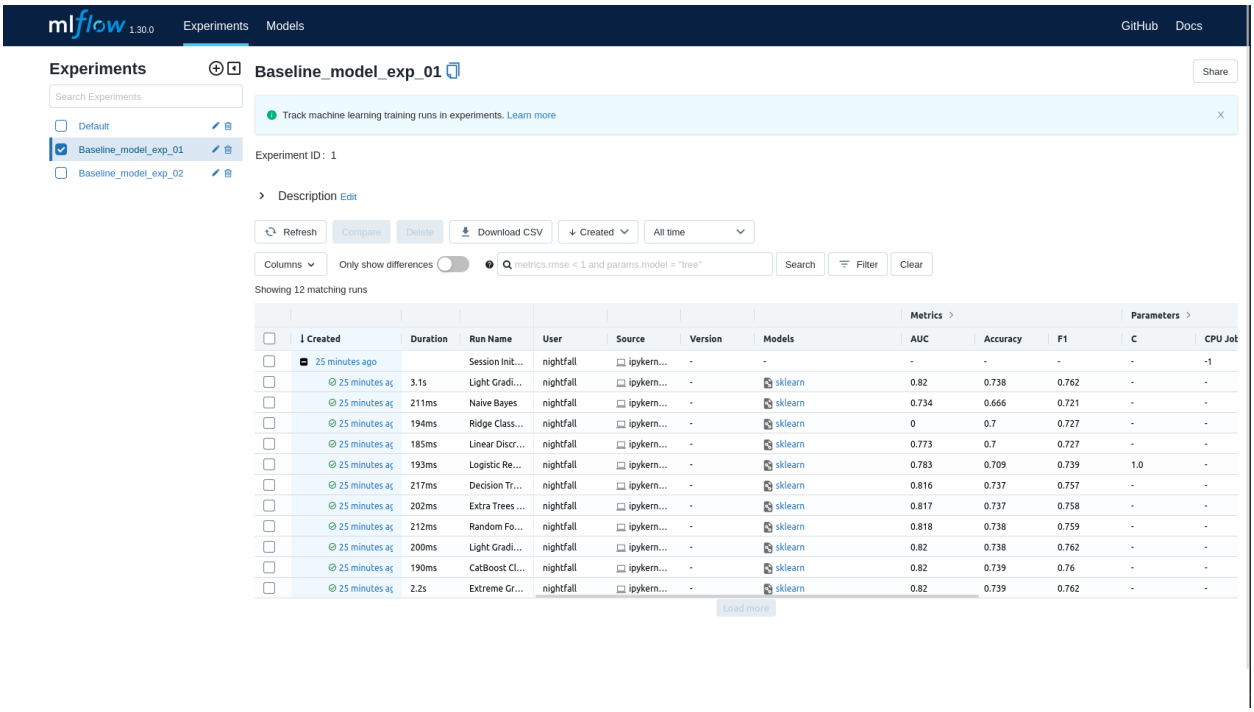
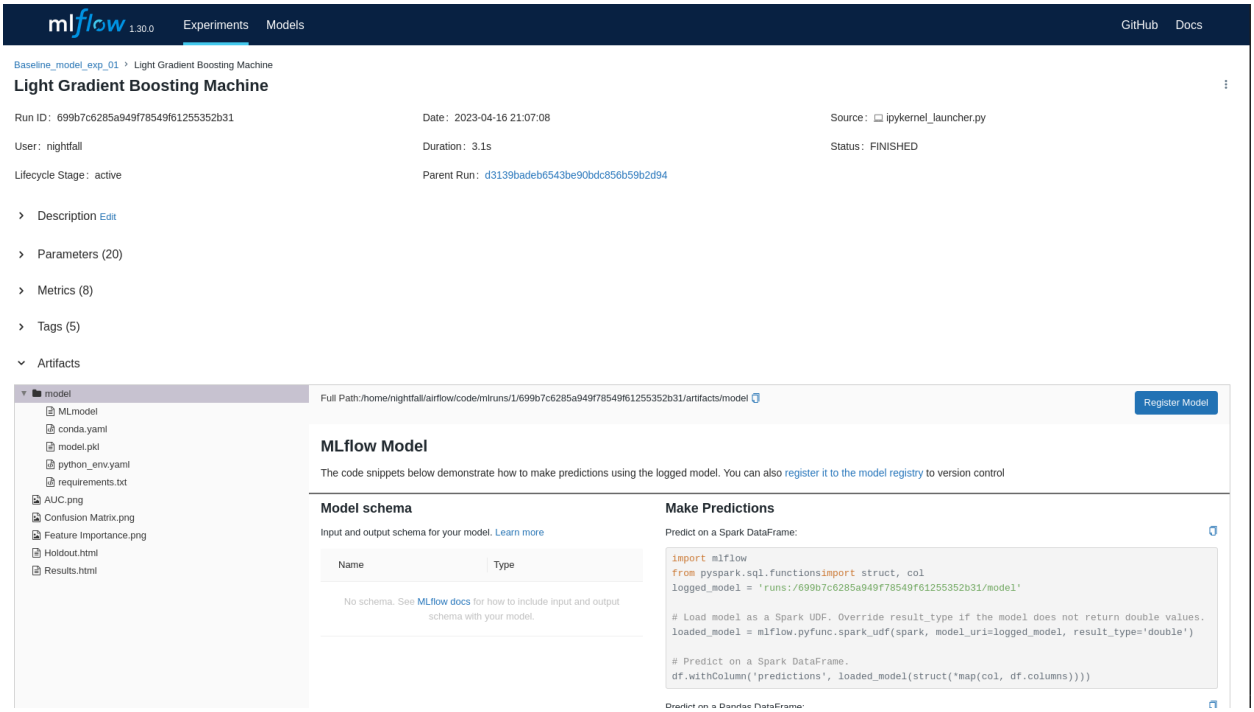
Model Experimentation – **CodePro**

Screenshot of MLflow UI

Screenshot of all the experiments

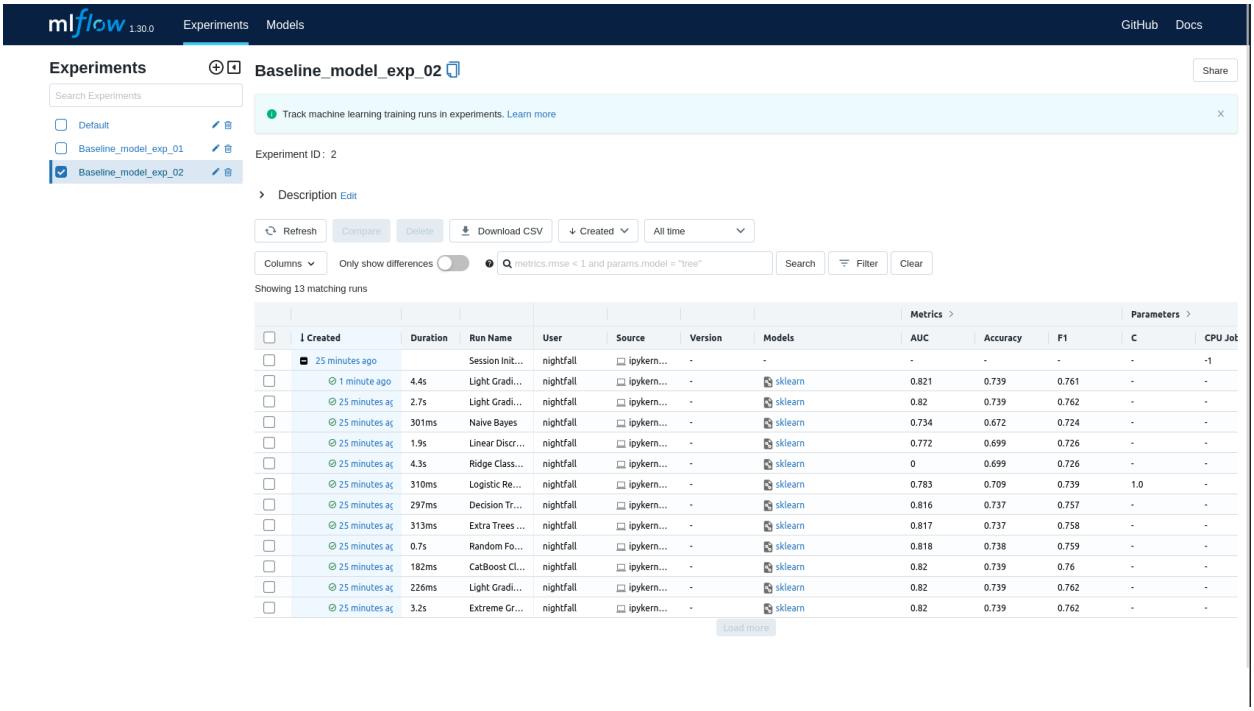


Screenshot of one experiment with all the artifacts visible

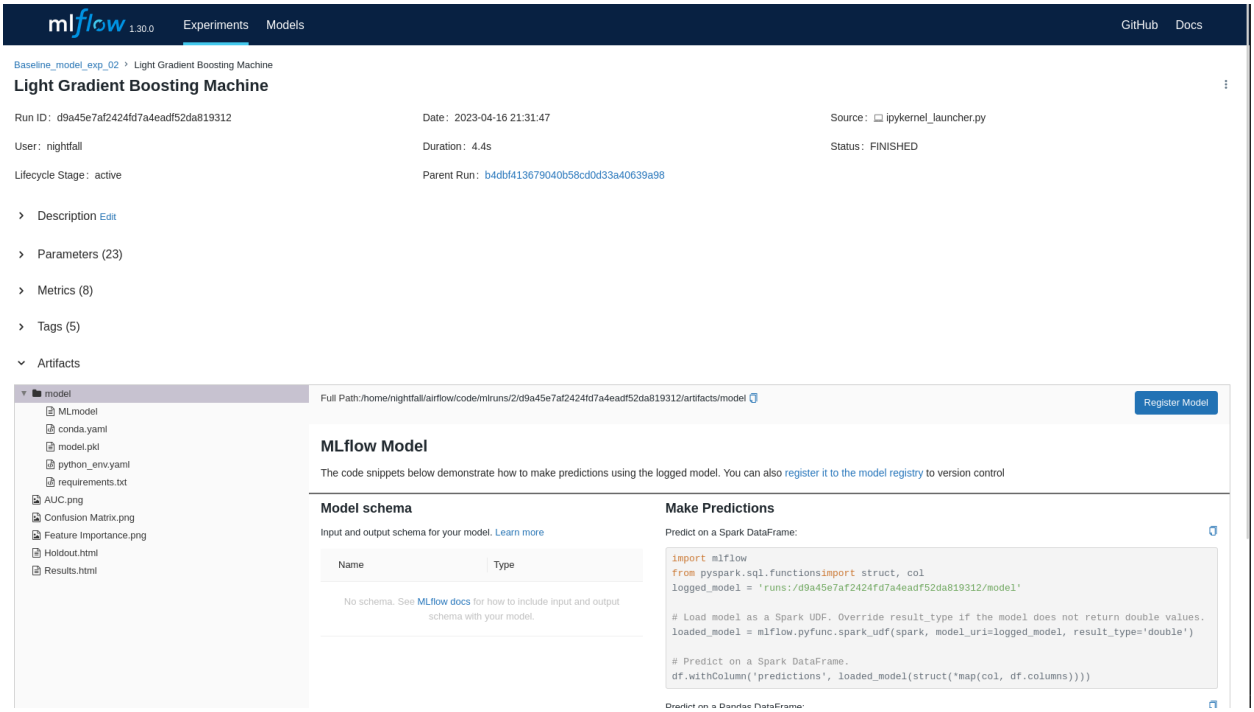

mlflow 1.30.0
Experiments
Models
GitHub
Docs
Baseline_model_exp_01 › Light Gradient Boosting Machine
Light Gradient Boosting Machine
Run ID: 699b7c6285a949f78549f61255352b31
Date: 2023-04-16 21:07:08
Source: ipykernel_launcher.py
User: nightfall
Duration: 3.1s
Status: FINISHED
Lifecycle Stage: active
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.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

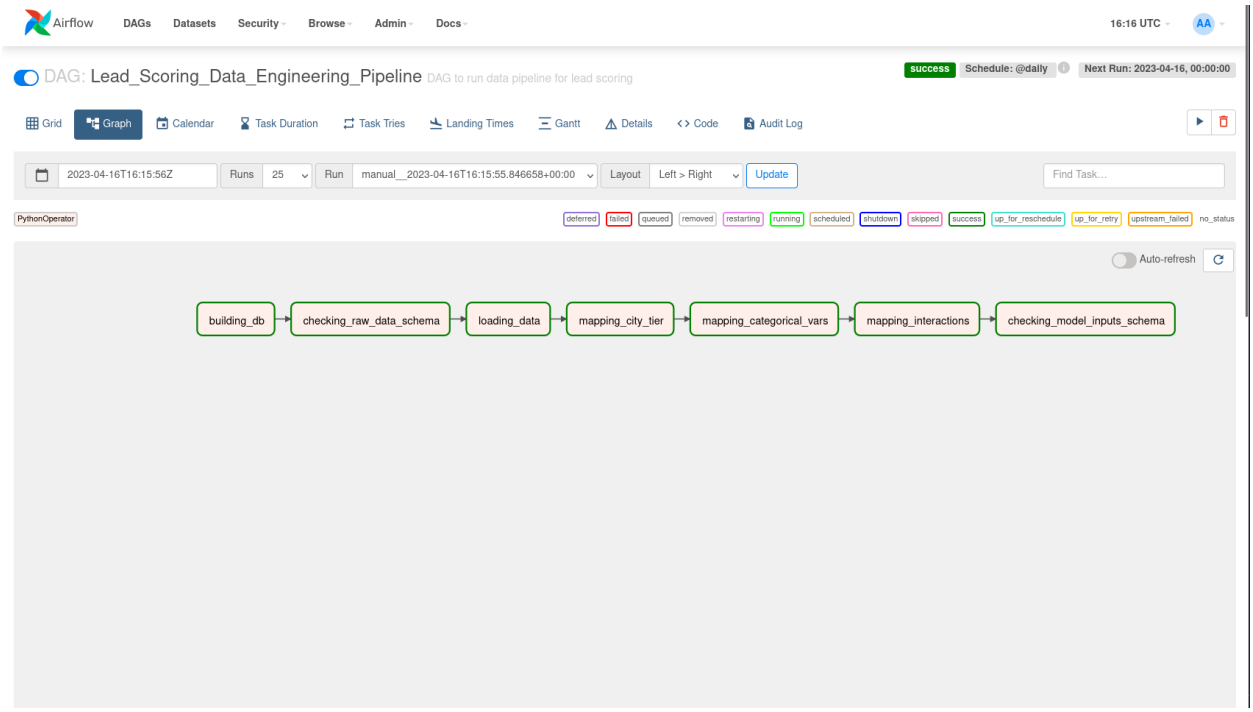


Screenshot of one experiment with all the artifacts visible

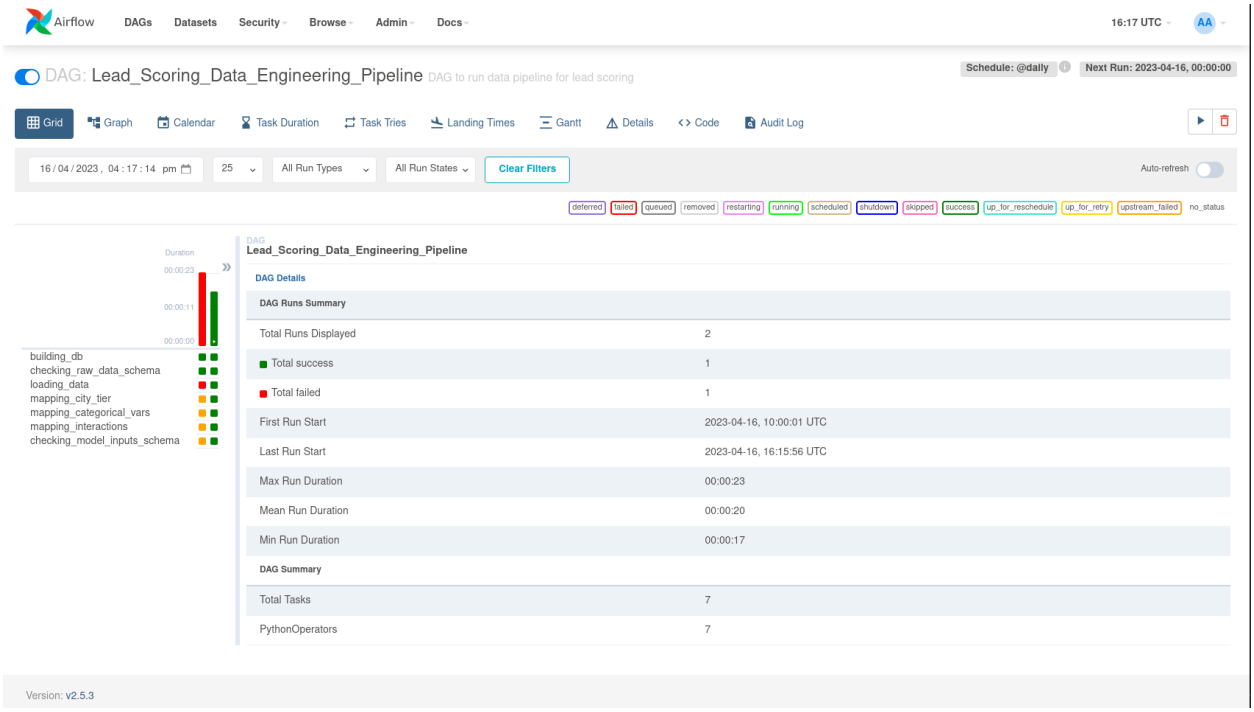

mlflow 1.30.0
Experiments
Models
GitHub
Docs
Baseline_model_exp_02 › Light Gradient Boosting Machine
Light Gradient Boosting Machine
Run ID: d9a45e7af2424fd7a4eadf52da819312
Date: 2023-04-16 21:31:47
Source: ipykernel_launcher.py
User: nightfall
Duration: 4.4s
Status: FINISHED
Lifecycle Stage: active
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.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

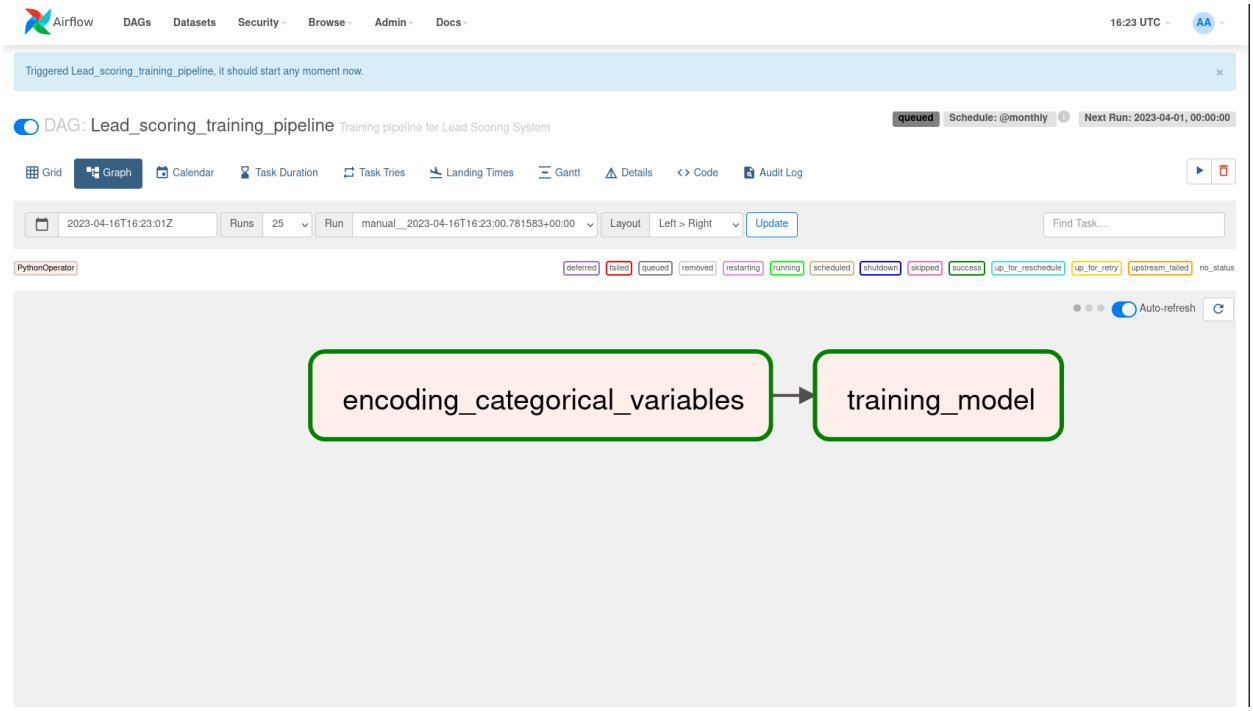


Screenshot of Airflow UI grid

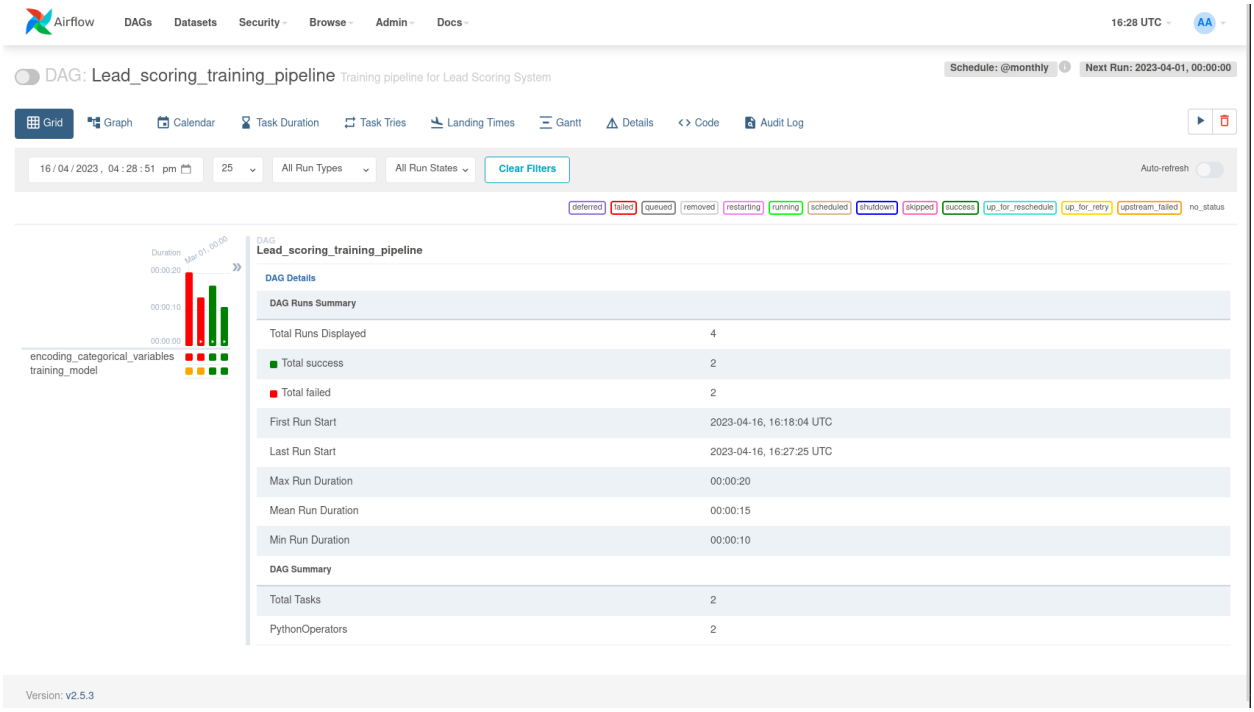

Airflow
DAGs
Datasets
Security
Browse
Admin
Docs
16:17 UTC
AA
DAG: Lead_Scoring_Data_Engineering_Pipeline
DAG to run data pipeline for lead scoring
Schedule: @daily
Next Run: 2023-04-16, 00:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
16 / 04 / 2023 , 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
Duration
00:00:23
00:00:11
00:00:00
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

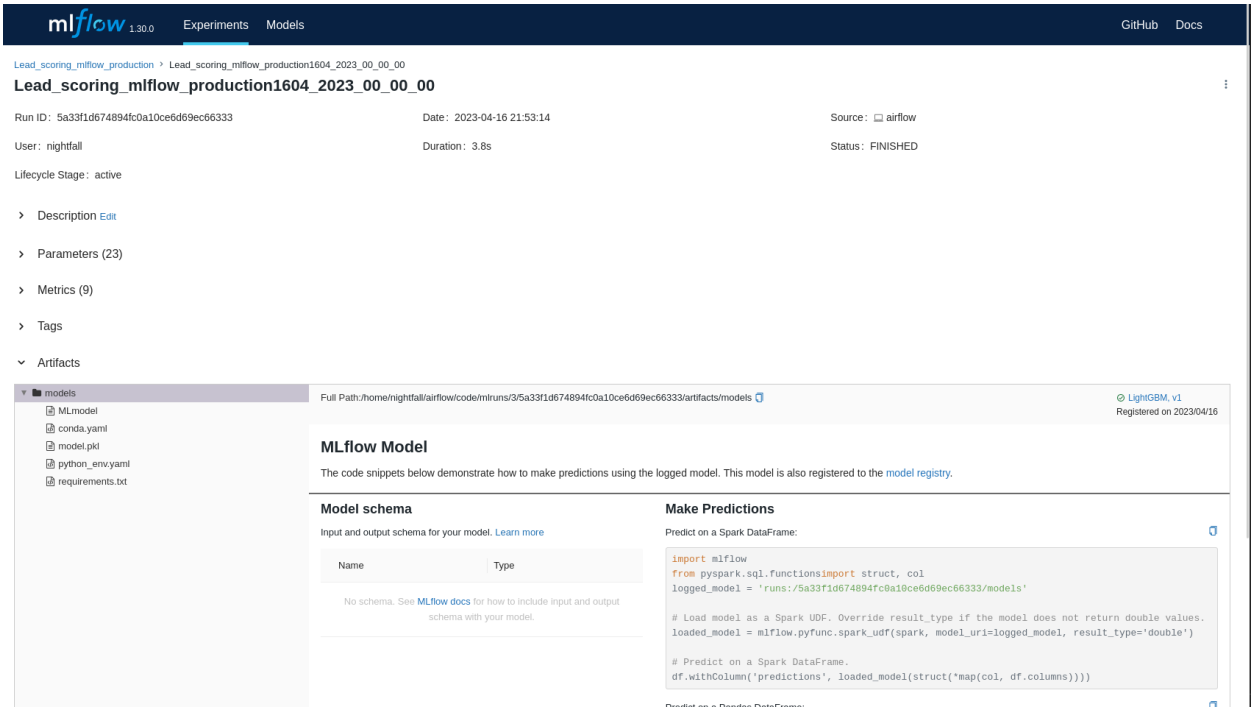


Screenshot of Airflow UI grid

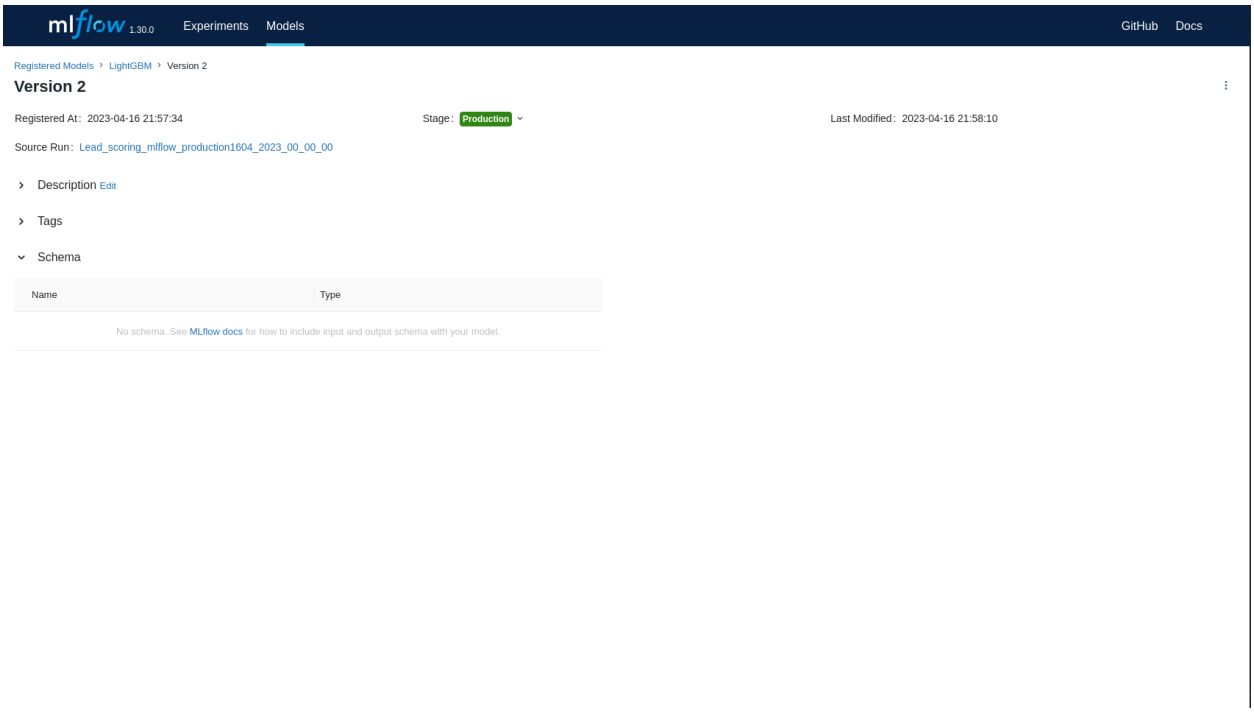

Airflow
DAGs
Datasets
Security
Browse
Admin
Docs
16:28 UTC
AA
DAG: Lead_scoring_training_pipeline
Training pipeline for Lead Scoring System
Schedule: @monthly
Next Run: 2023-04-01, 00:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
16 / 04 / 2023 , 04 : 28 : 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
Duration
Mar 01, 00:00
00:00:20
00:00:10
00:00:00
encoding_categorical_variables
training_model
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

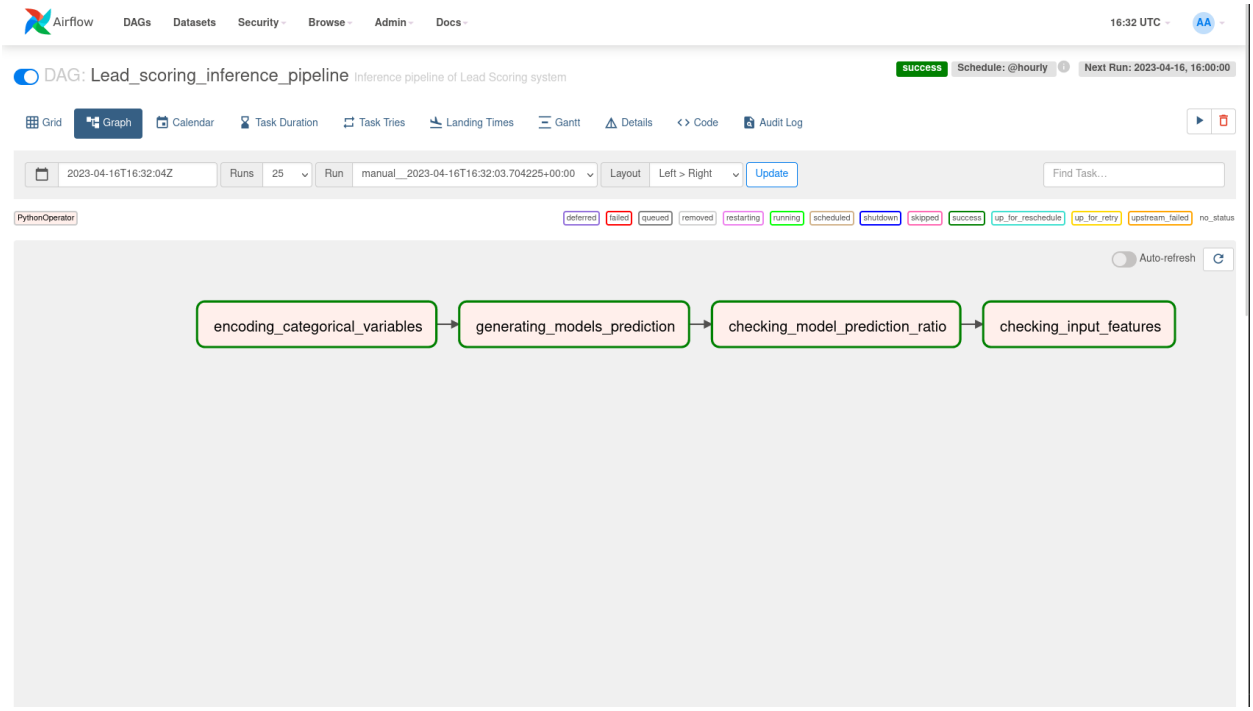


Screenshot of model registry with model name and stage as 'production'



Inference Pipeline

Screenshot of successful execution Airflow DAG in graph



Screenshot of Airflow UI grid