

# Resume Analyser and Job Recommendation System Based on NLP

Gautam Jaiswal

Applied Computer Science and  
Engineering

G.L. Bajaj Institute of Technology and  
Management, Greater Noida  
[cai20024@glbittm.ac.in](mailto:cai20024@glbittm.ac.in)

Aryan Uttam

Applied Computer Science and  
Engineering

G.L. Bajaj Institute of Technology and  
Management, Greater Noida  
[cai20003@glbittm.ac.in](mailto:cai20003@glbittm.ac.in)

Devesh Dhar Dubey

Applied Computer Science and  
Engineering

G.L. Bajaj Institute of Technology and  
Management, Greater Noida  
[cai20055@glbittm.ac.in](mailto:cai20055@glbittm.ac.in)

Dr. Pawan Kumar Mall

Applied Computer Science and Engineering

G.L. Bajaj Institute of Technology and Management, Greater Noida  
[pawankumar.mall@gmail.com](mailto:pawankumar.mall@gmail.com)

**Abstract**— The Resume Analyser is a streamlit-based web app that is used to analyse resumes which are in PDF format to extract information from it using different NLP techniques.

We have proposed a hybrid system which includes BERT for Name entity recognition and NLP pipeline to streamline the workflow. Our hybrid system excels in extracting key information from the resumes which include skills, experiences, and qualifications etc. This enables a more state of the art performance, achieving higher accuracy and faster execution time. SPACY's advanced linguistic features empower the system to discern not only the explicit qualifications but also the implicit nuances that play a crucial role in candidate-job alignment. It returns a resume score based on the presence of key elements such as a Declaration, qualifications, Experience etc. It also suggests jobs and internships based on the analysis of the resume and also helps in filtering resumes. This analyzer reduces the need for people to manually check their resumes. This analyzer works using natural language processing and text mining to deduce information from the resume. Finally, we were able to create such a hybrid model whose efficiency was approximately 95% and was able to perform its job on many different structures of resumes very efficiently.

**Keywords**— Natural Language Processing(NLP), Name Entity Recognition, streamlit, BiDirectional Encoder Representation and Transformers(BERT), Recommendation System.

## I. INTRODUCTION

In the ever-evolving landscape of talent acquisition and workforce optimization, the demand for advanced technologies to streamline the recruitment process has never been more pressing. This study introduces a groundbreaking solution at the intersection of Natural Language Processing (NLP) and state-of-the-art machine learning techniques — a Resume Analyser and Job Recommendation System. This innovative system is forged through a hybrid approach, synergizing the unparalleled capabilities of BERT (Bidirectional Encoder Representations from Transformers) for Named Entity Recognition (NER) and the robust linguistic understanding provided by SPACY NLP.

The proliferation of digital platforms and the deluge of resume data present both opportunities and challenges for recruiters and hiring managers. The old methods of Resume analysis was not only outdated but it lacked several features

which are necessary for today's analysis system. Older techniques were rule-based and less accurate and also lacked the ability to extract data from a variety of resumes. In the loop-hole to these challenges, our hybrid system seeks to revolutionize resume analysis techniques by combining the strengths of BERT's contextualized understanding of the language and SPACY's advanced language analysis techniques.

BERT, which is a type of transformer, is a state of the art machine which uses contextualized techniques to understand the meaning of each word in a sentence. By combining BERT-based NER techniques, our system achieves a very high level of precision and accuracy in identifying and categorizing entities within resumes, such as skills, experiences, and qualifications, that outperforms the traditional parsing methods.

Our research provides the basis of a crucial need for an intelligent, adaptive and an accurate system that not only parses resumes with but also facilitates dynamic and job recommendations. The hybrid approach presented here represents a leap forward in the fusion of cutting-edge technologies, promising to redefine the benchmarks for efficiency and accuracy in the realm of talent acquisition. As organizations strive to build agile and responsive teams, this hybrid Resume Analyser and Job Recommendation System stands as a beacon of innovation, poised to reshape the future of recruitment processes.

## II. LITERATURE SURVEY

In 2022, S. M. Chaware, Bhushan Kinge, Shrinivas Mandhare, and Pranali Chavan gave their research on Resume Screening Using ML and NLP.[1] They used different sets of approaches like SVM, Word2Vec, Knn etc on different categories of datasets and have attained an accuracy of 78% to 98% based on some specific conditions.

In [2] the authors used approaches like Cosine-Similarity and KNN algorithms to rank the resumes based on certain criteria. Their system performance was astonishing and average parsed accuracy was approx. 85% and ranking accuracy was close to 91%.

In [3] the authors used a BERT transformer and along with it used two different multilingual i.e. models having vocab of multiple languages. They did this in order to extract information in two different steps. Firstly, they extracted

information like names, previous employment etc. and used a second model to find skills, degrees etc.

In 2021, Agnieszka Wosiak researched Automatic information extraction from resumes in the IT sector. [4] This research was a result of a hybrid solution whose aim was to extract information from resumes specifically for the IT sector. But it failed to handle those resumes whose data was slightly different and thus it failed.

In [5] the authors used about 200+ different resume datasets to perform NLP operations on them. The ideology was based on Name Entity Recognition(NER).It mainly showcased the work of a candidate whose aim was to build a simple NER Model using Spacy's Library and NLTK stopwords.

In [6] the authors have used 500 resumes dataset for training purposes and 50 datasets for testing/evaluation purpose.A coding scheme called BILOU scheme was used here to train transformers. Convolutional Neural Networking Layer and Bidirectional-Long Short Term Memory(Bi-LSTM) layers are used to train the model. They have used this model to find only specific entities like the Name, Degree etc.

In [7] the authors were mainly focused on easing the recruitment process for the HR. In this, a vast set of resumes are parsed to the model from where the model will predict the best resumes which match the job description. It was basically a classification process which involved many different strategies.

In [8] the authors released their research paper on-Extraction of Resume based on NLP using Machine Learning. Since we know that a resume can have any specific format and not every resume is alike. This research was based on CV parsing which had customization options, development sheets and mapping limits. This system is used by Scout bundles to identify resumes. Since resumes can contain data in structured/unstructured format, many different scenarios arises during creation of models.

In [9] The authors have used a resume set containing more than 400 different resumes of many different formats like html,pdf,docs etc. PyDocx and PdfMiner are the two python libraries they have used to convert these unstructured resumes into text format or a format which can be fed into ML/DL models. They have used a dictionary of tags to filter out whether a particular piece of text is relevant or not. It was basically a classification based approach. Later they used fuzzy string matching to extract the relevant skills.The model they used is called XG Boost which gave an accuracy of approximately 90%.

In [10] this, the authors gave three different sets of modules to form on complete software or model. They have used Spacy's NLP pipelining system to attain this and also a semi-supervised ML model to achieve further accuracy. Once a candidate's resume was parsed through this model, it gave output as a score which depicts whether the resume is aligning with the job requirements or not.

In 2019, Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova gave their research paper named BERT:

Pre-training of Deep Bidirectional Transformers for Language Understanding.[11] This paper showed a new and improved model of BERT. Unlike the older version of BERT introduced in 2018 this was a more enhanced and improved model. Some of the major features included contextual understanding, Bi-Directional understanding of data which gave state-of-the-art performance and also improved the accuracy for tasks like sentence classification, token classification etc.

Later this model was developed into many other major models also like GPT.

In 2022, authors named Narendra G.O. and Hashwanth S. gave their research paper on the topic Named Entity Recognition based Resume Parser and Summarizer [12]. They have used spaCy, which is a natural language processing (NLP) library. Their study suggested a Machine Learning based approach to find matches in resumes, whose main focus was on the extraction of specific entities. While parsing any resume, it maintained track of the entities in order to classify these resumes based on the requirement. Entities which are important are extracted from the resume and then used for the classification purpose.

### III. METHODOLOGY

The presented methodology offers significant advantages to recruiters by expediting the candidate evaluation process aligned with specific job descriptions.

The Resume Analyser relies on an extensive array of Python modules, a strategic choice that greatly enhances the application's functionality and user experience. Among these modules, several noteworthy ones include Streamlit, PyMuPdf, NLTK (Natural Language Toolkit), SpaCy, BeautifulSoup, PyTorch, Transformers. By integrating these sophisticated libraries, the application achieves a seamless and efficient workflow. Streamlit facilitates intuitive and interactive user interfaces, ensuring a user-friendly experience. PyMuPdf empowers the application to handle PDF documents effectively, enabling comprehensive resume analysis. NLTK and SpaCy, powerful natural language processing libraries, enhance the system's language understanding capabilities, enabling nuanced content interpretation. Lastly, BeautifulSoup enables efficient parsing of HTML, contributing to the application's ability to extract valuable information from various sources.PyTorch is an open-source machine learning library for Python that provides a flexible and dynamic computational graph, making it particularly well-suited for deep learning tasks.The Transformers library, developed by Hugging Face, is a popular open-source library for natural language processing (NLP) tasks. It provides a collection of pre-trained models, including state-of-the-art transformer-based architectures such as BERT, GPT, and others. Transformers simplify the implementation and fine-tuning of these models for various NLP applications.

This meticulous selection of Python modules underscores the application's robustness, providing a solid foundation for streamlined resume processing and analysis. Each module plays a vital role in enhancing the application's functionality, ensuring that users can navigate, interpret, and process resumes with precision and ease.

Utilizing BERT (Bidirectional Encoder Representations from Transformers) for Named Entity Recognition (NER) has proven to be highly effective in advancing the accuracy and contextual understanding of information extraction tasks. BERT, a pre-trained transformer-based model, excels in capturing bidirectional contextual information, allowing it to grasp the intricate relationships and nuances present in natural language text. BERT generates contextualized word embeddings by considering the entire context of a word within a sentence. This contextual awareness is crucial for NER tasks where the meaning of an entity often depends on its surrounding words. Additionally, it incorporates an intelligent ranking mechanism, aligning resumes with the specific requirements outlined by the hiring organization. This strategic approach not only enhances the efficiency of candidate assessment but also ensures that the most relevant resumes are prioritized, facilitating a more targeted and informed recruitment decision-making process for recruiters and employers alike.

Job Recommendation in Resume Analyser represents a strategic approach aimed at optimizing the match between job seekers and employers, ensuring a seamless alignment of skills, qualifications, and aspirations. A notable aspect of this process involves employing cutting-edge technologies such as Beautiful Soup to scrape data from various online job portals. This meticulous data collection strategy ensures a comprehensive and diverse pool of job listings. By harnessing Beautiful Soup, the Resume Analyser acquires real-time, relevant job data from these portals. This collected information is then meticulously curated and presented to job seekers in a user-friendly format. By presenting curated and updated job listings, candidates gain access to a wealth of options that align precisely with their skill sets and career goals. Consequently, employers benefit from a pool of candidates whose profiles closely match their job requirements, facilitating a more efficient and effective hiring process.

#### IV. PROPOSED SOLUTION

The proposed solution for the development revolves around a meticulously structured approach that combines the strengths of BERT NER Techniques and SPACY NLP.

To begin, a comprehensive analysis of the system's requirements and objectives is undertaken, ensuring a clear understanding of the desired functionalities such as precise resume parsing, entity extraction, dynamic job recommendations.

Data, constituting a diverse array of resumes, is then collected and subjected to preprocessing. This involves Cleaning, Tokenization, Padding, Segment IDs, Special tokens and Label Encodings the documents to facilitate further analysis.

Figure 1. shows how the Input data looks like in BERT acceptable form.

This finalized data is fed into the BERT to fine-tune the model based on the requirements.

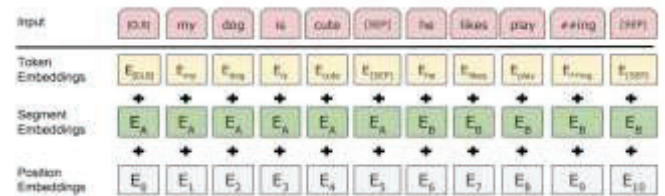


Fig. 1. BERT input representation. The input layer is the sum of token, segment and position embeddings.

The architecture of BERT is based on the Transformer model, which employs self-attention mechanisms to capture relationships between words in a sequence. What sets BERT apart is its bidirectional approach to contextualized embeddings. Unlike traditional models that process words in a left-to-right or right-to-left manner, BERT reads input text bidirectionally, enabling a richer understanding of context. BERT is pre-trained on massive amounts of text data and learns to predict missing words in sentences, allowing it to grasp intricate linguistic nuances. The model consists of multiple layers of attention-based transformers, creating a deep neural network capable of capturing complex language patterns. BERT's pre-trained knowledge can be fine-tuned for various downstream tasks, such as sentiment analysis, named entity recognition, and question answering, making it a versatile and powerful tool in natural language understanding.

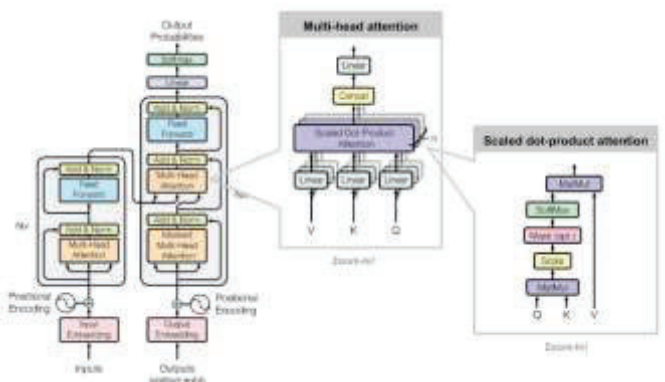


Fig. 2. BERT architecture showing its attention mechanism.

Evaluation of the system's performance is conducted rigorously, employing metrics such as precision, recall, and F1 score which continually enhance accuracy and relevance. Scalability considerations are addressed to optimize the system for handling large volumes of data, and deployment is executed in a user-friendly manner, whether in a cloud environment or on-premise infrastructure. Once the model is pre-trained on the custom dataset, it is ready for NER recognition on unknown data or resumes.

SPACY NLP is seamlessly integrated into the system to perform linguistic analysis on resumes. Leveraging SPACY's advanced linguistic features enhances the system's ability to capture syntactic and semantic elements, contributing to a more profound contextual understanding. It also helps in pipelining the task and also helps in creation of dataset for the fine-tuning of transformers.



Another major system lies in the development of a sophisticated job recommendation algorithm.

Once the entities are extracted from the resume then based on the skills of the candidate, we try to find the best matching job. Beautiful Soup is used to extract the jobs from different online job portals. The algorithm then calculates the similarity between the feature vectors of resumes and job descriptions, utilizing machine learning techniques such as cosine similarity to quantify the match. The outcome is a personalized and dynamic job recommendation system that changes in job requirements and candidate profiles.

## V. RESULT

In our research study, we have introduced an innovative approach to the parsing of resumes, focusing initially on the domains of Computer Science and Information Technology. Our current system, utilizing a deep learning model, has demonstrated a noteworthy accuracy level ranging from 94% to 95%. To advance our capabilities, we are committed to expanding our dataset to include resumes from diverse domains, thereby enhancing the model's applicability. Fine-tuning our knowledge base is another priority to ensure even greater accuracy. Furthermore, our ongoing efforts include the incorporation of various file formats and the expansion of entities within resumes such as Video files. These additions are anticipated to provide a more nuanced and comprehensive profile of candidates, contributing to a more sophisticated parsing outcome.

Notably, our system is offered free of charge to candidates, enabling them to familiarize themselves with the intricacies of Applicant Tracking System (ATS) software. Candidates are encouraged to incorporate relevant key terms for easy identification by prospective employers. Additionally, we envision implementing a feedback system for recruiting companies to contribute to a seamless and efficient hiring process.

## VI. CONCLUSIONS

In conclusion, the development and implementation of a Resume Analyser and Job Recommendation System, grounded in the innovative fusion of BERT NER Techniques and SPACY NLP, represent a significant stride forward in the realm of talent acquisition and recruitment processes. This hybrid approach, combining the contextualized understanding of BERT with the advanced linguistic analysis capabilities of SPACY, has demonstrated its efficacy in addressing the intricate challenges inherent in resume parsing, entity recognition, and job matching. Our study finds how good a model can perform to extract entities from the resumes and give it in a structured format. Deep learning models like BERT provide a very outstanding performance as they can work on different Resume formats and layouts. As a result of our research, the recruitment process will take less time and the selection will be more effective and efficient and unbiased for

any candidate. Overall, we showed how Deep Learning Algorithms can significantly increase and can intensify the recruitment process.

## FUTURE WORK

For Future work in this field, we will be adding some feature like MULTI-LANGUAGE SUPPORT to have a more broader user base, REAL-TIME JOB MARKET ANALYSIS which will allow our system to adapt with the current market trends, INTERVIEW PREPARATION ASSISTANCE which will provide interview related question based on the job and providing relevant tips to it.

## ACKNOWLEDGMENT

We are very thankful to our project guide Mr. Pawan Kumar Mall for providing us with timely guidelines. We are also thankful to Mr. Rajiv Kumar for his kind suggestions. Due to these guidelines and constant checks we are able to proceed further smoothly. We are also thankful for backing us in this project and showing confidence in us and project idea and very much thankful to our HOD Dr. Naresh Kumar and last but not the least our friends support us every time.

## REFERENCES

- [1] Kinge, Bhushan, et al. "Resume Screening Using Machine Learning and NLP: A Proposed System." (2022).
- [2] Tejaswini, K., et al. "Design and development of machine learning based resume ranking system." *Global Transitions Proceedings* 3.2 (2022): 371-375.
- [3] Vukadin, Davor, et al. "Information extraction from free-form CV documents in multiple languages." *IEEE Access* 9 (2021): 84559-84575.
- [4] Wosiak, Agnieszka. "Automated extraction of information from Polish resume documents in the IT recruitment process." *Procedia Computer Science* 192 (2021): 2432-2439.
- [5] Suresh, Yeresime, and A. Manusha Reddy. "A contextual model for information extraction in resume analytics using NLP's spacy." *Inventive Computation and Information Technologies: Proceedings of ICICIT 2020*. Singapore: Springer Singapore, 2021. 395-404.
- [6] Gaur, Bodhvi, et al. "Semi-supervised deep learning based named entity recognition model to parse education section of resumes." *Neural Computing and Applications* 33 (2021): 5705-5718.
- [7] Roy, Pradeep Kumar, Sarabjeet Singh Chowdhary, and Rocky Bhatia. "A machine learning approach for automation of resume recommendation system." *Procedia Computer Science* 167 (2020): 2318-2327.
- [8] Bhaliya, Nirali, Jay Gandhi, and Dheeraj Kumar Singh. "NLP based extraction of relevant resume using machine learning." (2020).
- [9] Gunaseelan, B., Supriya Mandal, and V. Rajagopalan. "Automatic Extraction of Segments from Resumes using Machine Learning." *2020 IEEE 17th India Council International Conference (INDICON)*. IEEE, 2020.
- [10] Amin, Sujit, et al. "Web application for screening resume." *2019 International Conference on Nascent Technologies in Engineering (ICNTE)*. IEEE, 2019.
- [11] Peters et al., 2019a. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" *arXiv:1810.04805v2 [cs.CL]* 24 May 2019
- [12] Narendra GO and Hashwant S, "Named Entity Recognition Based Resume Parser and Summarizer". *International Journal of Advanced Research in Science, Communication and Technology (IJARST)*