- Text summarization from basic to advanced approaches

-So what is summarization? Text summarization is the task of producing a concise and fluent summary of texts or documents while preserving key information.

There are two main strategies for summarization, which are extraction and abstraction. The goal of this project is to implement different Text summarization techniques on the same data set and compare them based on the same evaluation metrics.

-We use CNN and Daily Mail data set, which consists of online news articles paired with multiple summaries. This data is shown in the table.

The articles have an average of 781 tokens while the summaries have an average of 56 tokens.

- For our approaches,

First, our baseline model is a vanilla LSTM sequence-to-sequence model.

For our second model, we fine-tune the BART-base model. BART-base is a pre-trained transformer model trained on XSum which is a news data set. BART uses a standard sequence-to-sequence architecture with a bidirectional encoder similar to BERT and a left-to-right decoder similar to GPT.

For BART-large, we directly use the model without fine-tuning.

For T5, it is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and is trained on C4 data set. The goal of this model is to reframe all NLP tasks into a unified text-to-text-format where the input and output are always text strings.

-To evaluate the generated summaries, we use F1 from the ROUGE metric.

We select only three specific ROUGE metrics namely: ROUGE-1 and ROUGE-2 by measuring unigram and bigram similarity of reference summary and model-generated one. On the other hand, ROUGE-L measures the longest common subsequence between the two summaries.

-Here is our result.

From the table, we can see that vanilla LSTM model that was trained from scratch has a relatively disappointing performance in all ROUGE metrics compared with other models.

The performance of Transformer-based models like BART-large model and all three T5 models almost doubles the LSTM model. Although, they have very similar performance to each other in all metrics regardless of size of models and underlying datasets.

-Here are the examples of our model outputs compared with a reference summary.

We can see that the summary output from LSTM model is incorrect as seen in the word like Manchester United repeating itself in both subject and object.

In contrast, the outputs of other models are correct and understandable in terms of meaning and grammar.

-From our analysis, we conclude that

The accuracy of text summarization models can be increased by using transformer model.

Our current models can generate a correct and readable summary sentence in some cases.

Text summarization is still a challenging topic.

The overall accuracy of our models is still relatively low; and in other experiments, researchers can hardly generate a model with the accuracy higher than 0.4.

For Future work, we can incorporate Named-entity recognition to our model since large portion of summaries involve named entity.