

AIRFLOW HOMEPAGE

← → ↻ Not secure | notebooks.jarvislabs.ai:10964/home?status=active#

Airflow DAGs Security Browse Admin Docs 19:21 IST (+05:30) UU

Do not use **SQLite** as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use **SequentialExecutor** in production. [Click here](#) for more information.

DAGs

All 35 Active 3 Paused 32 Filter DAGs by tag Search DAGs

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions
<input checked="" type="checkbox"/> Lead_Scoring_Data_Engineering_Pipeline	airflow	10 3	@daily	2022-10-24, 15:21:08	2022-10-25, 05:30:00		
<input checked="" type="checkbox"/> Lead_scoring_inference_pipeline	airflow	3 1	@hourly	2022-10-25, 17:30:00	2022-10-25, 18:30:00		
<input checked="" type="checkbox"/> Lead_scoring_training_pipeline	airflow	10 4	@monthly	2022-10-24, 17:36:28	2022-11-01, 05:30:00		

Showing 1-3 of 3 DAGs

DATA PIPELINE

← → ↻ Not secure | notebooks.jarvislabs.ai:10964/dags/Lead_Scoring_Data_Engineering_Pipeline/grid

Airflow DAGs Security Browse Admin Docs 19:24 IST (+05:30) UU

☒ DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring Schedule: @daily Next Run: 2022-10-25, 05:30:00

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

25-10-2022 19:22:10 25 All Run Types All Run States Clear Filters

deferred failed queued running scheduled skipped success up_for_reschedule up_for_retry upstream_failed no_status

Auto-refresh

Duration 00:07:36 00:03:48 00:00:00

Oct 21, 01:21 Oct 23, 05:30

building_db
checking_raw_data_schema
loading_data
mapping_city_tier
mapping_categorical_vars
mapping_interactions
checking_model_inputs_schema

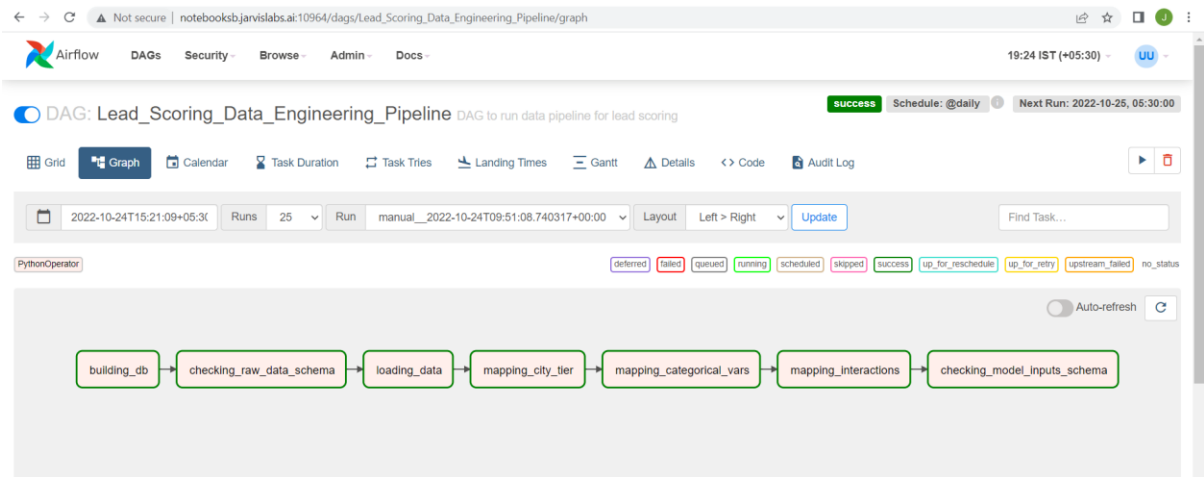
DAG Lead_Scoring_Data_Engineering_Pipeline

DAG Details

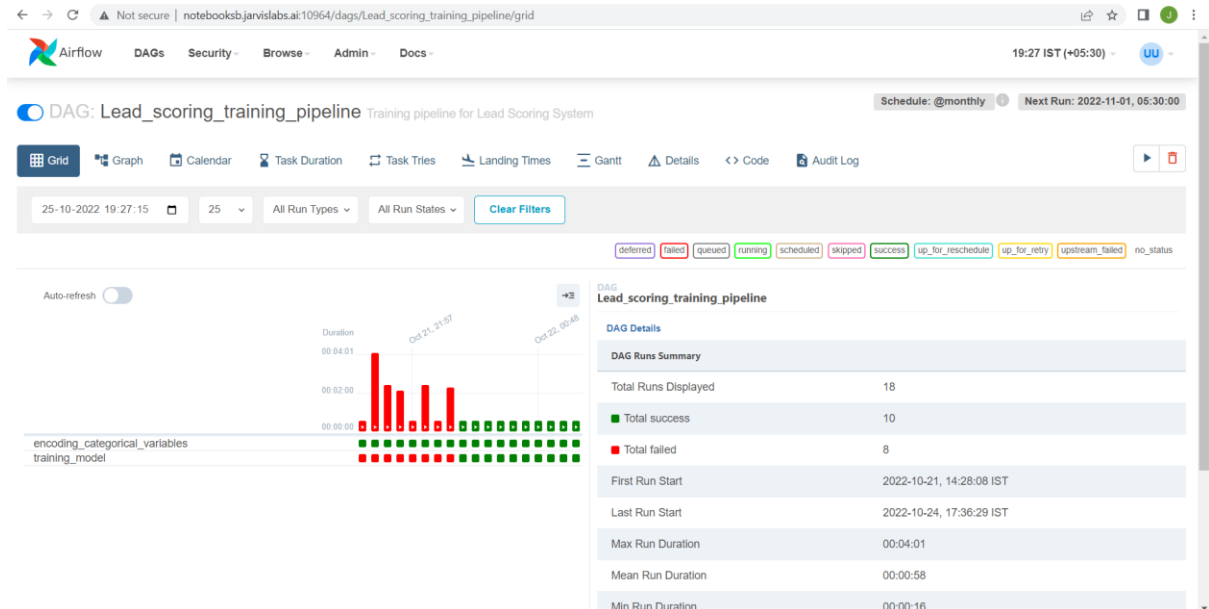
DAG Runs Summary

Total Runs Displayed	17
Total success	15
Total failed	2
First Run Start	2022-10-21, 01:15:25 IST
Last Run Start	2022-10-24, 15:21:09 IST
Max Run Duration	00:07:36
Mean Run Duration	00:01:19
Min Run Duration	00:00:00

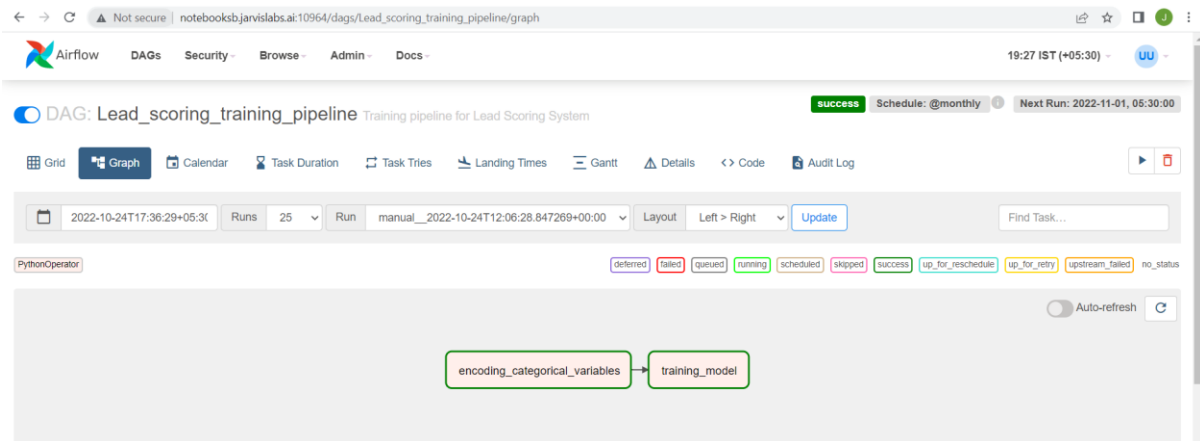
DATA PIPELINE – GRAPH



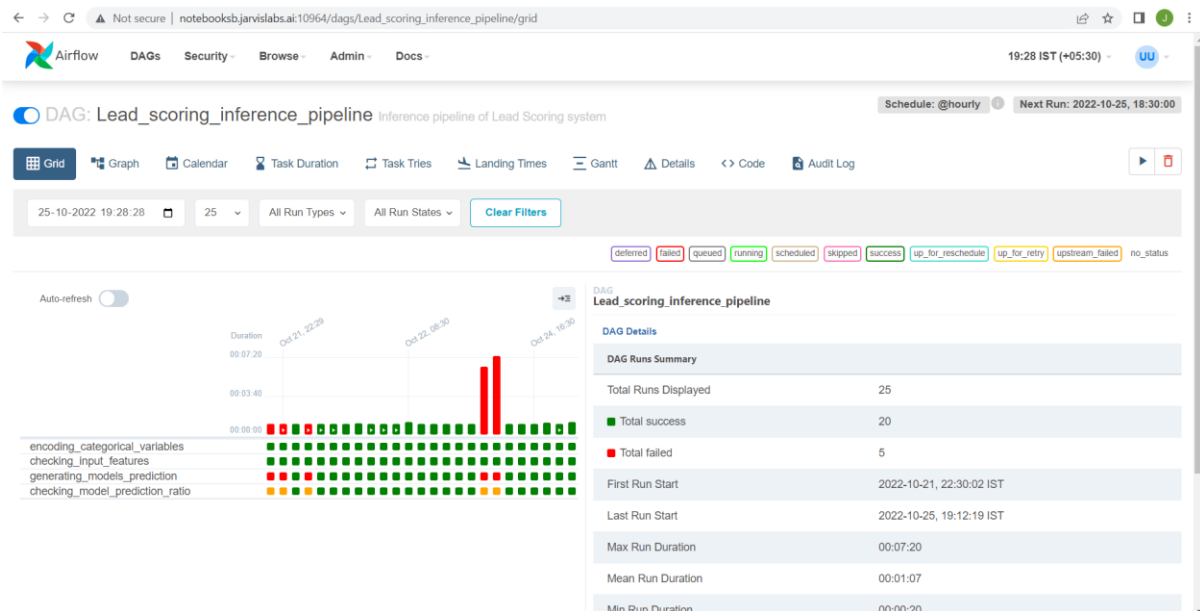
TRAINING PIPELINE



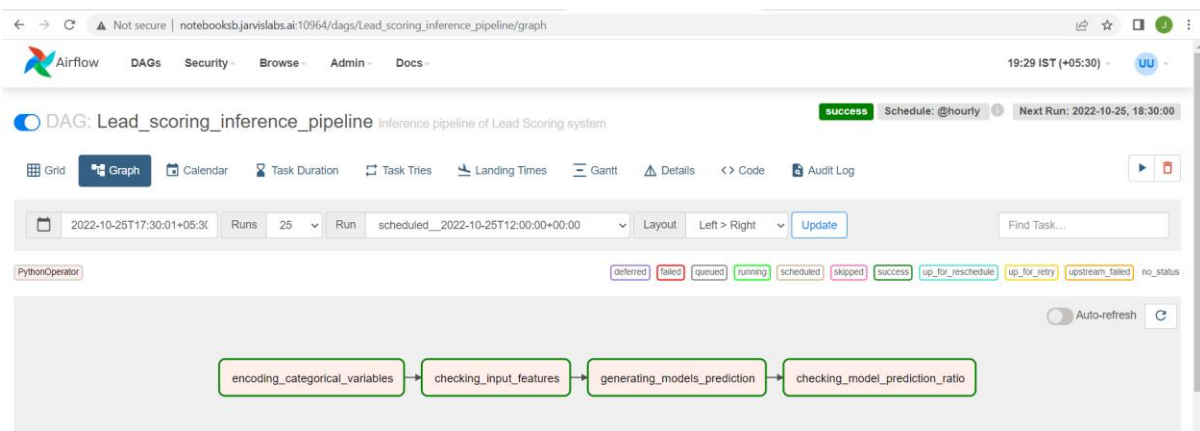
TRAINING PIPELINE – GRAPH



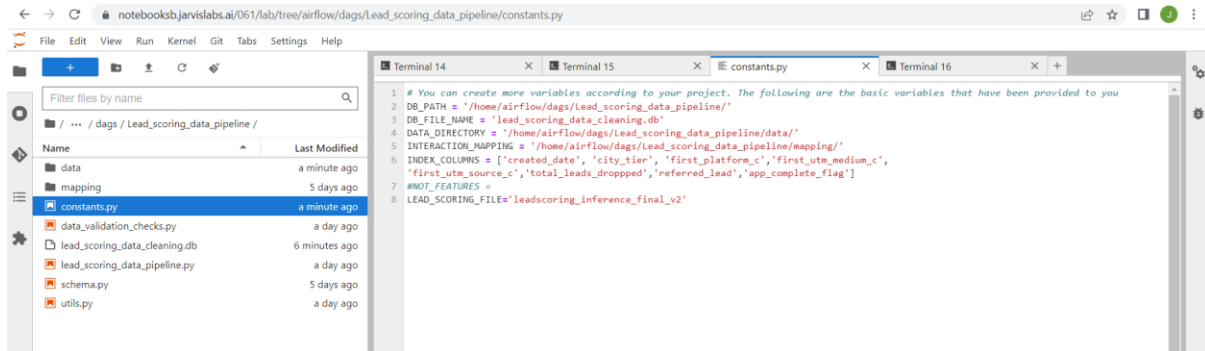
INFERENCE PIPELINE



INFERENCE PIPELINE – GRAPH

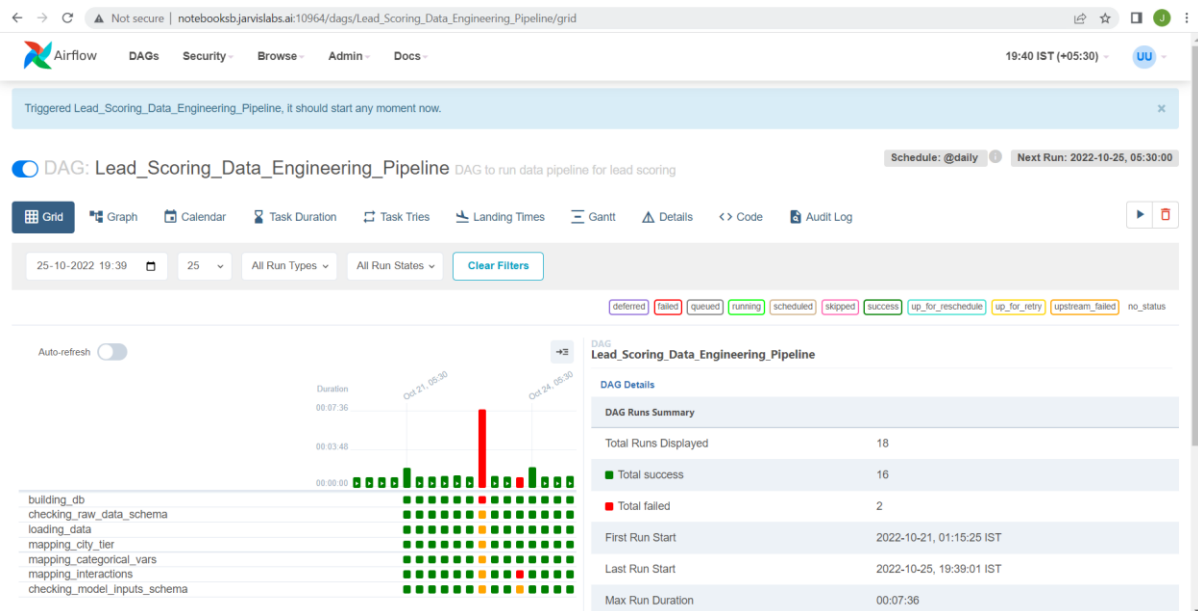


After Making changes in Data pipeline for Inference file

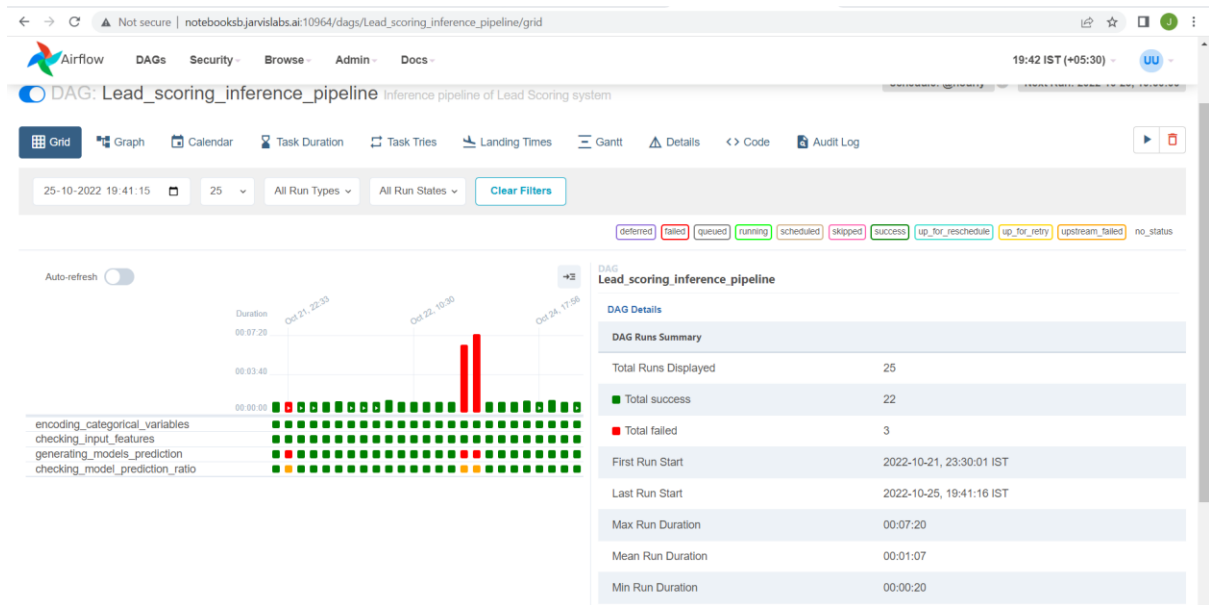


```
1 # You can create more variables according to your project. The following are the basic variables that have been provided to you
2 DB_PATH = '/home/airflow/dags/Lead_scoring_data_pipeline/'
3 DB_FILE_NAME = 'lead_scoring_data_cleaning.db'
4 DATA_DIRECTORY = '/home/airflow/dags/Lead_scoring_data_pipeline/data/'
5 INTERACTION_MAPPING = '/home/airflow/dags/Lead_scoring_data_pipeline/mapping/'
6 INDEX_COLUMNS = ['created_date', 'city_tier', 'first_platform_c', 'first_utm_medium_c',
7 'first_utm_source_c', 'total_leads_dropped', 'referred_lead', 'app_complete_flag']
8 #NOT_FEATURES =
9 LEAD_SCORING_FILE='leadscoring_inference_final_v2'
```

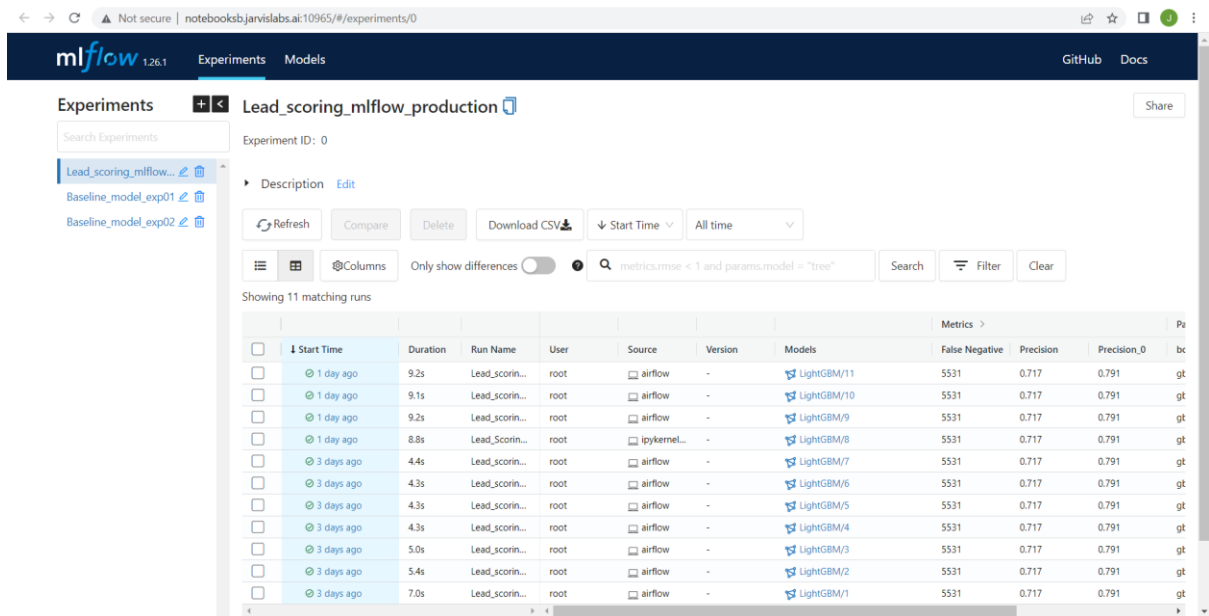
DATA PIPELINE



DATA INFERENCE



ML FLOW Server



← → ↻ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/1

mlflow 1.26.1 Experiments Models GitHub Docs

Experiments +< Baseline_model_exp01

Search Experiments

Lead_scoring_mlflow... Baseline_model_exp01 Baseline_model_exp02

Experiment ID: 1

Description Edit

Refresh Compare Delete Download CSV Start Time All time

Columns Only show differences metrics.rmse < 1 and params.model = "tree" Search Filter Clear

Showing 20 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1
<input type="checkbox"/>	1 day ago		Session Initialized 3067	root	ipykernel...	-	-	-	-	-
<input type="checkbox"/>	1 day ago		Light Gradient Boostin...	root	ipykernel...	-	sklearn	0.821	0.739	0.762
<input type="checkbox"/>	1 day ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.663	0.727
<input type="checkbox"/>	1 day ago		Linear Discriminant An...	root	ipykernel...	-	sklearn	0.773	0.701	0.728
<input type="checkbox"/>	1 day ago		Ridge Classifier	root	ipykernel...	-	sklearn	0	0.701	0.728
<input type="checkbox"/>	1 day ago		Logistic Regression	root	ipykernel...	-	sklearn	0.784	0.71	0.74
<input type="checkbox"/>	1 day ago		Random Forest Classifier	root	ipykernel...	-	sklearn	0.817	0.735	0.761
<input type="checkbox"/>	1 day ago		Decision Tree Classifier	root	ipykernel...	-	sklearn	0.817	0.736	0.758
<input type="checkbox"/>	1 day ago		Extra Trees Classifier	root	ipykernel...	-	sklearn	0.818	0.737	0.758
<input type="checkbox"/>	1 day ago		Light Gradient Boostin...	root	ipykernel...	-	sklearn	0.821	0.739	0.762
<input type="checkbox"/>	1 day ago		Extreme Gradient Boos...	root	ipykernel...	-	-	-	-	-
<input type="checkbox"/>	1 day ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.663	0.727

Registered Model

← → ↻ Not secure | notebooks.jarvislabs.ai:10965/#/models

mlflow 1.26.1 Experiments Models GitHub Docs

Registered Models

Share and manage machine learning models. Learn more

Create Model

Search by model name Search Filter Clear

Name	Latest Version	Staging	Production	Last Modified	Tags
LightGBM	Version 11	-	Version 7	2022-10-24 17:36:49	-

< 1 > 10 / page

ML Flow Model

Full Path: /home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models

LightGBM v7 Registered on 2022/10/22

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
logged_model = 'runs:/d190715a8eed42778a90ea59c21e5b56/models'

# 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()
```

Predict on a Pandas DataFrame:

```
import mlflow
logged_model = 'runs:/d190715a8eed42778a90ea59c21e5b56/models'

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)

# Predict on a Pandas DataFrame.
import pandas as pd
loaded_model.predict(pd.DataFrame(data))
```

← → ↻ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

mlflow 1.26.1 Experiments Models GitHub Docs

Lead_scoring_mlflow_production > Lead_scoring_mlflow_production

Lead_scoring_mlflow_production

Date: 2022-10-22 00:49:09 Source: airflow User: root
Duration: 4.4s Status: FINISHED Lifecycle Stage: active

► Description [Edit](#)

► Parameters (20)

► Metrics (11)

► Tags

▼ Artifacts

► models

Full Path: /home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models [🔗](#) lightGBM-v7 Registered on 2022/10/22

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
logged_model = 'runs:/d190715a8eed42778a90ea59c21e5b56/models'

# 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')
```

Metrics

← → ↻ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

mlflow 1.26.1 Experiments Models GitHub Docs

Lead_scoring_mlflow_production > Lead_scoring_mlflow_production

Lead_scoring_mlflow_production

Date: 2022-10-22 00:49:09 Source: airflow User: root
Duration: 4.4s Status: FINISHED Lifecycle Stage: active

► Description [Edit](#)

► Parameters (20)

▼ Metrics (11)

Name	Value
False Negative 🔗	5531
Precision 🔗	0.717
Precision_0 🔗	0.791
Precision_1 🔗	0.674
Recall 🔗	0.732
Recall_0 🔗	0.588
Recall_1 🔗	0.846
True Negative 🔗	20946
f1_0 🔗	0.675
f1_1 🔗	0.75
test_accuracy 🔗	0.717

Parameters

← → ↻ 🔒 Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56 🔍 📌 ⭐ 📄 🟢 ⋮

Parameters (20)

Name	Value
boosting_type	gbdt
class_weight	None
colsample_bytree	1.0
importance_type	split
learning_rate	0.1
max_depth	-1
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	100
n_jobs	-1
num_leaves	31
objective	None
random_state	42
reg_alpha	0.0
reg_lambda	0.0
silent	warn
subsample	1.0
subsample_for_bin	200000
subsample_freq	0

← → ↻ 🔒 Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56 🔍 📌 ⭐ 📄 🟢 ⋮

mlflow 1.26.1 Experiments Models GitHub Docs

Lead_scoring_mlflow_production > Lead_scoring_mlflow_production

Lead_scoring_mlflow_production

Date: 2022-10-22 00:49:09

Duration: 4.4s

Source: airflow

Status: FINISHED

User: root

Lifecycle Stage: active

Description Edit

Parameters (20)

Name	Value
boosting_type	gbdt
class_weight	None
colsample_bytree	1.0
importance_type	split
learning_rate	0.1
max_depth	-1
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	100
n_jobs	-1
num_leaves	31
objective	None
random_state	42
reg_alpha	0.0

Artifacts

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

models

MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt

Full Path:/home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models

LightGBM, v7
Registered on 2022/10/22

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
logged_model = 'runs:/d190715a8eed42778a90ea59c21e5b56/models'

# 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()
```

Predict on a Pandas DataFrame:

```
import mlflow
logged_model = 'runs:/d190715a8eed42778a90ea59c21e5b56/models'

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)

# Predict on a Pandas DataFrame.
import pandas as pd
loaded_model.predict(pd.DataFrame(data))
```

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

Artifacts

models

MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt

Full Path:/home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models/MLmodel

Size: 434B

```
artifact_path: models
flavors:
  python_function:
    env: conda.yaml
    loader_module: mlflow.sklearn
    model_path: model.pkl
    python_version: 3.8.12
  sklearn:
    code: null
    pickled_model: model.pkl
    serialization_format: cloudpickle
    sklearn_version: 0.23.2
mlflow_version: 1.26.1
model_uuid: 92b0df32f2ed4671a58ff5d18273bddd
run_id: d190715a8eed42778a90ea59c21e5b56
utc_time_created: '2022-10-21 19:19:10.329334'
```

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

Artifacts

models

MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt

Full Path:/home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models/conda.yaml

Size: 448B

```
channels:
- conda-forge
dependencies:
- python=3.8.12
- pip<=21.2.4
- pip:
- mlflow
- cffi==1.15.0
- cloudpickle==2.1.0
- colorama==0.4.4
- configparser==5.2.0
- cupy-cuda115==9.6.0
- cython==0.29.27
- defusedxml==0.7.1
- distributed==2021.11.2
- graphviz==0.20.1
- ipython==8.0.1
- lightgbm==3.3.2
- matplotlib==3.5.1
- pynvml==11.4.1
- pytest==6.2.5
- scikit-learn==0.23.2
- setproctitle==1.3.1
name: mlflow-env
```

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

▼ Artifacts

models

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

Full Path:/home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models/python_env.yaml

Size: 122B

python: 3.8.12
build_dependencies:
- pip==21.2.4
- setuptools==59.5.0
- wheel==0.37.0
dependencies:
- -r requirements.txt

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/0/runs/d190715a8eed42778a90ea59c21e5b56

▼ Artifacts

models

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

Full Path:/home/mlruns/0/d190715a8eed42778a90ea59c21e5b56/artifacts/models/requirements.txt

Size: 287B

mflow
cffi==1.15.0
cloudpickle==2.1.0
colorama==0.4.4
configparser==5.2.0
cupy-cuda115==9.6.0
cython==0.29.27
defusedxml==0.7.1
distributed==2021.11.2
graphviz==0.20.1
ipython==8.0.1
lightgbm==3.3.2
matplotlib==3.5.1
pyyaml==11.4.1
pytest==6.2.5
scikit-learn==0.23.2
setproctitle==1.3.1

Model: Light Gradient Boosting Machine

← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/1/runs/8a34f6a4dca94a3eb79d30eb72ccbc2b

▼ Artifacts

model

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

AUC.png

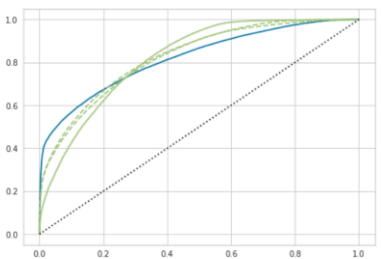
Confusion Matrix.png

Feature Importance.png

Holdout.html

Full Path:/home/mlruns/1/8a34f6a4dca94a3eb79d30eb72ccbc2b/artifacts/AUC.png

Size: 27.61KB



← → ↻ ⚠ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/1/runs/8a34f6a4dca94a3eb79d30eb72ccbc2b

▼ Artifacts

model

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

AUC.png

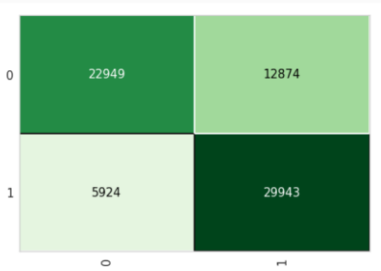
Confusion Matrix.png

Feature Importance.png

Holdout.html

Full Path:/home/mlruns/1/8a34f6a4dca94a3eb79d30eb72ccbc2b/artifacts/Confusion Matrix.png

Size: 7.47KB



	0	1
0	22949	12874
1	5924	29943



← → ↻ Not secure | notebooks.jarvislabs.ai:10965/#/experiments/1/runs/8a34f6a4dca94a3eb79d30eb72ccbc2b

Artifacts

Full Path: /home/mlruns/1/8a34f6a4dca94a3eb79d30eb72ccbc2b/artifacts/Holdout.html
Size: 775B

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Light Gradient Boosting Machine	0.7378	0.8198	0.8348	0.6993	0.7611	0.4755	0.4847