**Into (10 words – using 10 [limit 660])**

Welcome everyone; our group members are Grufan, Prerna, and Fernando.

**Problem Statement (20 words – using 21 – total 31)**

Headnotes are brief case summary statements for court cases created by commercial third parties and may be under copyright protection.

For our project we use Natural Language Processing summarization algorithms to reconstruct headnotes using court case opinions as our training dataset.

**EDA, Filters, and Data Prep (90 – 188 – 219)**

We use the North Carolina dataset from the Case Law Access Project.

We selected cases for the last ten years, since two- thousand eight. We filtered the data based on the following conditions.

**Overview of Research Done (60 – 60 - 279)**

We focused on two state-of-the-art models, PreSumm and MatchSum, based on the following criteria.

**Describe Focus Areas & PreSumm & MatchSum (120 – 206 - 485)**

Our first model, PreSumm, was from the paper Text Summarization with Pretrained Encoders by Yang Liu and Mirella Lapata. They use combine text extractive and abstractive summarization.

PreSumm’s extractive model is a novel document-level encoder based on BERT stacked by several inter-sentence Transformer layers to capture document-level features for extracting sentences.

The abstractive model adopts an encoder-decoder architecture, combining the same pre-trained BERT encoder with a randomly-initialized Transformer decoder. It uses a new fine-tuning schedule as a means of alleviating the mismatch between the two.

The second model, MatchSum, was from the paper Extractive Summarization as Text Matching by Ming Zhong and others. They formulate an extractive summarization task as a semantic text-matching problem, in which a source document and candidate summaries will be (extracted from the original text) matched in a semantic space.

MatchSum uses a Siamese-BERT architecture, leveraging the pre-trained BERT to derive semantically meaningful text embeddings, to compute the similarity between several candidate summaries to the source document and selects the best candidate summary.

The Siamese networks consist of two identical neural networks, each taking one of the two inputs. The last layers of the two networks are then fed to a contrastive loss function, which calculates the similarity between the two inputs.

**Describe Result Metric &** [**Rouge Score**](https://rxnlp.com/how-rouge-works-for-evaluation-of-summarization-tasks/#.Xra4Ey33lE4) **(120 – 173 - 658)**

The models use the three Rouge scores, standard metrics, shown here, to measure the overlap between the generated and actual summaries.

**Show Example of What Worked and What Didn’t (120 – 226 - )**

For PreSumm, we first generated both extractive and abstractive summaries of the opinions in our dataset using the pretrained model. We then trained the model using our dataset and generated the summaries again and compared the results.

For MathchSum, we generated extractive summaries using the BERT and RoBERTa models and compared the two results.

Here we share an actual and MatchSum generated headnote. We see three segments that are related. The first is about the prevention of the defendant from working for all of plaintiff's current or recent clients, regardless of location. Next, the generated summary references the extreme scope. Lastly, it references the relatively small number of clients with whom defendant worked with.

In this PreSumm headnote sample we see two similar segments. The first is about how the trial court lacked jurisdiction to extend defendants period of probation verbatim. The second is referencing the arrest judgment and to vacate the order modifying probation and imposing sentence.

We can see the abstractive quality of the PreSumm generated headnote by the fragmented sentences in the initial words.

The table shows the Rogue scores for the various models. Our models do very well compared to the pre-trained Rouge scores. The MatchSum RoBERTa produced the highest scores.

Due to time constraints, we were not able to train the MatchSum hence the missing pre-trained Rouge score.

**Future Work, Learning, etc. (90 – 91 - )**

These models limit the maximum sentence length to 512 and we may want to alter the length to improve our results.

Due to time constraints, we were not able to utilize our full data, including the various opinions, to train the models. For future work, using all the available cases and opinions to train the models could improve the performance of generating headnotes.

**Closing (10 – 10 - 862)**

In conclusion, we were able to achieve high ROUGE scores with the models we chose.

Thank you for your undivided attention.