Manual: How to Use the Classification Tool

This tool processes text data, applies preprocessing steps, trains a hybrid model using SMOTE, and evaluates it over repeated runs.

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

C:/Users/chibu/Documents/ISE-solution-main/Coursework/datasets/{project}.csv

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

3. **Text Preprocessing Includes:**

   o **Removing HTML tags and emojis**

   o **Removing stopwords (including custom ones)**

   o **Cleaning punctuation and casing**

4. **Vectorization:**

   o **TF-IDF (1-2 grams, max 1000 features)**

5. **Model Setup:**

   o **RandomForest + LogisticRegression (VotingClassifier)**

   o **Data balanced using SMOTE**

6. **Execution:**

   o **Run the script. It performs 30 training/evaluation repetitions and saves metric results.**

7. **Plotting (Optional): Use the plot_metric_trends(csv_path) function to visualize metrics over repetitions.**

8. **Outputs:**

   o **Preprocessed data: {project}_processed_data.csv**

   o **Repeated metrics: {project}_SMOTE_HybridModel_Metrics.csv**

   o **Final averaged metrics: {project}_SMOTE_HybridModel_Final_Metrics.csv**