

# ELEC-E550 Statistical Natural Language Processing

Pranav Arora, Yuvrajsinh Chudasama

February 2022

## 1 Abstract

The goal of this project is to implement a text summarization of academic articles to generate a summary of the article. Yale Scientific Article Summarization Dataset(ScisummNet) is used for the task as it is well suited for the task, contains the necessary human summary for comparison with the model generated summary. Different transformers models such as BERT,RoBERTa and GPT will be used for modelling. Basically, BERT is changed to make sentence embeddings for all the sentences in the article. This is achieved using the CLS tokens. The output from the BERT model is sentence vector. These vectors then passes through multiple layers to extract the summary of the document. Finally a a comparison is performed with the ground truth human summary using the Recall-Oriented Understudy for Gisting Evaluation(ROUGE) metric.