

Revised Timeline

Here's how the steps should look after reorganization:

Step 1: Preprocessing the Data

- 1.1 Load datasets, combine, shuffle.
- 1.2 Display dataset info.
- 1.3 Clean text.
- 1.4 Split into training and testing sets.

Step 2: Feature Engineering

- 2.1 Extract common features using TF-IDF.
- 2.2 Save extracted features to CSV.
- 2.3 Compare features (unique and overlapping).
- 2.4 Create visualizations (word clouds, bar charts, heatmaps).

Step 3: Model Training and Evaluation

- 3.1 Train and evaluate Logistic Regression.
- 3.2 Train and evaluate SVM.
- 3.3 Train and evaluate Random Forest.

Step 4: Hyperparameter Tuning

- 4.1 Perform hyperparameter tuning for Logistic Regression.
- 4.2 Evaluate the tuned model.

Step 5: Input-Based Prediction System

- 5.1 Accept user input.
- 5.2 Preprocess input text.
- 5.3 Predict using the best-trained model.
- 5.4 Display result and confidence level.