

Model Experimentation – CodePro

Screenshot of MLflow UI

Screenshot of all the experiments

The screenshot shows the MLflow Experiments page. The top navigation bar includes the MLflow logo, version 1.30.0, and tabs for Experiments and Models. The Experiments tab is active, and the experiment 'Baseline_model_exp_01' is selected. A sidebar on the left shows a list of experiments: 'Default', 'Baseline_model_exp_01' (selected), and 'Baseline_model_exp_02'. The main content area displays the experiment details for 'Baseline_model_exp_01', including a description, a table of runs, and a 'Load more' button.

Experiment ID: 1

Showing 12 matching runs

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

Load more

Screenshot of one experiment with all the artifacts visible

mlflow1.30.0

ExperimentsModels

GitHubDocs

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

model

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

AUC.png

Confusion Matrix.png

Feature Importance.png

Holdout.html

Results.html

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

Register 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.functions import 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))))
```

Deserializes as a Pandas DataFrame.

Screenshot of MLflow UI after dropping features

Screenshot of all the experiments

mlflow1.30.0

ExperimentsModels

GitHubDocs

Experiments

Search Experiments

Default

Baseline_model_exp_01

Baseline_model_exp_02

Baseline_model_exp_02

Share

Track machine learning training runs in experiments. [Learn more](#)

Experiment ID: 2

> Description [Edit](#)

Refresh

Compare

Delete

Download CSV

↓ Created

All time

Columns

Only show differences

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

Search

Filter

Clear

Showing 13 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
	1 minute ago	4.4s	Light Gradi...	nightfall	ipykern...	-	sklearn	0.821	0.739	0.761	-	-
	25 minutes ago	2.7s	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-
	25 minutes ago	301ms	Naive Bayes	nightfall	ipykern...	-	sklearn	0.734	0.672	0.724	-	-
	25 minutes ago	1.9s	Linear Discr...	nightfall	ipykern...	-	sklearn	0.772	0.699	0.726	-	-
	25 minutes ago	4.3s	Ridge Class...	nightfall	ipykern...	-	sklearn	0	0.699	0.726	-	-
	25 minutes ago	310ms	Logistic Re...	nightfall	ipykern...	-	sklearn	0.783	0.709	0.739	1.0	-
	25 minutes ago	297ms	Decision Tr...	nightfall	ipykern...	-	sklearn	0.816	0.737	0.757	-	-
	25 minutes ago	313ms	Extra Trees ...	nightfall	ipykern...	-	sklearn	0.817	0.737	0.758	-	-
	25 minutes ago	0.7s	Random Fo...	nightfall	ipykern...	-	sklearn	0.818	0.738	0.759	-	-
	25 minutes ago	182ms	CatBoost CL...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.76	-	-
	25 minutes ago	226ms	Light Gradi...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-
	25 minutes ago	3.2s	Extreme Gr...	nightfall	ipykern...	-	sklearn	0.82	0.739	0.762	-	-

Load more

Screenshot of one experiment with all the artifacts visible

mlflow1.30.0

ExperimentsModels

GitHubDocs

Baseline_model_exp_02>Light Gradient Boosting Machine

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

model

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

AUC.png

Confusion Matrix.png

Feature Importance.png

Holdout.html

Results.html

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

Register 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.functions import struct, col
logged_model = 'runs:/d9a45e7af2424fd7a4eadf52da819312/model'

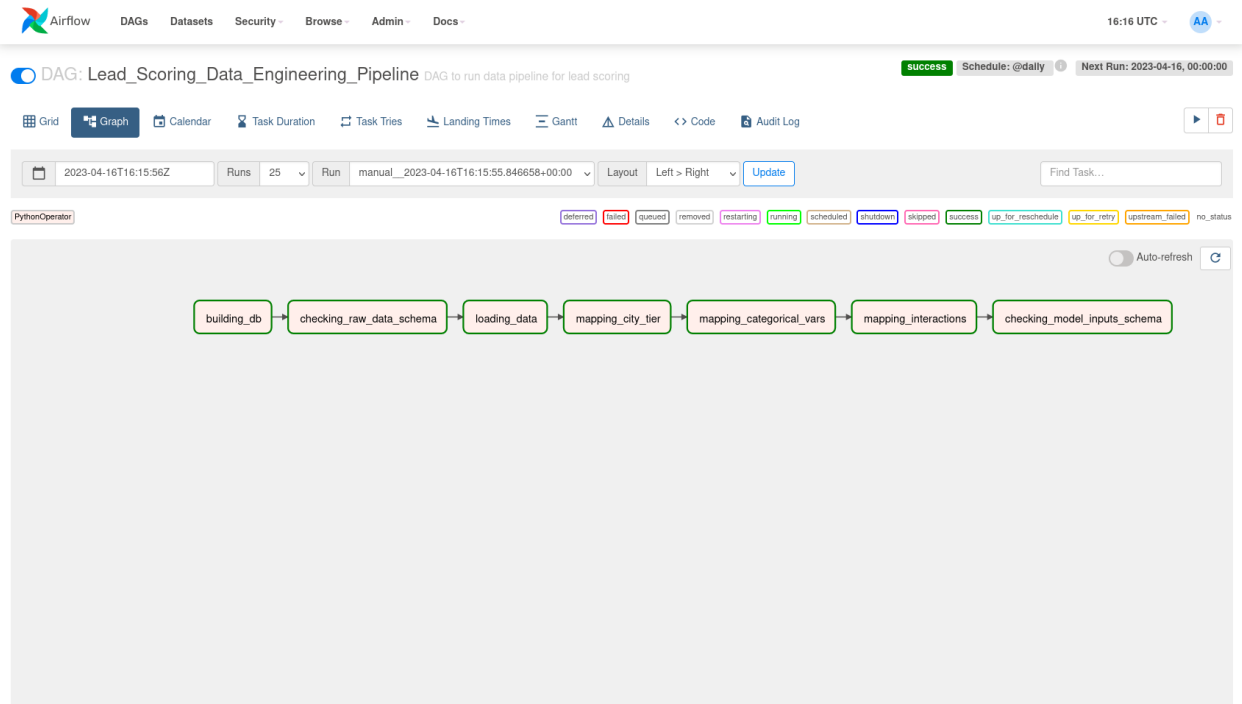
# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=logged_model, result_type='double')

# Predict on a Spark DataFrame.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

Deserializes as a Pandas DataFrame

Data Pipeline

Screenshot of successful execution Airflow DAG in graph



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

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

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

Auto-refresh

deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status



DAG Lead_Scoring_Data_Engineering_Pipeline

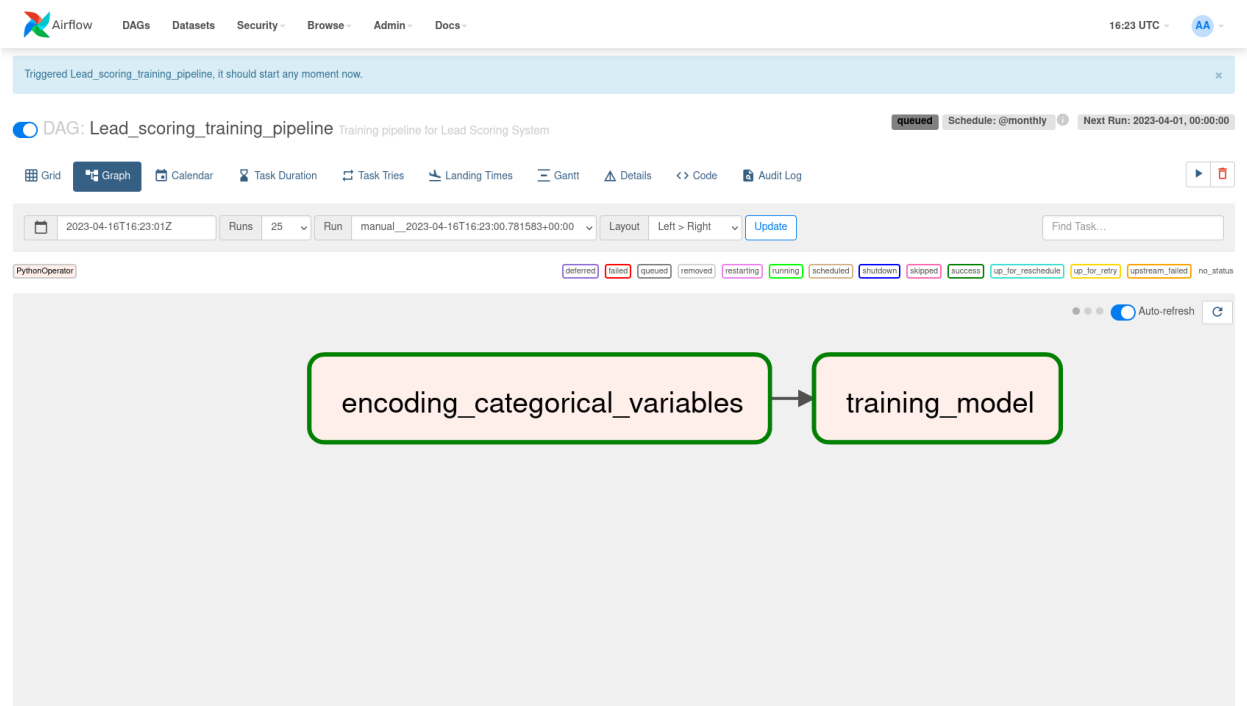
DAG Details

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

Training pipeline

Screenshot of successful execution Airflow DAG in graph



Screenshot of Airflow UI grid

DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System

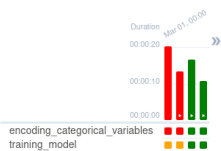
Schedule: @monthly Next Run: 2023-04-01, 00:00:00

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

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

Auto-refresh

deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status



DAG Lead_scoring_training_pipeline

DAG Details

DAG Runs Summary

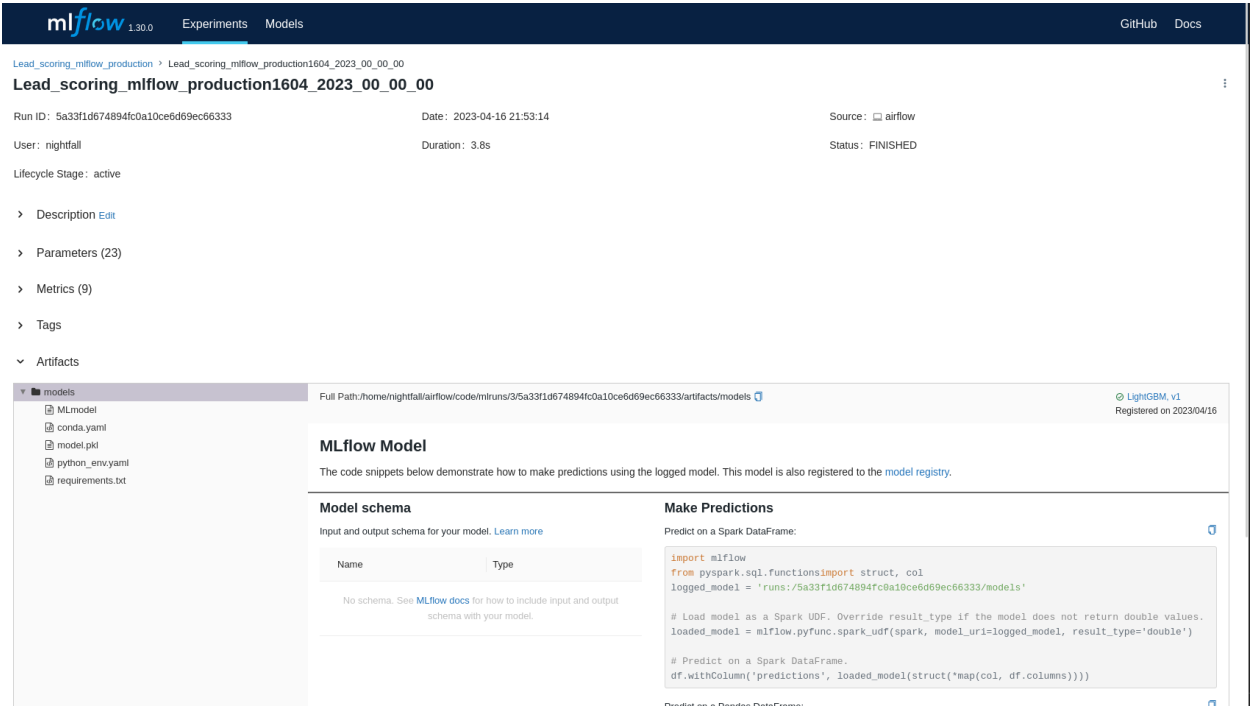
Total Runs Displayed	4
Total success	2
Total failed	2
First Run Start	2023-04-16, 16:18:04 UTC
Last Run Start	2023-04-16, 16:27:25 UTC
Max Run Duration	00:00:20
Mean Run Duration	00:00:15
Min Run Duration	00:00:10

DAG Summary

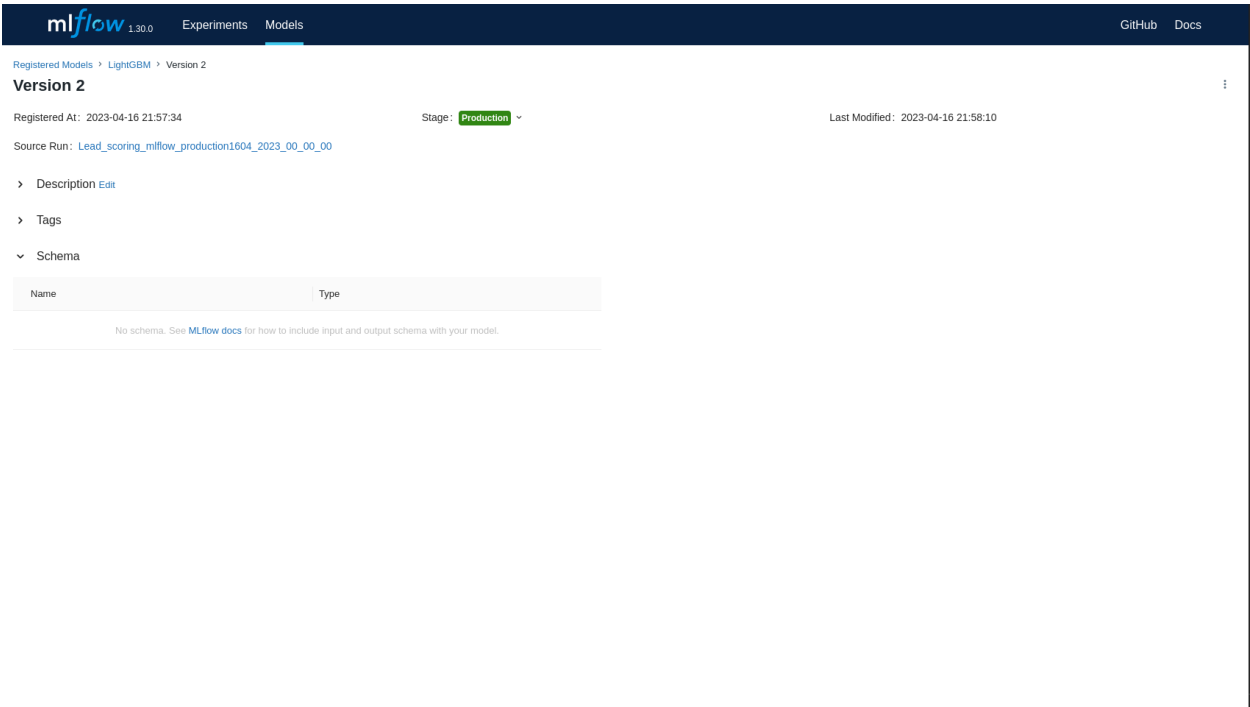
Total Tasks	2
PythonOperators	2

ML artifacts for training pipeline

Screenshot of experiments with all the artifacts visible

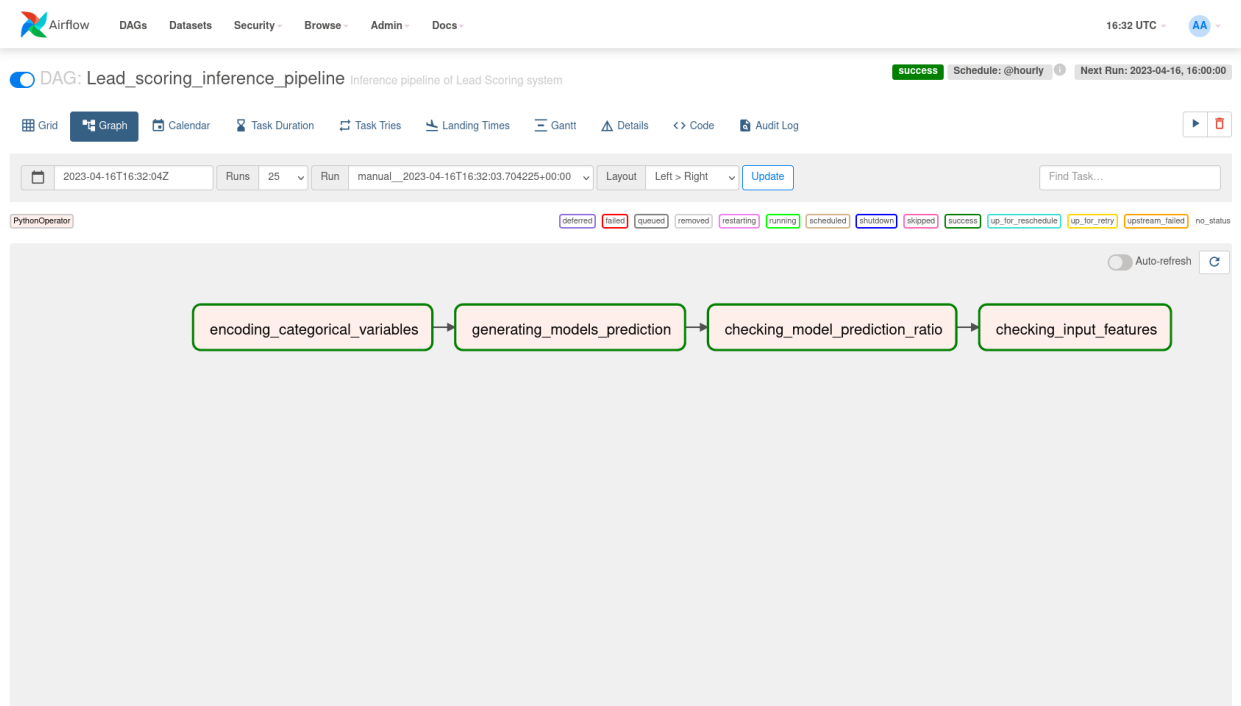


Screenshot of model registry with model name and stage as 'production'



Inference Pipeline

Screenshot of successful execution Airflow DAG in graph



Screenshot of Airflow UI grid

