

# Apr 14th, 2024

## Approaches tried from last interaction:

1. T5 Encoder + FAISS ((Facebook AI Similarity Search)
2. T5 Encoder + BiLSTM (BiDirectional LSTM)
3. T5 Encoder + BiLSTM + CRF (Conditional Random Field)
4. LegalBert uncased
5. LegalBert large

## Motivation behind each approach:

1. **T5 + FAISS:** This approach combines two different tools. T5 is good at understanding and representing text data, while FAISS is good at quickly searching through large amounts of data to find similar items. Together, T5 can understand the legal text and identify potential places to segment it, while FAISS can quickly find similar document segments that can guide the segmentation process.
2. **T5 + BiLSTM:** This approach uses T5 to understand the text, and BiLSTM (a type of neural network) to capture the context and relationships between different parts of the text. This is helpful because the context is important for determining correct class for the sentence.
3. **T5 + BiLSTM + CRF:** This builds on the previous approach by adding CRF (another type of machine learning model). CRFs are good at labeling sequences, such as identifying the different parts of a sentence or document. This can be useful for classifying legal documents, where the boundaries often correspond to specific legal concepts or sections.
4. **LegalBert:** LegalBert is a pre-trained model that has been specifically designed to work with legal text. This means it already understands a lot about the language and concepts used in legal documents. This approach uses LegalBert's understanding of legal text to classify documents more accurately, and may require less fine-tuning or additional training compared to models that weren't designed for legal text.

## Results:

Approach	Macro F1	Precision
T5 Encoder + FAISS	0.4806	0.3524
T5 Encoder + BiLSTM	0.4703	0.3923
<b>T5 Encoder + BiLSTM + CRF</b>	<b>0.8166</b>	<b>0.6161</b>
LegalBert uncased	0.8089	0.5873
LegalBert large	0.7919	0.5777

**Next 2-3 weeks:** Writing a research report summarising all the approaches.