

AI-Powered Model for Intelligent Resume Recommendation and Feedback

Abhishek Mishra

*Department of Computer Science and
Applications Amrita School of
Computing, Amrita Vishwa*

Vidyapeetham Amritapuri, Kerala, India

AM.EN.P2MCA22008@am.students.amrita.edu

Siddharth Singh

*Department of Computer Science and
Applications Amrita School of
Computing, Amrita Vishwa*

Vidyapeetham Amritapuri, Kerala, India

AM.EN.P2MCA22042@am.students.amrita.edu

Jisha R C

*Department of Computer
Science and Applications
Amrita School of Computing,
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India*

jisha@am.amrita.edu

Abstract—The research aims to develop a system that is capable of suggesting the most suitable resumes based on the job requirements provided by recruiters through uploaded files. The proposed System approach utilizes the Bert Model to enhance the precision and relevance of job recommendations, and ensure a seamless alignment between skills and job requirements. Furthermore, that system offers constructive feedback to candidates whose skills may not match the specified job requirements, going beyond the traditional job recommendation process. This feedback mechanism gives valuable insights for personal and professional growth, and it is also promoting a transparent and constructive interaction between candidates and the hiring process. In addition to providing instant resume recommendations and feedback, our model will also anticipate upcoming future job opportunities based on the candidate's skills set. This predictive feature is allowing candidate to strategically plan their career paths and stay ahead of evolving industry demands.

Index Terms—Bert Model, Resume Recommendation, Supervised learning, Feedback Mechanism

I. INTRODUCTION

In an ever-changing labor market, the problem of effectively matching candidates to suitable job opportunities has become crucial. With the increasing number of job applications and complexity of matching candidate skills with specific job requirements, it is necessary to integrate advanced technologies to streamline the hiring process. This research aims to contribute to this important task by introducing an innovative AI-powered model designed for intelligent job recommendations and feedback work. The main goal of this study is to develop a sophisticated system that will not only suggest the best resumes based on job requirements provided by recruiters, but also offer constructive feedback to candidates. By leveraging the transformation capabilities of Bert model and transformer architecture, this approach aims to improve the accuracy and relevance of job recommendations by understanding the fine grained job descriptions.

Through the use of natural language processing techniques, this model seeks to create a seamless connection between candidate profiles and the dynamic requirements of various job roles. This proposed system is going beyond to the traditional method by providing individualize feedback to candidates whose skills are not matching with the job

requirements. This feedback feature is not only increases the transparency of the recruitment process, but also gives a candidate valuable insights for their professional and personal growth.

In addition, the research expands its impact by incorporating a forward-looking aspect that involves predicting future job opportunities based on the candidate skill set. This foresight feature furnish candidates with the ability to proactively navigate their careers in line with the changing industry demands. In exploring the complexities of intelligent job recommendation and feedback, the incorporation of transformer-based models represents a significant advance in the use of artificial intelligence to address current challenges in the ever-changing labor market. The goal of this study is to make a valuable contribution to the ongoing discussion on the convergence of technology and recruitment leading to better, open candidate-centric strategies for career progression.

II. LITERATURE REVIEW

In [1] author proposed anticipated outcomes of the article include an overview of the current status of AI-based job recommendation systems and an analysis of the many strategies, technologies, and algorithms employed in this field. The investigation may focus on problems, patterns, and potential paths for these systems' creation and implementation. The writers most likely conducted a comprehensive assessment of the existing literature and research on this topic in order to offer insights and analysis. One way that the article could contribute is by pointing out areas in which research is lacking, offering potential fixes, and offering suggestions for future developments in AI-driven job recommendation systems.

The paper [2] by R. C. Tripathi and Chandramma discusses an intelligent system for analyzing resumes. This system aims to automatically analyze resumes and provide recommendations to improve them. The strategy or methodology mentioned in this study likely employs innovative algorithms or techniques for automatically analyzing resumes. Artificial intelligence (AI) techniques like machine learning or natural language processing (NLP) can be used to extract relevant information from resumes, such

as abilities, qualifications, and experience. Presumably, the authors propose a system that can intelligently compare resumes to job descriptions or criteria, selecting qualified applicants or providing resume improvements.

In [3] proposes an AI-powered platform to help older adults navigate career transitions. The platform leverages machine learning to match individuals with suitable job roles based on their skills, experience, and preferences. It also guides resume writing, interview preparation, and networking, aiming to empower older adults to re-enter the workforce or pursue new career paths. The system most likely uses machine learning algorithms to examine numerous criteria, such as talents, experience, and preferences, to recommend appropriate career routes and job prospects to users. The system's goal is to help individuals make informed career decisions and reach their full professional potential.

In [4] describes a system that uses machine learning to recommend careers to individuals. Based on the details provided, the system likely analyzes a person's characteristics and skills using machine learning algorithms to suggest job roles that would be a good fit. The system's objective is to give users personalized recommendations for jobs that fit their interests, strengths, and career objectives. In order to help people make educated decisions about their professional lives, the paper presumably outlines the methodology, algorithms, and possible advantages of this machine learning-based approach to work role fit and career suggestion.

In [5] offers an algorithm for automating the process of analyzing resumes and matching them to job specifications. Authors uses BERT, a cutting edge language model to evaluate resumes and job descriptions. Using BERT's contextual understanding of the language, the system can successfully extract relevant information from resumes and match them to job description requirements. This comprehensive approach promises to simplify the hiring process by automating candidate screening, which could potentially save hiring managers working time and there resources.

III. PROPOSED SYSTEM

The task involves selecting the most suitable CV [7] candidate from a collection using an AI-powered algorithm. The objective is to use artificial intelligence to effectively identify competitor resumes. The task aims to simplify the recruitment process by using sophisticated technology to identify the most suitable candidates, which improves the overall recruitment strategy. The objective is to increase the efficiency of CV evaluation by integrating artificial intelligence, leading to a more accurate and faster selection process to identify the most suitable contender in the talent pool.

We are creating an AI-based resume recommendation and feedback system with the primary goal of identifying the most suitable candidate's resume from a collection of resumes.

The dataset can be taken from various online sources including Kaggle, this dataset consists of three main components: ID, Profile, and Domain. Specifically, we are using the 'bert-base-uncased' model, which has been designed and optimized for this type of task. This approach relies heavily on BERT's ability to understand contextual nuances in textual material. By fine-tuning the dataset, the model generates embeddings that effectively capture small correlations between candidate skills and job prerequisites which improves the accuracy of recommendation.

A. Preprocessing

During the first phase of data preparation, a careful cleaning process was performed to remove the unusual and unnecessary features. This involved removing extraordinary individuals, numbers, and single-letter words, resulting in a more refined dataset free of uncommon phrases. The dataset was then tokenized using the Natural Language Toolkit (NLTK), which divides information into discrete tokens. The tokenized dataset underwent through necessary preparatory operations such as stop word removal, stemming, lemmatization, and de-duplication. This rigorous cleaning approach helped to create a structured and standardized data set that laid a solid foundation for subsequent studies and ensured data integrity and orderliness throughout the review process.

- 1) **Conversion strings to lowercase** : All words should be in lowercase to maintain consistency which avoids confusion caused by the different cases. This allows the person to focus on the meaning of the words rather than how they are spelled differently.
- 2) **Replace all single-letter words with an empty string**: Single words such as "a" or "I" have no meaning and often add noise to data collection without adding any value to it. These simple elements are crucial in the process of refining a dataset before editing. Removing single words simplifies the dataset, allowing ML and AI models to focus on more significant language patterns. This process improves the overall data quality and increases the efficiency and accuracy of further analysis and training.
- 3) **Removal of Stopwords** : These are the common words like "a", "the" or "is" that is often used in text but have no semantic meaning. Even though these terms are necessary for grammatical structure, machine learning models have a very narrow view of the text. Stop word elimination is a very basic preprocessing technique that reduces noise in a dataset and makes analysis more insightful. After removing these frequently used phrases, attention shifts to information-rich keywords, making it easier for the system to identify and extract meaningful patterns. This helps not only processing performance

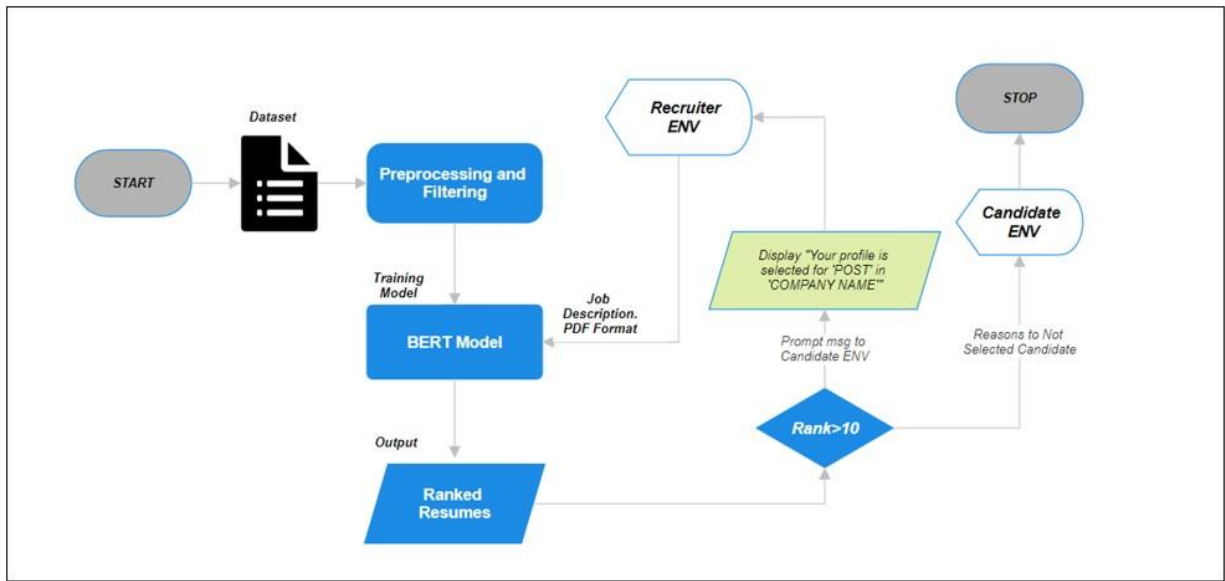


Fig. 1. System Flow Chat .

but also the model's ability to identify and evaluate the important text elements. Therefore, the removal of ignored words is essential to improve the textual data.

4) Stemming and Lemma formation :

To standardize words for natural language processing, stemming and lemmatization are text normalization techniques that reduce words to their most basic or root form. Lemmatization takes word meaning into account, whereas stemming distills words to their most basic form. By reducing word form variance, both techniques improve the model's ability to identify and extrapolate patterns.

Importance : Boosts model efficiency: These techniques assist the model in identifying recurring patterns in language by standardizing terminology. Minimizes multidimensionality: By reducing superfluous complexity and boosting processing efficiency, word variant reductions simplify the dataset.

B. Bert Model

BERT [6], or Bidirectional Encoder Representations from Transformers, aims to understand the contextual nuances in textual input using powerful natural language processing techniques. It is very good at capturing complex relationships between words by taking into account the context of the words that come before and after because it can understand the complex connections between applicant qualifications and job requirements, BERT contextual awareness makes it a valuable instrument in job design projects.

The primary advantage of using the BERT model is its ability to produce contextual embedding that contains

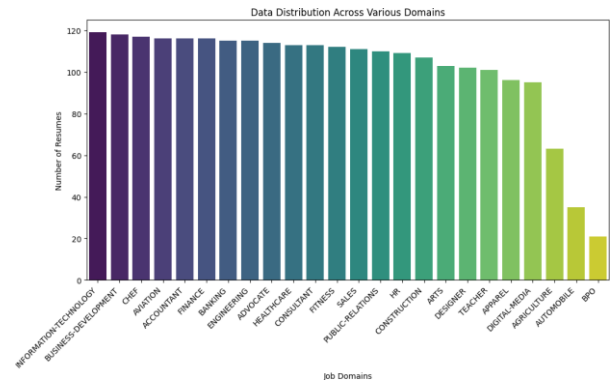


Fig. 2. Data Distribution across various domains.

accurate information about the semantic meaning of words in a given context. This is essential in job recommendation systems because the link between a candidate's skills and specific job criteria is nuanced and context-sensitive. BERT's contextual embedding helps in to more accurately display the complexities seen in the job descriptions and candidate profiles.

In addition, task-specific customization is possible thanks to BERT's fine-tuning capabilities, which ensure that the model is adapted at recognizing the subtleties of the job recommendation domain. The recommendation tool can provide more targeted and relevant recommendations using the BERT [10] model, which increases the standard of candidate matching as a whole. BERT's contextual knowledge captures the nuances of language and context, making job recommendations more targeted and informed. This greatly increases the efficiency of the research paper. Below we

defined a Bert training components

BERT Training Components

1) Word Embeddings in BERT :

XBERT=X WordPiece + X Positional

- 2) **Pre-training BERT** : Pre-training in the context of BERT (Bidirectional Encoder Representations from Transformers) includes the initial training of the model on a large corpus. The pre-training phase is often performed on a large dataset that allows BERT to get a compression perfect understanding of language patterns, grammar, and context. Pre-training helps BERT to go on a deep understanding of the language, making it a versatile and powerful tool for future specific applications. Now, We can pretrain the Bert Model in two ways are as follows:-

i) Masked Language Modeling (MLM):

In this method, a unique token known as "[MASK]" is replaced by several tokens that are randomly selected from a given input sequence. The model is then trained in a two-way approach to predict the original identity of masked tokens depending on the surrounding context. BERT learns to capture word dependencies and context bindings in previous and subsequent contexts by masking and predicting masked tokens. By iterating over large data sets, the model improves understanding of language nuances and complexities. MLM improves BERT's understanding of complex linguistic relationships and significantly increases its performance in various natural language understanding tasks such as question answering and sentiment analysis.

$$L_{MLM} = - \sum_i \log P(x_i | X_{BERT}, M)$$

ii) **Next Sentence Prediction (NSP)**: It is a training target used in large language models such as BERT. When using BERT, NSP needs to train a model to predict whether one phrase will follow the other phrase in two sentences. The model gains a knowledge of the relationship between successive sentences, which aids in the collection of contextual information and enhances performance on tasks like answering questions and making inferences from natural language. An essential component of Bert's training to improve comprehension and production of coherent textual sequences is the NSP task.

$$L_{NSP} = - \log P(\text{IsNext} | A, B) \quad \text{if IsNext,} \\ - \log P(\text{NotNext} | A, B) \quad \text{otherwise}$$

C. Forntend Portion

The user-friendly frontend platform will ensure that job description files may be posted and removed without any problems. Recruiters will be able to effortlessly manage all of their tasks with the help of this program. Given the significance of the frontend, we want to create an

interface that enhances the user experience overall and makes it easier for recruiters to carry out their duties. The approach begins with the recruiter connecting to the DRF-based application and posting job descriptions. Django acts as the backend, managing API requests, processing supplied documents, and storing them to a specific folder on the local system. The folder is then accessible in real time via Google Colab. To connect with Colab, use Python code within the Colab environment to get and alter Django-stored documents. This allows for simple collaboration and interaction with job descriptions directly within the Google Colab environment.

In essence, the process includes posting job descriptions via the Django Rest Framework, Django handling API calls and storing documents locally, and Google Colab reading and interacting with these documents in real time.

IV. RESULTS

The outcomes showed significant differences in these models' efficiency. The potential for accurate resume categorization was showcased by the Random Forest, Logistic Regression, and Linear SVM, which exhibited competitive skills in identifying patterns within the textual data. Notably, the BERT model surpassed the conventional machine learning techniques, obtaining the maximum performance in identifying multiple job areas from different CV's profiles by utilizing contextual embedding and deep learning techniques.

Moreover, the Naive Bayes model despite having certain shortcomings when compared to its competitors gave useful insights into the strength of simpler models for similar tasks. Its simplicity offers a fundamental comprehension that aids in the investigation of different approaches to data classification and pattern identification. These results underscore the necessity of deploying sophisticated NLP models like BERT for handling intricate and nuanced textual data, such as resumes. Additionally, this research reveals the essential complexities associated with model selection, highlighting the crucial need of careful weighing task-specific requirements to avoid model complexity.

In conclusion, the analysis of multiple models gives insight view on possible routes for improving resume categorization jobs. By using advanced approaches, such as BERT embedding can considerably improves the ability of automated systems in analyzing and classifying resumes across various job sectors.

Classifier	Accuracy
BERT	81.6%
Random Forest	65.2%
Logistic Regression	65.9%
Linear SVM	66.2%
Naive Bayes	53.3%

TABLE I: Accuracy Comparison of Different Models

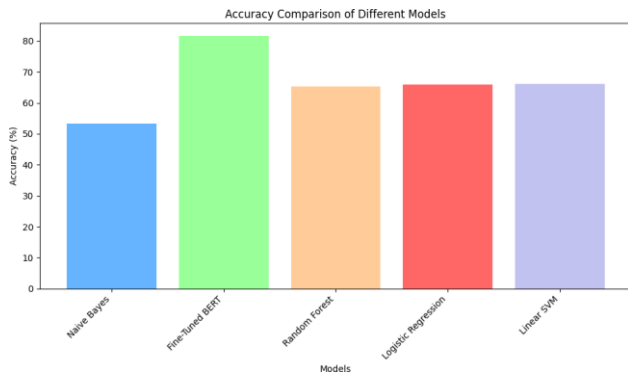


Fig. 3. Results from different models.

1) BERT (Bidirectional Encoder Representations from Transformers)

Google created this pre-trained deep learning model which performs exceptionally well on a range of natural language processing tasks. It contains contextual information by using the Transformer architecture and bidirectional context, which improves comprehension of linguistic subtleties. This model has made tremendous progress in language understanding for applications such as question answering and sentiment analysis.

Assumptions and Limitations :

For efficient language interpretation, this model depends on contextual embedding and pre-training on big datasets. Whereas the limitation is to handle non-linear interactions, extensive feature engineering is necessary, as it assumes linear correlations.

2) Linear SVM (Support Vector Machine)

For classification problems, supervised learning algorithms like (SVM) are employed. Its goal is to identify ideal hyperplane in the feature space for class separation. It also makes generalization better by optimizing the margin between classes.

Assumptions and Limitations : The incapacity of linear SVM is to model non-linear connections, and their susceptibility to feature scaling their quadratic training time with increasing sample size limit their ability to optimize the margin between linearly separable class.

3) Random Forest

By averaging the outputs of several decision trees, this model generally reduces over-fitting, handles big datasets, and enhances prediction accuracy. This resilient performance is achieved for both regression and classification problems.

Assumptions and Limitations : Assuming feature independence and ensemble learning enhances performance, these models are computationally

expensive, less effective on unstructured text without significant preprocessing, and more complex and difficult to read when additional trees are added.

4) Naive Bayes

Given the class label, the probabilistic classifier. This model generally assumes feature independence. Despite its naive assumptions, it is computationally efficient, handles huge datasets efficiently, and does well in text categorization.

Assumptions and Limitations : The Naive Bayes relies on the assumptions of feature independence and equal relevance due to which it performs badly on real-world data, where both assumptions frequently break down. It also performs poorly when handling strongly skewed distributions and poorly linked features, which limits its applicability to a variety of datasets. Use sophisticated models like Random Forest which address feature dependencies in order to get around Naive Bayes' drawbacks. Alternatively, tackle dependencies by developing new features that capture preexisting ones. Use feature selection techniques such as chi-squared tests to lessen sensitivity to irrelevant characteristics, and also employ regularization to prevent over fitting.

5) Logistic Regression

Logistic regression is a statistical model used in binary classification tasks. To represent the connection between input characteristic and the likelihood that an instance falls into a particular class, it employs a logistic function.

Assumptions and Limitations : Assumption of this model is having no significant col-linearity across features; linear relationships between input features and log-odds; observations are independent. Whereas limitation of this model dependence on linear relationships may not hold for complex datasets; requires extensive feature engineering to address non-linear correlations and interactions.

V. FUTURE SCOPE

Future areas in this research might look into adding other data sources to the AI-based resume suggestion and feedback system to give it a more thorough grasp of candidate profiles. Examining methods to integrate the current employment market characteristics into the recommendation system may enhance its precision and pertinence. Furthermore, investigating the possibility of integrating user feedback mechanisms may improve the adaptability and personalization of the system. By using AI-powered technologies that effectively match candidates with job roles and organizations may use these insights to improve their recruitment processes. This will save time and re-

sources while guaranteeing greater candidate-job fit. This study provides insightful information to companies looking to implement cutting edge technologies in order to maximize their hiring practices and raise overall efficacy. Overall, Organizations that use these insights have the potential to transform their hiring methods, making them more efficient, unbiased, and effective.

REFERENCES

- [1] A. Patil, D. Suwalka, A. Kumar, G. Rai and J. Saha, "A Survey on Artificial Intelligence (AI) based Job Recommendation Systems," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 730-737, doi: 10.1109/ICSCDS56580.2023.10104718.
- [2] R. C. Tripathi and Chandramma, "An Approach of Intelligent Automated Resume Analysis Recommendations," 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-5, doi: 10.1109/ICERECT56837.2022.10059603.
- [3] A. Jain et al., "Career Support Platform for Older Adults Powered by AI," 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 47-53, doi: 10.1109/CCWC57344.2023.10099112.
- [4] S. K. A. S., S. S., A. S. M and S. K. S., "Machine Learning based Ideal Job Role Fit and Career Recommendation System," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 64-67, doi: 10.1109/ICCMC56507.2023.10084315.
- [5] A. Pimpalkar, A. Lalwani, R. Chaudhari, M. Inshall, M. Dalwani and T. Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2023, pp. 1-5, doi: 10.1109/SCEECS57921.2023.10063010.
- [6] Bhatia, V., Rawat, P., Kumar, A. and Shah, R.R., 2019. End-to-end resume parsing and finding candidates for a job description using bert. arXiv preprint arXiv:1910.03089.
- [7] Yan Wang, Yacine Allouache, Christian Joubert. Analysing CV Corpus for Finding Suitable Candidates using Knowledge Graph and BERT. DBKDA 2021, The Thirteenth International Conference on Advances in Databases, Knowledge, and Data Applications, May 2021, Valencia, Spain. (hal-03325062).
- [8] T. S. Barros "Leveraging BERT for extractive text summarization on federal police documents," Knowledge and Information Systems, vol. 49, no. 5, pp. 1595-1613, 2023.
- [9] D. Zhang, H. Cao, Y. Li, B. Liu, and J. Hu, "Leveraging clinical BERT in multimodal mortality prediction models for COVID-19," in Proceedings of IEEE International Symposium on Computer-Based Medical Systems (CMBS), pp. 199-204, 2022.
- [10] Sudarshan, R., D. Sasikala, and S. Kalavathi. "Advancing Clinical Text Summarization through Extractive Methods using BERT-Based Models on the NBME Dataset." 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS). IEEE, 2023.
- [11] Sriharsha, C., et al. "Intelligent learning assistant using BERT and LSTM." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021.
- [12] K. S. Naveenkumar, R. Vinayakumar and K. P. Soman, "Amrita-CEN-SentiDB: Twitter Dataset for Sentimental Analysis and Application of Classical Machine Learning and Deep Learning," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019.
- [13] Prathibhamol, C. P., K. V. Jyothy, and B. Noora. "Multi label classification based on logistic regression (MLC-LR)." 2016 International conference on advances in computing, communications and informatics (ICACCI). IEEE, 2016.
- [14] A. Ajesh, J. Nair and P. S. Jijin, "A random forest approach for rating-based recommender system," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp. 1293-1297, doi: 10.1109/ICACCI.2016.7732225.
- [15] O. Balamurali, A. M. Abhishek Sai, M. Karthikeya and S. Anand, "Sentiment Analysis for Better User Experience in Tourism Chatbot using LSTM and LLM," 2023 9th International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2023, pp. 456-462, doi: 10.1109/ICSC60394.2023.10441148.
- [16] T. Aravind, B. S. Reddy, S. Avinash and J. G., "A Comparative Study on Machine Learning Algorithms for Predicting the Placement Information of Under Graduate Students," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 542-546, doi: 10.1109/I-SMAC47947.2019.9032654.
- [17] Kumar, C Maharana, Advait Murali, Srinath B., Premjith Kp, Soman. (2022). BERT-Based Sequence Labelling Approach for Dependency Parsing in Tamil. 1-8. 10.18653/v1/2022.dravidianlangtech-1.1.