Project Proposal

Problem statement:
Text Summarization - Condensing data but retaining content.

Motivation :
Automatic text summarization is the need of the hour to cope with ever-increasing amount of text data that is available online.
Recently, Deep learning approach in NLP has shown promising results for text summarization and I would like to explore it through a Capstone project.

Advantages of text summarization:
When textual data is summarized to shorter, more focused summaries:
1. It reduces reading time.
2. Selection process is easier, when researching.
3. Improves effectiveness of indexing.
4. Automatic text summarization is less biased than humans.
5. Useful in question-answering systems as they provide personalized information.

Dataset:
CNN News story dataset - a popular and free dataset for use in text summarization experiments with deep learning methods.This dataset contains more than 93,000 news articles where each article is stored in a single "*.story*" file.

Approach:
This is an supervised problem where the below approaches are employed for text summarization:

a. Extractive method :
Involves selection of phrases and sentences from the source document to make a new summary.

b. Abstractive method :
Involves generating entirely new phrases and sentences to capture the meaning of the source document.

The plan is to employ Deep Learning (abstractive approach) by framing the text summarization problem as a sequence-to-sequence learning problem with an improvement, using Pointer-Generator Networks.

Deliverable:
Application deployed as a web service with an API.

Computational resources:
GPU
32 GB Ram