Manual: How to Use the Classification Tool

This tool processes text data, applies preprocessing steps, trains a hybrid model using SMOTE, and evaluates the model based on classification metrics.

Step-by-step Usage:
1. Set project name:
   Update the project variable with the dataset name (e.g., 'tensorflow').

2. Ensure your data CSV is available in the path:
   Please ensure the location is pointing to the right directory. The main 'hybrid_br_classification.py' file should work instantly as long as the entire project is saved to the same directory. However, the experimental files will need to point ot a specific directory, eg:
   C:/Users/chibu/Documents/ISE-solution-main/Coursework/datasets/{project}.csv

The dataset must have columns: 'Title', 'Body', and 'Class'.

3. Preprocessing:
   - Removing HTML tags and emojis
   - Removing stopwords (including custom ones)
   - Cleaning punctuation and casing

4. Vectorization:
   - TF-IDF (1-2 grams, max 1000 features)

5. Model Setup:
   - RandomForest + LogisticRegression (both models)
   - Data balanced using SMOTE

6. Execution:
  Run the script. It performs 70/30 training/evaluation split and saves metric results.

7. Plotting(Optional): Use the plot_metrics function from utils.py to visualize metrics over iterations.

Output files generated for the main file:
- Processed data: {project}_processed_data.csv
- Repeated metrics: {project}_SMOTE_HybridModel_RepeatedMetrics.txt
- Final averaged metrics: {project}_SMOTE_HybridModel_Final_Metrics.csv