CS/CE 457/464 - Homework Assignment 10: NLP

Due Date: Monday, November 18 at 11:59 pm

Purpose:
Demonstrate understating of basic NLP techniques for text data

Points: 100

Deliverables: Submit ipynb code file along with your answer

- Perform NLP analysis on the following questions. Make sure to include interpretation of each result including text/visualizations/tables to support your answer

    1. Select any article on the website of your interest and perform Name Entity Relation (NER) analysis on it. Show all the necessary output and visualization and explain your findings. Also mention if you see any shortcoming or error in the algorithm output.
    2. Select any research paper of your interest and perform word frequency and keyword extraction analysis. Use at least 3 different techniques/libraries. Show all the necessary output and visualization and explain your findings. Compare different techniques and mention if you see any shortcoming or error in the algorithm output.
    3. Choose any keyword or twitter handler (page/account) of your choice and pull/scrap at least 20 (tweets) from Twitter. Manually put them in a text file. One tweet per line.
        a. Read the text file line by line.
        b. Perform Sentiment Analysis on the tweet and show total positive and negative sentiment counts.
        c. Display one Positive and one Negative sentiment and discuss if the algorithm is accurate.
    4. Pick any dataset and build recommendation engines using
        a. TF-IDF method
        b. Using word2vec (doc2vec) model (deep learning)
        c. Test both engines and compare the quality/accuracy of recommendations.