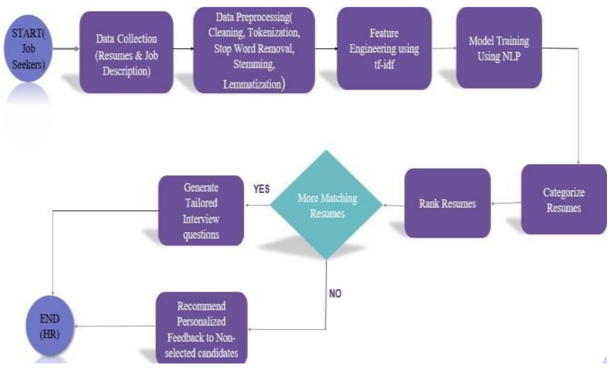

Fig 1.1 Flow chart of System architecture

A. Data Collection:

In the Data collection step, we took resume data of various formats like .pdf, .doc, .jpg, .png formats so that if resume is uploaded in any of the above formats, then the system will be able to process the resume and perform classification and ranking. We also used job description dataset which we took from Kaggle. The job description dataset contains columns like job ID, Job description etc. These job descriptions can also be either in .pdf or .doc or in normal text. The uploading of resume and job description are manual task.

B. Data Preprocessing:

In the data preprocessing step involved various steps like importing libraries, extracting data from various resume and job description formats. Then concatenation of all the extracted resume job data will be stored in single file names as resume concatenated data and similarly job description concatenated data. This way extracting data of all the resume files of different format into a single file where all data will be stored will ease the data cleaning process. similarly, data from all various job description file formats will be extracted and stored in a single csv file. This makes the cleaning process simpler to perform. Then the concatenated resume file and concatenated job description file was cleaned and the cleaned files will be saved for future use. Importing of various libraries is one important step in data preprocessing. Libraries like pandas for data manipulation NumPy for computation, nltk for natural language processing pyfpdf for extracting data from pypdf are some libraries that were imported. Data cleaning of both resume and job description involves converting all text to lower case, removal of emails, URL's, digits, punctuation. Removal of stop words and performed stemming, lemmatization are important step in data cleaning process. After cleaning tokenization will be performed where the works are separated as tokens and the file will be stored as cleaned resume and cleaned job descriptions. All these will be performed by the software automatically.

C. Feature Engineering:

The Feature Engineering step in necessary extract important features from both resume and job description to see the relevance. It also converts the text data into numerical features which is necessary for model training. In this step libraries like pandas NumPy, sklearn, spacy, transformers etc are used. This feature engineering used distill BERT is used. It has functions like extract keywords, word embeddings, contextual embeddings match skills, reduce dimensionality and feature engineering pipeline. extract keyword's function which is used for extracting important keywords from the file using TF-IDF (Term Frequency Inverse Document Frequency). Word embedding's function is used to state the relationship between words which is necessary to understand the context. Contextual embedding function uses BERT and itsvariants to understand the context more clearly and precisely. Match skills function is used to match the skills in resume with that of job description. Reduce dimensionality function is used reduce the dimensionality like 3d to 2d etc. Then the feature engineering pipeline is the function where all the function call has taken place by passing input parameters. From the feature engineering step resume features and job description features files are stored as output. These files will be used for further process.

D. Model training:

The model training involves training the model to categorize the resume with relevance to job description using BERT model. Here we create the loader for data for both validation and training. Then we loaded the pre trained BERT model. Then we have defined the AdamW and crossEntropyloss function. Using the BERT model we trained with our dataset. then we evaluated the model performance.

Several optimization techniques were used to enhance the performance of our model. Hyperparameter tuning was the first step where we sought the optimal values for the learning rate, batch size, and sequence length. Further, fine-tuning the pre-trained BERT model on our particular dataset helped bridge the gap between the requirements of the domain and the nature of the model. Techniques such as cross-validation, dimensionality reduction, and feature scaling also served to minimize processing time while losing no accuracy.

E. *Ranking resumes:*

In this step the cleaned resume and cleaned job description was loaded. Then the pretrained BERT model (BERT-base- uncased) and tokenizer were loaded.

This model extracted the text from cleaned resume and job description and converted into numerical data. cosine similarity is calculated to know how much each resume matches with job description. Based on this score the resume was ranked.

F. *Interview question generation:*

In this part the libraries used are pandas and scikit learn. In these first keywords will be extracted from both resume and job description. The keywords that is present in both resume and job description will have high tf-idf value. There will be a set of general templates will be created manually. In those templates according to each resume keywords which has high tf-idf value will be replaced in that general template in order to generate tailored interview question for each resume.

G. *Sending feedback:*

This function takes two arguments non top resumes and job description. For sending emails to non-top resumes candidates, we use smtp which simple mail transfer protocol to send mail. There is general template created for all mails. In general template in lacked skill place what skill lacks in their resume from job description will be replaced and that mail is sent to particular mail id.

V. IMPLEMENTATION

We developed a website where stakeholders are HR and job seekers. In our website the HR can upload resumes of various file formats. Once he clicks upload then all those resumes will be stored under resume folder. Once if the job description also uploaded by HR, then these resume and job description data will be extracted and categorizing and ranking will be performed. Once if the HR inputs the number of top resumes, he wants then those resumes will be stored in top resume folder in HR system in the same order of rankings. If HR clicks generate interview questions, then interview question set of interview question for each resume will be stored in separate file in interview question folder. If HR clicks send feedback, then email will be sent to all non- top rank resume candidates stating what lacks in their resume compared to job description.

VI. RESULT

The resume classification and ranking system has been successfully implemented with the accuracy of 98%, f1 score of 0.97 in classifying resumes with relevance to job description. Our system was able to processes 2000 resumes of different formats which includes .pdf, .doc, .docx, .png,.jpg. and able to classify these resumes with relevance to job description within a span of 20 minutes. The system was reliable as cross validation and dropout techniques were used to prevent overfit In feature engineering we used techniques like TF-IDF vectorization, word embeddings, contextual embeddings which highly contributed in the accuracy of the model by identifying the important and domain specific phrases or nuances in the resume and job description. In addition to that our system automatically generated interview questions tailored according to each person resume which will be useful for HR while conducting interview. This improves work efficiency for HR while taking interview. Also, our system sent automatic feedback for candidates whose resume did not match much with the job description stating what lacks in their resume when compared to job description which will help them to understand like what skills they need to learn further.

VII. EXPERIMENTAL RESULT

Test Accuracy = 0.9867
Test Classification Report:

	precision	recall	f1-score	support
0	0.95	1.00	0.97	37
1	1.00	0.98	0.99	113
accuracy			0.99	150
macro avg	0.97	0.99	0.98	150
weighted avg	0.99	0.99	0.99	150

Fig.1.2 Accuracy & f1 score of tested resumes

Upon collecting 2000 resume files and 700 job description files then preprocessing it which involves extracting text from various file formats and those text will undergo cleaning which include removal special character, numbers, symbols and then stemming and lemmatization will be performed on those text and will be stored. and after doing feature selection then training model with BERT algorithm, we got accuracy as shown in Fig.1.2.

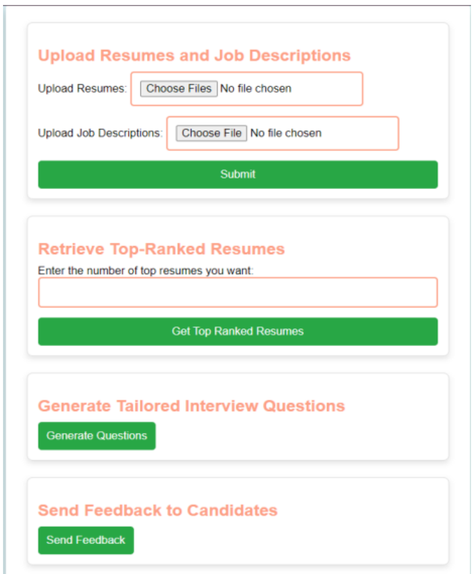

Fig. 1.3 webpage

Fig. 1.4 resume files of different formats

Fig.1.3 is the image of the webpage of our project. In this web page there is placeholder to upload resumes where multiple resumes of different formats as shown in Fig.1.4 can be uploaded at a time and 1 job description will be uploaded. Upon uploading once, the files are submitted the HR will receive a prompt stating submitted successfully.

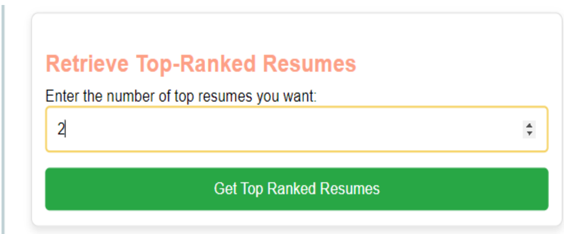

Fig. 1.5 number of top resumes HR's needs

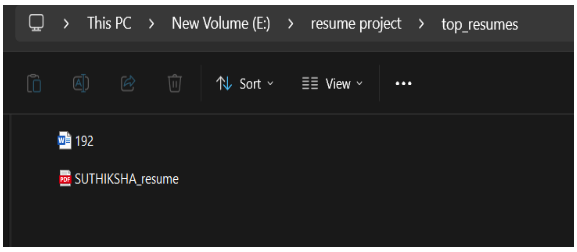

Fig. 1.6 top resumes in HR's folder

Fig.1.5 shows the image of the placeholder where the HR inputs the number of top resumes he needed and those resumes will be provided by our system to HR by storing those top resumes in the order of their ranking in HR local folder as xshown in Fig1.6.

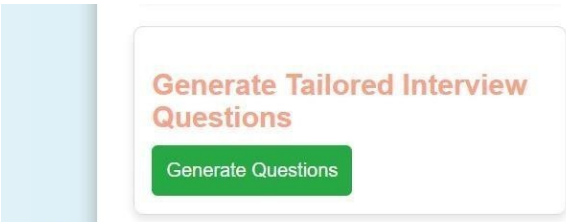

Fig. 1.7 question generation button

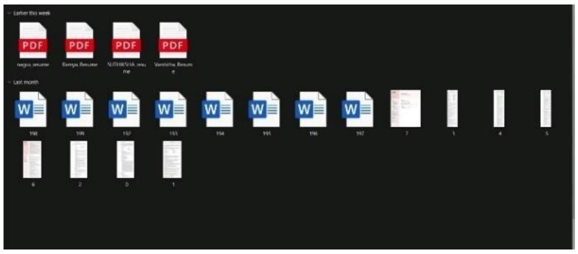

Fig.1.8 generated interview questions

Fig.1.7 shows the button from our web page for interview question generation. When the HR clicks the generate questions button then the interview questions will be generated according to each top candidates resumes and will be stored in word file as shown in Fig 1.8 in HR's local folder with filename as rank number.

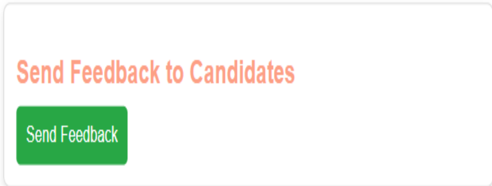

Fig.1.9 Feedback button

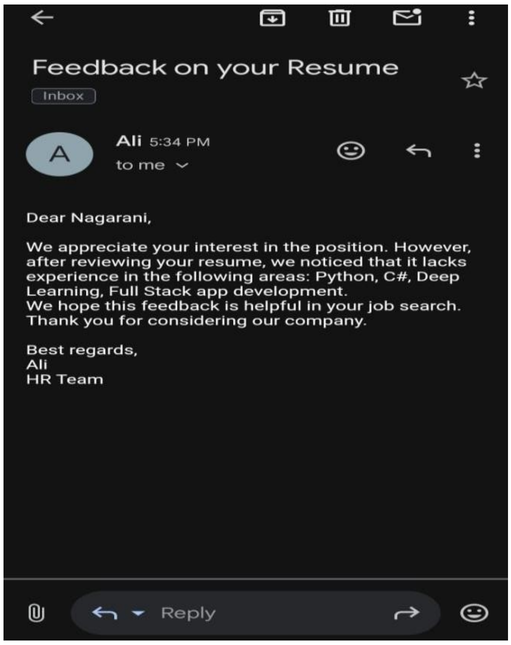

Fig.1.10 candidate's Resume feedback

Fig.1.9 shows the image of send feedback button from our webpage through which HR can send automatic feedback to all those candidates whose resume was not matched much with job description. Fig.1.10 shows the image of mail that is received by the candidate that states what lacks in their resume when compared to job description.

VIII. Conclusion

In conclusion, our project successfully implemented an automated resume categorizing and ranking system that streamlines the job selection process by categorizing and ranking resumes with relevance to job description. This system has reduced the manual work drastically like automating the categorizing and ranking work, interview generation, feedback sending which will have positive impact in productivity. This system also helps the candidates whose resume was not matched much with the job description to know what skills they lacked and help them improve their awareness towards what skills and technologies the industry needs. Also, the system makes the recruitment process more transparent with the help of automated system. This transparency in the system will make the candidates believe that the recruitment process conducted fairly. The future work may focus on adding Multilanguage support so that system can used across various countries where English is official language which will help improve recruitment globally.

References

[1] Heakl, Ahmed & Mohamed, Youssef & Mohamed, Noran & Sharkaway, Ali & Zaky, Ahmed. (2024). ResumeAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models.

[2] S. Ambareesh, N. K. Thakur, U. Bhattarai, S. K. Yadav, J. N. Thakur and A. K. Mahato, "Resume Shortlisting Using NLP," *2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)*, Bangalore, India, 2024, pp. 1-5, doi: 10.1109/ICDECS59733.2023.10502580.

[3] K. Thangaramya, G. Logeswari, S. Gajendran, J. Deepika Roselind and N. Ahirwar, "Automated Resume Parsing and Ranking using Natural Language Processing," *2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT)*, Vellore, India, 2024, pp. 1-6, doi: 10.1109/AIIoT58432.2024.10574696.

[4] B. Lalitha, Sirisha Kadiyam and Ritisha Varma Kalidin, 2023. Applicant Screening System using NLP. International Conference on Innovation Data Communication Technologies and Application. ISBN:979-8-3503-9720-8

[5] Thi-Thuy-Quynh Trinh, Yu-Chi Chung, R.J. Kuo, A domain adaptation approach for resume classification using graph attention networks and natural language processing, Knowledge-Based Systems, Volume 266, 2023, 110364, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2023.110364.

[6] D. Pant, D. Pokhrel and P. Poudyal, "Automatic Software Engineering Position Resume Screening using Natural Language Processing, Word Matching, Character Positioning, and Regex," *2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, Hammamet, Tunisia, 2022, pp. 44-48, doi: 10.1109/IC_ASET53395.2022.9765916.

[7] Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and development of machine learning based resume ranking system, Global Transitions Proceedings, Volume 3, Issue 2, 2022, Pages 371-375, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.10.002.

[8] E. Abdollahnejad, M. Kalman and B. H. Far, "A Deep Learning BERT-Based Approach to Person-Job Fit in Talent Recruitment," *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2021, pp. 98-104, doi: 10.1109/CSCI54926.2021.00091.

[9] Bhatia, Vedant & Rawat, Prateek & Kumar, Ajit & Shah, Rajiv Ratn. (2019). End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT. 10.48550/arXiv.1910.03089.

[10] Pradeep Kumar Roy, Sarabjeet Singh Chowdhary, Rocky Bhatia, A Machine Learning approach for automation of Resume Recommendation system, Procedia Computer Science, Volume 167, 2020.

[11] Heakl, Ahmed & Mohamed, Youssef & Mohamed, Noran & Sharkaway, Ali & Zaky, Ahmed. (2024). ResumeAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models.

[12] Turmula mani Harsha, Gangaraju Sai Moukthika and Dudipalli Siva Sai, Automated Resume Screener using Natural Language Processing, International Conference,ISBN- 978-1-6654-8328-5

[13] Jiahao Liu, Yifan Shen, Yijie Zhang, and Sujatha krishnamoorthy. 2021. Resume Parsing based on Multi-label Classification using Neural Network models.

[14] S. Ramraj, V. Sivakumar and K. Ramnath G., "Real-Time Resume Classification System Using LinkedIn Profile Descriptions," 2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy.

[15] S. Nasser, C. Sreejith and M. Irshad, "Convolutional Neural Network with Word Embedding Based Approach for Resume Classification," 2018 International Conference on Emerging Trends and Innovations in Engineering and Technological Research (ICETIETR), Ernakulam, India, 2018, pp. 1-6, doi: 10.1109/ICETIETR.2018.8529097.