Feature Engineering Lab

Romain BLANCHOT 67540460063

May 6, 2025

Feature Engineering Lab

CPE393 MLOps Lab – Feature Engineering Pipeline

GitHub Repository

Link to the GitHub repository : https://github.com/LeMarechalDeFer/MLOps_CPE393-FeatureEngineering
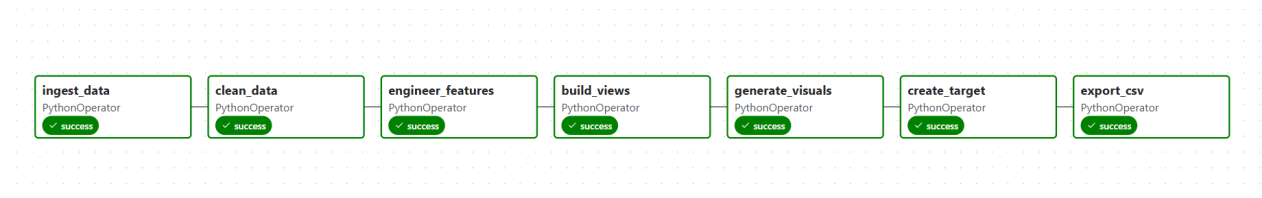
Pipeline Overview

Figure 1: Pipeline Diagram

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=ingest_data/attempt=1.log"]
2 [2025-05-06, 06:07:42] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3 [2025-05-06, 06:07:42] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4 [2025-05-06, 06:07:45] INFO - Ingesting data: source="root"
5 [2025-05-06, 06:07:45] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=clean_data/attempt=1.log"]
2 [2025-05-06, 06:07:46] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3 [2025-05-06, 06:07:46] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4 [2025-05-06, 06:07:48] INFO - Cleaning data: source="root"
5 [2025-05-06, 06:07:48] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=engineer_features/attempt=1.log"]
2 [2025-05-06, 06:07:49] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3 [2025-05-06, 06:07:49] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4 [2025-05-06, 06:07:50] INFO - Engineering features: source="root"
5 [2025-05-06, 06:07:50] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=build_views/attempt=1.log"]
2 [2025-05-06, 06:07:51] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3 [2025-05-06, 06:07:51] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4 [2025-05-06, 06:07:52] INFO - Building views: source="root"
5 [2025-05-06, 06:07:53] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"

Welcome

⊙ Health

✓ MetaDatabase   ✓ Scheduler   ✓ Triggerer   ✓ Dag Processor

▢ Links

❌ **Failed Dags** ›   🔵 **Running Dags** ›   ⚡ **Active Dags** ›

▦ History

📅 | Last 24 hours ▾ | 2025-05-05, 06:11:05 - 2025-05-06, 06:11:05

📊 3 **Dag Runs**

▷ 0 Queued	0.00%

⩗ 0 Running	0.00%

✓ 3 Success	100.00%

❌ 0 Failed	0.00%

📄 21 **Task Instances**

✓ 21 Success	100.00%

Figure 2: Menu Screenshot

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=generate_visuals/attempt=1.log"]

```
2  [2025-05-06, 06:07:53] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3  [2025-05-06, 06:07:53] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4  [2025-05-06, 06:07:55] INFO - Generating visuals: source="root"
5  [2025-05-06, 06:07:55] INFO - Available columns in daily: ['close', 'ma_close_30d', 'rsi_14']: source="root"
6  [2025-05-06, 06:07:56] INFO - Available columns in spikes: ['volatility_abs', 'is_spike']: source="root"
7  [2025-05-06, 06:07:57] INFO - Available columns in weekly: ['close_first', 'close_last', 'close_mean', 'volume_sum', 'volatility_abs_max']: source="root"
8  [2025-05-06, 06:07:58] INFO - Available columns in daily (RSI): ['close', 'ma_close_30d', 'rsi_14']: source="root"
9  [2025-05-06, 06:07:59] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"
```

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=create_target/attempt=1.log"]

```
2  [2025-05-06, 06:08:00] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3  [2025-05-06, 06:08:00] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4  [2025-05-06, 06:08:01] INFO - Creating target: source="root"
5  [2025-05-06, 06:08:02] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"
```

▶ Log message source details: sources=["/opt/airflow/logs/dag_id=tsmc_improved_prediction_pipeline/run_id=manual__2025-05-05T23:07:37.995260+00:00/task_id=export_csv/attempt=1.log"]

```
2  [2025-05-06, 06:08:02] INFO - DAG bundles loaded: dags-folder: source="airflow.dag_processing.bundles.manager.DagBundlesManager"
3  [2025-05-06, 06:08:02] INFO - Filling up the DagBag from /opt/airflow/dags/tsmc_improved_pipeline.py: source="airflow.models.dagbag.DagBag"
4  [2025-05-06, 06:08:04] INFO - Exporting CSV: source="root"
5  [2025-05-06, 06:08:04] INFO - Done. Returned value was: None: source="airflow.task.operators.airflow.providers.standard.operators.python.PythonOperator"
```