

Mlflow screen shots

Mlflow screenshot showing experiment with all features to determine top model considerations.

The screenshot displays the Mlflow web interface for an experiment named "Experiment_with_all_features". The interface includes a sidebar with a search bar and a list of experiments. The main panel shows the experiment details, including a description, a list of runs, and a table of metrics.

Experiment ID: 1

Description: [Edit](#)

Actions: Refresh, Compare, Delete, Download CSV, Start Time (All time)

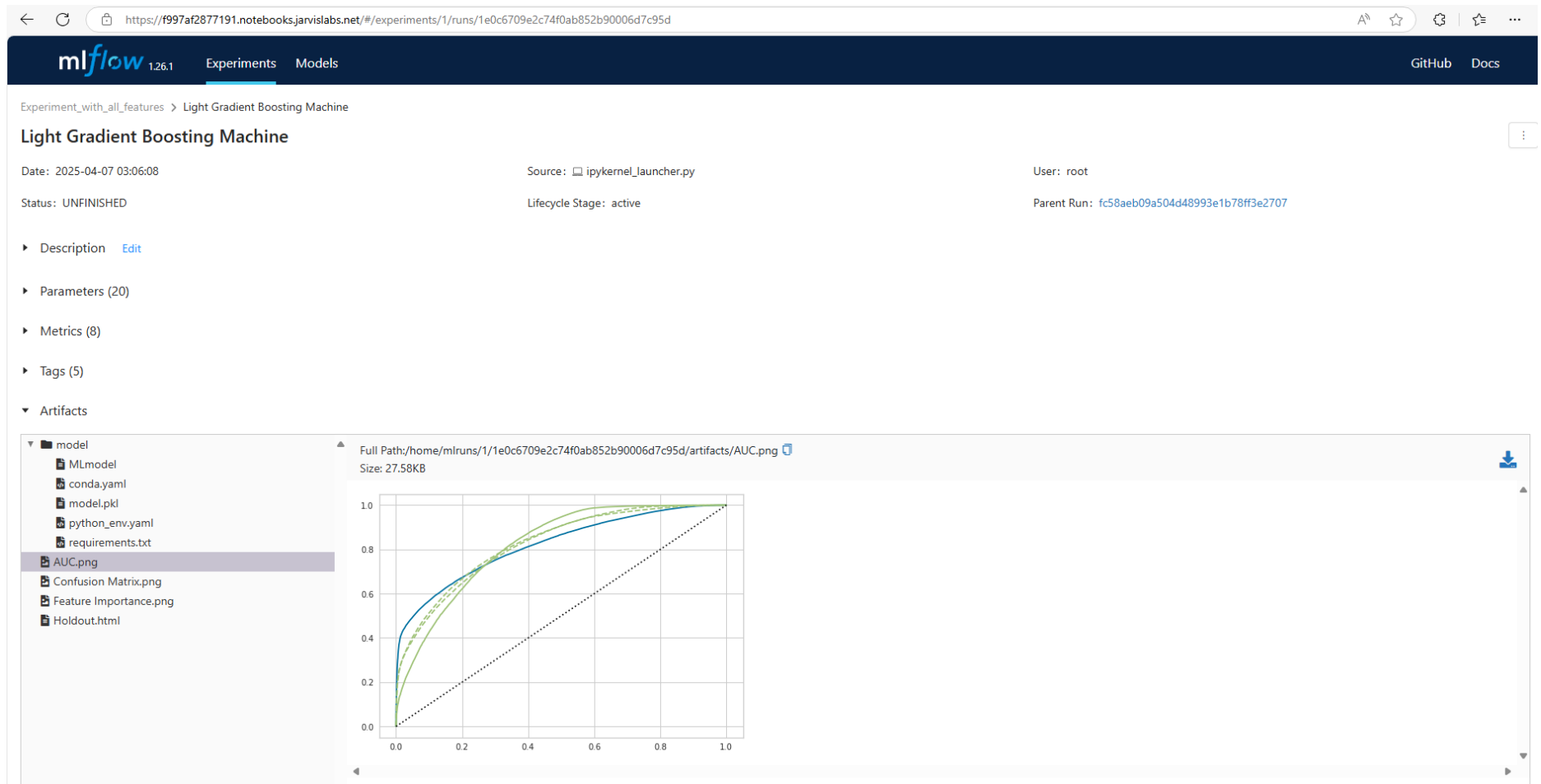
Columns: Only show differences (toggle on), Search, Filter, Clear

Search: metrics.rmse < 1 and params.model = "tree"

Showing 34 matching runs

		Duration	Run Name	User	Source	Models	Metrics <								Parameters >	
	Start Time						AUC	Accuracy	F1	Kappa	MCC	Prec.	Recall	TT	C	
<input type="checkbox"/>	4 minutes ago		Session Initi...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-	
<input type="checkbox"/>	3 minutes ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	0.476	0.484	0.703	0.831	1.37	-	
<input type="checkbox"/>	3 minutes ago		Ridge Classif...	root	ipykernel...	sklearn	0	0.701	0.728	0.401	0.408	0.67	0.797	0.09	-	
<input type="checkbox"/>	3 minutes ago		Naive Bayes	root	ipykernel...	sklearn	0.734	0.657	0.725	0.313	0.357	0.608	0.897	0.092	-	
<input type="checkbox"/>	3 minutes ago		Linear Discr...	root	ipykernel...	sklearn	0.773	0.701	0.728	0.401	0.408	0.67	0.797	0.29	-	
<input type="checkbox"/>	3 minutes ago		Logistic Reg...	root	ipykernel...	sklearn	0.784	0.71	0.74	0.419	0.43	0.673	0.823	1.674	1.0	
<input type="checkbox"/>	3 minutes ago		Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	0.472	0.479	0.704	0.82	0.182	-	
<input type="checkbox"/>	3 minutes ago		Extra Trees C...	root	ipykernel...	sklearn	0.818	0.737	0.758	0.472	0.479	0.704	0.82	1.092	-	
<input type="checkbox"/>	3 minutes ago		Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	0.474	0.482	0.704	0.825	0.886	-	
<input type="checkbox"/>	3 minutes ago		Extreme Gra...	root	ipykernel...	sklearn	0.821	0.738	0.761	0.476	0.483	0.703	0.829	3.314	-	
<input type="checkbox"/>	3 minutes ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	0.476	0.484	0.703	0.831	0.562	-	

Mlflow screen of all artifacts for top performing model (LGBM) when using all features.



Mlflow screenshot showing experiments with selected features

The screenshot shows the Mlflow web interface. The top navigation bar includes the Mlflow logo (1.26.1), 'Experiments', and 'Models' tabs. On the right, there are links for 'GitHub' and 'Docs'. The main header shows 'Experiments' with a '+ <' icon and the selected experiment 'Experiment_with_reduced_features' with a 'Share' button.

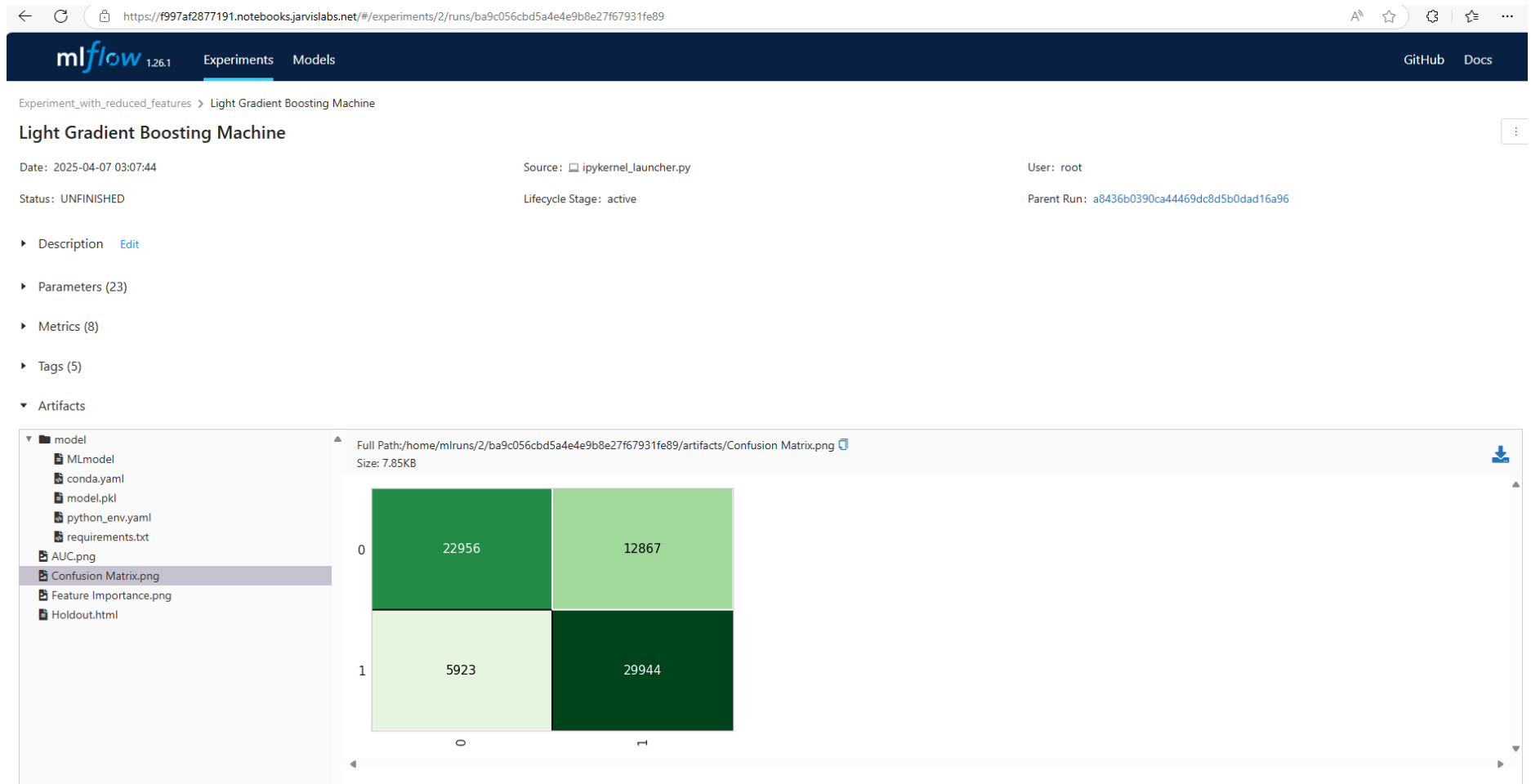
Below the header, there is a search bar for 'Search Experiments' and a list of experiments: 'Experiment_with_all_f...', 'Experiment_with_red...' (selected), and 'Lead_Scoring_Trainin...'. A light blue banner states: 'Track machine learning training runs in experiments. Learn more'.

The experiment details section shows 'Experiment ID: 2'. Below this, there are buttons for 'Refresh', 'Compare', 'Delete', 'Download CSV', and a 'Start Time' dropdown set to 'All time'. There is also a 'Columns' button, a toggle for 'Only show differences', and a search bar containing 'metrics.rmse < 1 and params.model = "tree"'. 'Search', 'Filter', and 'Clear' buttons are also present.

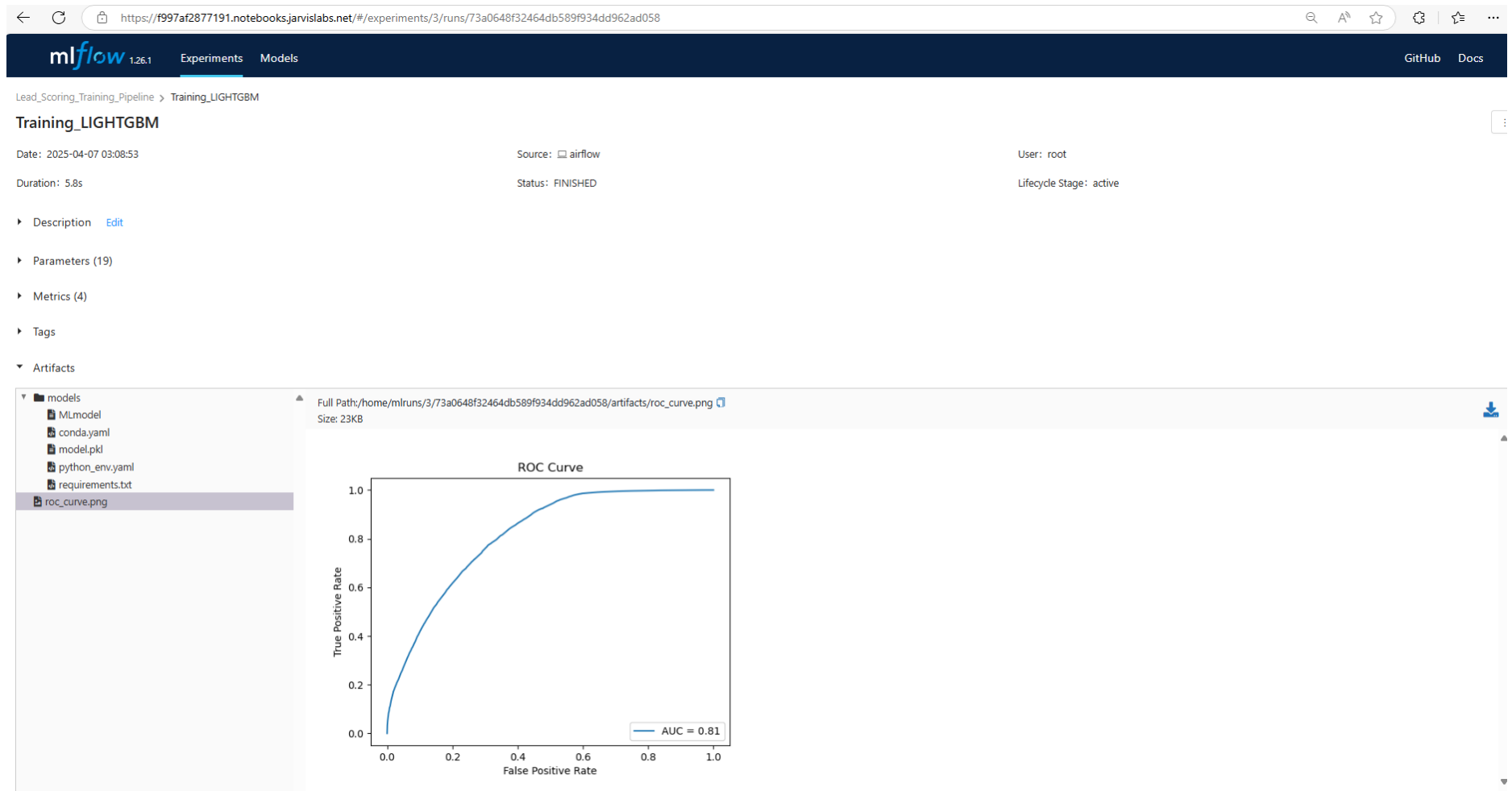
The table below shows 'Showing 23 matching runs'. The table has columns for 'Start Time', 'Duration', 'Run Name', 'User', 'Source', 'Version', 'Models', and various metrics (AUC, Accuracy, F1, Kappa, MCC, Prec., Recall). The first row is 'Session Init...' with a start time of '6 minutes ago'. The subsequent rows show runs for 'Light Gradie...', 'Decision Tre...', 'Extra Trees C...', 'Random For...', 'Extreme Gra...', and 'Light Gradie...'.

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	Kappa	MCC	Prec.	Recall
<input type="checkbox"/>	6 minutes ago		Session Init...	root	ipykernel...	-	-	-	-	-	-	-	-	-
<input type="checkbox"/>	5 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.82	0.738	0.762	0.476	0.485	0.703	0.832
<input type="checkbox"/>	6 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	0.476	0.485	0.702	0.833
<input type="checkbox"/>	6 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	0.472	0.479	0.704	0.82
<input type="checkbox"/>	6 minutes ago		Extra Trees C...	root	ipykernel...	-	sklearn	0.818	0.736	0.758	0.472	0.479	0.704	0.82
<input type="checkbox"/>	6 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.819	0.737	0.759	0.473	0.48	0.704	0.823
<input type="checkbox"/>	6 minutes ago		Extreme Gra...	root	ipykernel...	-	sklearn	0.821	0.738	0.761	0.475	0.484	0.702	0.831
<input type="checkbox"/>	6 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	0.476	0.485	0.702	0.833

Mlflow screenshot showing artifacts for best performing model with selected features



Mlflow screenshot showing model training logged by training pipeline



Mlflow screenshot showing trained model being promoted to “production”

The screenshot displays the MLflow web interface in a browser. The address bar shows the URL: `https://f997af2877191.notebooks.jarvislabs.net/#/models/LightGBM/versions/24`. The MLflow logo and version 1.26.1 are in the top left, with navigation tabs for Experiments and Models. The right side of the header has links for GitHub and Docs.

The main content area shows the breadcrumb path: Registered Models > LightGBM > Version 24. Below this, the title "Version 24" is displayed. The "Registered At" timestamp is 2025-04-07 03:08:58, and the "Last Modified" timestamp is 2025-04-07 03:08:59. The "Source Run" is identified as Training_LIGHTGBM. The "Stage" is set to "Production", indicated by a green badge. There are expandable sections for Description (with an Edit link), Tags, and Schema.

Under the Schema section, there is a table with two columns: Name and Type. The table is currently empty, and a message below it states: "No schema. See [MLflow docs](#) for how to include input and output schema with your model."

Name	Type
------	------

Airflow screenshots

Airflow screenshot showing all DAGs

DAGs

All 5 Active 4 Paused 1

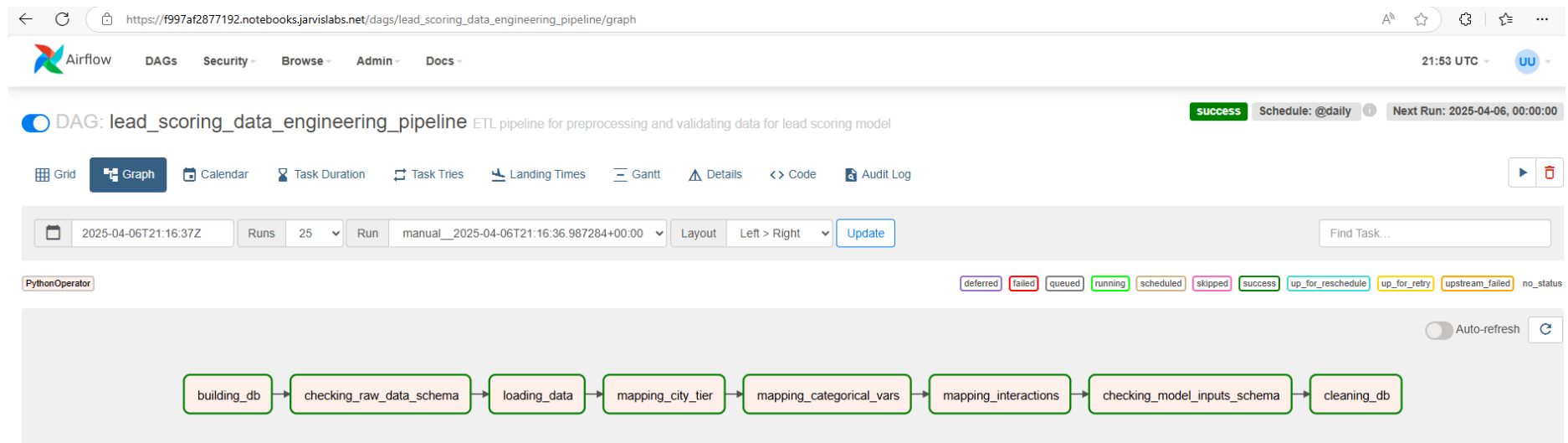
Filter DAGs by tag

Search DAGs

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions
<input checked="" type="checkbox"/> lead_scoring_data_engineering_pipeline data_pipeline data_validation ETL lead_scoring	airflow	<input type="radio"/> 4 <input type="radio"/> <input type="radio"/> <input type="radio"/>	@daily	2025-04-06, 21:16:36	2025-04-06, 00:00:00	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 8 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	▶ 🗑️
<input checked="" type="checkbox"/> lead_scoring_inference_pipeline inference lead_scoring pipeline	airflow	<input type="radio"/> 5 <input type="radio"/> <input type="radio"/> <input type="radio"/>	@daily	2025-04-06, 21:57:43	2025-04-06, 00:00:00	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 11 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	▶ 🗑️
<input checked="" type="checkbox"/> lead_scoring_training_pipeline model training	airflow	<input type="radio"/> 5 <input type="radio"/> <input type="radio"/> <input type="radio"/>	@daily	2025-04-06, 21:38:29	2025-04-06, 00:00:00	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 3 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	▶ 🗑️
<input type="checkbox"/> master_pipeline_dag master unit tests	airflow	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	None			<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	▶ 🗑️
<input checked="" type="checkbox"/> unit_test_runner testing validation	airflow	<input type="radio"/> 1 <input type="radio"/> <input type="radio"/> <input type="radio"/>	None	2025-04-06, 20:44:02		<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> 4 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	▶ 🗑️


Showing 1-5 of 5 DAGs


Airflow screenshot showing successful data pipeline in graph view





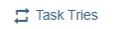
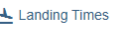








Airflow screenshot showing successful data pipeline in grid view

← ↻ 🔒 https://f997af2877192.notebooks.jarvislabs.net/dags/lead_scoring_data_engineering_pipeline/grid?root= 🔊 ☆ ⚙️ ⚙️ ⋮

 **DAGs** Security Browse Admin Docs 21:53 UTC UU

 **DAG: lead_scoring_data_engineering_pipeline** ETL pipeline for preprocessing and validating data for lead scoring model Schedule: @daily ⓘ Next Run: 2025-04-06, 00:00:00

 **Grid**  Graph  Calendar  Task Duration  Task Times  Landing Times  Gantt  Details  Code  Audit Log  

04/06/2025 09:53:45 PM 📅 25 ▾ All Run Types ▾ All Run States ▾ [Clear Filters](#)

deferred failed queued running scheduled skipped success up_for_reschedule up_for_retry upstream_failed no_status

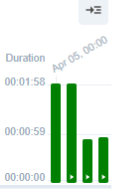
Auto-refresh ☐

Duration









00:01:58

00:00:50

00:00:00



Apr 05, 00:00


building_db	
checking_raw_data_schema	
loading_data	
mapping_city_tier	
mapping_categorical_vars	
mapping_interactions	
checking_model_inputs_schema	
cleaning_db	

DAG

lead_scoring_data_engineering_pipeline

DAG Details

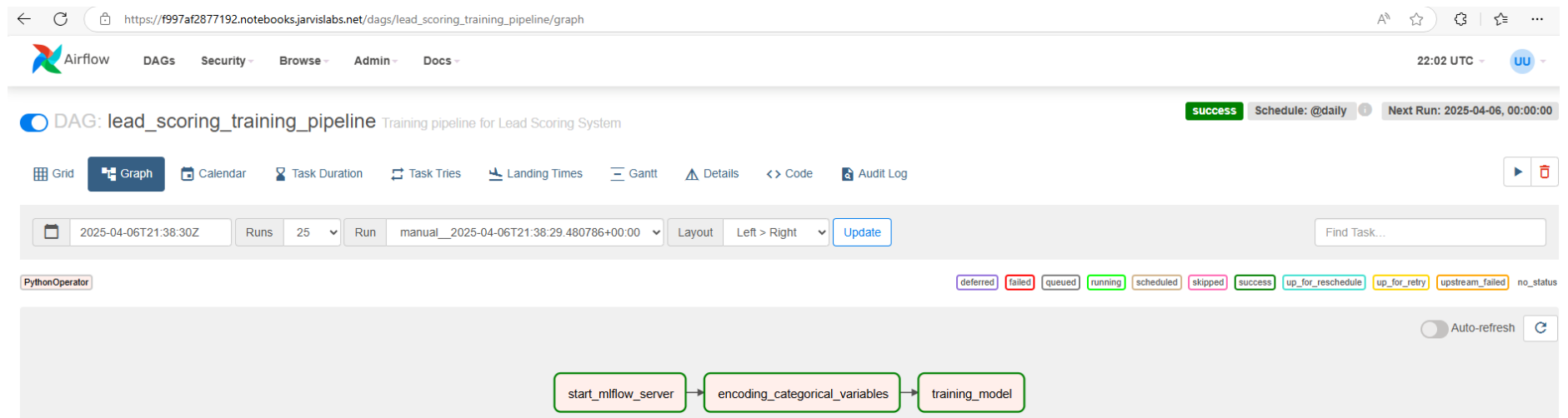
DAG Runs Summary

Total Runs Displayed	4
 Total success	4
First Run Start	2025-04-06, 20:34:13 UTC
Last Run Start	2025-04-06, 21:16:37 UTC
Max Run Duration	00:01:58
Mean Run Duration	00:01:25
Min Run Duration	00:00:52

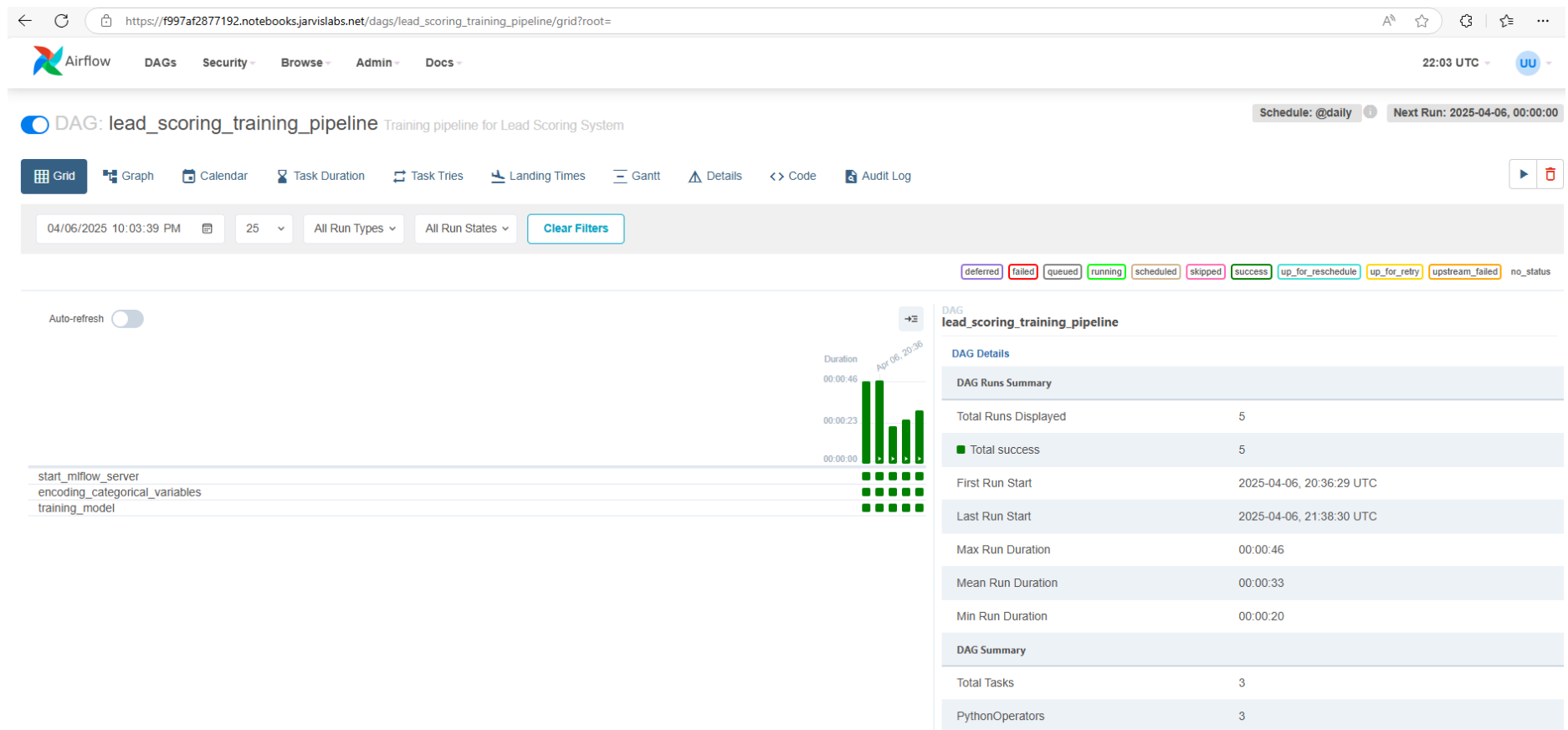
DAG Summary

Total Tasks	8
PythonOperators	8

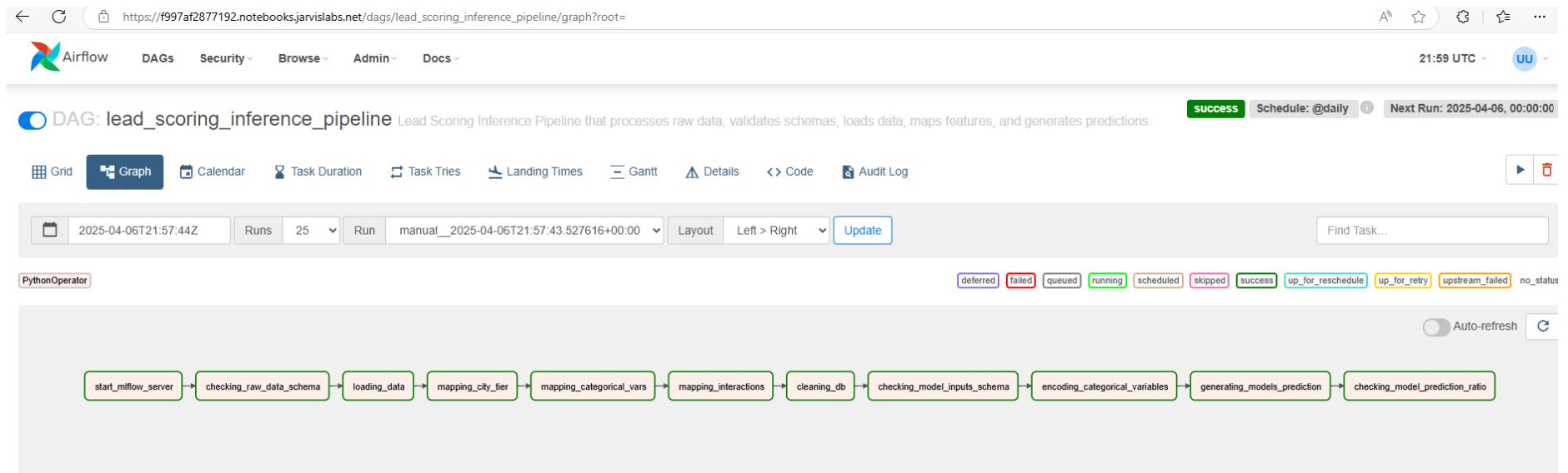
Airflow screenshot showing successful training pipeline in graph view



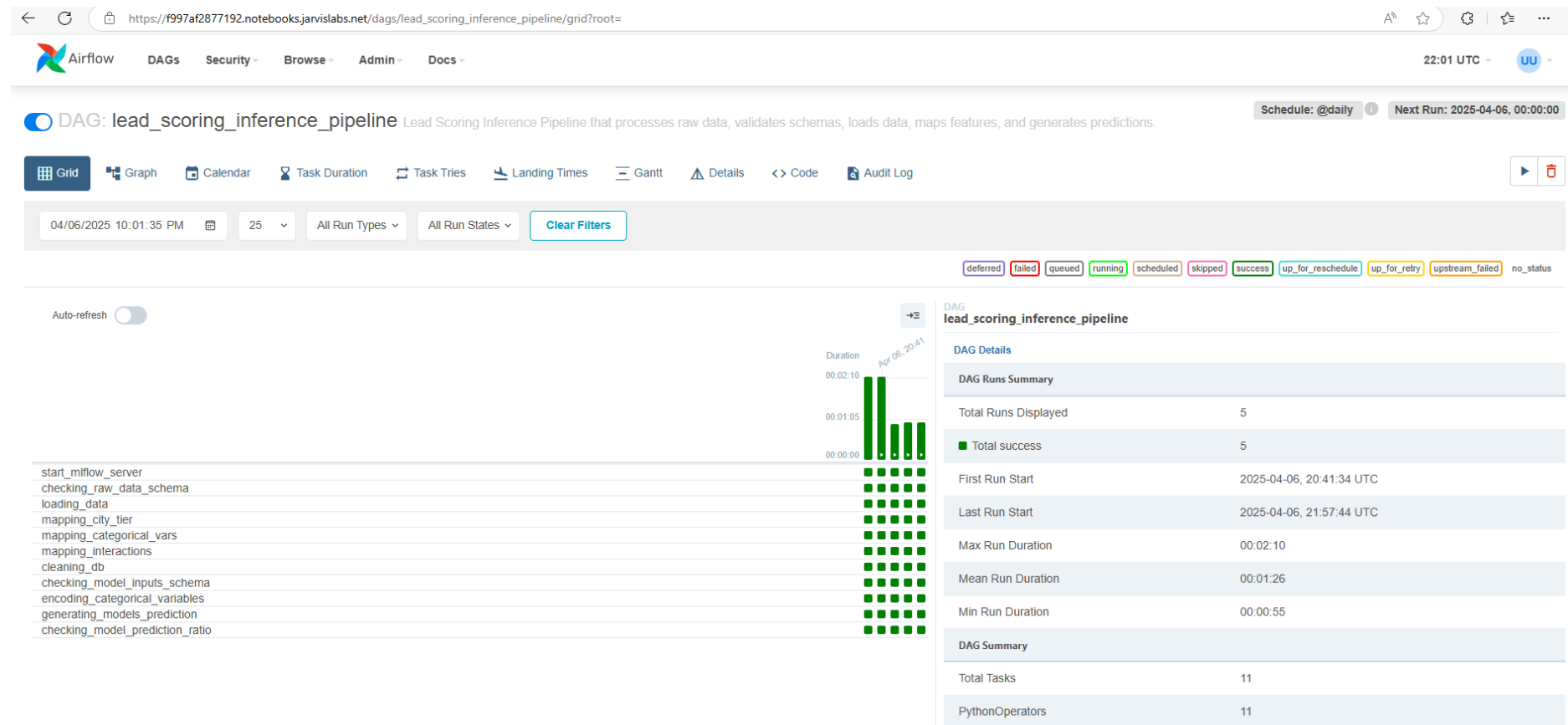
Airflow screenshot showing successful training pipeline in grid view



Airflow screenshot showing successful inference pipeline in graph view



Airflow screenshot showing successful inference pipeline in grid view



Folder structure

```
airflow
|-- airflow.cfg
|-- dags
|   |-- __init__.py
|   |-- lead_scoring_data_pipeline
|   |   |-- __init__.py
|   |   |-- constants.py
|   |   |-- data
|   |   |   |-- leadscoring.csv
|   |   |   |-- leadscoring_inference.csv
|   |   |-- data_validation_checks.py
|   |   |-- lead_scoring_data_cleaning.db
|   |   |-- lead_scoring_data_pipeline.py
|   |   |-- mappings
|   |   |   |-- city_tier_mapping.py
|   |   |   |-- interaction_mapping.csv
|   |   |   |-- significant_categorical_level.py
|   |   |-- schema.py
|   |   |-- utils.py
|   |-- lead_scoring_inference_pipeline
|   |   |-- __init__.py
|   |   |-- constants.py
|   |   |-- lead_scoring_inference_pipeline.py
|   |   |-- prediction_distribution.txt
|   |   |-- schema.py
|   |   |-- utils.py
```

```
airflow
|-- dags
|   |-- lead_scoring_training_pipeline
|   |   |-- __init__.py
|   |   |-- constants.py
|   |   |-- lead_scoring_training_pipeline.py
|   |   |-- utils.py
|   |-- master_pipeline_dag.py
|   |-- unit_test
|   |   |-- __init__.py
|   |   |-- constants.py
|   |   |-- leadscoring_test.csv
|   |   |-- test_runner_dag.py
|   |   |-- test_with_pytest.py
|   |   |-- unit_test_cases.db
|-- notebooks
|   |-- 01.data_cleaning.ipynb
|   |-- 02.model_experimentation.ipynb
|   |-- data
|   |   |-- cleaned_data.csv
|   |   |-- leadscoring.csv
|   |-- mappings
|   |   |-- city_tier_mapping.py
|   |   |-- interaction_mapping.csv
|   |   |-- significant_categorical_level.py
|   |-- profile_reports
|   |   |-- cleaned_data_report.html
|   |   |-- raw_data_report.html
|-- screenshots.pdf
|-- webserver_config.py
```