

Model Experimentation

Screenshot of MLflow UI

Screenshot of all the experiments

The screenshot shows the MLflow Experiments page. On the left, there's a sidebar with 'Experiments' and a search bar. Below it, a list of experiments is shown, with 'Baseline_model_exp_01' selected. The main area displays the details for 'Baseline_model_exp_01', including a description, a table of runs, and a 'Load more' button. The table shows 12 matching runs with columns for Created, Duration, Run Name, User, Source, Version, Models, Metrics (AUC, Accuracy, F1), and Parameters (C, CPU Jot).

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

Screenshot of one experiment with all the artifacts visible

The screenshot shows the MLflow UI for a specific experiment, 'Light Gradient Boosting Machine'. It displays the run ID, date, source, user, duration, status, lifecycle stage, and parent run. Below this, there's a list of artifacts, including 'model', 'AUC.png', 'Confusion Matrix.png', 'Feature Importance.png', 'Holdout.html', and 'Results.html'. The 'model' artifact is expanded, showing the 'MLflow Model' schema and the code snippets for making predictions on a Spark DataFrame.

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

Artifacts

- model
 - MLmodel
 - conda.yaml
 - model.pkl
 - python_env.yaml
 - requirements.txt
 - AUC.png
 - Confusion Matrix.png
 - Feature Importance.png
 - Holdout.html
 - Results.html

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/model1'

# 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))))
```

Download on a Databricks Workspace

Screenshot of MLflow UI after dropping features

Screenshot of all the experiments

The screenshot shows the MLflow Experiments page for 'Baseline_model_exp_02'. The interface includes a search bar, a list of experiments (Default, Baseline_model_exp_01, Baseline_model_exp_02), and a table of runs. The table columns are: Created, Duration, Run Name, User, Source, Version, Models, Metrics (AUC, Accuracy, F1), and Parameters (C, CPU Job). The table shows 13 matching runs, with the first run being 'Session Init...' and the last run being 'Extreme Gr...'. A 'Load more' button is visible at the bottom of the table.

Created	Duration	Run Name	User	Source	Version	Models	Metrics	Parameters
25 minutes ago		Session Init...	nightfall	ipykern...	-	-	-	-
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	-

Screenshot of one experiment with all the artifacts visible

The screenshot shows the MLflow UI for a specific experiment, 'Baseline_model_exp_02 > Light Gradient Boosting Machine'. The experiment details include: Run ID: d9a45e7af2424fd7a4eadf52da819312, Date: 2023-04-16 21:31:47, Source: ipykernel_launcher.py, User: nightfall, Duration: 4.4s, Status: FINISHED, Lifecycle Stage: active, and Parent Run: b4dbf413679040b58cd0d33a40639a98.

The 'Artifacts' section is expanded, showing a list of artifacts: MLmodel, conda.yaml, model.pkl, python_env.yaml, requirements.txt, AUC.png, Confusion Matrix.png, Feature Importance.png, Holdout.html, and Results.html. The 'MLflow Model' section is also visible, showing the code snippets for making predictions using the logged model.

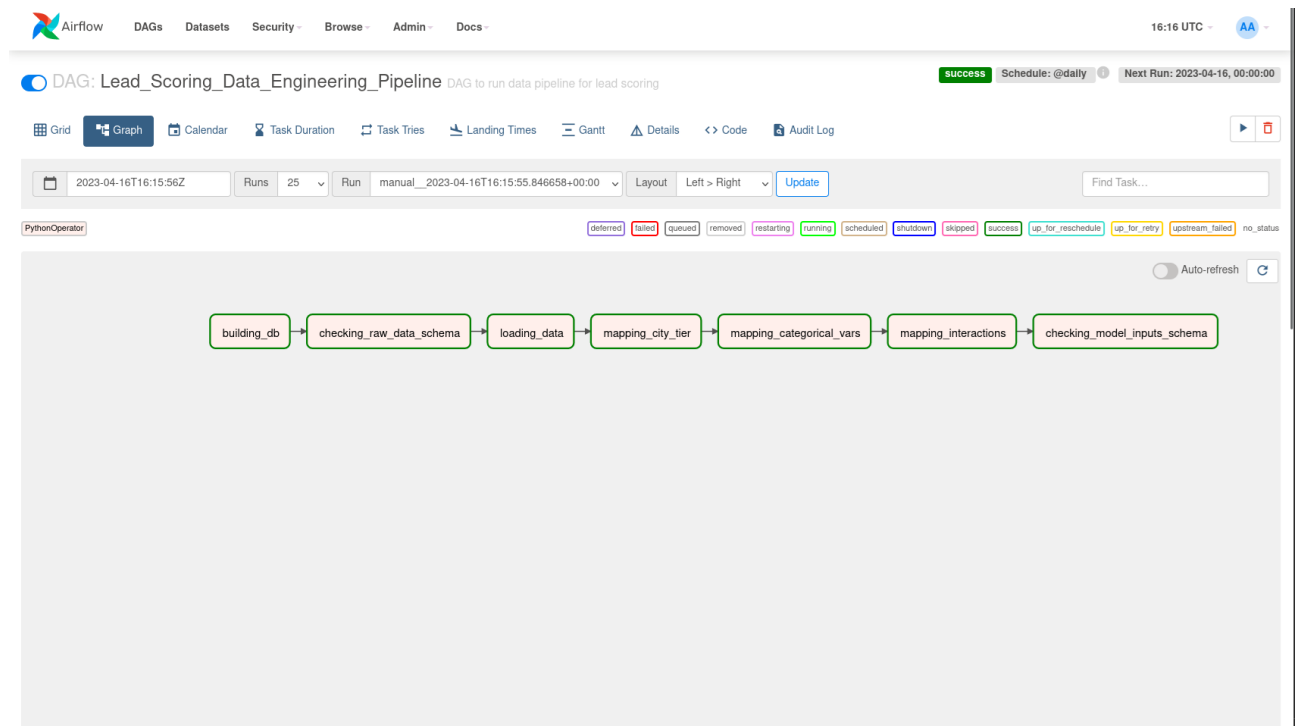
```
import mlflow
from pyspark.sql.functions import struct, col
logged_model = 'runs:/d9a45e7af2424fd7a4eadf52da819312/model'

# 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')

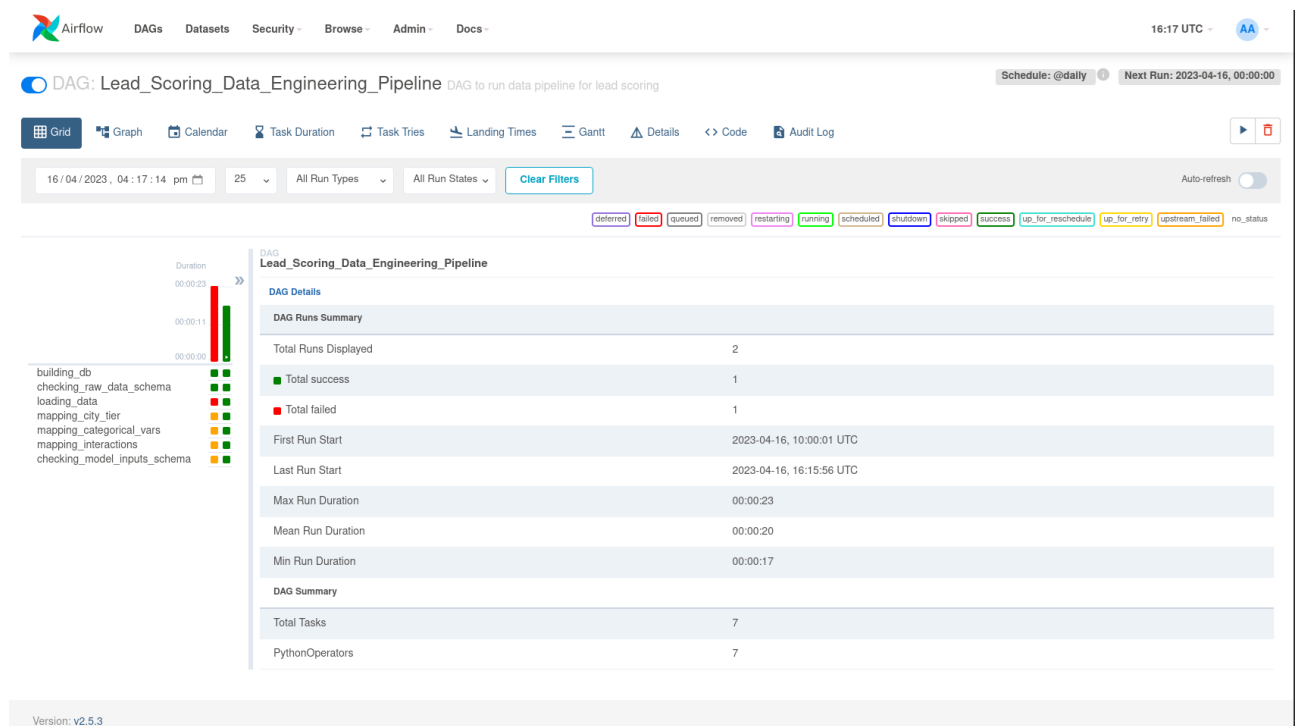
# Predict on a Spark DataFrame.
df.withColumn('predictions', loaded_model(struct(*map(col, df.columns))))
```

Data Pipeline

Screenshot of successful execution Airflow DAG in graph

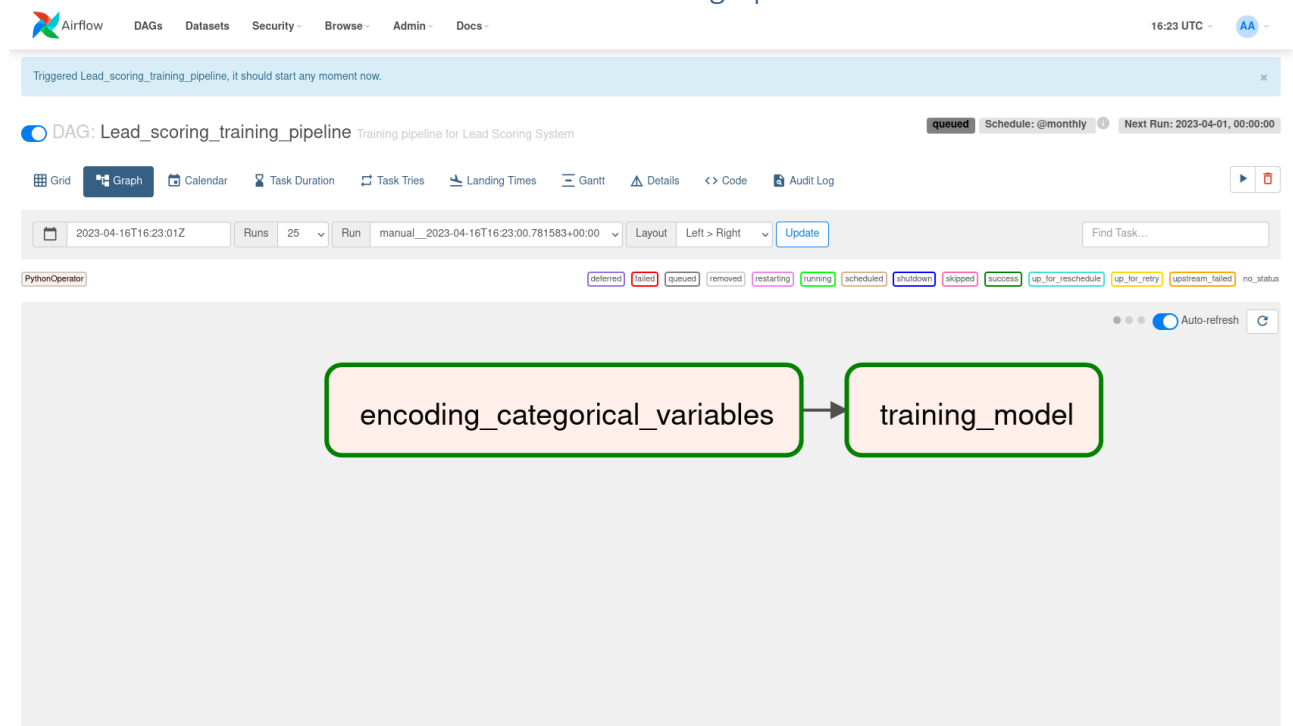


Screenshot of Airflow UI grid

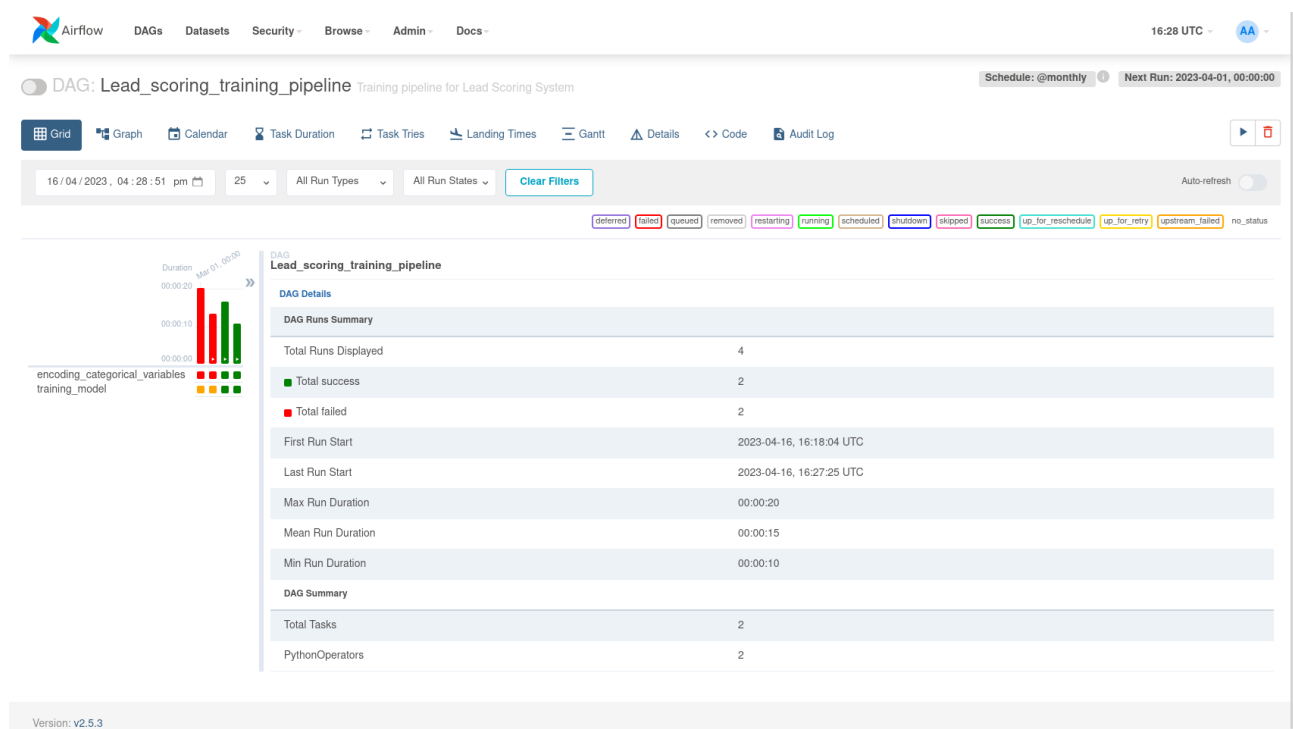


Training pipeline

Screenshot of successful execution Airflow DAG in graph



Screenshot of Airflow UI grid



ML artifacts for training pipeline

Screenshot of experiments with all the artifacts visible

The screenshot shows the MLflow web interface. At the top, there's a navigation bar with 'mlflow 1.30.0', 'Experiments', and 'Models' tabs. Below the navigation bar, the breadcrumb trail is 'Lead_scoring_mlflow_production > Lead_scoring_mlflow_production1604_2023_00_00_00'. The main title is 'Lead_scoring_mlflow_production1604_2023_00_00_00'. Below the title, there's a summary section with 'Run ID: 5a33f1d674894fc0a10ce6d69ec66333', 'Date: 2023-04-16 21:53:14', 'Source: airflow', 'User: nightfall', 'Duration: 3.8s', and 'Status: FINISHED'. Below this, there's a 'Lifecycle Stage: active' section. A sidebar on the left contains a list of artifacts: 'models', 'MLmodel', 'conda.yaml', 'model.pkl', 'python_env.yaml', and 'requirements.txt'. The main content area shows the 'models' artifact, which is an 'MLflow Model'. It includes a 'Model schema' section with a table for input and output schema, and a 'Make Predictions' section with a code snippet for loading and predicting with the model.

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.functions import struct, col
logged_model = 'runs:/5a33f1d674894fc0a10ce6d69ec66333/model1s'

# 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'

The screenshot shows the MLflow web interface for the model registry. At the top, there's a navigation bar with 'mlflow 1.30.0', 'Experiments', and 'Models' tabs. Below the navigation bar, the breadcrumb trail is 'Registered Models > LightGBM > Version 2'. The main title is 'Version 2'. Below the title, there's a summary section with 'Registered At: 2023-04-16 21:57:34', 'Stage: Production', and 'Last Modified: 2023-04-16 21:58:10'. Below this, there's a 'Source Run: Lead_scoring_mlflow_production1604_2023_00_00_00' section. A sidebar on the left contains a list of artifacts: 'Description', 'Tags', and 'Schema'. The main content area shows the 'Schema' artifact, which is a table with columns 'Name' and 'Type'. Below the table, there's a message: 'No schema. See [MLflow docs](#) for how to include input and output schema with your model.'

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

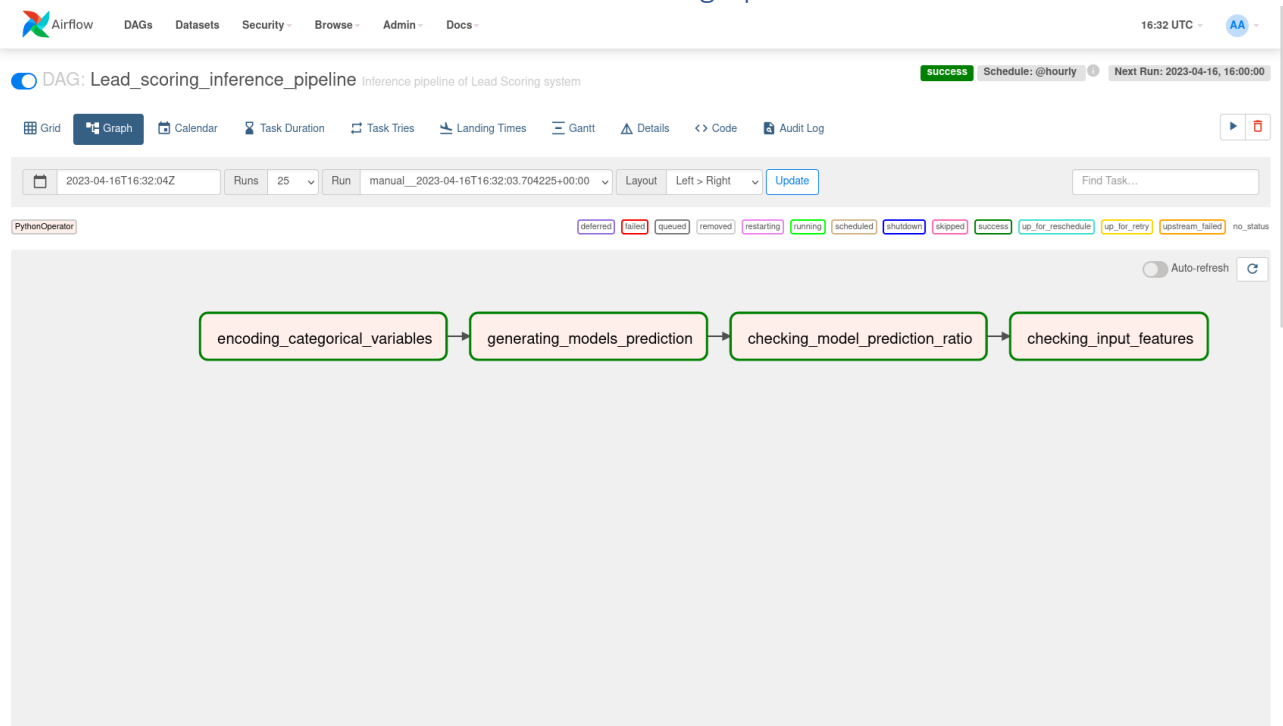
Tags

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



Screenshot of Airflow UI grid

