**Project 2: Text Data Classification**

**Due Date:** Nov. 4, 2024 (Monday), by 11:59 p.m.   
**Submission Format:** Individual

For this project, you will implement three machine learning classifiers: **Decision Tree, Random Forest (RF), and Naïve Bayes** using scikit-learn. These classifiers will be applied to a text dataset located in the folder cleanRaw. Please ensure the cleanRaw folder is in your working directory.

**Instructions:**

* Use the provided code template (Project2\_template.ipynb or Project2\_template.py) and make necessary modifications as you complete each task below.

**Tasks:**

1. **Classify Text Data**
   * Implement and apply the three classifiers (Decision Tree, Random Forest, and Naïve Bayes) to classify the text dataset provided.
2. **Evaluate Classifier Performance**
   * Compare the accuracy of each classifier on the training data to determine the best-performing model.
3. **Test Data Prediction with Best Classifier**
   * Using the best classifier identified in Step 2:
     + Predict the labels for the test dataset.
     + Display a confusion matrix for the predictions.
     + Report precision, recall, and F1-score metrics on the test data.
4. **Automate Best Classifier Selection**
   * Extend the code to automatically select the best classifier based on performance on the training data.
   * Use the selected model to predict the labels for the test dataset and display the corresponding evaluation metrics as in Step 3.

**Submission:**  
Submit your completed code in a Python file or Jupyter Notebook format.