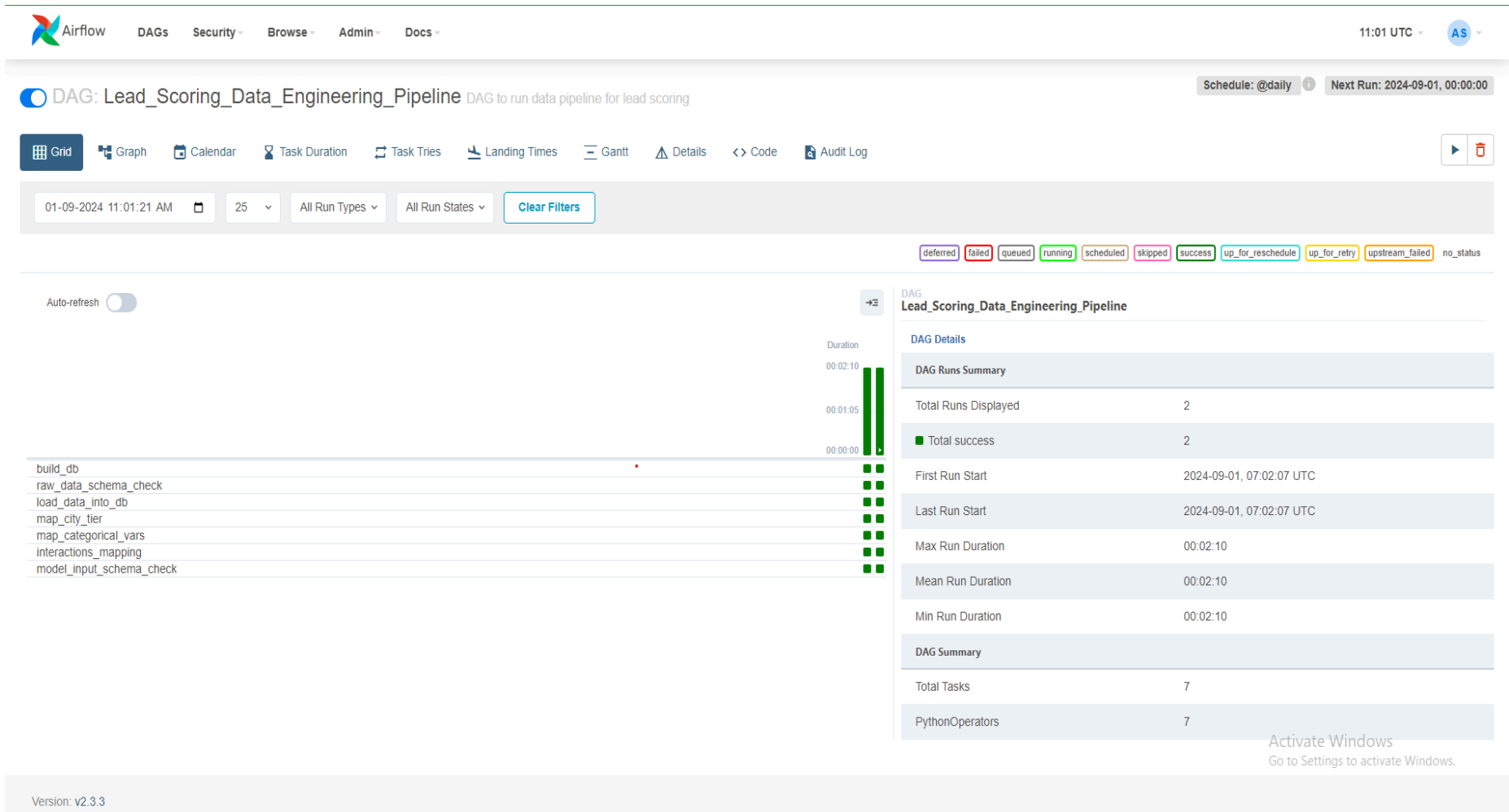





# Airflow Data Engineering Pipeline









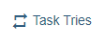
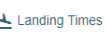






# Airflow Data Training Pipeline





 [DAGs](#) [Security](#) [Browse](#) [Admin](#) [Docs](#)

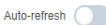

17:41 UTC 


Triggered Lead\_scoring\_training\_pipeline, it should start any moment now. 

 **DAG: Lead\_scoring\_training\_pipeline** Training pipeline for Lead Scoring System Schedule: @monthly  Next Run: 2024-09-01, 00:00:00


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

01-09-2024 05:39:16 PM  25  All Run Types  All Run States 

 Auto-refresh 




Task	Duration
encode_features	00:00:00
get_trained_model	00:00:00



**DAG**  
**Lead\_scoring\_training\_pipeline**

**DAG Details**

**DAG Runs Summary**

Total Runs Displayed	1
 Total success	1
First Run Start	2024-09-01, 17:39:16 UTC
Last Run Start	2024-09-01, 17:39:16 UTC
Max Run Duration	00:00:20
Mean Run Duration	00:00:20
Min Run Duration	00:00:20

**DAG Summary**

Total Tasks	2	Activate Windows Go to Settings to activate Windows.
PythonOperators	2	

# MLFlow Experiments

mlflow1.26.1

ExperimentsModels

Lead\_scoring\_mlflow\_production > Run 2c33ae5c3480427ca87eaa6a2f9a648f

Run 2c33ae5c3480427ca87eaa6a2f9a648f

Date: 2024-09-01 23:09:32

Source: airflow

User: root

Status: UNFINISHED

Lifecycle Stage: active

DescriptionEdit

Parameters (23)

Metrics (7)

Tags

Artifacts

LightGBM

MLmodelconda.yamlmodel.pklpython\_env.yamlrequirements.txt

Full Path/home/mlruns/2/2c33ae5c3480427ca87eaa6a2f9a648f/artifacts/LightGBM

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
logged_model = 'runs:/2c33ae5c3480427ca87eaa6a2f9a648f/LightGBM'

# 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()
```

Activate Windows  
Go to Settings to activate Windows.

Lead\_scoring\_mlflow\_production &gt; Run 2c33ae5c3480427ca87eaa6a2f9a648f

## Run 2c33ae5c3480427ca87eaa6a2f9a648f

Date: 2024-09-01 23:09:32

Source:  airflow

User: root




Status: UNFINISHED

Lifecycle Stage: active

▸ Description [Edit](#)

▸ Parameters (23)

▼ Metrics (7)

Name	Value
AUC 	0.611
Accuracy 	0.597
F1 	0.71
Kappa 	0.076
MCC 	0.082
Prec. 	0.637
Recall 	0.802

▸ Tags

▼ Artifacts

▼  LightGBMFull Path:/home/mlruns/2/2c33ae5c3480427ca87eaa6a2f9a648f/artifacts/LightGBM [Register Model](#)

Activate Windows  
Go to Settings to activate Windows.

# Airflow Inference Pipeline

