

## Replication Instructions

To replicate the experimental results reported:

Dataset:

Use the same CSV file used during development (e.g., pytorch.csv)

Ensure it is placed in the D:/datasets/ directory

Preprocessing:

Includes HTML tag removal, emoji removal, stopwords filtering (via NLTK), and string cleaning

Feature Extraction:

TF-IDF vectorizer with bigrams: `ngram_range=(1, 2)` and `max_features=1000`

Classification:

Two classifiers: LinearSVC (with SMOTE), and GaussianNB (with GridSearchCV)

Evaluation:

Metrics averaged over 10 runs: Accuracy, Precision, Recall, F1-score, AUC

Repetition:

Each run uses a different random seed (0 to 9)

Expected Files:

Output CSV will log average results and variation over runs