Assignment 4 Report

Cleaning and Preprocessing:

- Dataset is in CSV format.
- Loaded dataset using Pandas and converted the dataset into dataframe
- Removed Html tags.
- Normalized text
- Removed punctuations
- Removed stop words.
- Lemmatization.

Model Training:

- Used GPT2Tokenizer and GPT2LMHeadModel for tokenization and sequence modeling:
- Used Dataset class (GPT2summatyDataset) for handle tokenization and formatting.
- Used only 2 percent of dataset and split it into training and testing data.

Model Evalution:

- Model is evaluated using ROUGE scoring system. This compares generated and actual summaries based on overlap of n-grams.
- ROUGE-1, ROUGE-2, ROUGE-L.

```
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
```

```
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2':
Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL':
Score(precision=1.0, recall=1.0, fmeasure=1.0)}
```

{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}
{'rouge1': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rouge2': Score(precision=1.0, recall=1.0, fmeasure=1.0), 'rougeL': Score(precision=1.0, recall=1.0, fmeasure=1.0)}