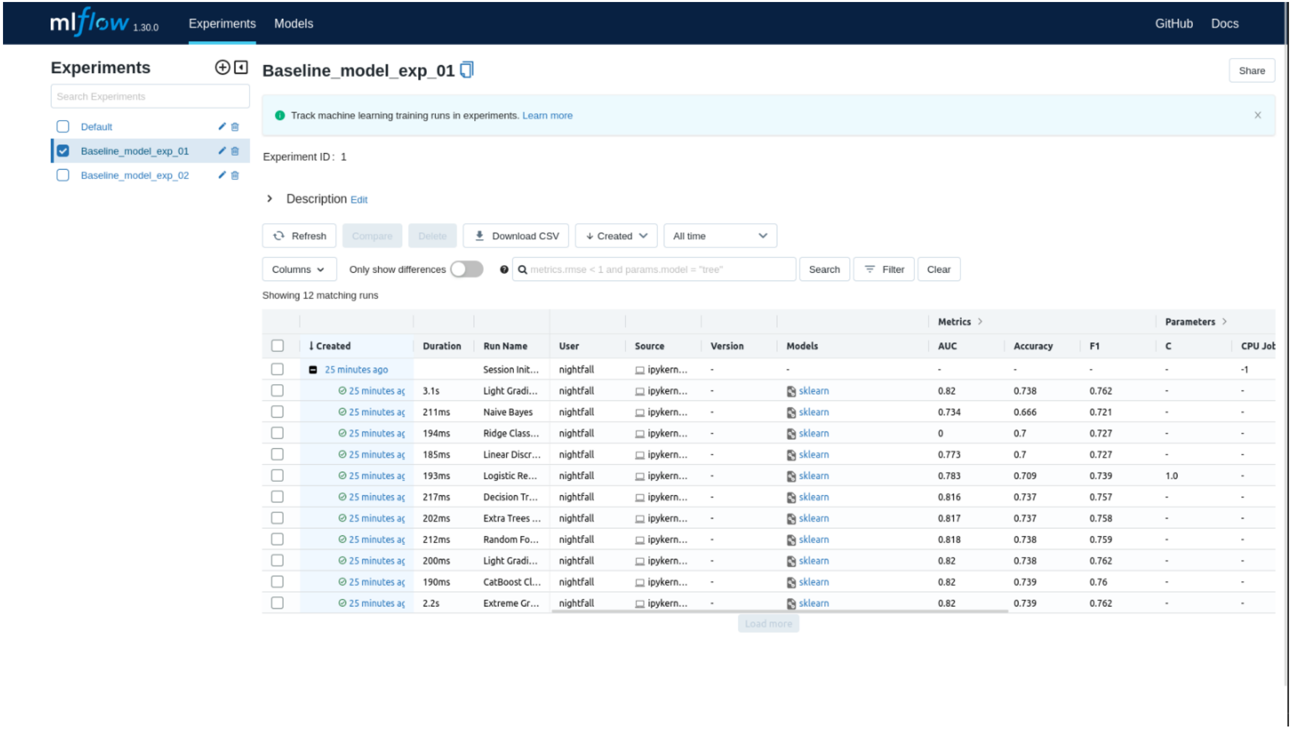
Model Experimentation by Parias

Screenshot of all the experiments



☐ Default

☑ Baseline_model_exp_01

☐ Baseline_model_exp_02

Baseline_model_exp_01

Track machine learning training runs in experiments. Learn more

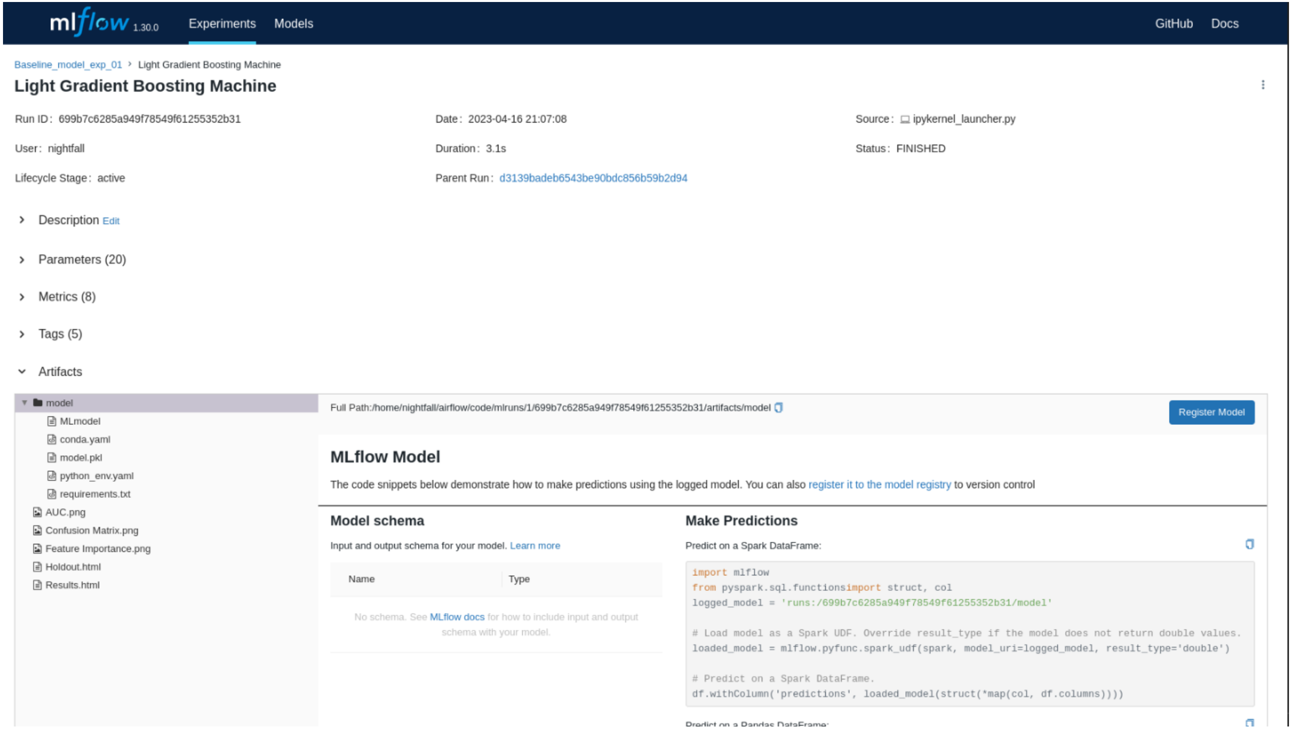
Experiment ID: 1

› Description Edit

Showing 12 matching runs

								Metrics ›			Parameters ›	
☐	↓ Created	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jot
☐	25 minutes ago		Session Init...	nightfall	ipykern...	-	-	-	-	-	-	-1
☐	25 minutes ag	3.1s	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.738	0.762	-	-
☐	25 minutes ag	211ms	Naive Bayes	nightfall	ipykern...	-	sklearn	0.734	0.666	0.721	-	-
☐	25 minutes ag	194ms	Ridge Class...	nightfall	ipykern...	-	sklearn	0	0.7	0.727	-	-
☐	25 minutes ag	185ms	Linear Discr...	nightfall	ipykern...	-	sklearn	0.773	0.7	0.727	-	-
☐	25 minutes ag	193ms	Logistic Re...	nightfall	ipykern...	-	sklearn	0.783	0.709	0.739	1.0	-
☐	25 minutes ag	217ms	Decision Tr...	nightfall	ipykern...	-	sklearn	0.816	0.737	0.757	-	-
☐	25 minutes ag	202ms	Extra Trees ...	nightfall	ipykern...	-	sklearn	0.817	0.737	0.758	-	-
☐	25 minutes ag	212ms	Random Fo...	nightfall	ipykern...	-	sklearn	0.818	0.738	0.759	-	-
☐	25 minutes ag	200ms	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.738	0.762	-	-
☐	25 minutes ag	190ms	CatBoost Cl...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.76	-	-
☐	25 minutes ag	2.2s	Extreme Gr...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-

Screenshot of one experiment with all the artifacts visible



Baseline_model_exp_01 › Light Gradient Boosting Machine

Light Gradient Boosting Machine

Run ID: 699b7c6285a949f78549f61255352b31

Date: 2023-04-16 21:07:08

Source: ipykernel_launcher.py

User: nightfall

Duration: 3.1s

Status: FINISHED

Lifecycle Stage: active

Parent Run: d3139badeb6543be90bdc856b59b2d94

› Description Edit

› Parameters (20)

› Metrics (8)

› Tags (5)

⌄ Artifacts

Full Path:/home/nightfall/airflow/code/mlruns/1/699b7c6285a949f78549f61255352b31/artifacts/model

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. You can also register it to the model registry to version control

Model schema

Input and output schema for your model. Learn more

Name	Type

No schema. See MLflow docs for how to include input and output schema with your model.

Make Predictions

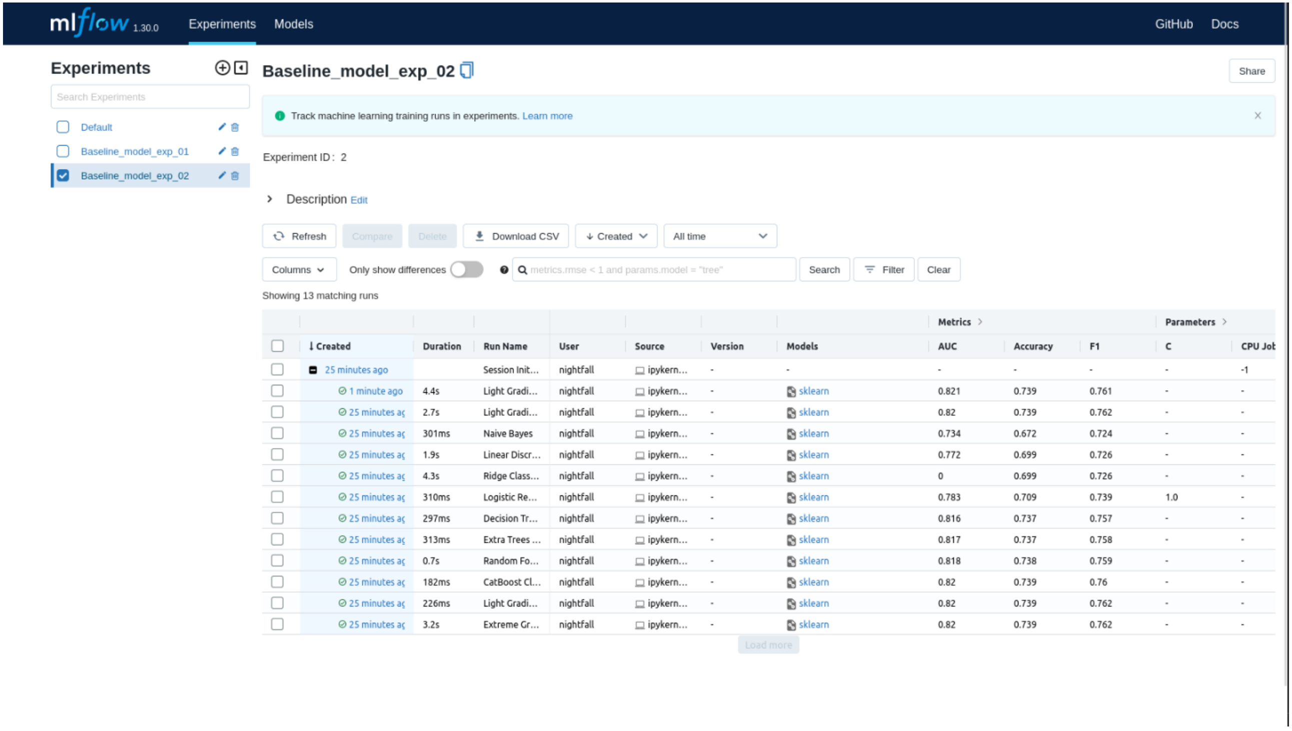
Predict on a Spark DataFrame:

```
import mlflow
from pyspark.sql.functionsimport struct, col
logged_model = 'runs:/699b7c6285a949f78549f61255352b31/model'

# 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.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

Predict on a Pandas DataFrame:

Screenshot of MLflow UI after dropping features

Screenshot of all the experiments



Baseline_model_exp_02

Track machine learning training runs in experiments. Learn more

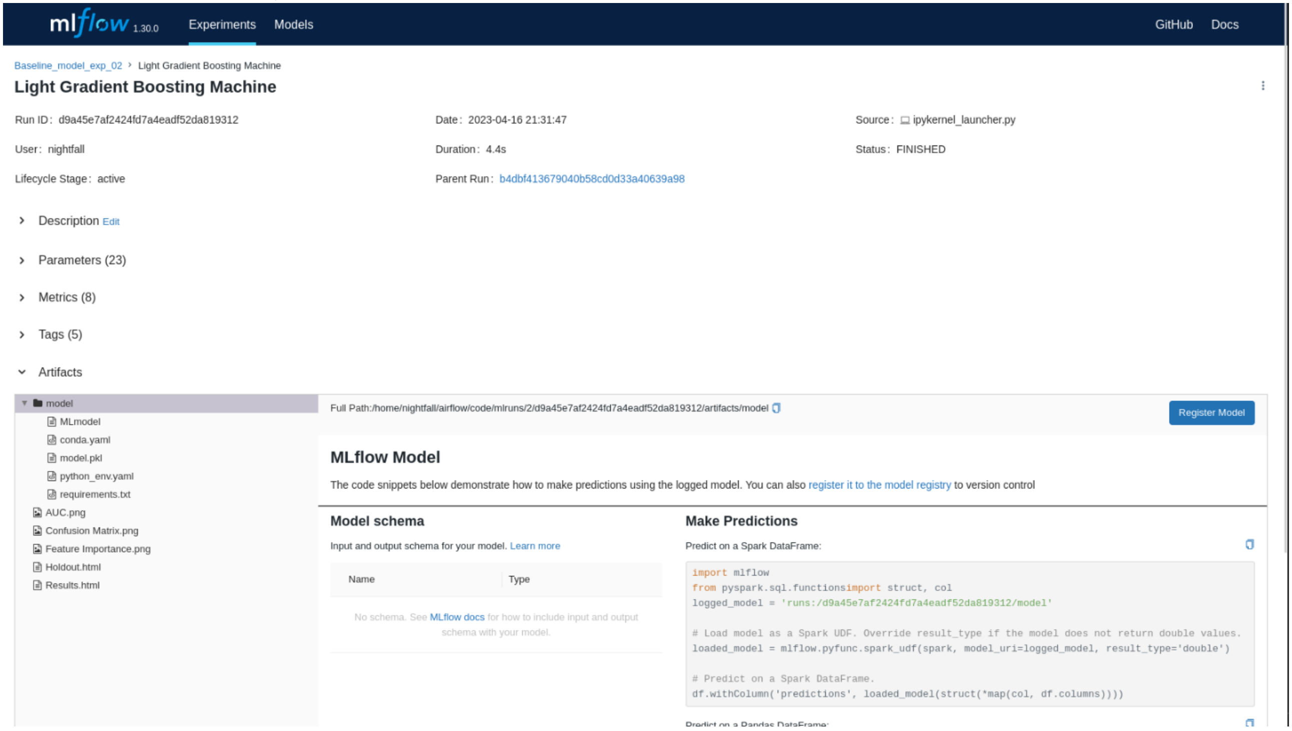
Experiment ID: 2

› Description Edit

Showing 13 matching runs

□	↓ Created	Duration	Run Name	User	Source	Version	Models	Metrics › AUC	Accuracy	F1	Parameters › C	CPU Jot
□	25 minutes ago		Session Init...	nightfall	ipykern...	-	-	-	-	-	-	-1
□	1 minute ago	4.4s	Light Gradi...	nightfall	ipykern...	-	sklearn	0.821	0.739	0.761	-	-
□	25 minutes ag	2.7s	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-
□	25 minutes ag	301ms	Naive Bayes	nightfall	ipykern...	-	sklearn	0.734	0.672	0.724	-	-
□	25 minutes ag	1.9s	Linear Discr...	nightfall	ipykern...	-	sklearn	0.772	0.699	0.726	-	-
□	25 minutes ag	4.3s	Ridge Class...	nightfall	ipykern...	-	sklearn	0	0.699	0.726	-	-
□	25 minutes ag	310ms	Logistic Re...	nightfall	ipykern...	-	sklearn	0.783	0.709	0.739	1.0	-
□	25 minutes ag	297ms	Decision Tr...	nightfall	ipykern...	-	sklearn	0.816	0.737	0.757	-	-
□	25 minutes ag	313ms	Extra Trees ...	nightfall	ipykern...	-	sklearn	0.817	0.737	0.758	-	-
□	25 minutes ag	0.7s	Random Fo...	nightfall	ipykern...	-	sklearn	0.818	0.738	0.759	-	-
□	25 minutes ag	182ms	CatBoost Cl...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.76	-	-
□	25 minutes ag	226ms	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-
□	25 minutes ag	3.2s	Extreme Gr...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-

Screenshot of one experiment with all the artifacts visible



Light Gradient Boosting Machine

Run ID: d9a45e7af2424fd7a4eadf52da819312

Date: 2023-04-16 21:31:47

Source: ipykernel_launcher.py

User: nightfall

Duration: 4.4s

Status: FINISHED

Lifecycle Stage: active

Parent Run: b4dbf413679040b58cd0d33a40639a98

› Description Edit

› Parameters (23)

› Metrics (8)

› Tags (5)

⌄ Artifacts

Full Path:/home/nightfall/airflow/code/mlruns/2/d9a45e7af2424fd7a4eadf52da819312/artifacts/model

MLflow Model

The code snippets below demonstrate how to make predictions using the logged model. You can also register it to the model registry to version control

Model schema

Input and output schema for your model. Learn more

Name	Type

No schema. See MLflow docs for how to include input and output schema with your model.

Make Predictions

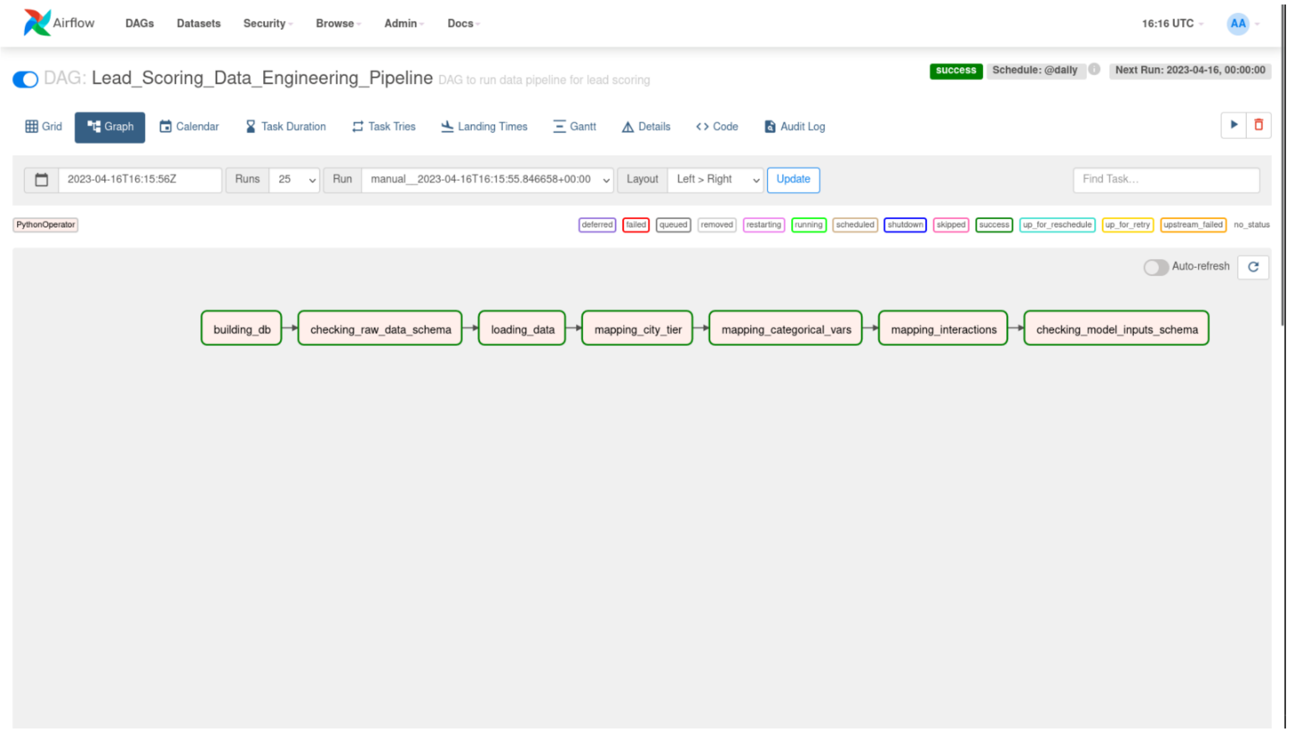
Predict on a Spark DataFrame:

```
import mlflow
from pyspark.sql.functionsimport struct, col
logged_model = 'runs:/d9a45e7af2424fd7a4eadf52da819312/model'

# 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.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

Predict on a Pandas DataFrame:

Data Pipeline

Screenshot of successful execution Airflow DAG in graph



DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring

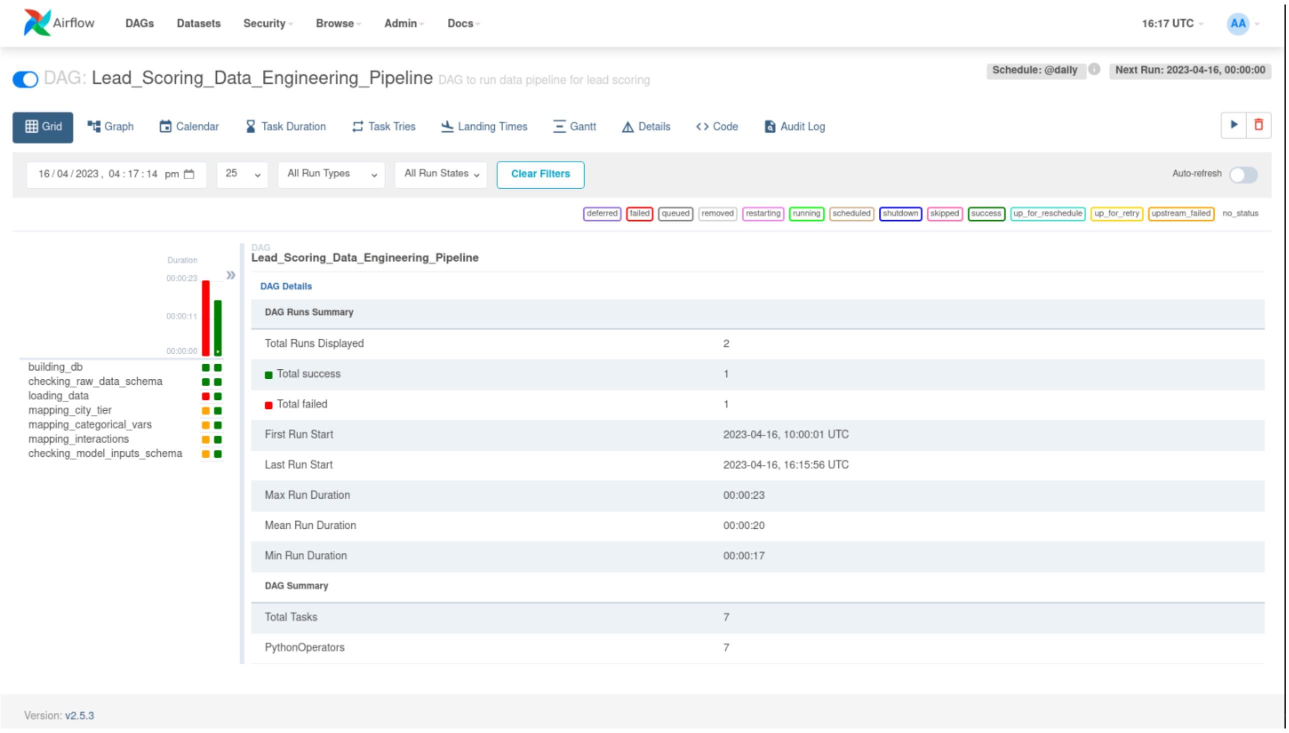
success Schedule: @daily Next Run: 2023-04-16, 00:00:00

2023-04-16T16:15:56Z Runs 25 Run manual__2023-04-16T16:15:55.846658+00:00 Layout Left > Right

PythonOperator

building_db → checking_raw_data_schema → loading_data → mapping_city_tier → mapping_categorical_vars → mapping_interactions → checking_model_inputs_schema

Screenshot of Airflow UI grid



DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring

Schedule: @daily Next Run: 2023-04-16, 00:00:00

16 / 04 / 2023 , 04 : 17 : 14 pm 25 All Run Types All Run States

building_db	🟩	🟩
checking_raw_data_schema	🟩	🟩
loading_data	🟥	🟩
mapping_city_tier	🟧	🟩
mapping_categorical_vars	🟧	🟩
mapping_interactions	🟧	🟩
checking_model_inputs_schema	🟧	🟩

DAG
Lead_Scoring_Data_Engineering_Pipeline

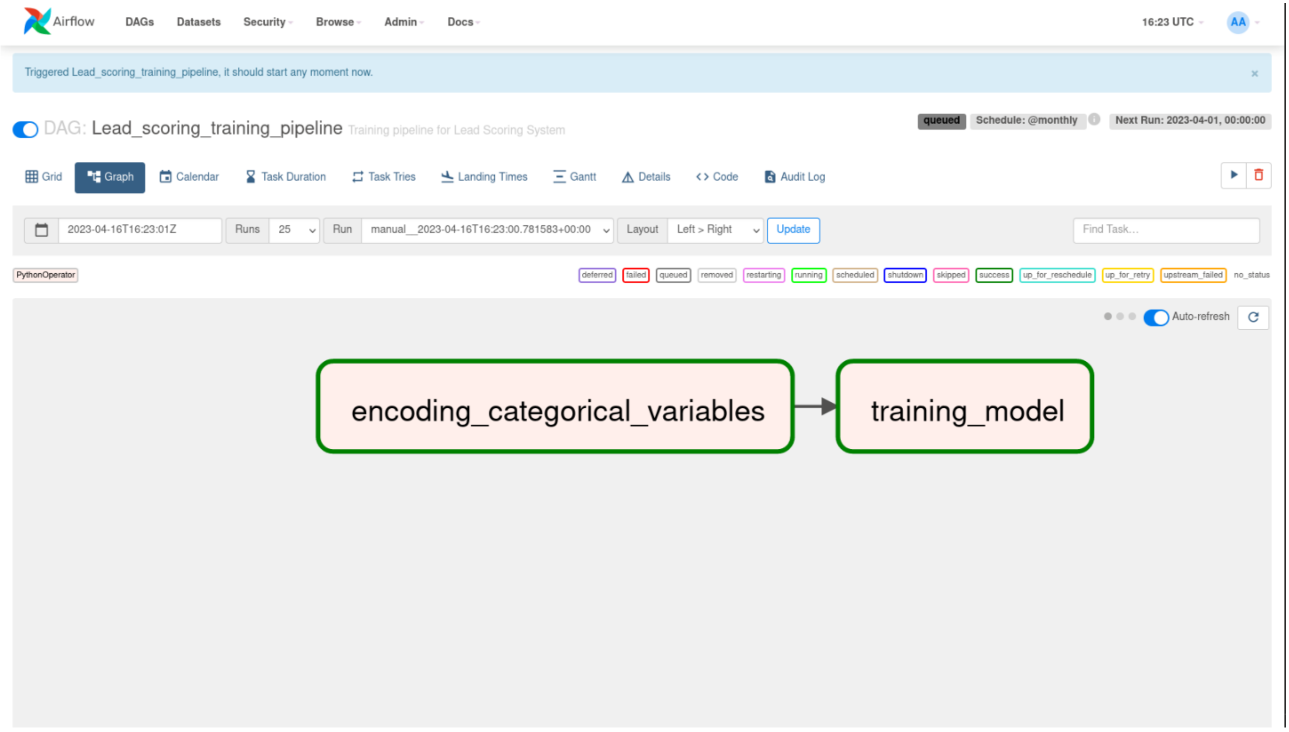
DAG Details

DAG Runs Summary	
Total Runs Displayed	2
🟩 Total success	1
🟥 Total failed	1
First Run Start	2023-04-16, 10:00:01 UTC
Last Run Start	2023-04-16, 16:15:56 UTC
Max Run Duration	00:00:23
Mean Run Duration	00:00:20
Min Run Duration	00:00:17
DAG Summary	
Total Tasks	7
PythonOperators	7

Version: v2.5.3

Training pipeline

Screenshot of successful execution Airflow DAG in graph



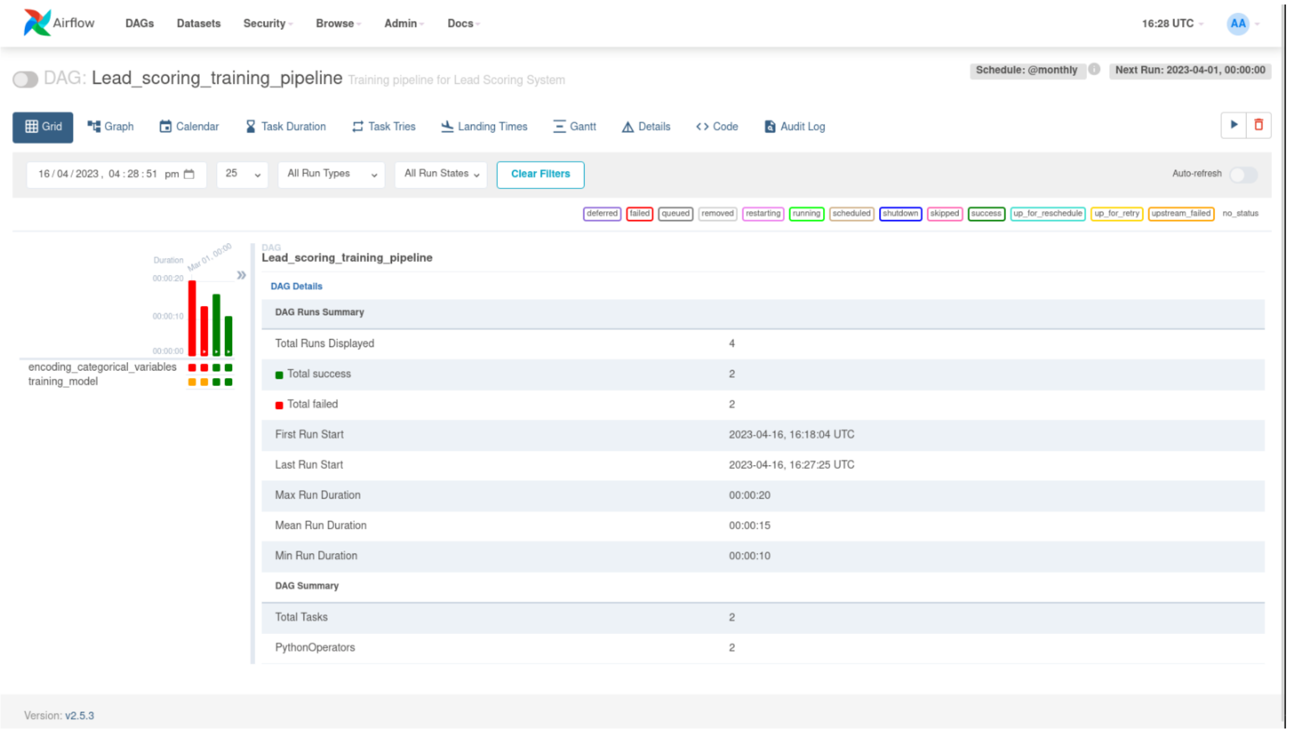
Triggered Lead_scoring_training_pipeline, it should start any moment now.

DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System

queued Schedule: @monthly Next Run: 2023-04-01, 00:00:00

2023-04-16T16:23:01Z Runs 25 Run manual__2023-04-16T16:23:00.781583+00:00 Layout Left > Right Update

Screenshot of Airflow UI grid



DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System

Schedule: @monthly Next Run: 2023-04-01, 00:00:00

16 / 04 / 2023 , 04 : 28 : 51 pm 25 All Run Types All Run States Clear Filters

DAG
Lead_scoring_training_pipeline

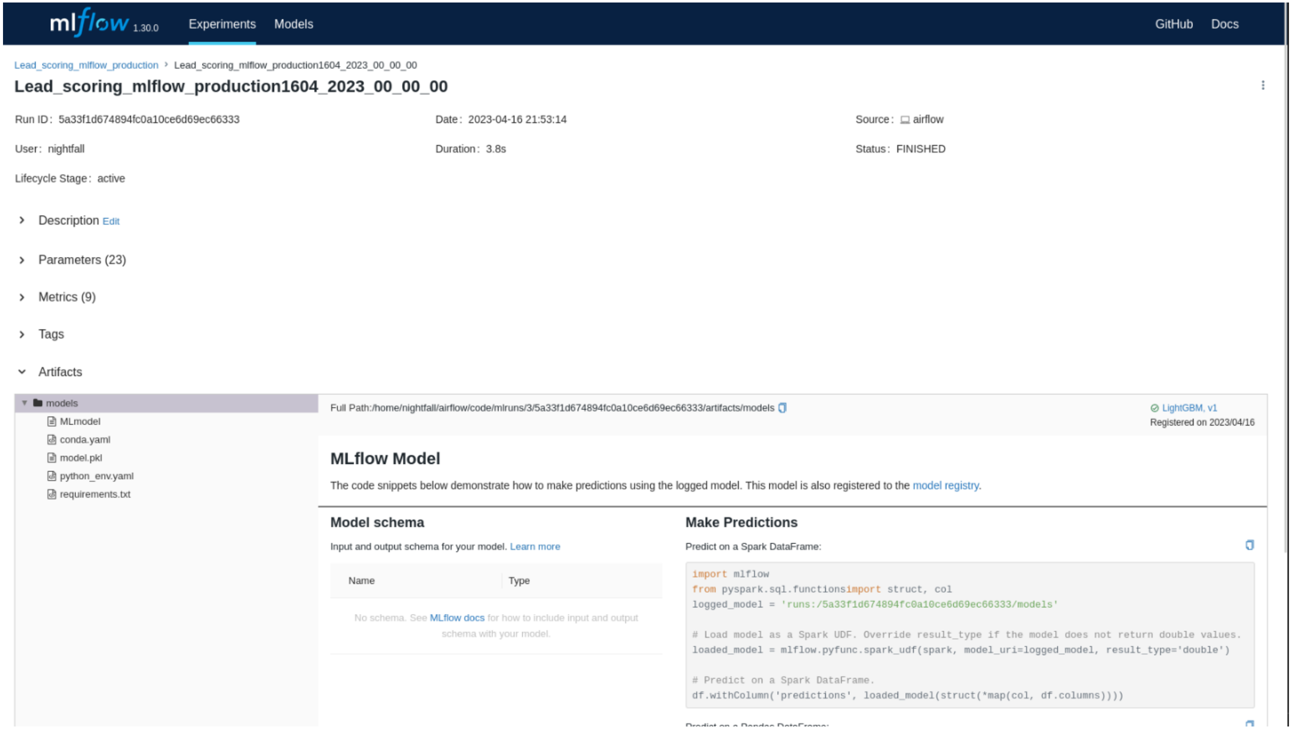
DAG Details

DAG Runs Summary	
Total Runs Displayed	4
🟩 Total success	2
🟥 Total failed	2
First Run Start	2023-04-16, 16:18:04 UTC
Last Run Start	2023-04-16, 16:27:25 UTC
Max Run Duration	00:00:20
Mean Run Duration	00:00:15
Min Run Duration	00:00:10
DAG Summary	
Total Tasks	2
PythonOperators	2

Version: v2.5.3

ML artifacts for training pipeline

Screenshot of experiments with all the artifacts visible



Lead_scoring_mlflow_production › Lead_scoring_mlflow_production1604_2023_00_00_00

Lead_scoring_mlflow_production1604_2023_00_00_00

Run ID: 5a33f1d674894fc0a10ce6d69ec66333
Date: 2023-04-16 21:53:14
Source: airflow
User: nightfall
Duration: 3.8s
Status: FINISHED
Lifecycle Stage: active

› Description Edit

› Parameters (23)

› Metrics (9)

› Tags

⌄ Artifacts

- models
  - MLmodel
  - conda.yaml
  - model.pkl
  - python_env.yaml
  - requirements.txt

Full Path:/home/nightfall/airflow/code/mlruns/3/5a33f1d674894fc0a10ce6d69ec66333/artifacts/models

LightGBM, v1
Registered on 2023/04/16

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 MLflow docs for how to include input and output schema with your model.

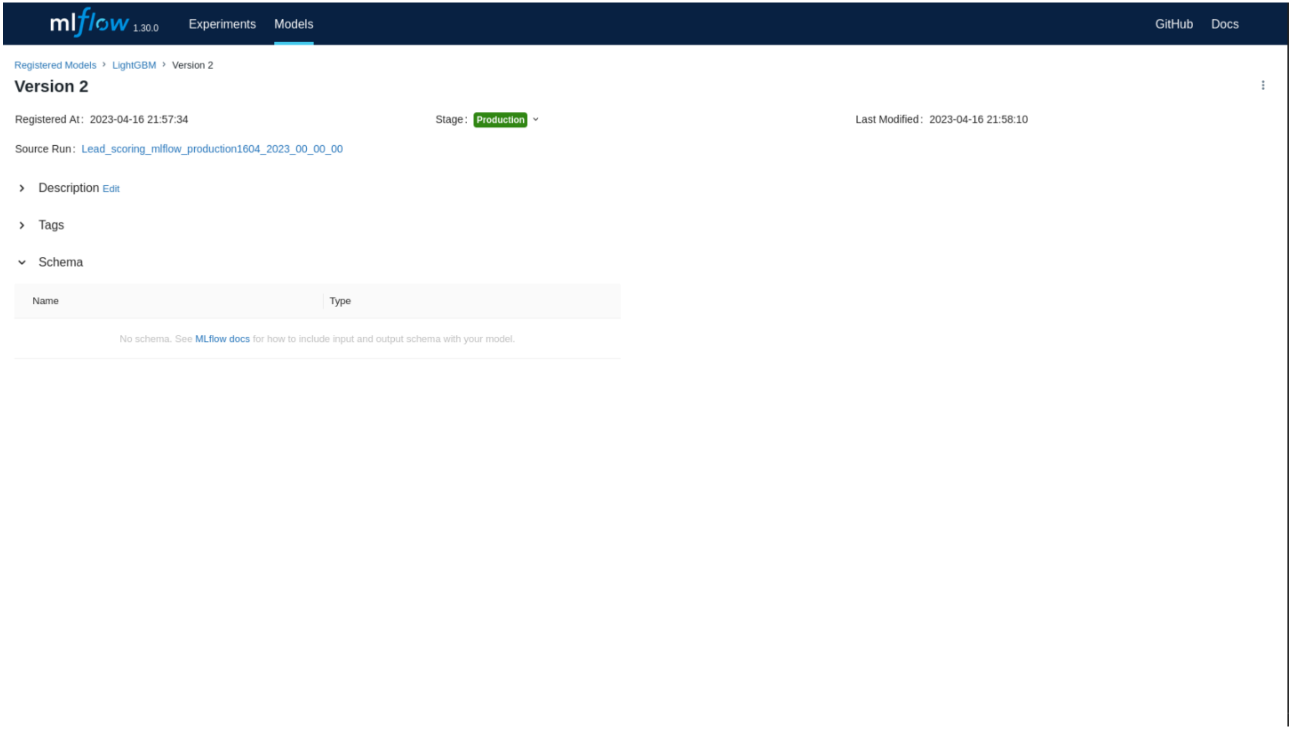
Make Predictions

Predict on a Spark DataFrame:

```
import mlflow
from pyspark.sql.functionsimport struct, col
logged_model = 'runs:/5a33f1d674894fc0a10ce6d69ec66333/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.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

Predict on a Pandas DataFrame:

Screenshot of model registry with model name and stage as 'production'



Registered Models › LightGBM › Version 2

Version 2

Registered At: 2023-04-16 21:57:34
Stage: Production
Last Modified: 2023-04-16 21:58:10
Source Run: Lead_scoring_mlflow_production1604_2023_00_00_00

› Description Edit

› Tags

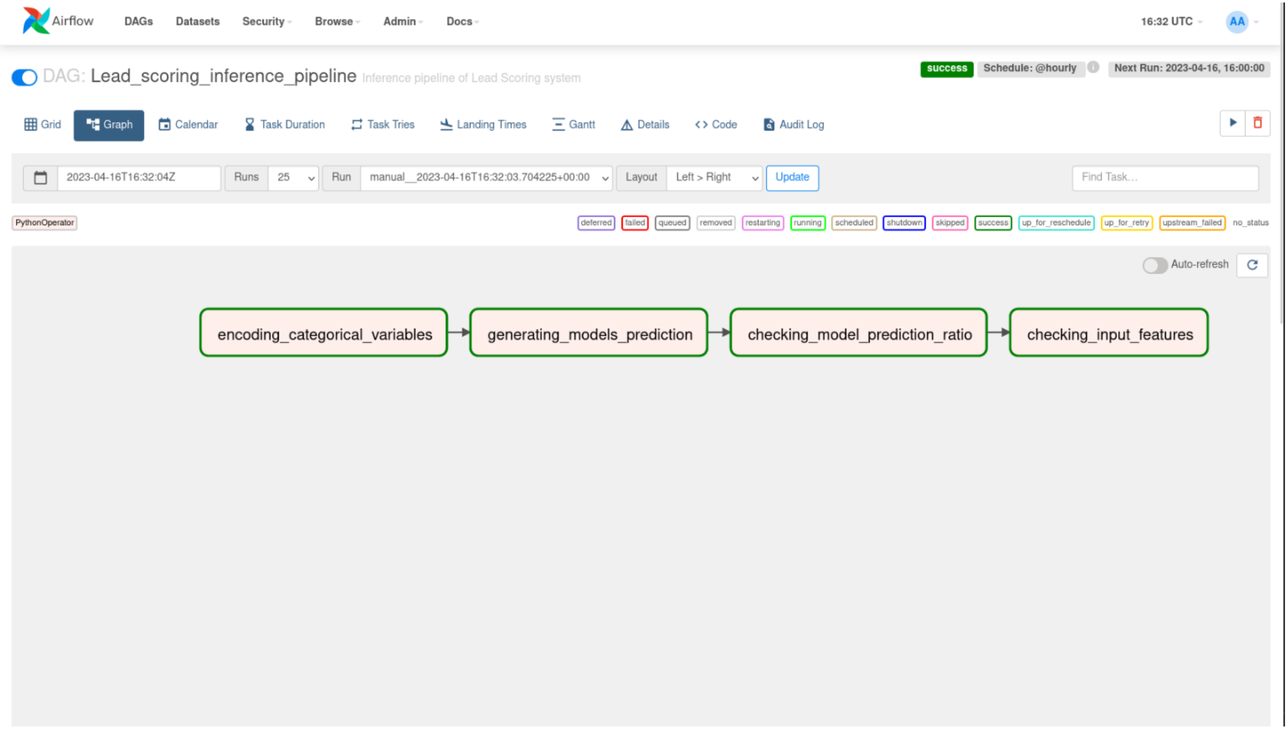
⌄ Schema

Name	Type

No schema. See MLflow docs for how to include input and output schema with your model.

Inference Pipeline

Screenshot of successful execution Airflow DAG in graph



DAG: Lead_scoring_inference_pipeline Inference pipeline of Lead Scoring system

success Schedule: @hourly Next Run: 2023-04-16, 16:00:00

2023-04-16T16:32:04Z

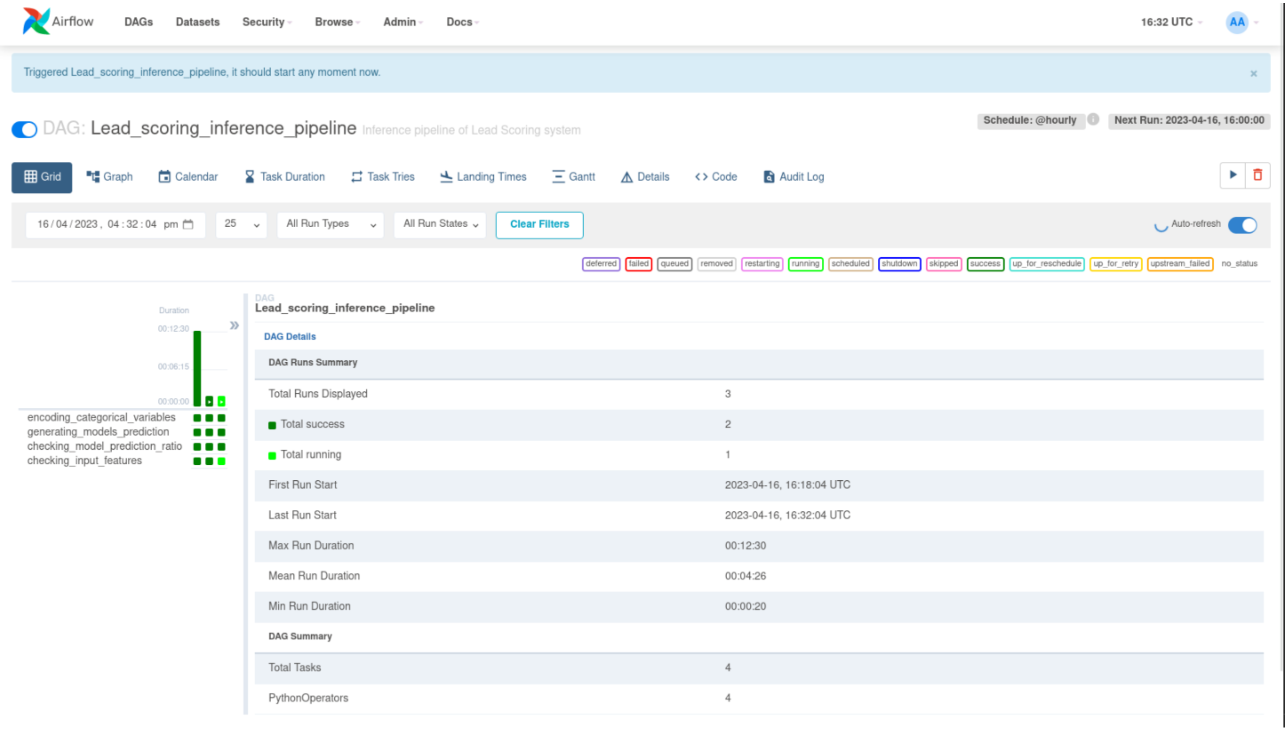
Runs 25

Run manual__2023-04-16T16:32:03.704225+00:00

Layout Left > Right

encoding_categorical_variables → generating_models_prediction → checking_model_prediction_ratio → checking_input_features

Screenshot of Airflow UI grid



Triggered Lead_scoring_inference_pipeline, it should start any moment now.

DAG: Lead_scoring_inference_pipeline Inference pipeline of Lead Scoring system

Schedule: @hourly Next Run: 2023-04-16, 16:00:00

16 / 04 / 2023 , 04 : 32 : 04 pm

25

All Run Types

All Run States

DAG
Lead_scoring_inference_pipeline

DAG Details

DAG Runs Summary	
Total Runs Displayed	3
🟩 Total success	2
🟩 Total running	1
First Run Start	2023-04-16, 16:18:04 UTC
Last Run Start	2023-04-16, 16:32:04 UTC
Max Run Duration	00:12:30
Mean Run Duration	00:04:26
Min Run Duration	00:00:20
DAG Summary	
Total Tasks	4
PythonOperators	4