CS 7319 - Homework 2 - Bonus

Sentiment Analysis Sentiment Analysis with NLP and Real Data

Extra points toward the term project
Due: September 9, 2025

Notes:

Natural Language Processing is not a software architecture by itself. It's an application domain or technology.

But every NLP application (chatbot, sentiment analyzer, translation system, etc.) must be built on top of a well-thought-out software architecture to ensure it works efficiently, scales, and can be maintained.

NLP is the functionality, while software architecture is the structure that delivers that functionality.

Steps:

1. Watch the YouTube Video
   - Video: "Natural Language Processing in Python"
https://www.youtube.com/watch?v=xvqsFTUsOmc

   - Take notes on the key concepts of NLP (tokenization, stopwords, stemming, sentiment analysis, etc.).

2. Apply NLP on Real Web Data
   - Instead of using the provided text datasets, collect real-world text data from a public website or API (examples: news articles, product reviews, or social media posts).
   - Use any programming language or library (e.g., Python with NLTK) to:
     • Preprocess the text (tokenization, lowercasing, stopword removal).
     • Perform sentiment classification (positive, negative, neutral).
     • Compare your results against the keyword-based method in Parts A and B.

3. Deliverables

1. Screenshots of your program output.
2. Source code in a zip file.

3. A short write-up describing:
   • Which website or dataset you used.
   • Which NLP library you chose.
   • How your NLP-based sentiment results differed from the keyword-based analysis.

Learning Objective of Bonus

- To expand students' perspective beyond keyword matching.
- To introduce real-world data complexity (slang, sarcasm, varied word usage).
- To highlight how architectural choices (single-pass, MapReduce, NLP pipeline) affect implementation, scalability, and accuracy.