

Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use `SequentialExecutor` in production. [Click here](#) for more information.

## DAGs

All 35	Active 3	Paused 32	Filter DAGs by tag	Search DAGs				
DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
Lead_Scoring_Data_Engineering_Pipeline	airflow		daily	2023-01-16, 18:48:19	2023-01-16, 00:00:00			...
Lead_scoring_inference_pipeline	airflow		hourly	2023-01-16, 18:49:44	2023-01-16, 19:00:00			...
Lead_scoring_training_pipeline	airflow		monthly	2023-01-16, 18:43:32	2023-01-01, 00:00:00			...
example_bash_operator	airflow		8-***		2023-01-15, 00:00:00			...

The screenshot displays the Airflow web interface for the DAG 'Lead\_Scoring\_Data\_Engineering\_Pipeline'. The top navigation bar includes links for DAGs, Security, Browse, Admin, and Docs. The top right shows the current time as 19:11 UTC and the next run scheduled for 2023-01-16, 00:00:00. The main area shows a Gantt chart for the DAG's execution history, with a legend indicating 'Total success'. The chart is filtered for 'All Run Types' and 'All Run States'. The bottom left shows the DAG's task list, including 'building\_db', 'checking\_raw\_data\_schema', 'loading\_data', 'mapping\_city\_tier', 'mapping\_categorical\_vars', 'mapping\_interactions', and 'checking\_model\_inputs\_schema'.

Airflow DAGs Security Browse Admin Docs 19:12 UTC VS

DAG: Lead\_scoring\_training\_pipeline Training pipeline for Lead Scoring System success Schedule: @monthly Next Run: 2023-01-01, 00:00:00

Grid Graph Calendar Task Duration Task Tiles Landing Times Gantt Details <> Code Audit Log

2023-01-16T18:43:33Z Runs 25 Run manual\_\_2023-01-16T18:43:32.600494+00:00 Layout Left > Right Update Find Task...

PythonOperator

deferred failed queued running scheduled skipped success up\_for\_reschedule up\_for\_retry upstream\_failed no\_status

Auto-refresh

```

graph LR
    encoding_categorical_variables[encoding_categorical_variables] --> training_model[training_model]
  
```

Airflow DAGs Security Browse Admin Docs 19:13 UTC VS

DAG: Lead\_scoring\_inference\_pipeline Inference pipeline of Lead Scoring system Schedule: @hourly Next Run: 2023-01-16, 19:00:00

Grid Graph Calendar Task Duration Task Tiles Landing Times Gantt Details <> Code Audit Log

01/16/2023 07:13:08 PM 25 All Run Types All Run States Clear Filters

Auto-refresh

Lead\_scoring\_inference\_pipeline

DAG Details

DAG Runs Summary

Total Runs Displayed	14
Total success	9
Total failed	5
First Run Start	2023-01-16, 22:00:51 UTC
Last Run Start	2023-01-16, 18:49:45 UTC
Max Run Duration	00:24:59

encoding\_categorical\_variables  
checking\_input\_features  
generating\_models\_prediction  
checking\_model\_prediction\_ratio

Airflow DAGs Security Browse Admin Docs 19:13 UTC VS

DAG: Lead\_scoring\_inference\_pipeline Inference pipeline of Lead Scoring system success Schedule: @hourly Next Run: 2023-01-16, 19:00:00

Grid Graph Calendar Task Duration Task Tiles Landing Times Gantt Details <> Code Audit Log

2023-01-16T18:49:45Z Runs 25 Run manual\_\_2023-01-16T18:49:44.007180+00:00 Layout Left > Right Update Find Task...

PythonOperator

deferred failed queued running scheduled skipped success up\_for\_reschedule up\_for\_retry upstream\_failed no\_status

Auto-refresh

```

graph LR
    encoding_categorical_variables[encoding_categorical_variables] --> checking_input_features[checking_input_features]
    checking_input_features --> generating_models_prediction[generating_models_prediction]
    generating_models_prediction --> checking_model_prediction_ratio[checking_model_prediction_ratio]
  
```

mlflow1.26.1ExperimentsModels

Experiments

Lead\_scoring\_mlflow\_production

Share

Search Experiments

Default

Lead\_scoring\_mlflow...

Lead\_Scoring\_Trainin...

Track machine learning training runs in experiments. Learn more

Experiment ID: 1

Description

RefreshCompareDeleteDownload CSVStart TimeAll time

ColumnsOnly show differencesmetrics.mean < 1 and param.model = "tree"SearchFilterClear

Showing 3 matching runs

								Metrics			Parameters		
	Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision_0	boosting_type	class_weight	colsample_bytree
	7 minutes ago	6.2s	Lead_scorin...	root	airflow	-	LightGBM/24	3996	0.737	0.792	gbdt	None	1.0
	4 days ago		Lead_scorin...	root	airflow	-		-	-	-	-	-	-
	4 days ago		Lead_scorin...	root	airflow	-		-	-	-	-	-	-

Load more

mlflow1.26.1ExperimentsModels

Experiments

Lead\_Scoring\_Training\_Pipeline

Share

Search Experiments

Default

Lead\_scoring\_mlflow...

Lead\_Scoring\_Trainin...

Track machine learning training runs in experiments. Learn more

Experiment ID: 2

Description

RefreshCompareDeleteDownload CSVStart TimeAll time

ColumnsOnly show differencesmetrics.mean < 1 and param.model = "tree"SearchFilterClear

Showing 30 matching runs

								Metrics			Parameters		
	Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision_0	boosting_type	class_weight	colsample_bytree
	20 hours ago	4.5s	Lead_scorin...	root	airflow	-	LightGBM/23	3996	0.737	0.792	gbdt	None	1.0
	1 day ago	5.2s	Lead_scorin...	root	airflow	-	LightGBM/22	3996	0.737	0.792	gbdt	None	1.0
	1 day ago	5.1s	Lead_scorin...	root	airflow	-	LightGBM/21	3996	0.737	0.792	gbdt	None	1.0
	1 day ago	4.4s	Lead_scorin...	root	ipykernel	-	LightGBM/20	-	-	-	gbdt	None	1.0
	1 day ago	4.6s	Lead_scorin...	root	ipykernel	-	LightGBM/19	-	-	-	gbdt	None	1.0
	1 day ago	4.3s	Lead_scorin...	root	airflow	-	LightGBM/18	-	-	-	gbdt	None	1.0
	1 day ago	4.6s	Lead_scorin...	root	airflow	-	LightGBM/17	-	-	-	gbdt	None	1.0
	1 day ago	4.1s	Lead_scorin...	root	ipykernel	-	LightGBM/15	-	-	-	gbdt	None	1.0
	1 day ago	4.0s	Lead_scorin...	root	ipykernel	-	LightGBM/14	-	-	-	gbdt	None	1.0
	1 day ago	3.9s	Lead_scorin...	root	ipykernel	-	LightGBM/13	-	-	-	gbdt	None	1.0
	1 day ago	3.9s	Lead_scorin...	root	ipykernel	-	LightGBM/12	-	-	-	gbdt	None	1.0
	1 day ago	3.9s	Lead_scorin...	root	ipykernel	-	LightGBM/11	-	-	-	gbdt	None	1.0
	1 day ago	3.8s	Lead_scorin...	root	ipykernel	-	LightGBM/10	-	-	-	gbdt	None	1.0
	1 day ago	3.8s	Lead_scorin...	root	ipykernel	-	LightGBM/9	-	-	-	gbdt	None	1.0

mlflow1.26.1ExperimentsModels

Registered Models > LightGBM > Version 24

Version 24

Registered At: 2023-01-16 18:43:57

Stages: Production

Last Modified: 2023-01-16 18:45:20

Source Run: Lead\_scoring\_mlflow\_production\_run\_16\_01\_2023\_00\_00\_00

Description

Tags

Schema

mlflow1.26.1ExperimentsModels

Lead\_scoring\_mlflow\_production > Lead\_scoring\_mlflow\_production\_run\_16\_01\_2023\_00\_00\_00

Lead\_scoring\_mlflow\_production\_run\_16\_01\_2023\_00\_00\_00

Date: 2023-01-16 18:43:52

Source: airflow

User: root

Duration: 6.2s

Status: FINISHED

Lifecycle Stage: active

Description

Parameters (20)

Metrics (13)

Tags

Artifacts

mlflow1.26.1ExperimentsModels

Lead\_scoring\_mlflow\_production > Lead\_scoring\_mlflow\_production\_run\_16\_01\_2023\_00\_00\_00

1

Date: 2023-01-16 18:43:52

Source: airflow

User: root

Duration: 6.2s

Status: FINISHED

Lifecycle Stage: active

Description [Edit](#)

Parameters (20)

Metrics (13)

Name	Value
False Negative <a href="#">lat</a>	3996
Precision <a href="#">lat</a>	0.737
Precision_0 <a href="#">lat</a>	0.792
Precision_1 <a href="#">lat</a>	0.7
Recall <a href="#">lat</a>	0.746
Recall_0 <a href="#">lat</a>	0.64
Recall_1 <a href="#">lat</a>	0.833
True Negative <a href="#">lat</a>	1529
F1 score <a href="#">lat</a>	0.735
f1_0 <a href="#">lat</a>	0.708
f1_1 <a href="#">lat</a>	0.761
roc_auc <a href="#">lat</a>	0.746
test_accuracy <a href="#">lat</a>	0.737

mlflow1.26.1ExperimentsModels

Lead\_scoring\_mlflow\_production > Lead\_scoring\_mlflow\_production\_run\_16\_01\_2023\_00\_00\_00

1

Date: 2023-01-16 18:43:52

Source: airflow

User: root

Duration: 6.2s

Status: FINISHED

Lifecycle Stage: active

Description [Edit](#)

Parameters (20)

Metrics (13)

Tags

Artifacts

spynb\_checkpoints

models

mlmodel

conda.yaml

model.pkl

python\_env.yaml

requirements.txt

Full Path: /Assignment/mlruns/1/5c2ad71d52864f65be1cd6ce8827718c/artifacts/models

v3.4.0 Registered on 2023-01-16

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. This model is also registered to the [model registry](#).

Model schema

Input and output schema for your model. [Learn more](#)

Name	Type
No schema. See <a href="#">MLflow docs</a> for how to include input and output schema with your model.	

Make Predictions

Predict on a Spark DataFrame

```
import mlflow
logged_model = "runs:/5c2ad71d52864f65be1cd6ce8827718c/models"

# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=logged_model, result_type='double')

# Predict on a Spark DataFrame.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
```

Predict on a Pandas DataFrame

```
import mlflow
logged_model = "runs:/5c2ad71d52864f65be1cd6ce8827718c/models"

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)
```

The screenshot shows the JupyterLab file browser interface. The top menu bar includes File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. Below the menu is a toolbar with icons for creating a new file, opening a file, uploading, refreshing, and deleting. A search bar labeled "Filter files by name" is present. The current directory is "/ Assignment /". A table lists the files and folders in this directory:

Name	Last Modified
Lead_scoring_data_pipeline	12 days ago
Lead_scoring_inference_pipeline	20 hours ago
Lead_scoring_training_pipeline	5 days ago
mlruns	6 days ago
unit_test	a day ago

The screenshot shows the JupyterLab file browser interface with the directory changed to "/ Assignment / unit\_test /". The search bar is empty. A table lists the files and folders in this directory:

Name	Last Modified
city_tier_mapping.py	5 days ago
constants.py	5 days ago
data_validation_checks.py	5 days ago
dummy.ipynb	a day ago
interaction_mapping.csv	5 days ago
lead_scoring_data_pipeline.py	5 days ago
leadscoring_test.csv	5 days ago
schema.py	5 days ago
significant_categorical_level.py	5 days ago
test_with_pytest.py	5 days ago
unit_test_cases.db	5 days ago
utils_output.db	5 days ago
utils.py	5 days ago

