Large Language Models for Indian Legal Text Summarisation

Hemanth Kumar M, Jayanth P and Anand Kumar M
Artificial Intelligence, Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India

2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) | 979-8-3503-8592-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONECCT62155.2024.10677065

***Abstract*—Summarizing legal case judgments is a complex task in Legal Natural Language Processing (NLP), with a gap in understanding how various summarization models, including extractive and abstractive approaches and analysing the perform within the domain of legal documents. Since there are around 4 crore pending cases in the Indian court system, this study addresses the challenge of laborious task of manually summarizing legal documents. It introduces both supervised and unsupervised models for both extractive and abstractive summarization, showcasing their effective performance through evaluations using ROUGE metrics and BERT score. BART, T5, PEGASUS, ROBERTA, Legal-PEGASUS, Legal-BERT models are used for abstractive summarisation. TextRank, LexRank, LSA, Summarizer BERT, KL-Summ are used in case of extractive summarisation. Longformer, Bert - Legal Pegasus are also considered for the task of Summarisation. In the domain of legal document summarization, we used GPT-4 and LLAMA-2, employing prompt engineering with both Zero-shot and One-shot prompts to extract summaries. As far of our knowledge, this is the first paper that used Large Language Models like GPT-4 and LLama-2 for the task of Legal Text summarisation. Along with that a user-friendly chatbot has been developed utilizing the Llama model and specifically designed to respond for queries related to legal texts. Additionally, a web application has been created, allowing users to upload legal documents for summarization. An option is given to users to select from various languages including Telugu, Tamil, Kannada, Malayalam, and Hindi. As a result the summarised text is converted into respective language.**

***Index Terms*—Legal ,Summarisation, NLP, Large language models, GPT-4, LLAMA-2**

I. INTRODUCTION

Introducing the concept of text summarization, this process involves condensing lengthy documents into concise summaries, offering a quick overview of the content [14]. There exist two methodologies for text summarization: extractive summarization and abstractive summarization. Extractive summarization involves pinpointing crucial sentences or phrases within the text and extracting them to formulate a summary. Widely adopted, this method is comparatively straightforward to implement. Nevertheless, it occasionally results in summaries that may seem fragmented or disconnected. On the other hand, abstractive summarization entails creating a novel summary that represents a fluid and coherent rephrasing of the original text. While this approach is more complicated, it has the potential to yield summaries that are rich in information and more captivating. Document summarization is not a new problem. There are many document summarizations like News document summarization, Scientific article summarization.But, there is a difference in Legal document summarization. In News document summarization, the document is short and usually, first paragraph summarises the document. In scientific article summarization, the article is long, but the text in the article is segmented into para headings. Where as, a court case or legal document is long and there are no segments and no para headings. This makes the legal document summarization bit more complicated.

Legal documents, known for their length and unique abbreviations, often require labor-intensive manual summarization. The emergence of Artificial Intelligence (AI) and Machine Learning (ML) has paved the way for automatic summarization, potentially saving significant time and effort [12]. This innovation holds particular promise in the legal domain, not only benefiting professionals but also facilitating understanding for beginners and the general public. With an overwhelming backlog of over 4.70 crore pending cases in various Indian courts [2], the adoption of automatic summarization came as a potential solution. The remaining sections of the paper are described as follows. Section - II shows the prior studies and literature survey. Section - III shows Dataset Explanation. Section - IV shows Experiments, Section - V shows the Deployment and Section - VI shows Results and Discussions and at last Section- VII shows Conclusion and Future Works.

II. LITERATURE SURVEY

There exist many prior works related to both extractive and abstractive summarization tasks. For extractive summarization, domain independent models (Supervised : BERT etc and Unsupervised : Lex Rank, Text Rank and LSA etc.) and domain dependent models (Gist etc.) are available. For Abstractive summarization, domain independent models (BART, T5 , Pegasus) and domain dependent models (Legal Pegasus) are available. Still there is a gap in abstractive summarization.

Satyajit et al. in [5] introduced an innovative text normalization approach for the summarization of legal documents within the Indian legal system. The proposed method demonstrates significant improvements in ROUGE scores when compared to raw text, showcasing enhanced precision, recall, and F-scores across various ROUGE categories, such as ROUGE-1 (precision: 0.46, recall: 0.55, F-score: 0.48). Apart from the current emphasis on widely used evaluation metric ROUGE scores, it's worth noting that for evaluating text summarisation, metrics like BERTScore [11] are recognized as valuable.

Jain et.al in [6] addressed the critical need for in-depth research in legal text processing considering the large pool of incomplete legal cases. It focuses on developing systems for efficient summarization, benefiting both legal professionals and the general public. Anand et.al in [3] explored document summarization methods by specifically focusing on two approaches: abstractive Fig 2 and extractive Fig 1. The inherent complexity and intricacies embedded within these documents make comprehensive understanding a time-consuming and labor-intensive endeavor [8].

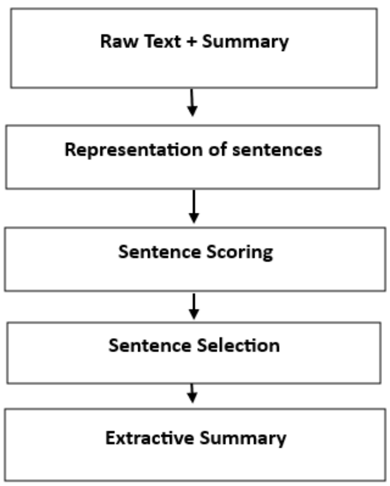

Fig. 1: Extractive Summarisation

A recent investigation by Saloni Sharma et.al [11] delved into the efficiency of seven machine learning-based summarization models on judgment report datasets sourced from the Indian national legal portal.

Within summary-based methodologies, a legal document's condensed version can be produced through automated summarization techniques or by noticing specific paragraphs serving as a comprehensive summary of the entire document [10]. Particularly prevalent in legal text document summarization, extractive summarization emerges as a favored approach, motivating to develop systems capable of autonomously generating summaries for extensive and intricate legal datasets [4].

One of the critical advantages of automatic summarization [7] lies in its potential to democratize legal understanding, allowing individuals without extensive legal training to grasp the essential points within complex documents.

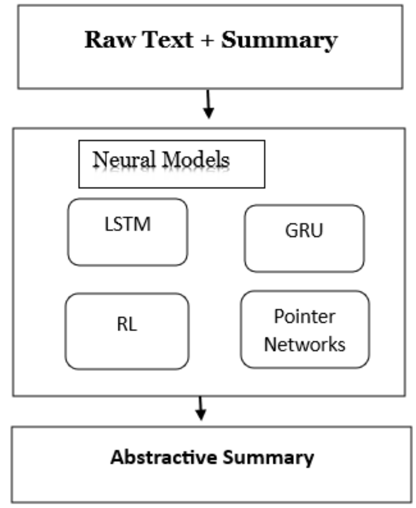

Fig. 2: Abstractive Summarization

The research objectives for this study are:

- To build a Hybrid Approach for legal document summarization. As the legal documents are large , if Hybrid approach is used, it can reduce the complexity of the document.
- While extensive literature exists on extractive summarization, a notable gap is still maintained in the domain of Abstractive Summarization, which increased research attention.
- The literature survey highlights a substantial research gap in the development of evaluation metrics tailored to the legal domain for summarization tasks [9]. Existing metrics such as ROUGE are insufficient in evaluating the quality of system-generated summaries. In response, leveraging BERT for paragraph similarity calculation emerges as a viable approach to address existing limitations.

III. Dataset Explanation

A scarcity of publicly available datasets specifically designed for legal case document summarization, particularly in the English language, is evident in the current landscape. The dataset [1] , we collected contains 7130 documents for Abstractive summarization. Out of which, 5624 are used for training and 1406 are used as validation, the remaining 100 documents are used for testing. There are 50 documents for Extractive summarization task.

IV. Experiments

We proposed experiments is as shown in Fig 3.we divided our experiments into four sections.

- Section-1 : Abstractive methods
- Section-2 : Extractive methods
- Section-3 : Hybrid Models

[1] https://zenodo.org/records/7152317.Yz6mJ9JByC0

- Section-4 : Large Language Models

In section-1, we considered two different types experiments. one is Domain Independent (BART, T5, Pegasus) and other is Domain Dependent (Legal Pegasus, Legal BERT, Encoder-Decoder). In section-2, various Extractive techniques are considered which includes, Text Rank, Lex rank, LSA, BERT and KL Summ. In section-3, we considered Hybrid model. At first, BERT is used for extractive summarization , then Legal Pegasus is used for abstractive summarization. In section-4, we considered Large Language Models like GPT-4 and Llama-2 for the task of Legal Text summarization using text generation pipeline.

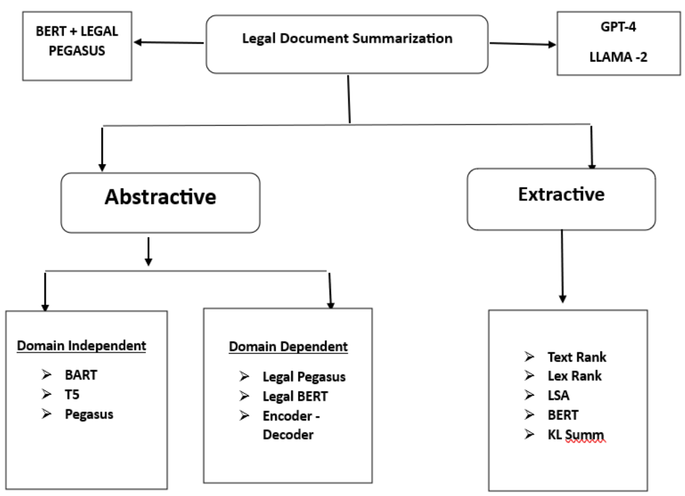

Fig. 3: Proposed Experiments

A. Abstractive Summarization

These summary and judgement text are stored in .txt format. Initially these text files are converted into csv format, with columns Text, Summary and Id, where Text column describes judgement,Summary column describes summary and Id column describes, the file name. This is done for three sets i.e., train,test and validation.

1) Domain Independent

The Models that are considered for the task of summarization are BART , T5 , Pegasus and Roberta.

2) Domain Dependent

The models that are considered for legal document summarization which are domain dependent are Legal Pegasus, Legal BERT, Encoder- Decoder.

B. Extractive Summarization

The Models that are considered for the task of Extractive summarization are TextRank, LexRank, LSA, Summarizer BERT, KL-Summ.

C. Hybrid Model

The Hybrid Models that are considered for this task are Longformer, Bert-Pegasus.

- Longformer, a transformer based model specifically designed for processing long documents. Longformer-base-4096, which is inspired from BERT architecture initialized from the RoBERTa checkpoint.

- BERT-Legal PEGASUS, which is designed as a hybrid model utilizing both BERT for extractive summarization by capturing essential details directly and Legal PEGASUS for abstractive summarization, generating fluent summaries with better context understanding.

D. Large Language Models

In our exploration of advancing legal document summarization, we employed large language models, namely GPT-4 and LLAMA-2. we used Prompt Engineering Techniques for the task of Legal document summarization using GPT-4 and LLAMA-2. We considered Text Generation Pipeline for this task. Zero shot and One shot prompts are considered for using Langchain.

Prompt engineering involves carefully creating input prompts to understand the behavior of language models during the summarization process. In our approach, we experimented with both Zero-shot and One-shot prompts. The Zero-shot prompts allow the models to generate summaries without explicit training on legal examples and based on their pre-existing knowledge and general language understanding. On the other hand, One-shot prompts involve providing the models with a single example or prompt that explicitly instructs them on the expected summarization output.

V. Deployment

A. Chatbot

we created the legal chatbot as shown in Fig 4,where user can ask his/her queries related to summary generated, so that the legal chatbot can answer those. A chat bot designed for legal queries and document summarization with the LangChain library, Hugging Face embeddings, and the Llama language model. This chatbot incorporated a custom prompt template for question-answering retrieval, effectively blending contextual information and user queries. we created a vector database for text documents(dataset) using LangChain.

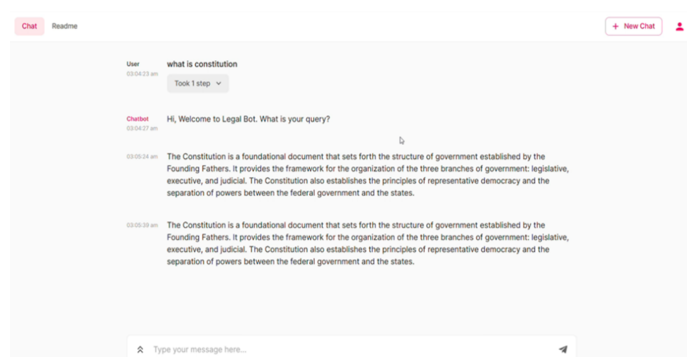

Fig. 4: Legal Text Chat Bot

TABLE I: Performance of the models for Abstractive summarization

Model	ROUGE - 1	ROUGE - 2	ROUGE - L	BERT SCORE
BART	0.32	0.09	0.18	0.69
T5	0.12	0.02	0.093	0.57
Pegasus	0.30	0.06	0.17	0.75
Roberta	0.34	0.04	0.19	0.64
Legal Pegasus	0.43	0.12	0.27	0.77
Legal BERT	0.19	0.05	0.27	0.62
Encoder - Decoder	0.21	0.08	0.17	0.68

TABLE II: Performance of the models for Extractive summarization

Model	ROUGE - 1	ROUGE - 2	ROUGE - L	BERT SCORE
LSA	0.30	0.07	0.14	0.71
Lex Rank	0.42	0.11	0.16	0.82
Text Rank	0.44	0.12	0.18	0.85
KL Summ	0.20	0.06	0.19	0.63
BERT	0.28	0.07	0.14	0.59

TABLE III: Performance of Hybrid models for the task Legal Text summarization

Metric	Longformer	Hybrid
Rouge-1	0.24	0.25
Rouge-2	0.11	0.07
Rouge-L	0.19	0.14
BERT Score	0.81	0.73

TABLE IV: Performance of LLama for the task Legal Text summarisation (Abstractive)

Metric	Zero-shot	One-shot
Rouge-1	0.301	0.228
Rouge-2	0.098	0.066
Rouge-L	0.250	0.179
BERT Score	0.779	0.795

TABLE V: Performance of LLama for the task Legal Text summarization(Extractive)

Metric	Zero-shot	One-shot
Rouge-1	0.368	0.335
Rouge-2	0.115	0.048
Rouge-L	0.216	0.176
BERT Score	0.65	0.68

TABLE VI: Performance of GPT-4 for the task Legal Text summarization (Abstractive)

Metric	Zero-shot	One-shot
Rouge-1	0.18	0.31
Rouge-2	0.05	0.12
Rouge-L	0.16	0.29
BERT Score	0.78	0.82

B. Deployment

We created a website for legal document summarization as shown in Fig 5. Users initiate the process by uploading their legal documents directly through the interface. The back-end Legal Pegasus, then deploys advanced natural language processing techniques to summarise. This results in the generation of a clear and concise summary. This also allows users to select their preferred language for the translation of the summary. The Languages the user can select includes Telugu, Tamil, Kannada, Malayalam, Hindi.

VI. RESULTS AND DISCUSSIONS

In abstractive summarization performance, Legal Pegasus outperforms other models. It achieved the highest scores in ROUGE-1 (0.43), ROUGE-2 (0.12), and ROUGE-L (0.27) as shown in Table I , indicating its superior ability to generate summaries with overlap in unigrams, bigrams, and longer sequences. Legal Pegasus also excels in BERT Score (0.77), for abstractive summarization. While BART and Roberta demonstrate competitive performance in certain metrics, Legal Pegasus consistently emerges as the top-performing model, showing its effectiveness in abstractive summarization tasks.

In extractive summarization, Text Rank stands out as the leading model . It attains the highest scores in ROUGE-1 (0.44), ROUGE-2 (0.12), and ROUGE-L (0.18) as shown in Table II, show casing its effectiveness in selecting sentences

TABLE VII: Performance of GPT-4 for the task Legal Text summarization(Extractive)

Metric	Zero-shot	One-shot
Rouge-1	0.31	0.51
Rouge-2	0.047	0.105
Rouge-L	0.167	0.217
BERT Score	0.69	0.79

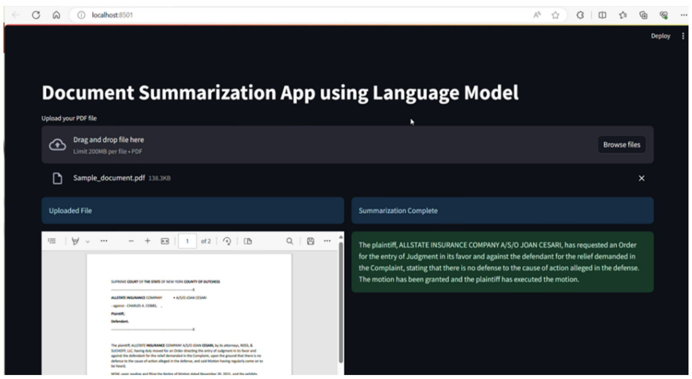

Fig. 5: Document Summarizer Web Deploy

that align well with reference summaries. Lex Rank also performs well, excelling in ROUGE-1 (0.42) and BERT Score (0.82), showcasing its proficiency in capturing important information for extraction. While KL Summ, LSA, and BERT shows reasonable performance, Text Rank emerges as the most robust model for extractive summarization tasks.

A Hybrid BERT-PEGASUS model performs better than Longformer when using Rouge-1(0.25) as shown in Table III . However, Longformer outperforms the Hybrid model in Rouge-2, Rouge-L and BertScore. So, in simpler terms, if we focus on Rouge-1 and Rouge-L, the Hybrid model is better, but for Rouge-2 and BERT Score, Longformer takes the lead proving its ability to handle longer sequences that ensures better semantic knowledge and similarity between the generated and reference summaries.

In case of Large Language Models, in the zero-shot scenario, LLama outperforms GPT-4 in Rouge-1 (0.301 vs. 0.18), Rouge-2 (0.098 vs. 0.05), and Rouge-L (0.25 vs. 0.16) scores as shown in Table IV and Table VI, showcasing its superiority in legal text summarization according to these Rouge metrics. However, in the one-shot scenario, GPT-4 exhibits better performance, surpassing LLama in Rouge-1 (0.31 vs. 0.228), Rouge-2 (0.12 vs. 0.066), and Rouge-L (0.29 vs. 0.179) scores. Interestingly, when considering BERT Score, a metric used for text summarization evaluation, GPT-4 consistently achieves higher scores in both zero-shot (0.78) and one-shot (0.82) scenarios compared to LLama (0.779 and 0.795). The results for Extractive summarization using GPT-4 and Llama is shown in Table VII and Table V respectively.

VII. Conclusion & Future Works

In summary, our work in legal text summarization introduced a novel approach in Indian legal context for both extractive and abstractive, using BART, T5, Pegasus, Roberta models and few legal domain specific models such as Legal-Pegasus, Legal-BERT. The proposed Legal Pegasus model obtained best results via rouge as well as bert score in abstractive summarization because of the advantage of pretraining in Legal domain. Hybrid Longformer methodology significantly improves ROUGE scores, emphasizing the effectiveness of domain-specific adaptations.

As a future scope, we aspire to look into the domain of summarizing legal text by employing instruction fine-tuning along with large language models, such as Llama-2 and GPT-4. Our aim is to create Automated Summarizer using the large language models like Llama-2 and GPT-4.

References

[1] Sharma, Saloni, et al. ”A Comprehensive Analysis of Indian Legal Documents Summarization Techniques.” SN Computer Science 4.5 (2023): 614.

[2] PTI, “Over 4.70 crore cases pending in various courts: Govt,” 3 2022.

[3] Anand, Deepa, and Rupali Wagh. ”Effective deep learning approaches for summarization of legal texts.” Journal of King Saud University-Computer and Information Sciences 34.5 (2022): 2141-2150.

[4] Satwick Gupta, M. N., et al. ”Extractive summarization of Indian legal documents.” Edge Analytics: Select Proceedings of 26th International Conference—ADCOM 2020. Singapore: Springer Singapore, 2022.

[5] Ghosh, Satyajit, Mousumi Dutta, and Tanaya Das. ”Indian Legal Text Summarization: A Text Normalization-based Approach.” 2022 IEEE 19th India Council International Conference (INDICON). IEEE, 2022.

[6] Jain, Deepali, Malaya Dutta Borah, and Anupam Biswas. ”Summarization of legal documents: Where are we now and the way forward.” Computer Science Review 40 (2021): 100388.

[7] Bhattacharya, Paheli, et al. ”Incorporating domain knowledge for extractive summarization of legal case documents.” Proceedings of the eighteenth international conference on artificial intelligence and law. 2021.

[8] Bhattacharya, Paheli, et al. ”Incorporating domain knowledge for extractive summarization of legal case documents.” Proceedings of the eighteenth international conference on artificial intelligence and law. 2021.

[9] Jain, Deepali, Malaya Dutta Borah, and Anupam Biswas. ”Summarization of legal documents: Where are we now and the way forward.” Computer Science Review 40 (2021): 100388.

[10] Trivedi, Anya, et al. ”Similarity Analysis of Legal Documents: A Survey.” ICT Analysis and Applications: Proceedings of ICT4SD 2020, Volume 2. Springer Singapore, 2021.

[11] Zhang, Tianyi, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. ”Bertscore: Evaluating text generation with bert.” arXiv preprint arXiv:1904.09675 (2019).

[12] Surden, Harry. ”Artificial intelligence and law: An overview.” Georgia State University Law Review 35 (2019): 19-22.

[13] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. ”Attention is all you need.” Advances in neural information processing systems 30 (2017)

[14] N. Andhale and L. A. Bewoor, ”An overview of Text Summarization techniques,” 2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, India, 2016, pp. 1-7, doi: 10.1109/ICCUBEA.2016.7860024.