LegalEase: Summarization of legal documents and query-answering chatbot

Aman Ranjan Megha Kunal Sapra Ishan
2021376 2021337 2021330 2021465

IIIT Delhi

Abstract

The project defines the issues consumers face when dealing with complex legal documents and getting trapped in unknown agreements.LegalEase aims to turn these documents into clear, comprehensible summaries and offer an interactive chatbot for customized inquiries.We have tried some famous Pretrained summary generation models from hugging faces like T5-transformer,Facebook-BART,Legal Pegasus and Pegasus.These are for general summary generation.We had fine tuned these models on our India legal dataset . We have also discussed various approaches to train our model.In the final prototype, we have integrated the summary generation with a chatbot for specific user query, 2 types of output that is text and audio and also in hindi language.

1 Introduction

We started with identifying the very common problem in the legal domain, i.e, understanding these lengthy documents and not missing on any important information/point.To proceed with this, we read some domain related research papers and their proposed solutions. We get to know many existing technologies that are used for this task and what are their limitations and how to improve upon them using fine-tuning and some IR techniques like Mean cosine similarity, etc.

2 Problem Statement

It takes a lot of time and confusion to sort through the complex legal documents, or websites terms and conditions. Most people don't have enough time or get irritated while going through complex terminology. In particular, when it comes to contracts, loans, and real estate transactions, current solutions fall short of providing user-friendly experiences. This lack of simplicity and clarity is a significant obstacle to making confident, well-informed decisions in essential transactions.

- Time-consuming
- Confusing
- Contracts
- Loans
- Unapproachable
- Real estate
- User-friendly
- Informed decisions
- Lack of simplicity
- Lack of clarity

3 Motivation

Most of us have faced issues related to this problem. Let's see what Kunal faced "A few years ago, my family bought an apartment. We were told that everyone would share the terrace. This was important to us. Later, the builders decided to add another floor and give the terrace to the top floor. When we complained, they showed us the agreement. It was50 page long and we needed to read it all carefully. It said that the builders could make these changes. We were disappointed but couldn't do anything. This experience made us realize how hard it is to understand legal documents." A wide range of people are impacted by the challenge of understanding legal documents, including small businesses without dedicated legal departments and individuals engaged in personal transactions.These documents often contain intricate language and terminology,making it difficult for the common individual to understand fully.Failing to comprehend the terms and conditions fully can have negative consequences, such as monetary losses and legal disputes. Not only that, nowadays most of the websites ask users to agree to their terms conditions, we found it important to be aware of all the terms to ignore scam etc. As a result, making it easier to understand these documents is vital for financial stability, legal safety, and convenience.

4 Novelty

Exiting products like AI, GPTs, and others are primarily trained in foreign databases and they are generic. But in the Indian context there are less similar things. Also, several services and tools, such as chatbots provide legal advice and document summarization tools, aim to demystify legal documents. These solutions frequently need to offer interactive, document-specific guidance and customized, thorough summaries. Most of the current available tools are generic and need to adjust to the particular complexities of individual documents. Moreover, they change the real meaning of the legal thing. Some research is being done in this domain, on how to improve the output and not miss any important information, but there's a lack of good prototypes using these solutions. Some students have worked on documents related to the judgment of Indian courts and made a summarization tool.However, we would like to train our model on other legal domains as well and integrate a query asking chatbot for specific doubts.

5 Solution

LegalEase is an innovative tool designed to simplify complex legal documents. It transforms legal judgments,contracts, loans, and real estate agreements into concise, easy-to-understand summaries.Users receive a clear breakdown of the key points and terms by uploading PDFs or scanning images of legal documents. We will also create an interactive chatbot tailored to each document, allowing users to ask specific questions and receive instant, understandable answers. This approach saves time and gives users confidence and clarity. In the future, we also plan to integrate it using a website extension. It also have feature of translation from one language to other(currently English to Hindi) and Text to voice .

6 Literature Review

6.1 Indian Legal Text Summarization: A Text Normalization-based Approach

Satyajit Ghosh, Mousumi Dutta, and Tanaya Das' research work addresses the difficulty of legal text summary in India. Due to the lack of summarized Indian legal documents dataset, they proposed a domain-independent approach. The steps involve text extraction and cleaning, text normalization, then fragmentation and then these fragments are given as input to the model. Finally, all the outputs are merged. They are using the BART model as PEGASUS performs poorly. Even without domain-specific training data, their experiments reveal that summarization efficacy improves significantly. *Conclusion:* Not much useful for abstractive summarization, this method loses semantic meaning. Read more here.

6.2 The Right to Remain Plain: Summarization and Simplification of Legal Documents

This Stanford paper discusses the existing work in the domain of text summarization and how to build upon these technologies using fine-tuning within the legal domain. It also further opens the possibilities of training the models on more specific domains in law. It uses BART as an example. *Stanford Paper.*

6.3 Legal Case Document Summarization: Extractive and Abstractive Methods and their Evaluation

The paper written by Abhay Shukla and others addresses challenges in legal case summarization. This paper compares the extractive and abstractive models based on their performance on legal documents. The challenge with the abstractive method is the token limit. In abstraction models, the Chunking method is used to process large amounts of data into chunks and then pass through the model. Legal Pegasus gave the best result followed by BART. But for segment-wise evaluation, Legal-LED performed better. In extractive methods, DSDR and SummaRuNNer are better. *Conclusion:* Abstractive summarization method has limited tokens. Although chunking overcomes this, we still end up getting incomplete and unorganized output. Extractive methods represent final judgment and status better but miss important precedents and arguments. Read more here.

6.4 Summarization of Legal Documents: Where are We Now and the Way Forward

The model employs various machine-learning models, including gradient boosting, multilayer perceptrons (MLPs), and deep learning methods with long short-term memory (LSTM) units. These models were utilized to tackle the problem of generating gist statements from judgment documents as a sentence classification issue. The performance of the models was assessed using precision, recall, and F1 measure. The best result achieved was an F1 measure of 0.9372, indicating high accuracy in selecting sentences that constitute the gist of legal documents. The techniques involved in the models' development include the use of legal linguistic statistical information, different word embedding methods for feature extraction, and the application of ensemble methods to combine predictions from multiple models. The research also explored the importance of contextual information surrounding sentence segments, employing LSTM units to capture such context effectively. *Conclusion:* Extractive summarization may miss the broader context of legal documents which can lead to wrong conclusions. The complex language and terminology is also an issue. Also, some other evaluation metrics are needed, to capture quality. And it is country specific. Read more here.

6.5 Extracting the Gist of Chinese Judgments of the Supreme Court

It takes lengthy legal documents as input, preprocesses them by tasks like tokenization and sentence segmentation, then utilizes various summarization techniques, including extractive and abstractive methods, to generate concise summaries. Extractive summarization involves selecting the most important sentences or passages from the original document based on criteria like relevance and importance, while abstractive summarization may involve generating new sentences that capture the essence of the text. The generated summaries are evaluated using metrics like ROUGE scores to assess their quality compared to reference summaries. The paper discusses both domain independent and domain-specific legal document summarization techniques. Despite the challenges of unique characteristics of legal documents and diverse structure and heavy usage of citations, the paper mentions the existence of tools like CaseSummarizer, which utilize word frequency and domain-specific knowledge to produce summaries tailored for legal professionals. Additionally, it presents two case studies focusing on automatic summarization of legal documents from different countries, providing a comparative analysis of various summarization techniques. *Conclusion:* Deep learning methods face challenges such as reliance on extensive data, overfitting, and opaque decision-making processes.It suggested some areas of improvement and advance NLP and ML research that is continuously going. Read more here.

7 Methodology

7.1 Approaches for Generating Summaries

In our project, we employed two primary approaches for generating summaries of legal documents:

1. **Extractive Summarization:** It involves selecting and extracting important sentences or phrases directly from the original document to form a summary. It involves top-k ranked sentences.The issue is it is unable to capture the semantic meaning of documents and may miss out important information.Also, as the sentences are extracted independently there is no correlation .

2. **Abstractive Summarization:** The system generates a summary by understanding the meaning and context of the original document and then formulating concise and coherent sentences to convey the main ideas.It requires advanced NLP and generation techniques .It has potential to generate novel insights.

Keeping in mind the pros and cons, we proceed with the Abstractive Summarization Method *Abstractive Summarization Method*. The limitation with existing abstractive summarization methods is the tokenization limit of 1024 tokens. To overcome this and fine-tune our model on these pre-trained models, we followed the methods outlined below:

7.2 Implementation Steps

1. **Data Collection:** From a variety of online sources, we collected summaries of Supreme Court rulings.

2. **Data Preprocessing:** To ensure consistency and cleanliness in our dataset, we preprocessed the data by performing operations such as eliminating white spaces, converting text to lowercase, managing missing values (NaN), eliminating special characters, and tokenizing the text.

3. **Chunking:** To effectively manage the Pegasus token limit while processing the text, we split the judgment files into smaller portions.

4. **Chunk Comparison:** We used mean cosine similarity to compare each chunk of the judgment to the given summary. This step was essential for determining the critical sections and evaluating how each chunk related to the summary.

5. **New Summary Generation for Judgment Chunks:** Alongside comparing each judgment chunk to the provided summary, we also generated new summaries for each chunk. This allowed us to create concise, informative summaries tailored to each judgment section.

6. **Fine-tuning Pegasus:** We fine-tuned the Pegasus-large model using the judgment chunks and their corresponding summary chunks. This process helped the model learn the nuances of legal language and summarization.

7. **Summary Generation for New Texts:** To generate a summary for a new legal text, we proposed a straightforward approach: break the document into chunks, generate a summary for each chunk using the fine-tuned model, and then concatenate these chunk summaries to form the full summary of the document.

7.3 Terms & Conditions Summarization

We have used T5-Transformer for generating a summary.The token limit was 512, so again we have used a chunking method. Summary for every chunk is extracted ,then that is combined and a new summary is found from this combined summary.In this we have directly called the model and skipped fine-tuning.

1. Gathered TC Texts

2. Preprocessed Data

3. Gathered TC Texts

4. Preprocessed Data

5. Chunking

6. Called T5 Transformer

7. Generated Summaries

8. Combined Summaries

9. Output final result

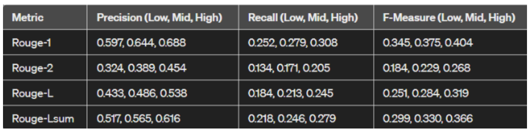
7.4 Chatbot for Legal Document Interaction

We used Google generative AI API for specific queries of users regarding any document.Along with pdf reader for inputting data.

8 Evaluation

8.1 ROUGE Score

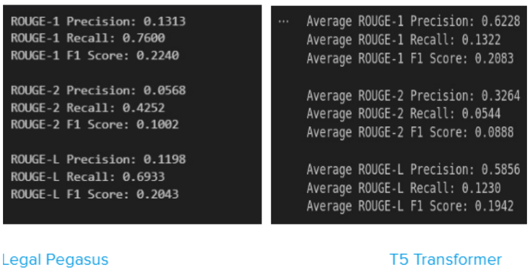
After fine-tuning, we evaluated the model's performance using Rouge scores, a standard metric in natural language processing for evaluating the quality of generated summaries compared to reference summaries. This section can be used to describe the technical steps needed to achieve the stated aim of the research proposal. **Our fine tuned model ROUGE Scores:**



Metric	Precision (Low, Mid, High)	Recall (Low, Mid, High)	F-Measure (Low, Mid, High)
Rouge-1	0.597, 0.644, 0.688	0.252, 0.279, 0.308	0.345, 0.375, 0.404
Rouge-2	0.324, 0.389, 0.454	0.134, 0.171, 0.205	0.184, 0.229, 0.268
Rouge-L	0.433, 0.486, 0.538	0.184, 0.213, 0.245	0.251, 0.284, 0.319
Rouge-Lsum	0.517, 0.565, 0.616	0.218, 0.246, 0.279	0.299, 0.330, 0.366

Figure 1: Fine tuned model ROUGE Scores.

Our fine tuned model ROUGE Scores:



```
ROUGE-1 Precision: 0.1313
ROUGE-1 Recall: 0.7600
ROUGE-1 F1 Score: 0.2240

ROUGE-2 Precision: 0.0568
ROUGE-2 Recall: 0.4252
ROUGE-2 F1 Score: 0.1002

ROUGE-L Precision: 0.1198
ROUGE-L Recall: 0.6933
ROUGE-L F1 Score: 0.2043
```

```
Average ROUGE-1 Precision: 0.6228
Average ROUGE-1 Recall: 0.1322
Average ROUGE-1 F1 Score: 0.2083

Average ROUGE-2 Precision: 0.3264
Average ROUGE-2 Recall: 0.0544
Average ROUGE-2 F1 Score: 0.0888

Average ROUGE-L Precision: 0.5856
Average ROUGE-L Recall: 0.1230
Average ROUGE-L F1 Score: 0.1942
```

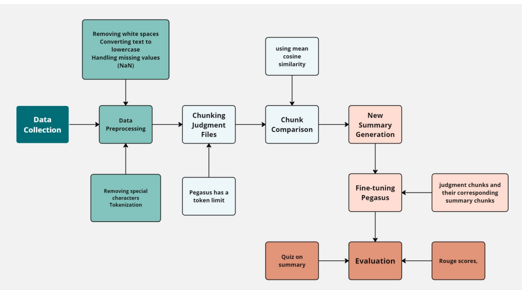
Legal Pegasus T5 Transformer

Figure 2: Fine tuned model ROUGE Scores.

8.2 Quiz answering evaluation:

We asked some participants to interact with the model by uploading a large document and understanding the summary. After the model gave a summary of the document, we took a quiz of the participants, based on their understanding of the summary. This helped us in improving upon the required aspects.

9 Model Pipeline



10 Technologies Used

Integrate information retrieval techniques for data generation .Then fine tunned model on this generated data. These technologies behind LegalEase :

10.1 Machine learning models

For summarization used fine-tune Pegasus and T5-Transformer. The chatbot uses Google generative AI API call, along with a pdf reader.

10.2 For Website

APIs, Flask, Python, HTML, CSS, Javascript, ML, Web development, Backend Server,streamlit.

10.3 Future work

Creation of website extension Optical character recognition (OCR) technology creates machine-readable text from scanned documents and photos.

11 Contributions

Aman Ranjan: Research paper study for getting approach ,Data filtering and Model fine tuning and model training (T5-Transformer,legal-pegasus),Text to audio conversion ,Text translation to other languages ,PPT and report

Megha: Research paper study,Legal-pegasus model training,Chatbot integration (frontend and backend),Terms conditions summarizer using T5-transformer,PPT and report

Ishan: Research paper study,Backend website development and integration model and various features

Kunal Sapra: Research paper study,Facebook BART and pegasus model fine tuning and training ,Frontend website development

12 References

https://arxiv.org/pdf/2206.06238.pdf https://web.stanford.edu/cl //aclanthology.org/2022.aacl − main.77.pdf https : //www.sciencedirect.com/science/article/abs/pii/S15740137210 //dl.acm.org/doi/10.1145/3322640.3326715