main

February 16, 2025

1 import

```
[1]:  # --- iPython Config --- #
      from IPython import get_ipython
      if 'IPython.extensions.autoreload' not in get_ipython().extension_manager.
       ↪loaded:
          get_ipython().run_line_magic('load_ext', 'autoreload')
      else:
          get_ipython().run_line_magic('reload_ext', 'autoreload')
      %autoreload 2

      # --- System and Path --- #
      import os
      import sys
      repo_path = os.path.dirname(os.getcwd())
      if repo_path not in sys.path:
          sys.path.append(repo_path)

      # --- Standard Libraries --- #
      import warnings
      warnings.filterwarnings('ignore')

      # --- Custom Modules --- #
      from src.data import dataloader
      from src.data.data_processor import DataProcessor
      from src.models import MultiTrainer, LogisticRegressionModel, XGBoostModel
```

```
[2]:  # Load Data
      file_path = os.path.join(repo_path, "data", "raw", "dataset.parquet")
      df_dataset = dataloader.load_data(file_path)
```

```
Memory usage: Before=134.49MB -> After=65.62MB, Decreased by 51.2%
Data loaded successfully.
```

```
[3]:  # Data Processing
      data_processor = DataProcessor()
      df_train, df_test = data_processor.process(df_dataset, target="Class")
```

```python
# Training
# custom model
models = {"LogisticRegression": LogisticRegressionModel(),
          "XGBoost": XGBoostModel()}
# Initialize trainer
multi_trainer = MultiTrainer(
    df_train,
    df_test,
    target="Class",
    models=models,
    main_metric="recall",
    output_dir=os.path.join(repo_path, "models"),
)
multi_trainer.train_all_models(tune_params=False)
multi_trainer.evaluate_all_models()
```

```
Training LogisticRegression…
Model saved to /Users/pupipatsingkhorn/Developer/repositories/fraud-detection-
european-credit-card-
transactions-2023/models/2025-02-16-2340-LogisticRegressionModel.pkl
Training time LogisticRegression: 1.53 seconds.
Training XGBoost…
Model saved to /Users/pupipatsingkhorn/Developer/repositories/fraud-detection-
european-credit-card-transactions-2023/models/2025-02-16-2340-XGBoostModel.pkl
Training time XGBoost: 1.64 seconds.
TRAIN | LogisticRegression recall: 0.9517
TRAIN | XGBoost recall: 1.0000
TEST | LogisticRegression recall: 0.9511
TEST | XGBoost recall: 1.0000
```

```
[4]: {'train': {'LogisticRegression': {'precision': 0.9778115953085538,
        'recall': 0.9516929546200622,
        'f1': 0.9645754981353962,
        'accuracy': 0.9650485474835228,
        'roc_auc': 0.9935839474162925},
      'XGBoost': {'precision': 0.99997655242487,
        'recall': 1.0,
        'f1': 0.9999882760749862,
        'accuracy': 0.9999882759146252,
        'roc_auc': 0.9999999995876375}},
     'test': {'LogisticRegression': {'precision': 0.9765972046661129,
        'recall': 0.951109704196124,
        'f1': 0.9636849607982894,
        'accuracy': 0.9641594710092679,
        'roc_auc': 0.993271341982931},
      'XGBoost': {'precision': 0.9994726850875343,
        'recall': 1.0,
        'f1': 0.9997362730101799,
```

```
'accuracy': 0.9997362080790673,
'roc_auc': 0.9999849842306959}}}
```