

REPLICATION

Author: Javier Rodríguez Fernández (Jxr497@student.bham.ac.uk)

Professor: Tao Chen

School of Computer Science, University of Birmingham, B15 2TT, UK

1 DATASET PREPARATION

- Download the datasets: tensorflow.csv, pytorch.csv, keras.csv, incubator-mxnet.csv, and caffe.csv.
- Place all files in the **project root directory**, same as the code files.

2 RUNNING THE SOLUTION AND THE BASELINE CODE

For each dataset (e.g., TensorFlow):

1) Open both “Baseline_Code.py”, and “Solution_Javier.py”, and set:

- project = 'tensorflow' (**Options:** 'tensorflow', 'pytorch', 'keras', 'incubator-mxnet', 'caffe')
- REPEAT = 30 (Do not modify this value).

2) Execute the code in the terminal:

- “python Solution_Javier.py”

Output: A CSV file named “{project}_SVM.CSV” (e.g. {project}= “tensorflow”) with metrics like F1, Accuracy, etc. This file will be stored in the project root directory.

- “python Baseline_Code.py”

Output: A CSV file named “tensorflow_NB.CSV” (e.g. {project}= “tensorflow”) with metrics like F1, Accuracy, etc. This file will be stored in the project root directory.

Both files will be necessary for the statistical tests in 3.

3 STATISTICAL ANALYSIS

For each dataset (e.g., TensorFlow):

1) **Open** “StatisticalAnalysis_Javier.py” and set:

- Project = “tensorflow” (must match the dataset name used in step 3)

2) **Execute** the code “pyhton StatisticalAnalysis_Javier”

3) **Outputs:**

- **2D Pareto Front plot:** F1 vs. Accuracy (Figures 7-11 in my report) at “pareto_front_F1_vs_AUC_{timestamp}.png”
- **Mean Metric plot:** SVM vs Naive Bayes models at “mean_points_plot_{timestamp}.png”
- **P values plot:** Results of the Wilcoxon test ($p < 0.05$ confirms significance) at “p_values_plot_{timestamp}.png”

4 EXPECTED RESULTS

After running all datasets:

- **CSV files:** 5 files for each model (e.g. tensorflow_SVM/NB.csv, pytorch_SVM/NB.csv, etc.) with metrics averaged over 30 runs.
- **Pareto Plots:** expected Pareto dominance in F1/Accuracy (SVM clusters in the top-right quadrant).
- **Statistical Significance:** p-values < 0.05 for F1 and Accuracy across all datasets.

5 NOTES FOR REPRODUCIBILITY

- 1) **Hyperparameters:** The SVM “param_grid” that I used to carry out my statistical test was:

```
param_grid = {
    'C': [ 1, 10],
    'gamma': [ 0.01, 0.1],
    'kernel': ['rbf']
}
```

These parameters were optimized via GridSearchCV over 30 splits.

- 2) **Baseline Comparison:** Ensure {project}_NB.csv (Naive Bayes results) is generated for statistical tests.
- 3) **Fixed Random States:** The code uses random_state=0–29 for reproducibility.