Task Description

Apply Text analysis Lifecycle on spam dataset to detect if the email is spam or ham.

The dataset contains 5572 rows × 2 columns (v1, v2), use this dataset to build a prediction model that will accurately classify which texts are spam.

- Apply the most appropriate preprocessing steps (Tokenization, stemming, lemmatization, etc.)
- Apply Feature Generation& Feature Extraction
- Apply the model (select the most suitable classifier)
- Evaluate the selected model (Accuracy, F1-score, Precision, Recall)
- Use Python and needed libraries like(nltk, pandas, sklearn)

The dataset file is attached to this file.

Deadline& Delivery:

- The project deadline will be on Wednesday 16/4/2025.
- Only one member can deliver and discuss the task with the assigned TA.

Teams:

- Each group between (3-6) members.
- This is the registration form and it will be close on Sunday 16/3/2025.
  Form: Registration Form