Final Project Description

Throught the course, you and your teammates will work together to use a real dataset and show an application of a text analytics approach in a real business, healthcare, policy, or legal setting. The scope and topic is up to you, but the project will be vealuated based on:

1-Impact it has (1) has anyone already done this type of analysis? 2) what is the expected business/economic/ or societal impact? 3)why the topic is of importance?)

2-Completeness of material:

- A maximum of two page document, epxlaining the problem, your analytics approach, and the expected impact.
- R or/and Python code.
- The data file(s) used.
- A video recording of the code running over the data and showing example results
- A 15- minute presentation.

The grading is done comparatively—that is, the top 10% projects (the outstanding ones) receive an A, next 20% receive A-, next 20% receive B+, next 20% receive a B, next 20% receive a B-, and the last 10% a grade between C+ and F according to the objective quality of the work turned in.

Here are examples of topics you can work on:

- **Named Entity Recognition for Legal/business Texts:**

  - Develop a system for identifying and classifying named entities (e.g., persons, organizations, locations) in legal documents.
  - Evaluate the accuracy and efficiency of the system on legal/business text datasets.

- **Fake News Detection:**

  - Build a model to detect fake news articles by analyzing the language, style, and context of the content.
  - Explore the use of natural language processing (NLP) techniques to identify misleading information.

- **Cross-Language Text Analysis:**

  - Perform text analytics on multilingual datasets, addressing challenges such as language translation, cultural nuances, and differences in text structure.
  - Investigate techniques for improving cross-language text analysis accuracy.

- **Healthcare Text Mining:**

- Analyze electronic health records using text mining techniques to extract valuable information such as patient outcomes, treatments, and disease patterns.
- Explore the challenges and opportunities in processing healthcare-related texts.

User Review Analysis for Product Improvement:

- Analyze user reviews for a set of products to identify common issues and areas for improvement.
- Utilize sentiment analysis and opinion mining techniques to extract insights for product enhancement.

Text-based Fraud Detection:

- Apply text analytics to detect fraudulent activities by analyzing textual data, such as transaction descriptions or user communication.
- Investigate the integration of text-based features with traditional fraud detection methods.