# CISB5123 Text Analytics
# Sem 2 2024/2025

## Lab Assignment 2

Sentiment Analysis is the process of classifying the content of documents as positive, negative and/or neutral. In this assignment, you will explore sentiment classification using the Amazon Fine Food Review dataset.

Dataset Description

This dataset consists of reviews of fine foods from Amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

The dataset consists of 10 columns:

1. Id – Row id
2. ProductId – Unique identifier for the product
3. UserId – Unique identifier for the user
4. ProfileName – Profile name of the user
5. #Helpfulness Numerator – Number of users who found the review helpful
6. #Helpfulness Denominator – Number of users who indicated whether they found the review helpful or not
7. Score – rating between 1 and 5
8. Time – Timestamp for the review
9. Summary – Brief summary of the review
10. Text – Text of the review

Source:
https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews?select=Reviews.csv

Instructions

1. Data Preprocessing:
   - Load the dataset and perform necessary preprocessing steps.
2. Feature Extraction:
   - Utilize appropriate techniques (e.g., Bag-of-Words, TF-IDF) to convert text data into numerical features.
3. Model Selection:
   - Experiment with different methods for sentiment classification from the lexicon-based and machine-learning based approaches.
4. Model Evaluation:
   - Evaluate the performance of each model using appropriate evaluation metrics.
   - Compare the performance of different models and analyze the results.
5. Discussion:
   - Discuss the strengths and weaknesses of the selected models for sentiment classification.

Deliverables

- Python code implementing the preprocessing, feature extraction, model training, and evaluation.
- In the script, include your name and ID a short paragraph on the discussion about the strengths and weaknesses of the selected models for sentiment classification.
- Upload the Python script (.ipynb file) AND the extracted data (.csv file) to your GitHub repository, and submit the link to the Python script in Brighten.

Note: Late submissions may incur penalties, and no modifications are allowed after the due date.