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

For this 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 – 189 – total 200)**

We are using a dataset of selected cases from the state of North Carolina provided by the Harvard Law School Library Case Law Access Project. There are ninety- seven thousand six hundred cases dating back to seventeen seventy- eight. There are thirty- two columns and we focused on the case body data, specifically the opinions and headnotes. Cases can have multiple or no opinions and headnotes. We observed up to six opinions in some cases.

We selected cases since two- thousand eight with headnotes having a length of more that one hundred fifty and majority opinions longer than the headnotes, around three thousand seven hundred cases.

We removed the return characters, extracted, and tokenized the opinions and headnotes from the case body data. During which we utilized the pre-processing steps from each model to label the opinions. Finally, we split the data into training, validation, and test sets.

For labeling, we compared sentences from each opinion with their headnotes, extracting those opinion sentences that increased the Rouge score to create the extractive summary. We then labeled the sentences that were included in the summary with one and zero otherwise.

**Overview of Research Done (60 words)**

**Describe Focus Areas & PreSumm & MatchSum (120 words)**

**Describe Result Metric & Rouge Score (120 words)**

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

**Future Work, Learning, etc. (90 words)**