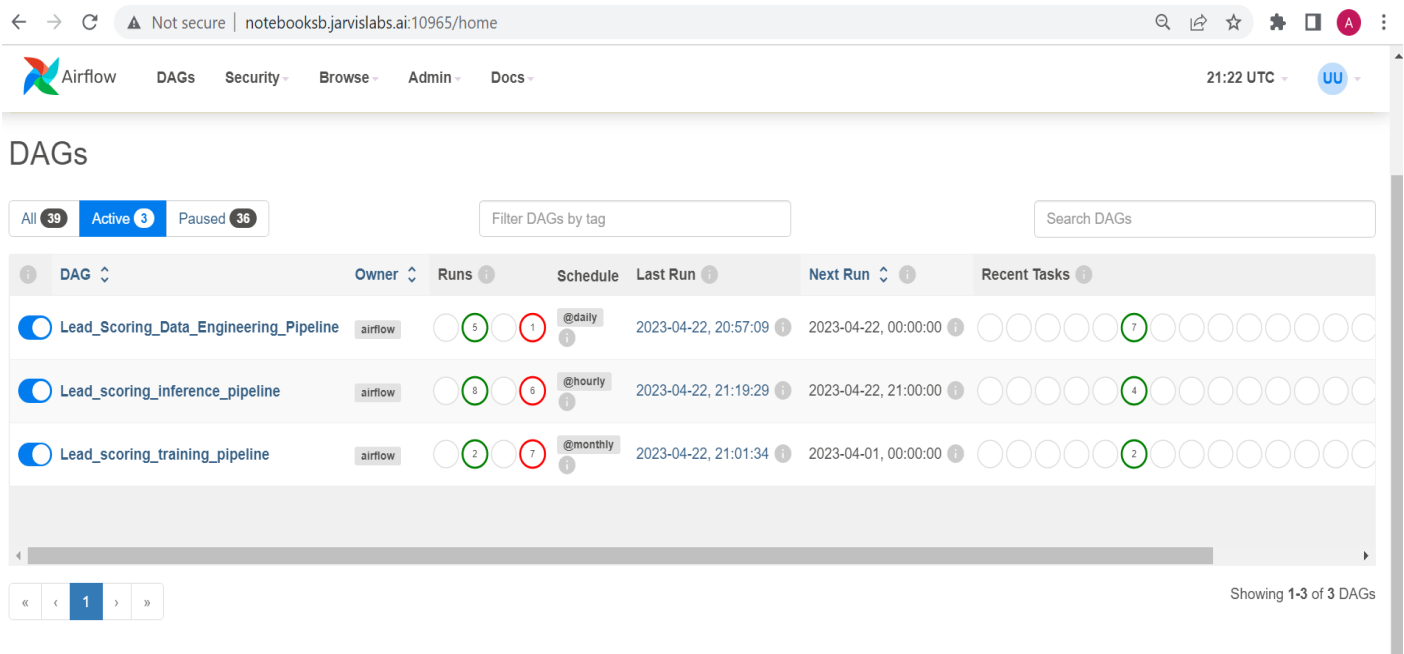
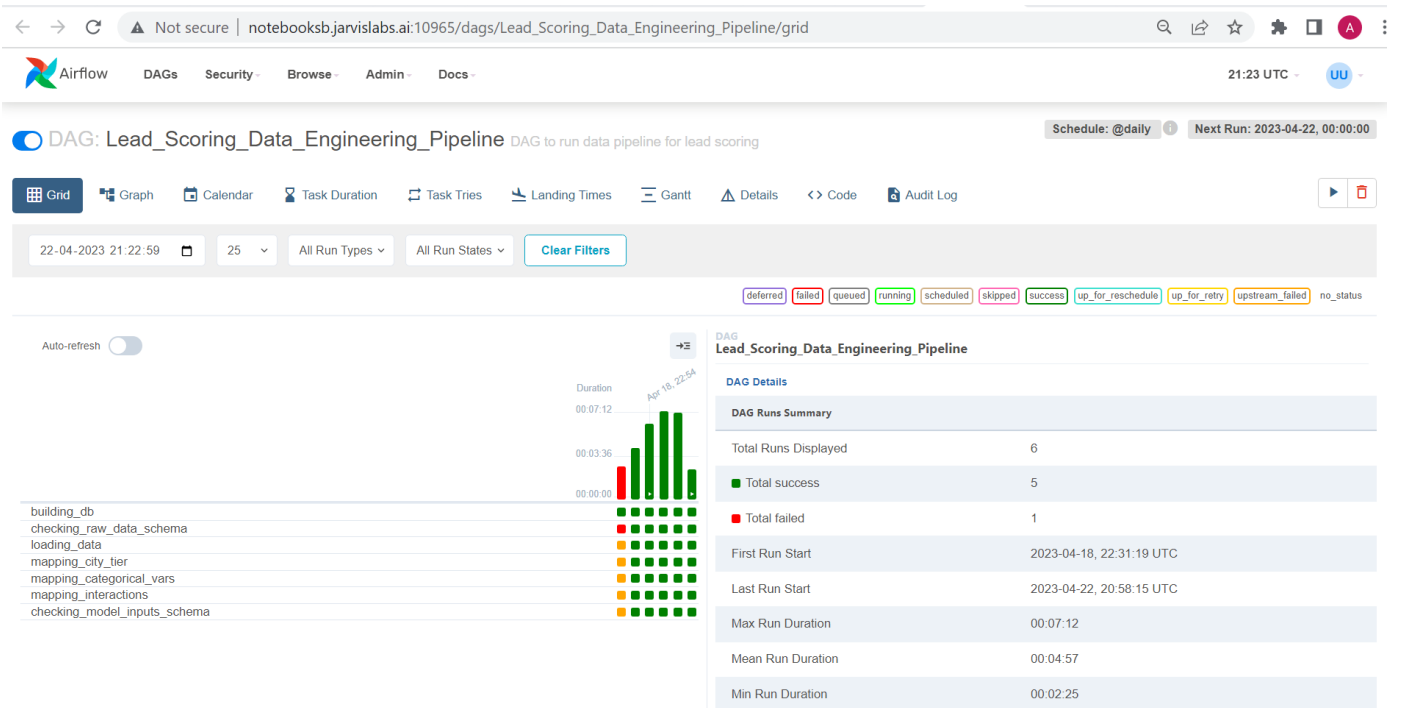


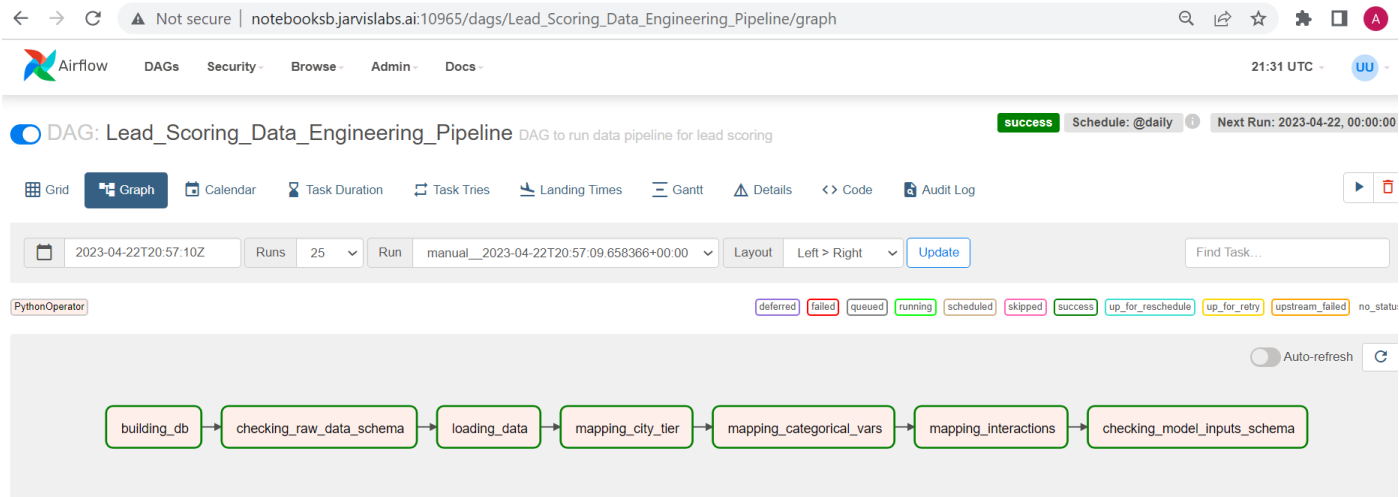
Airflow DAGs Dashboard

Showing DAGs for the assignment

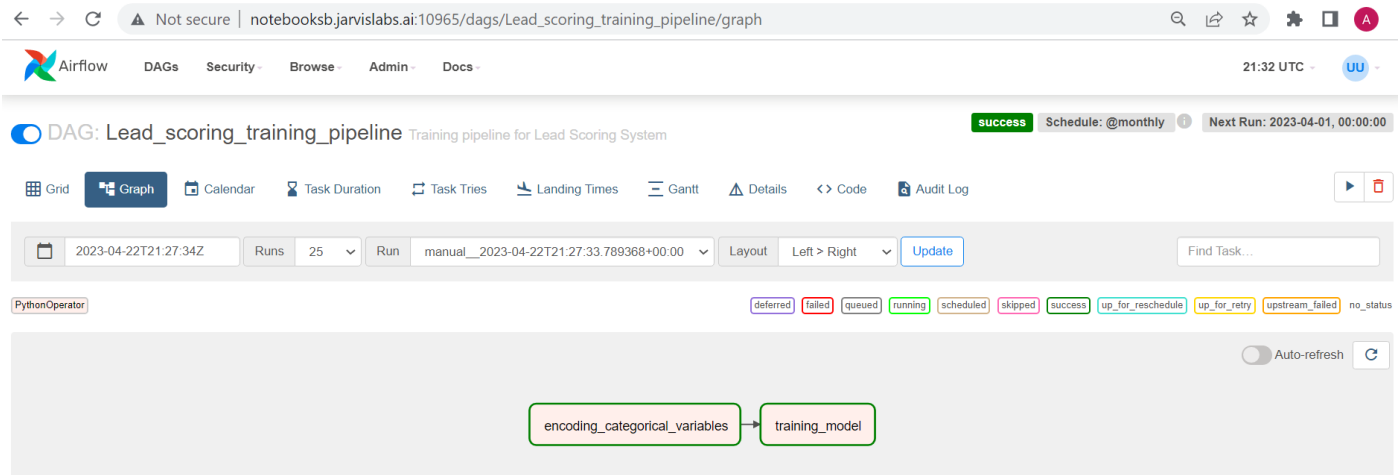
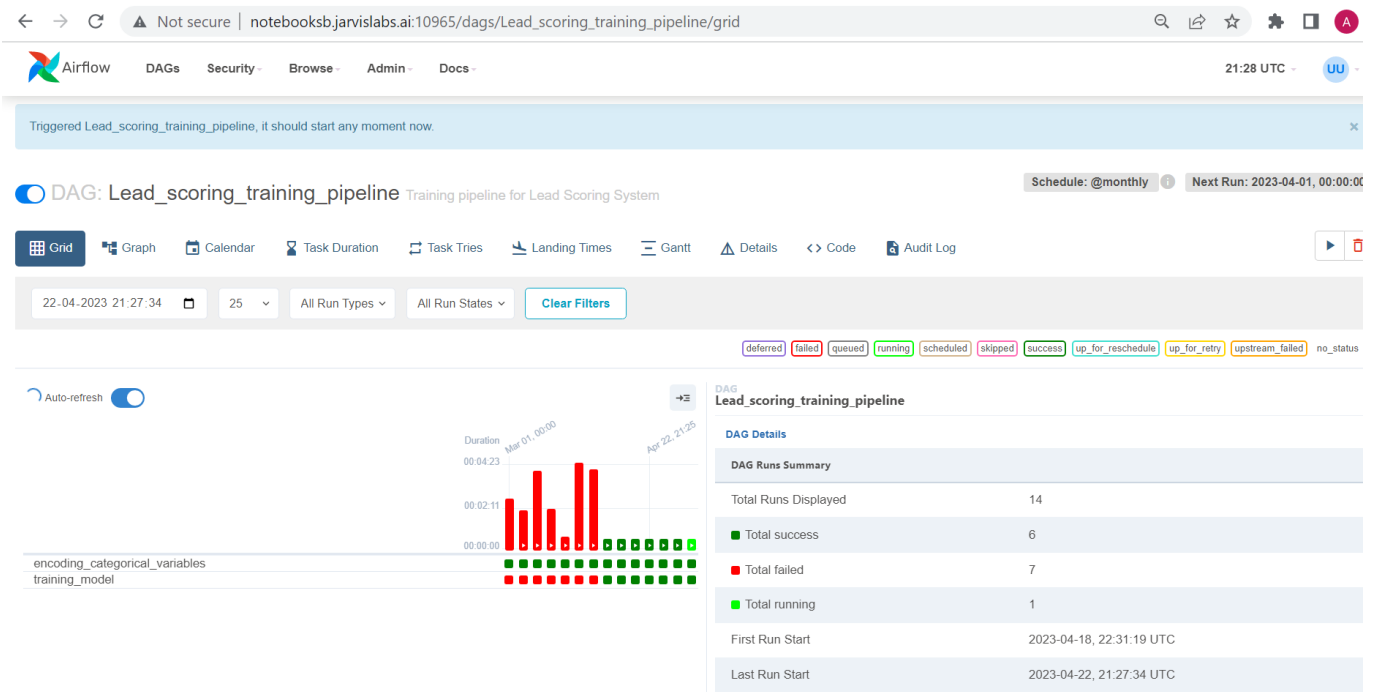


Data Pipeline

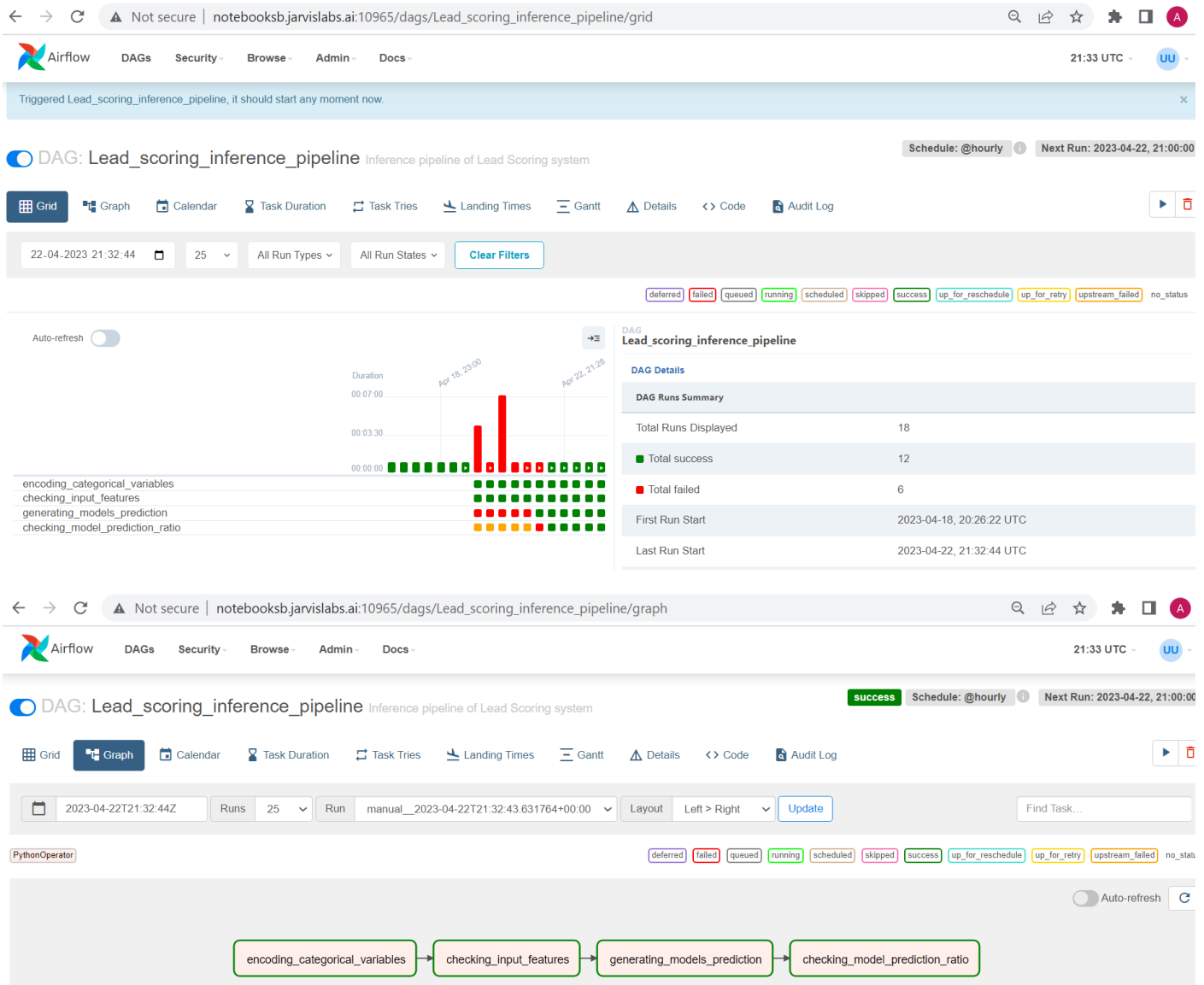




Training pipeline



Inference pipeline



Changes done in Data pipeline for inference file

(As per Automation Workflow of Inference Pipeline, 11th bullet point)

Lead Scoring Data Pipeline -> Constants file

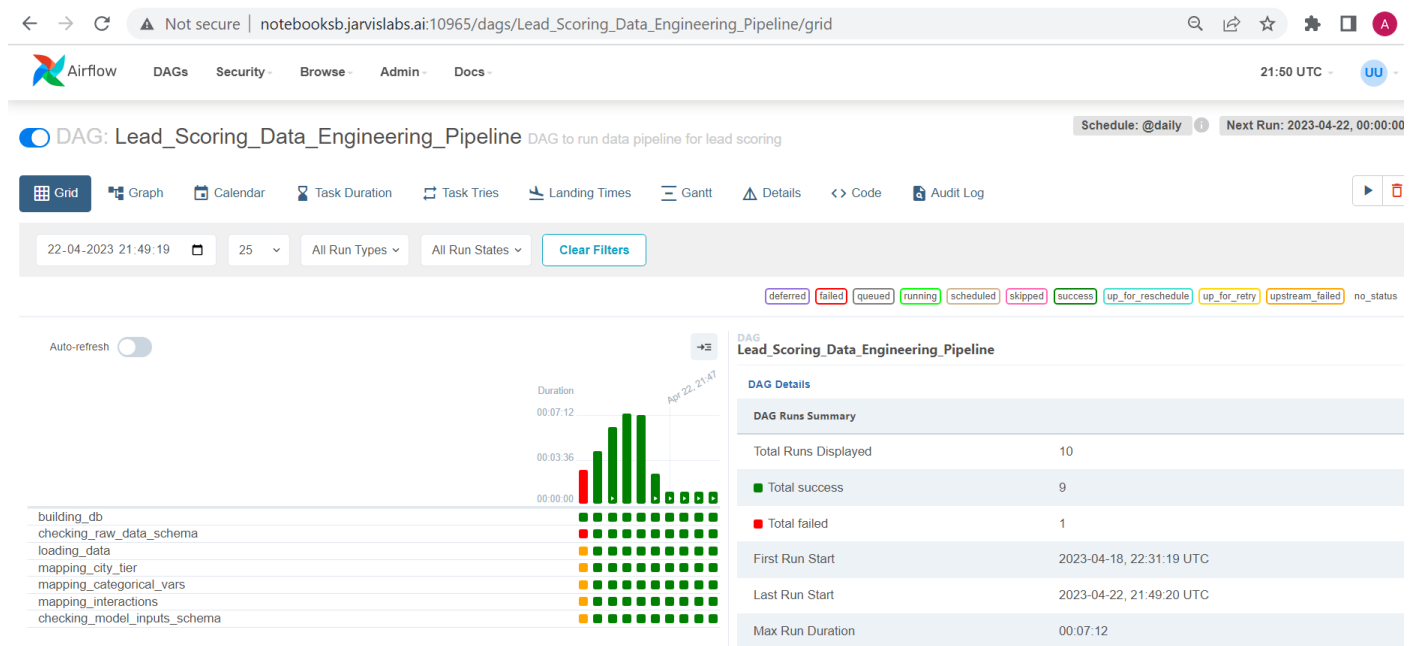
The image shows a Jupyter Notebook interface with the file 'constants.py' open. The file contains the following code:

```

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/Assignment/01_data_pipeline/scripts/'
3 DB_PATH = '/home/dags/Lead_scoring_data_pipeline/'
4 DB_FILE_NAME = 'Lead_scoring_data_cleaning.db'
5 UNIT_TEST_DB_FILE_NAME = ''
6 DATA_DIRECTORY = '/home/dags/Lead_scoring_data_pipeline/data/' # './data/'
7 LEAD_SCORING_FILE = 'leadscoring_inference' #changes done as inference pipeline 11 point #leadscoring' #.csv file
8 INTERACTION_MAPPING = '/home/dags/Lead_scoring_data_pipeline/mapping/'
9 #INDEX_COLUMNS_TRAINING = ['created_date', 'city_tier', 'first_platform_c', 'first_utm_medium_c',
10 # 'first_utm_source_c', 'total_leads_dropped', 'referred_lead', 'app_complete_flag']
11
12 #After making change for inference, directly using this variable, #changes done as inference pipeline 11 point
13 INDEX_COLUMNS_TRAINING = ['created_date', 'city_tier', 'first_platform_c', 'first_utm_medium_c',
14 'first_utm_source_c', 'total_leads_dropped', 'referred_lead']
15
16 INDEX_COLUMNS_INFERENCE = []
17 NOT_FEATURES = []
18
19

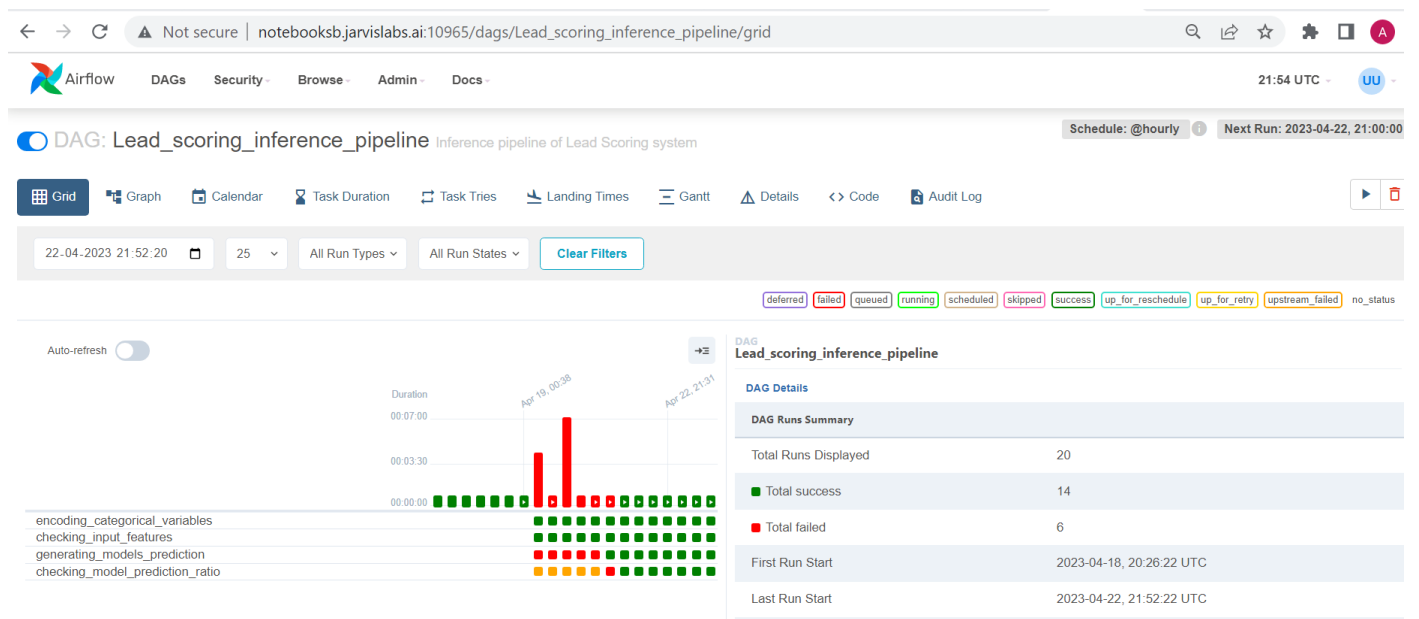
```

Data Pipeline



Once this is done first trigger the Data pipeline manually and then trigger the inference pipeline once the data pipeline's run is complete.

Inference pipeline



ML Flow server

← → ↻ 🔒 Not secure | notebooks.jarvislabs.ai:10964/#/experiments/0 🔍 📄 ☆ ⚙️ 🖨️ 📱

mlflow 1.26.1 Experiments Models GitHub Docs

Experiments + < Default 📄 Share

Search Experiments

Default 📄 🗑️

Model_Building_Pip... 📄 🗑️

Model_Building_Pip... 📄 🗑️

Model_Building_Pip... 📄 🗑️

Model_Building_Pip... 📄 🗑️

Track machine learning training runs in experiments. [Learn more](#) ✕

Experiment ID: 0

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 7 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision_0	Parameters
<input type="checkbox"/>	26 minutes ago	4.3s	Lead_scorin...	root	airflow	-	LightGBM/7	3956	0.716	0.831	gbdt
<input type="checkbox"/>	27 minutes ago	4.3s	Lead_scorin...	root	airflow	-	LightGBM/6	3956	0.716	0.831	gbdt
<input type="checkbox"/>	27 minutes ago	4.4s	Lead_scorin...	root	airflow	-	LightGBM/5	3956	0.716	0.831	gbdt
<input type="checkbox"/>	28 minutes ago	4.3s	Lead_scorin...	root	airflow	-	LightGBM/4	3956	0.716	0.831	gbdt
<input type="checkbox"/>	29 minutes ago	4.3s	Lead_scorin...	root	airflow	-	LightGBM/3	3956	0.716	0.831	gbdt
<input type="checkbox"/>	52 minutes ago	4.8s	Lead_scorin...	root	airflow	-	LightGBM/2	3956	0.716	0.831	gbdt
<input type="checkbox"/>	3 days ago	7.3s	Lead_scorin...	root	airflow	-	LightGBM/1	15511	0.669	0.641	gbdt

← → ↻ 🔒 Not secure | notebooks.jarvislabs.ai:10964/#/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 7	-	Version 2	2023-04-23 02:58:00	-

< 1 > 10 / page

Models

← → ↻ 🔒 Not secure | notebooks.jarvislabs.ai:10964/#/experiments/0/runs/0b12c9f66942403bb2b59c890b7ecab9 🔍 📄 ☆ ⚙️ 🖨️ 📱

Artifacts

models

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

Full Path: /home/miruns/0/0b12c9f66942403bb2b59c890b7ecab9/artifacts/models

LightGBM v7 Registered on 2023/04/23

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:/0b12c9f66942403bb2b59c890b7ecab9/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:/0b12c9f66942403bb2b59c890b7ecab9/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))
```

Metrics

← → ↺

⚠ Not secure | notebooks.jarvislabs.ai:10964/#/experiments/0/runs/0b12c9f66942403bb2b59c890b7ecab9

mlflow1261

ExperimentsModels

Default > Lead_scoring_mlflow_production

Lead_scoring_mlflow_production

Date: 2023-04-23 02:57:56

Source: airflow

User: root

Duration: 4.3s

Status: FINISHED

Lifecycle Stage: active

▶ Description [Edit](#)

▶ Parameters (20)

▼ Metrics (12)

Name	Value
False Negative	3956
Precision	0.716
Precision_0	0.831
Precision_1	0.661
Recall	0.746
Recall_0	0.542
Recall_1	0.89
True Negative	19502
f1_0	0.656
f1_1	0.759
roc_auc	0.716
test_accuracy	0.716

▶ Tags

▼ Artifacts

models

MLmodel

conda.yaml

model.pkl

Full Path:/home/mlruns/0/0b12c9f66942403bb2b59c890b7ecab9/artifacts/models

MLflow Model

Parameters

← → ↺

⚠ Not secure | notebooks.jarvislabs.ai:10964/#/experiments/0/runs/0b12c9f66942403bb2b59c890b7ecab9

▶ 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
reg_lambda	0.0
silent	warn
subsample	1.0
subsample_for_bin	200000
subsample_freq	0

▼ Metrics (12)

File details:

- Complete assignment code is kept at git repo shared in AssignmentCodeAtGitRepo.txt (https://github.com/akashpd/MLOps_Assignment_PipeLine_MLFlow_and_AirFlow)
- All DAGs which were run are kept in DAGs_4_AP.zip
- inference_test.ipynb file is also kept in unit_test folder which was used to test inference pipeline.

My Jarvis lab folder structure highlighted

notebooksb.jarvislabs.ai/061/lab/tree/Assignment/03

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name	Last Modified
airflow	seconds ago
Assignment	6 days ago
dags	4 days ago
dashboard	6 days ago
data	6 days ago
database	4 days ago
mlruns	5 days ago
notebooks	5 days ago
scripts	6 days ago
Assignemnt_1_AP.zip	5 days ago
Assignemnt_2_AP.zip	5 days ago
Assignemnt_3_AP.zip	4 days ago
Assignemnt_4_AP.zip	25 minutes ago
Assignment.zip	6 days ago
DAGs_4_AP.zip	25 minutes ago
lead_scoring_data_cleaning.db	4 days ago

My DAGs folder location

notebooksb.jarvislabs.ai/061/lab/tree/airflow/airflow.cfg

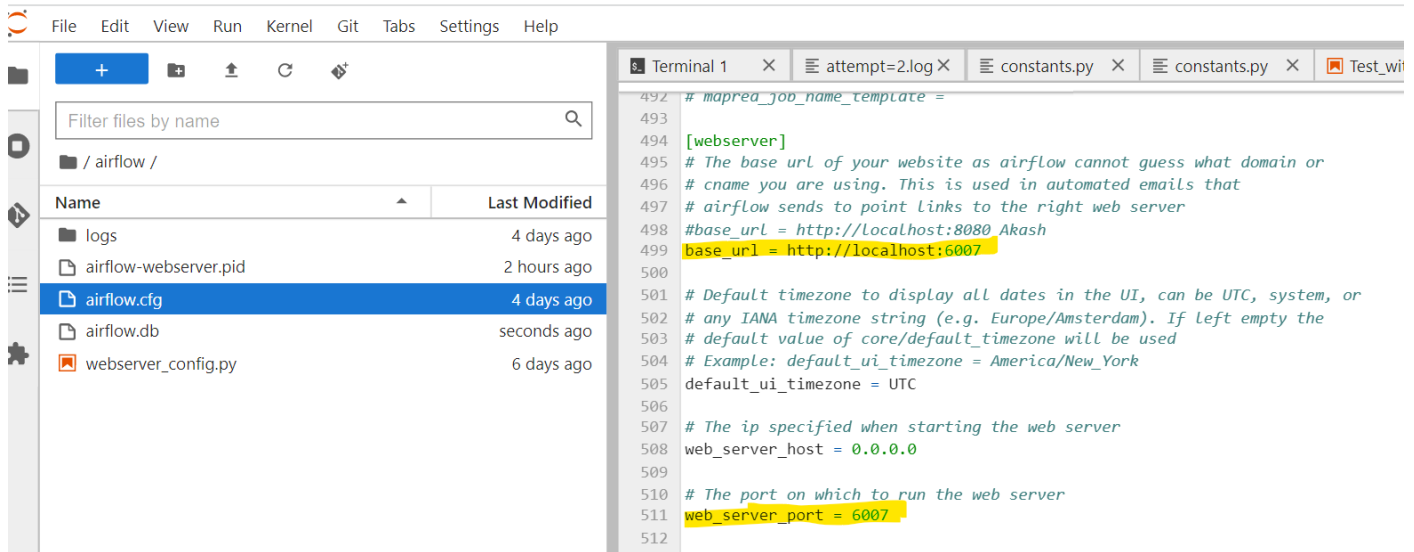
Filter files by name

Name	Last Modified
logs	4 days ago
airflow-webserver.pid	2 hours ago
airflow.cfg	4 days ago
airflow.db	seconds ago
webserver_config.py	6 days ago

```

1 [core]
2 # The folder where your airflow pipelines live, most likely a
3 # subfolder in a code repository. This path must be absolute.
4 #dags_folder = /home/airflow/dags
5 dags_folder = /home/dags
6
7 # Hostname by providing a path to a callable, which will resolve the hostname.
8 # The format is "package.function".
9 #
10 # For example, default value "socket.getfqdn" means that result from getfqdn() of "socket"
11 # package will be used as hostname.
12 #
13 # No argument should be required in the function specified.
14 # If using IP address as hostname is preferred, use value ``airflow.utils.net.get_host_ip_address``
15 hostname_callable = socket.getfqdn

```

Airflow Web server port configuration

After all code is done and working correctly then run below steps in summary to run airflow and DAGs (in order)

➤ **on Terminal 1**

airflow db init
airflow webserver

➤ **on Terminal 2**

airflow scheduler

➤ **Run DAG Lead_scoring_data_pipeline from Airflow DAGs dashboard**

➤ **On Terminal 3**

cd dags

mlflow server --backend-store-uri='sqlite:///./Lead_scoring_training_pipeline/Lead_scoring_mlflow_production.db' --default-artifact-root="/home/mlruns" --port=6006 --host=0.0.0.0

➤ **Run DAG Lead_scoring_training_pipeline from Airflow DAGs dashboard**

➤ **Register new version Model from mlflow/models**

➤ **Run DAG Lead_scoring_inference_pipeline from Airflow DAGs dashboard**