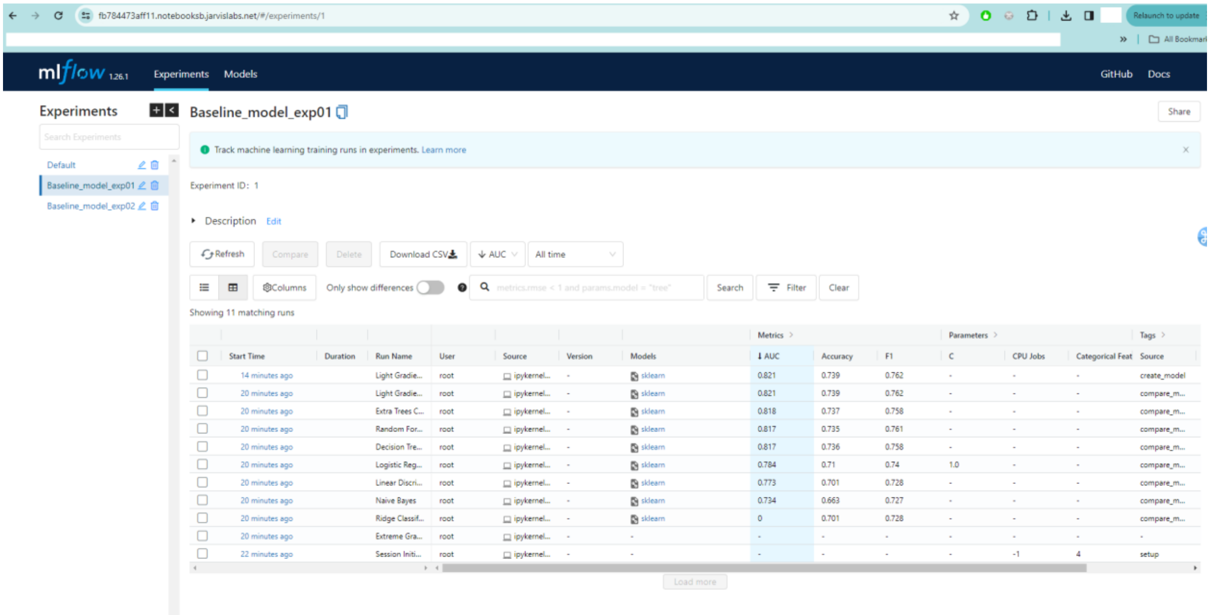
Model Experimentation → screenshot of all the experiments → Baseline_model_exp01



Baseline_model_exp01

Track machine learning training runs in experiments. Learn more

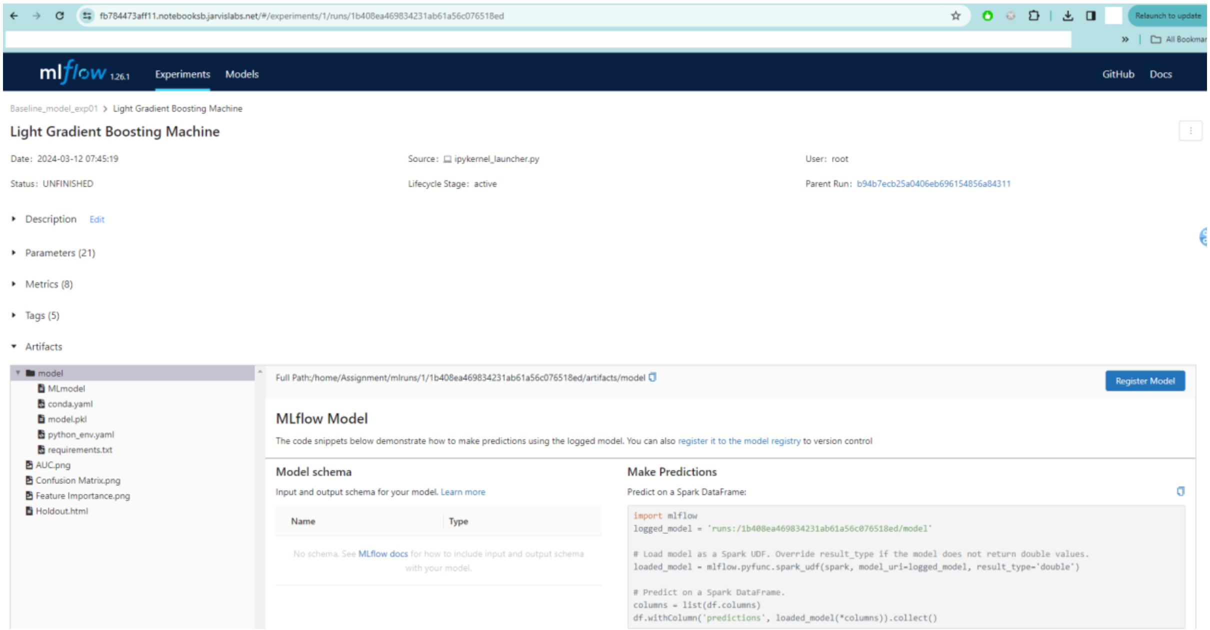
Experiment ID: 1

Description Edit

Showing 11 matching runs

								Metrics >			Parameters >			Tags >
☐	Start Time	Duration	Run Name	User	Source	Version	Models	↓ AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source
☐	14 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	create_model
☐	20 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	compare_m...
☐	20 minutes ago		Extra Trees C...	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	-	compare_m...
☐	20 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.817	0.735	0.761	-	-	-	compare_m...
☐	20 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...
☐	20 minutes ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.784	0.71	0.74	1.0	-	-	compare_m...
☐	20 minutes ago		Linear Discri...	root	ipykernel...	-	sklearn	0.773	0.701	0.728	-	-	-	compare_m...
☐	20 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.663	0.727	-	-	-	compare_m...
☐	20 minutes ago		Ridge Classif...	root	ipykernel...	-	sklearn	0	0.701	0.728	-	-	-	compare_m...
☐	20 minutes ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-
☐	22 minutes ago		Session Initi...	root	ipykernel...	-	-	-	-	-	-	-1	4	setup

Model Experimentation → screenshot of one experiments with all the artifacts visible → Light Gradient Boosting Artifacts under Baseline_model_exp01



Light Gradient Boosting Machine

Date: 2024-03-12 07:45:19

Source: ipykernel_launcher.py

User: root

Status: UNFINISHED

Lifecycle Stage: active

Parent Run: b94b7ecb25a0406eb696154856a84311

Full Path:/home/Assignment/mlruns/1/1b408ea469834231ab61a56c076518ed/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

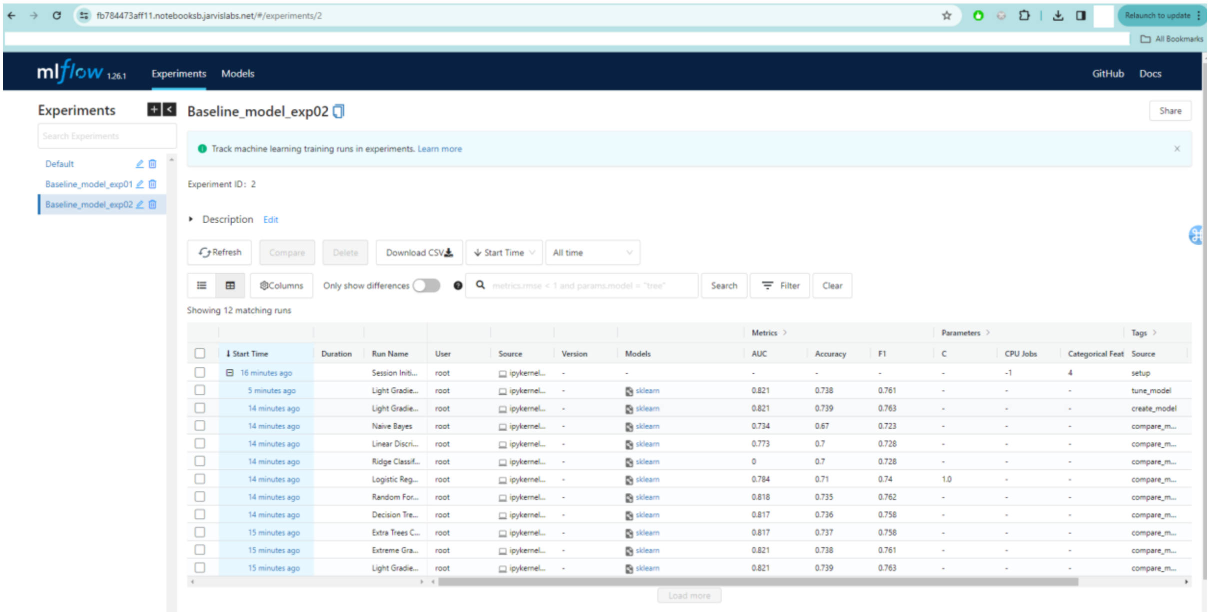
Predict on a Spark DataFrame:

```
import mlflow
logged_model = 'runs:/1b408ea469834231ab61a56c076518ed/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.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
```

Model Experimentation → Screenshot of mlflow ui after dropping features → Baseline_model_exp02



Baseline_model_exp02

Track machine learning training runs in experiments. Learn more

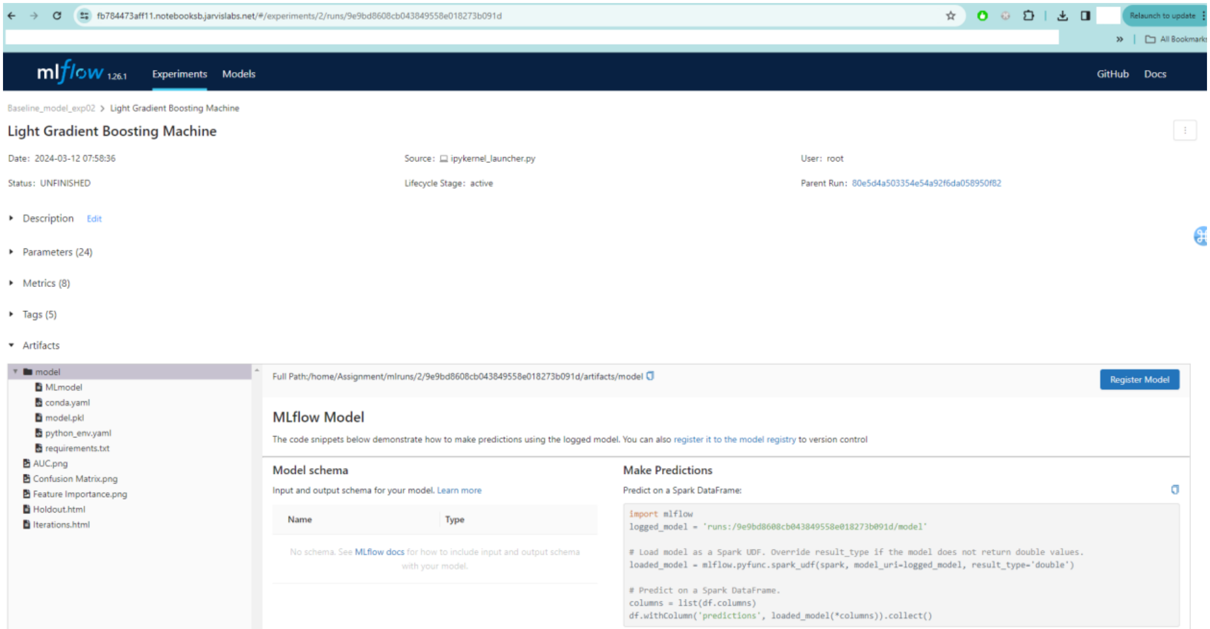
Experiment ID: 2

Description Edit

Showing 12 matching runs

							Metrics			Parameters			Tags
Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source
16 minutes ago		Session Initi...	root	ipykernel...	-	-	-	-	-	-	-1	4	setup
5 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.761	-	-	-	tune_model
14 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.763	-	-	-	create_model
14 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.67	0.723	-	-	-	compare_m...
14 minutes ago		Linear Discri...	root	ipykernel...	-	sklearn	0.773	0.7	0.728	-	-	-	compare_m...
14 minutes ago		Ridge Classif...	root	ipykernel...	-	sklearn	0	0.7	0.728	-	-	-	compare_m...
14 minutes ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.784	0.71	0.74	1.0	-	-	compare_m...
14 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.818	0.735	0.762	-	-	-	compare_m...
14 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...
15 minutes ago		Extra Trees C...	root	ipykernel...	-	sklearn	0.817	0.737	0.758	-	-	-	compare_m...
15 minutes ago		Extreme Gra...	root	ipykernel...	-	sklearn	0.821	0.738	0.761	-	-	-	compare_m...
15 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.763	-	-	-	compare_m...

Model Experimentation → Screenshot of mlflow ui after dropping features → Light Gradient Boosting Artifacts under Baseline_model_exp02


fb784473aff11.notebooksb.jarvislabs.net/#/experiments/2/runs/9e9bd8608cb043849558e018273b091d
Relaunch to update
All Bookmarks
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Baseline_model_exp02 > Light Gradient Boosting Machine
Register Model

Light Gradient Boosting Machine

Date: 2024-03-12 07:58:36

Source: ipykernel_launcher.py

User: root

Status: UNFINISHED

Lifecycle Stage: active

Parent Run: 80e5d4a503354e54a92f6da058950f82

Description Edit

Parameters (24)

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
- Iterations.html

Full Path:/home/Assignment/mlruns/2/9e9bd8608cb043849558e018273b091d/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

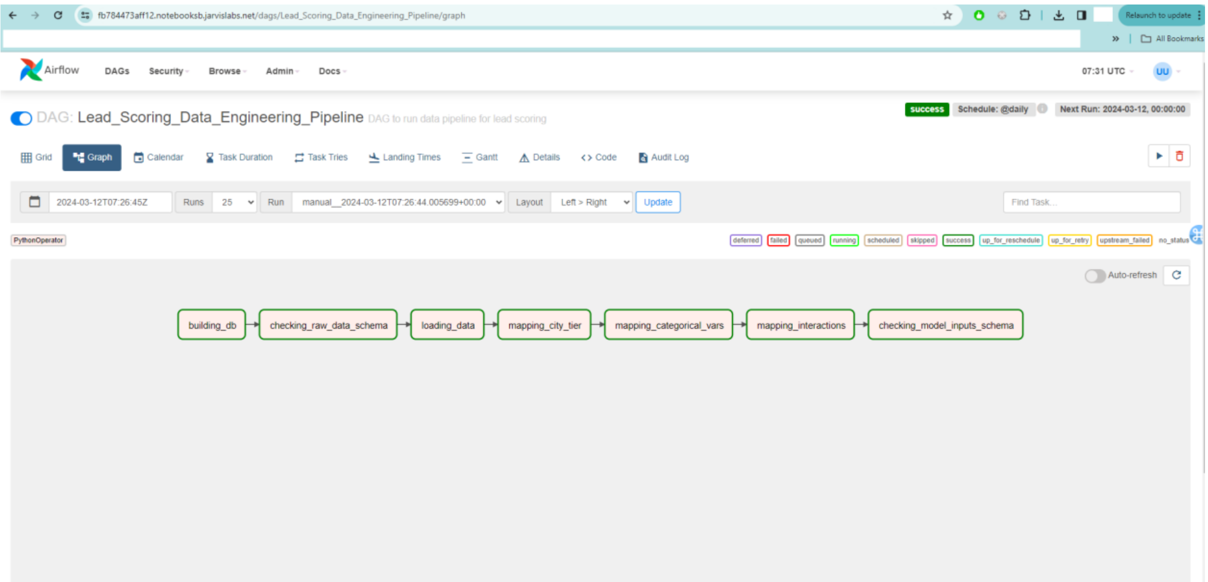
Predict on a Spark DataFrame:

```
import mlflow
logged_model = 'runs:/9e9bd8608cb043849558e018273b091d/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.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
```

Data Pipeline - Create an airflow dag python file for data pipeline - Screenshot of sucessful execution Airflow DAG in graph



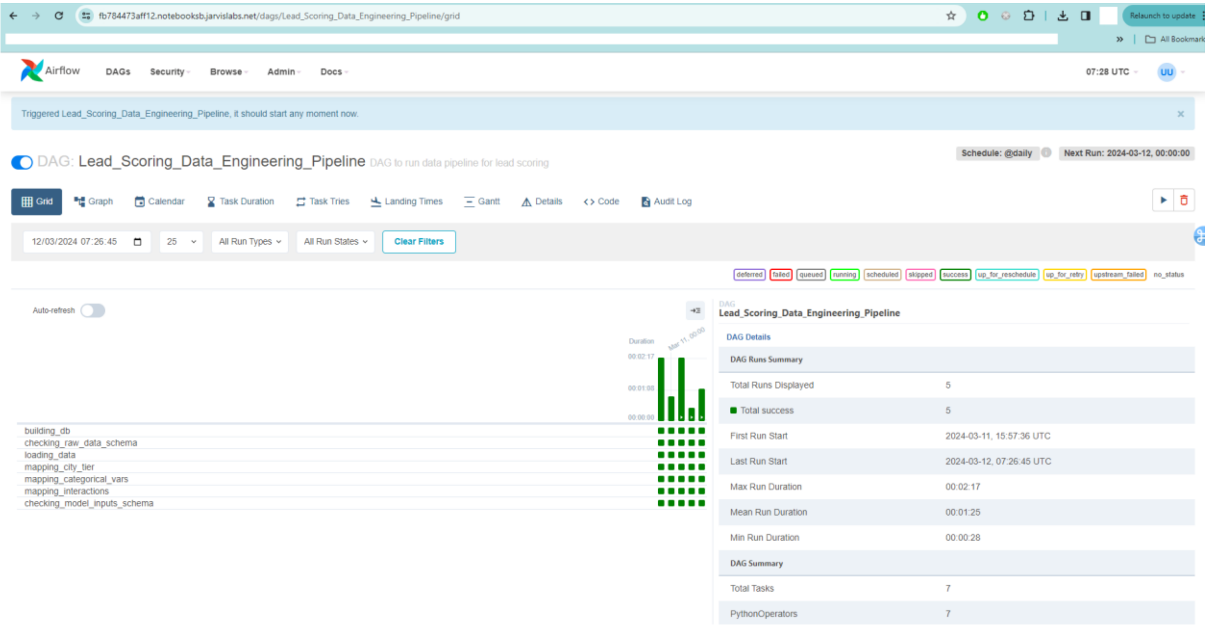
fb784473aff12.notebooksb.jarvislabs.net/dags/Lead_Scoring_Data_Engineering_Pipeline/graph

DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring

success Schedule: @daily Next Run: 2024-03-12, 00:00:00

2024-03-12T07:26:45Z Runs 25 Run manual__2024-03-12T07:26:44.005699+00:00 Layout Left > Right

Data Pipeline - Create an airflow dag python file for data pipeline - Screenshot of Airflow UI grid



fb784473aff12.notebooksb.jarvislabs.net/dags/Lead_Scoring_Data_Engineering_Pipeline/grid

Triggered Lead_Scoring_Data_Engineering_Pipeline, it should start any moment now.

DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring

Schedule: @daily Next Run: 2024-03-12, 00:00:00

12/03/2024 07:26:45 25

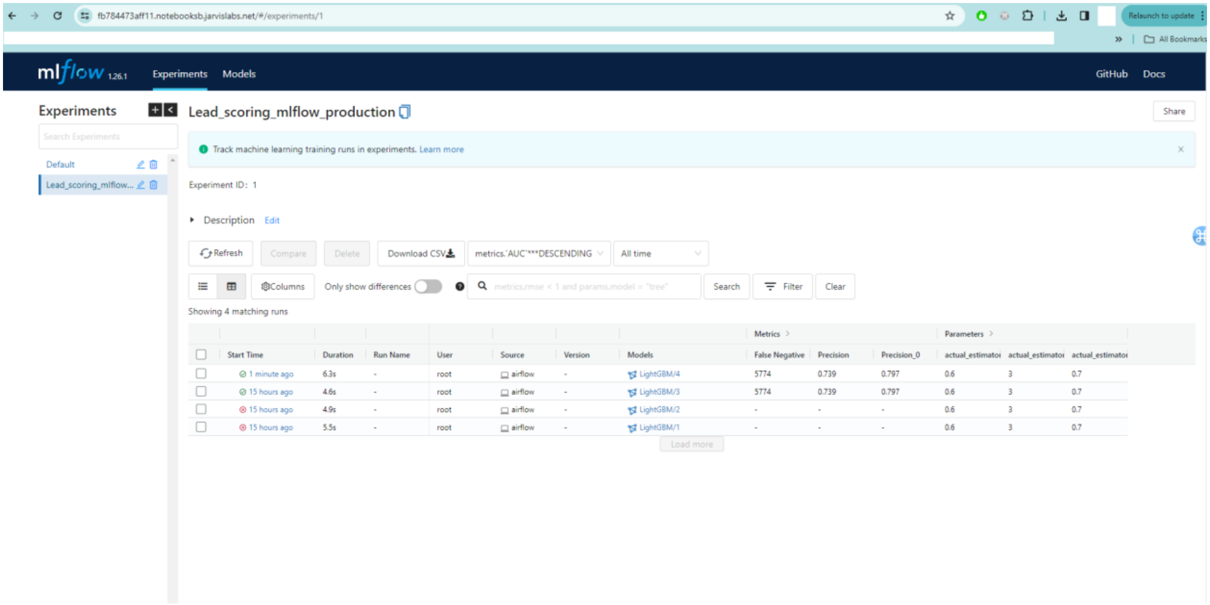
DAG
Lead_Scoring_Data_Engineering_Pipeline

DAG Details

DAG Runs Summary	
Total Runs Displayed	5
🟩 Total success	5
First Run Start	2024-03-11, 15:57:36 UTC
Last Run Start	2024-03-12, 07:26:45 UTC
Max Run Duration	00:02:17
Mean Run Duration	00:01:25
Min Run Duration	00:00:28
DAG Summary	
Total Tasks	7
PythonOperators	7

Small timelines are from LeadScoring_Inference.csv dataset
High ones are on training dataset LeadScoring.csv

Training Pipeline - Create functions for training pipeline (10%) + setup mlflow in utils.py→screenshot of experiments



Lead_scoring_mlflow_production

Track machine learning training runs in experiments. Learn more

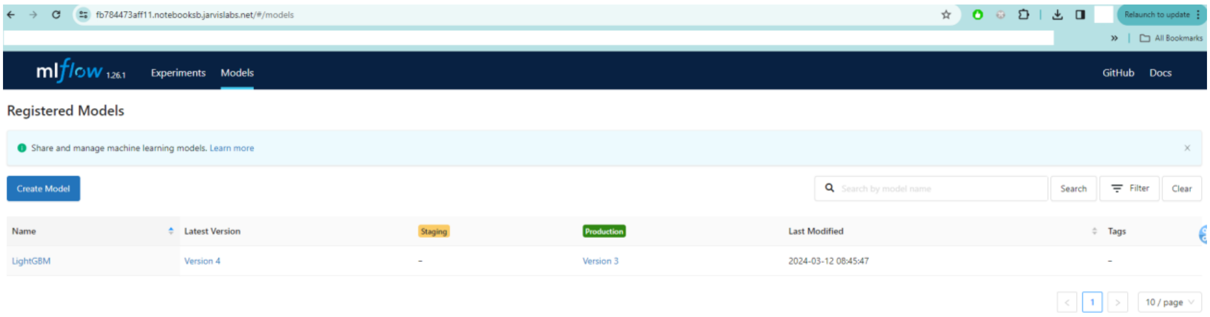
Experiment ID: 1

Description Edit

Showing 4 matching runs

							Metrics			Parameters		
Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision_0	actual_estimator	actual_estimator	actual_estimator
1 minute ago	6.3s	-	root	airflow	-	LightGBM/4	5774	0.739	0.797	0.6	3	0.7
15 hours ago	4.6s	-	root	airflow	-	LightGBM/3	5774	0.739	0.797	0.6	3	0.7
15 hours ago	4.9s	-	root	airflow	-	LightGBM/2	-	-	-	0.6	3	0.7
15 hours ago	5.5s	-	root	airflow	-	LightGBM/1	-	-	-	0.6	3	0.7

Training Pipeline - Create functions for training pipeline (10%) + setup mlflow in utils.py→ screenshot of model registry with model name and stage as 'production'

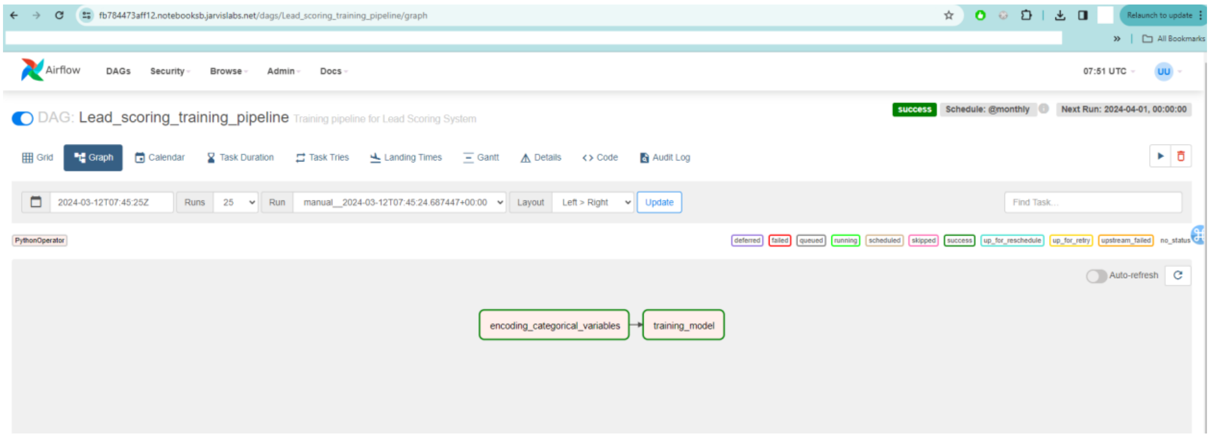


Registered Models

Share and manage machine learning models. Learn more

Name	Latest Version	Staging	Production	Last Modified	Tags
LightGBM	Version 4	-	Version 3	2024-03-12 08:45:47	-

Training Pipeline - Create an airflow dag python file for training pipeline→ Screenshot of sucessful execution Airflow DAG in graph

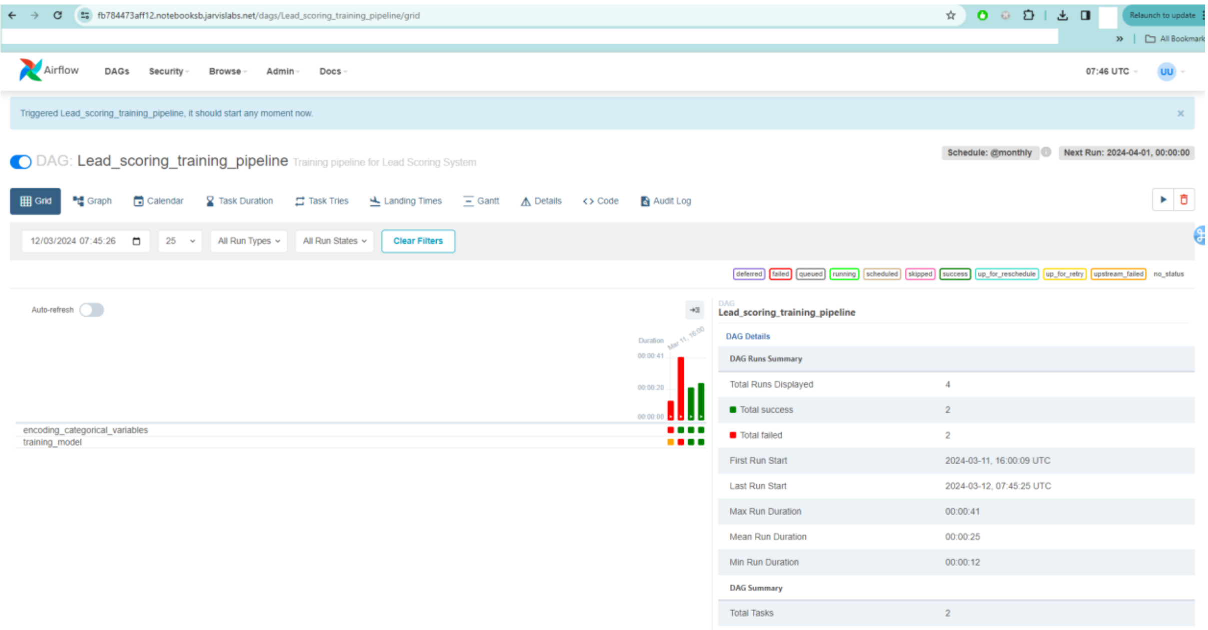


DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System

success Schedule: @monthly Next Run: 2024-04-01, 00:00:00

2024-03-12T07:45:25Z Runs 25 Run manual__2024-03-12T07:45:24.687447+00:00 Layout Left > Right Update

Training Pipeline - Create an airflow dag python file for 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

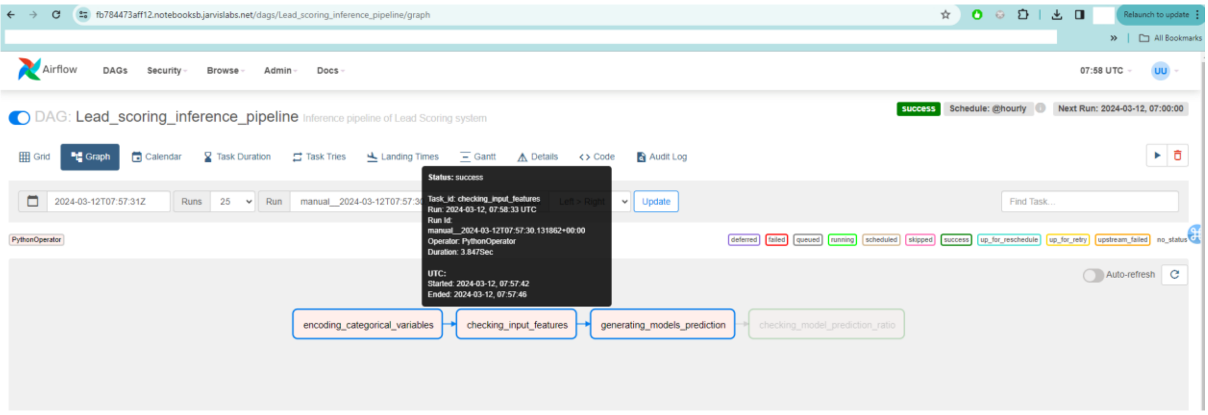
Schedule: @monthly Next Run: 2024-04-01, 00:00:00

DAG
Lead_scoring_training_pipeline

DAG Details

DAG Runs Summary	
Total Runs Displayed	4
🟩 Total success	2
🟥 Total failed	2
First Run Start	2024-03-11, 16:00:09 UTC
Last Run Start	2024-03-12, 07:45:25 UTC
Max Run Duration	00:00:41
Mean Run Duration	00:00:25
Min Run Duration	00:00:12
DAG Summary	
Total Tasks	2

Inference Pipeline - Create an airflow dag python file for inference pipeline → Screenshot of sucessful execution Airflow DAG in graph



DAG: Lead_scoring_inference_pipeline Inference pipeline of Lead Scoring system

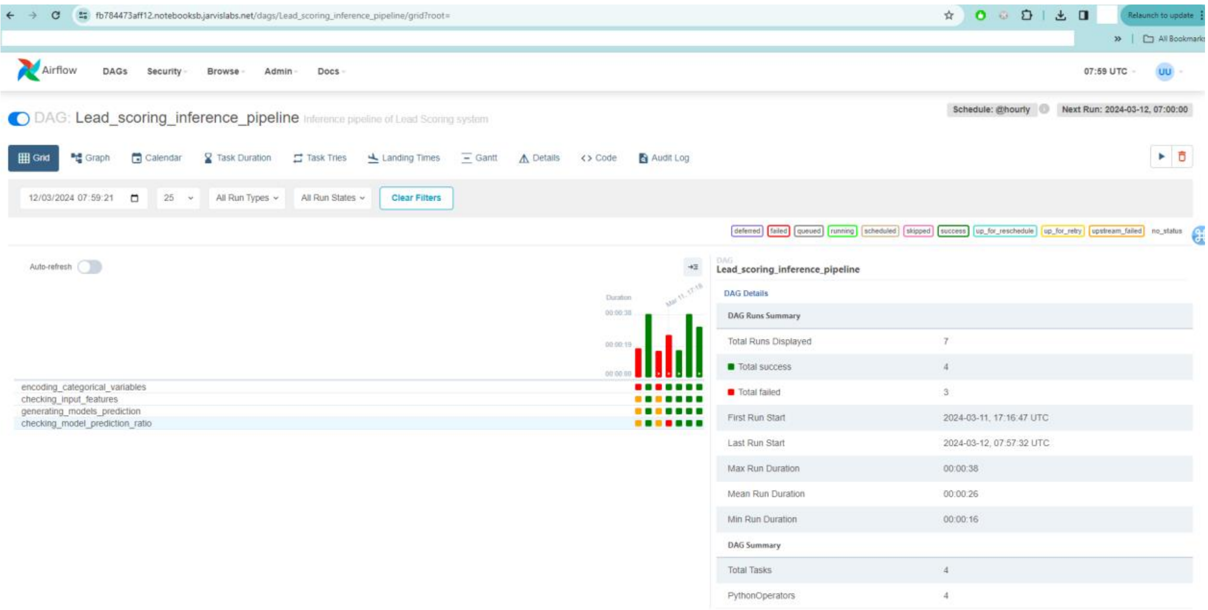
success Schedule: @hourly Next Run: 2024-03-12, 07:00:00

Status: success

Task_id: checking_input_features
Run: 2024-03-12, 07:58:33 UTC
Run Id:
manual__2024-03-12T07:57:30.131862+00:00
Operator: PythonOperator
Duration: 3.847Sec

UTC:
Started: 2024-03-12, 07:57:42
Ended: 2024-03-12, 07:57:46

Inference Pipeline - Create an airflow dag python file for inference pipeline → Screenshot of Airflow UI grid



DAG: Lead_scoring_inference_pipeline Inference pipeline of Lead Scoring system

Schedule: @hourly

Next Run: 2024-03-12, 07:00:00

12/03/2024 07:59:21

encoding_categorical_variables
checking_input_features
generating_models_prediction
checking_model_prediction_ratio

DAG
Lead_scoring_inference_pipeline

DAG Details

DAG Runs Summary	
Total Runs Displayed	7
🟩 Total success	4
🟥 Total failed	3
First Run Start	2024-03-11, 17:16:47 UTC
Last Run Start	2024-03-12, 07:57:32 UTC
Max Run Duration	00:00:38
Mean Run Duration	00:00:26
Min Run Duration	00:00:16
DAG Summary	
Total Tasks	4
PythonOperators	4