Assignment 1: Text Classification

Objective

This assignment aims to help you understand the basics of text preprocessing, classification, and evaluation.

Instructions

Complete the tasks below. Each task specifies the marks assigned. Submit your code, outputs, and a brief explanation for each step.

Dataset: text_class

Tasks

Task 1: Data Exploration (2 Marks)

- **Instructions**:
    1. Load the sample dataset provided in **text_class.csv**.
    2. Display the first 5 rows of the dataset.
    3. Print the total number of rows and the count of unique labels in the dataset.
    4. Complete any necessary step which is required before preprocessing.

Task 2: Preprocessing Text Data (3 Marks)

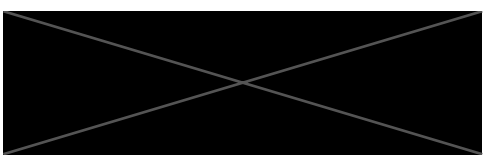
- **Instructions**:
    1. Convert all text to lowercase.
    2. Remove punctuation and special characters.
    3. Tokenize the text and remove stopwords. Provide the processed version of the first 5 rows.

Task 3: Train a Classifier (3 Marks)

- **Instructions**:
    1. Split the data into training and test sets (80% training, 20% testing).
    2. Train a simple logistic regression model.
    3. Predict the labels on the test set and calculate accuracy. Provide the accuracy score and a brief comment on the result.

Task 4: Evaluate the Model (2 Marks)

- **Instructions**:
    1. Evaluate the performance of the model.
    2. Write one or two sentences on how the confusion matrix helps analyze the results.
- Submit your Python code and output in a single Jupyter Notebook or script file.



- Include a brief explanation for each task in markdown cells or comments.