CIS 621 - Big Data and Parallel Database Processing Systems

Fall 2022

Professor: Dr. Sunnie S. Chung

PROJECT REPORT

***“News Summarizer using BERT model”***

# Couse of action

* Explore the BERT language model and the underlying pipeline like transformers, encoders, and decoders. Along with exploring the additional linguistic features layer like:
  + NER: Name entity label
  + POS: The part-of-speech tag
  + DEP: Syntactic dependency, i.e., the relation between tokens
  + STOP: Is the token part of a stop list.
* Reading about huggingface transformers.
* Identifying the type of input data required.
* Pre-processing the raw data into BERT readable format.
* Analyzing the data to increase accuracy.
* Deciding on an evaluation metrics for error analysis
* Listing various input data sources.
* Extracting data from news publications’ websites.
* Understanding the implementation of existing projects built on BERT.
* Developing a RESTful API to host the model by wrapping it in a python Flask web application.
* Designing and building a front-end application to demonstrate the API we created.

# Current Status

We have distributed the tasks among all the teammates and every member of the group of progressing towards their assigned task. Below is a list of members and their task:

1. Anshul Maske (2830914)
   1. Reading about BERT (base) model.
   2. Trying Flask framework.
   3. Learning about RESTful APIs.
   4. Finding various news articles and publications
   5. Understanding the structure of news articles for web scrapping.