



CISB5123 Text Analytics Sem 2 2023/2024

Lab Assignment 3 Due Date: **31 May 2024**

This is a pair assignment. Please submit in a group of **TWO (2)** students.

For this assignment, you will be working with 'news_dataset.csv' that have no labelled topics and attempting to find 4 topics to assign the column 'text' to.

Based on the 'news_dataset.csv' dataset, perform the following tasks:

- Import the necessary libraries
- Read the data (use only the 'text' column)
- Perform text pre-processing (i.e. remove null values, remove stopwords, stemming, lemmatization)
- Perform LDA using Gensim
- Evaluate the LDA model using Coherence score
- Interpret the result

Deliverables

- Python code implementing the above tasks.
- In the script, include the name and ID of the group members and a short paragraph on the interpretation of the coherence score.
- Save the file as LabAssignment3_Group MembersID.ipynb (replace "GroupMembersID" with the actual IDs of the group members).
- Each group **MUST** upload the Python script (.ipynb file) to either one of the group member's 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.