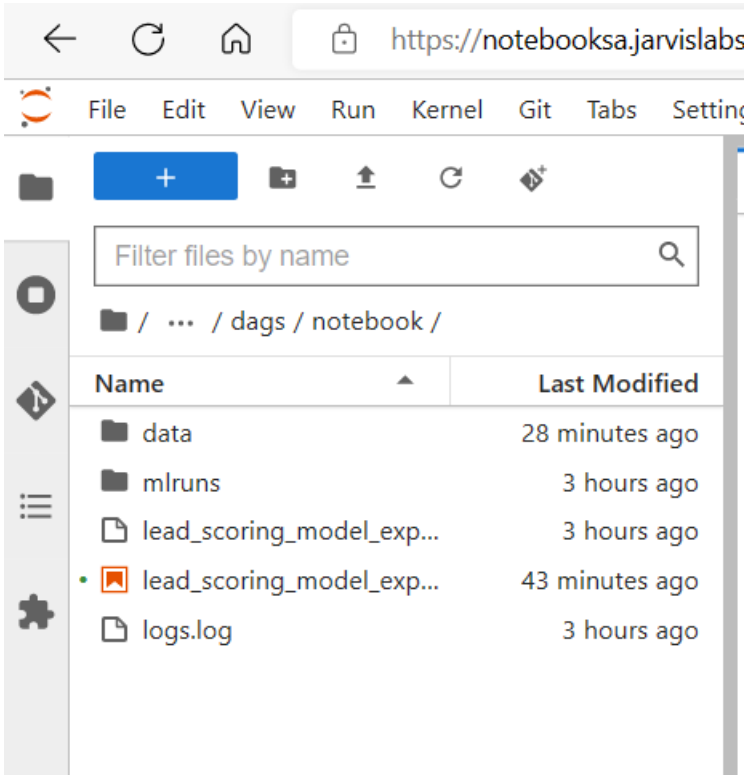
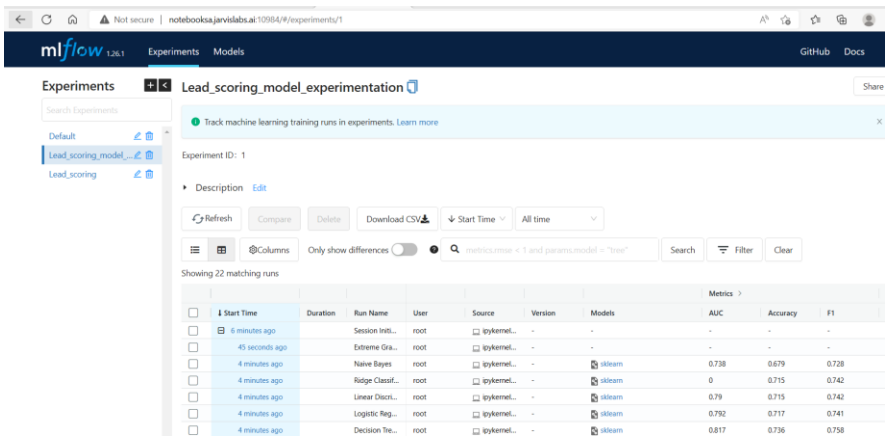
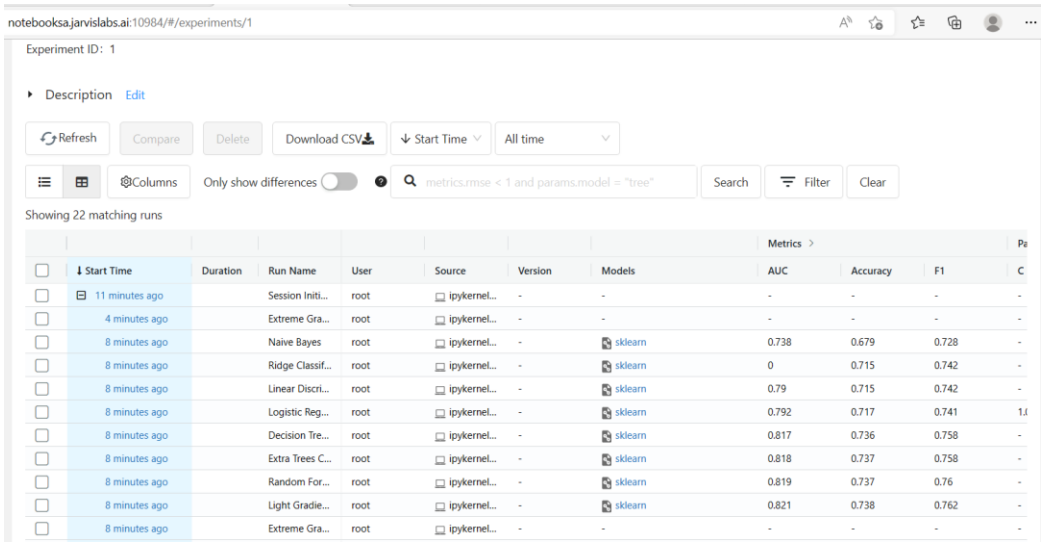
1) **Model notebook and data folder**

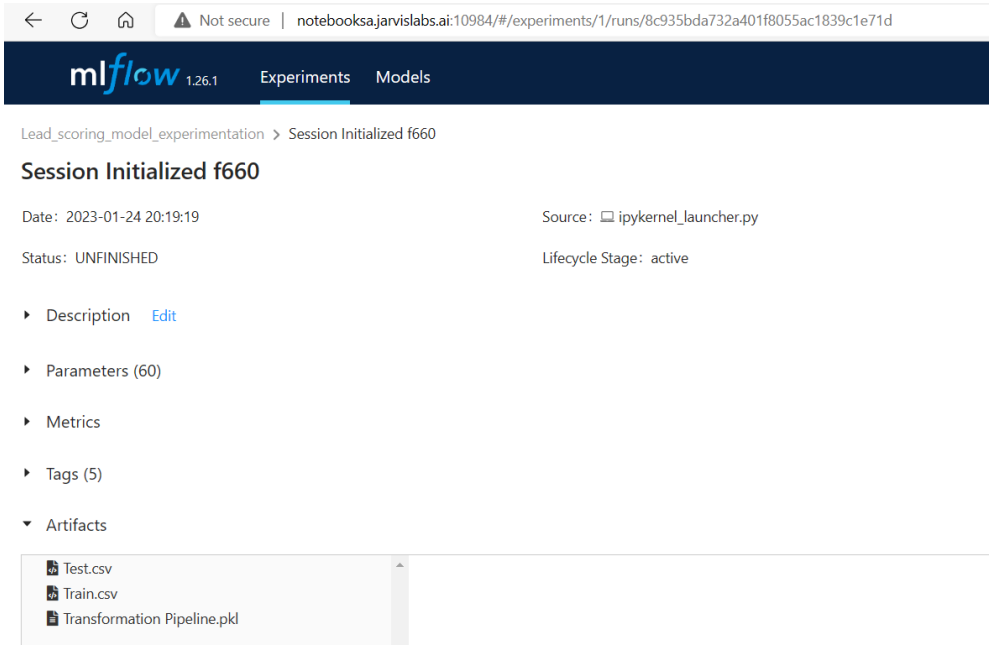


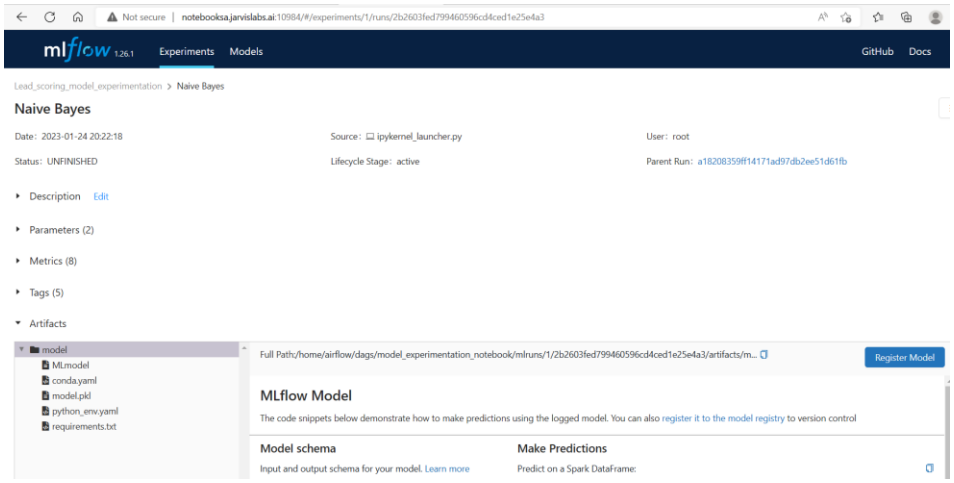

```
root@b3de51368704:~/airflow/dags/model_experimentation_notebook# tree
.
├── data
│   ├── cleaned_data.csv
│   └── leadscoring.csv
├── lead_scoring_model_experimentation.db
├── lead_scoring_model_experimentation.ipynb
├── logs.log
└── mlruns
```

Screenshot of mlflow ui

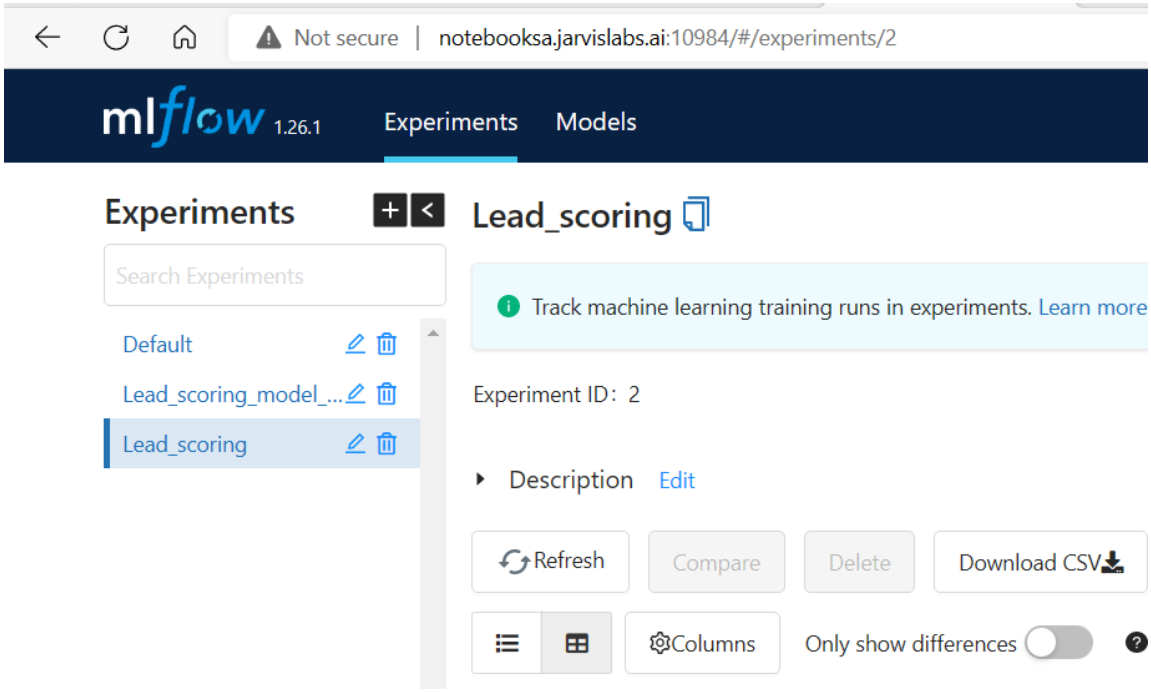


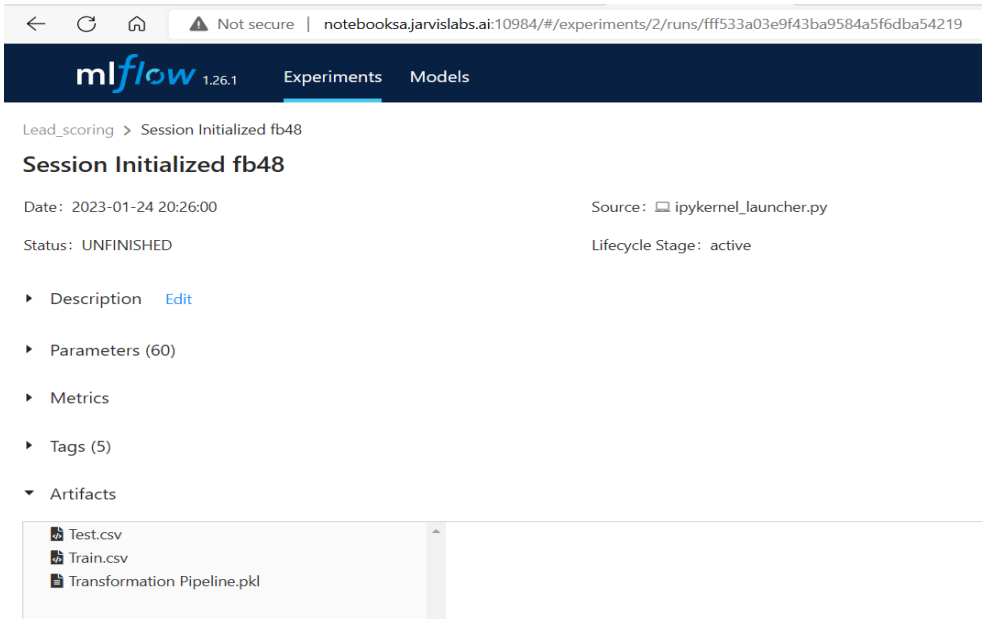

notebooksa.jarvislabs.ai:10984/#/experiments/1
Experiment ID: 1
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
metrics.rmse < 1 and params.model = "tree"
Search
Filter
Clear
Showing 22 matching runs
Metrics
Start Time
Duration
Run Name
User
Source
Version
Models
AUC
Accuracy
F1
11 minutes ago Session Initi... root ipykernel... - - - - - -
4 minutes ago Extreme Gra... root ipykernel... - - - - - -
8 minutes ago Naive Bayes root ipykernel... - sklearn 0.738 0.679 0.728 -
8 minutes ago Ridge Classif... root ipykernel... - sklearn 0 0.715 0.742 -
8 minutes ago Linear Discri... root ipykernel... - sklearn 0.79 0.715 0.742 -
8 minutes ago Logistic Reg... root ipykernel... - sklearn 0.792 0.717 0.741 -
8 minutes ago Decision Tre... root ipykernel... - sklearn 0.817 0.736 0.758 -
8 minutes ago Extra Trees C... root ipykernel... - sklearn 0.818 0.737 0.758 -
8 minutes ago Random For... root ipykernel... - sklearn 0.819 0.737 0.76 -
8 minutes ago Light Gradie... root ipykernel... - sklearn 0.821 0.738 0.762 -
8 minutes ago Extreme Gra... root ipykernel... - - - - - -

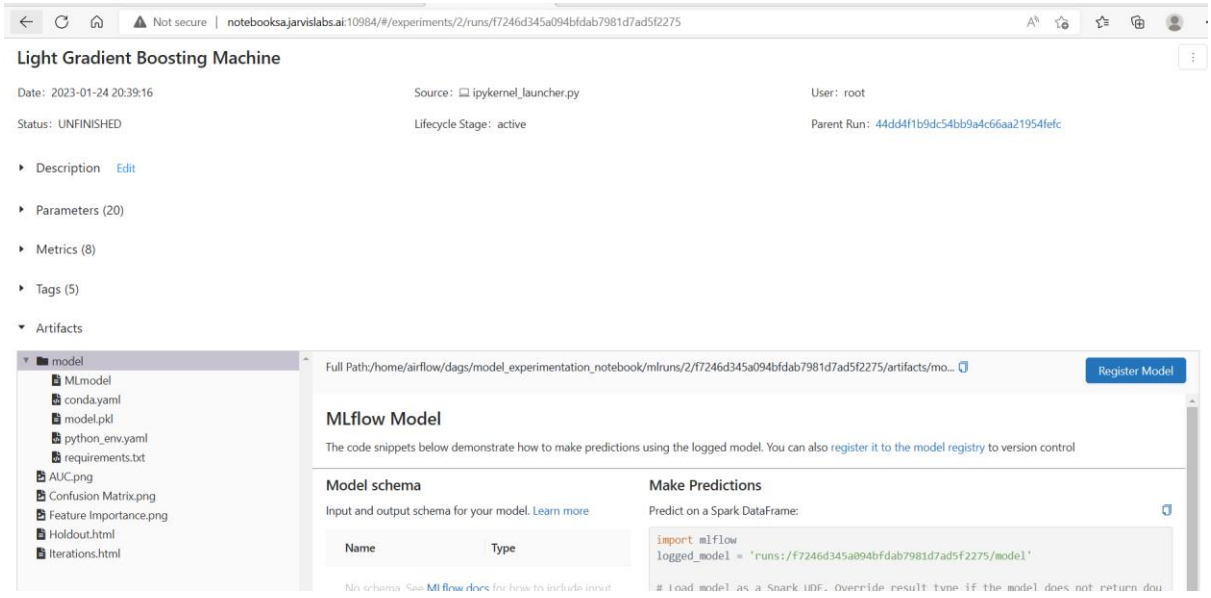

Not secure | notebooksa.jarvislabs.ai:10984/#/experiments/1/runs/8c935bda732a401f8055ac1839c1e71d
mlflow 1.26.1
Experiments
Models
Lead_scoring_model_experimentation > Session Initialized f660
Session Initialized f660
Date: 2023-01-24 20:19:19
Source: ipykernel_launcher.py
Status: UNFINISHED
Lifecycle Stage: active
Description Edit
Parameters (60)
Metrics
Tags (5)
Artifacts
Test.csv
Train.csv
Transformation Pipeline.pkl

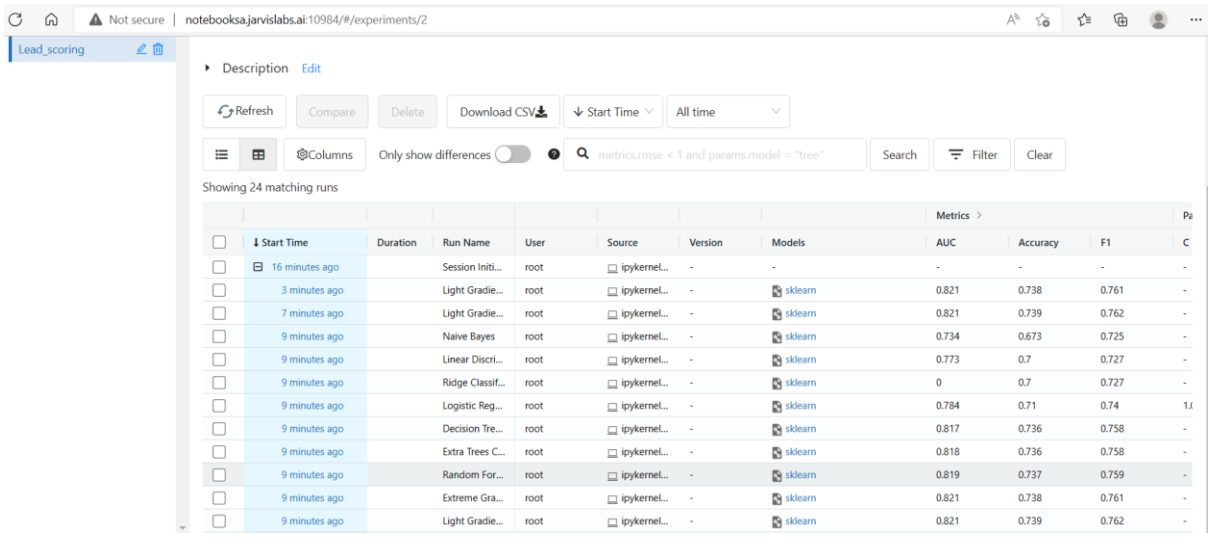


Screenshot of mlflow ui after dropping features

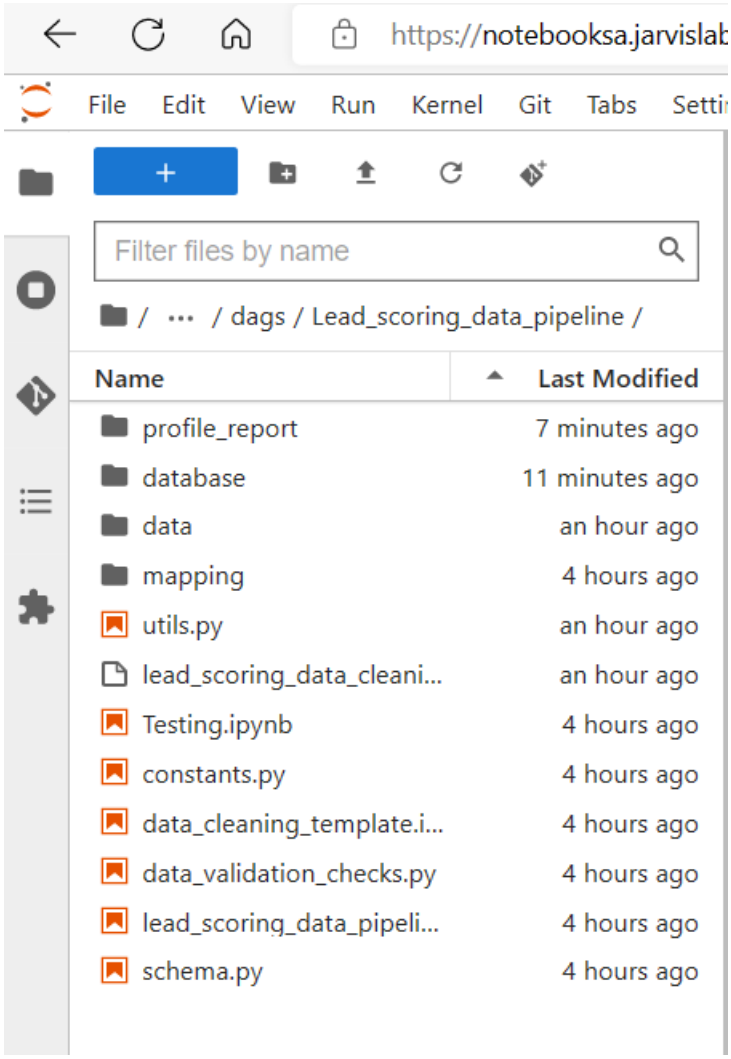



Not secure | notebooksa.jarvislabs.ai:10984/#/experiments/2/runs/fff533a03e9f43ba9584a5f6dba54219
mlflow 1.26.1
Experiments
Models
Lead_scoring > Session Initialized fb48
Session Initialized fb48
Date: 2023-01-24 20:26:00
Source: ipykernel_launcher.py
Status: UNFINISHED
Lifecycle Stage: active
Description Edit
Parameters (60)
Metrics
Tags (5)
Artifacts
Test.csv
Train.csv
Transformation Pipeline.pkl


Not secure | notebooksa.jarvislabs.ai:10984/#/experiments/2/runs/f7246d345a094bfdab7981d7ad5f2275
Light Gradient Boosting Machine
Date: 2023-01-24 20:39:16
Source: ipykernel_launcher.py
User: root
Status: UNFINISHED
Lifecycle Stage: active
Parent Run: 44dd4f1b9dc54bb9a4c66aa21954fefc
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
Iterations.html
Full Path:/home/airflow/dags/model_experimentation_notebook/mlruns/2/f7246d345a094bfdab7981d7ad5f2275/artifacts/mo...
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
Make Predictions
Predict on a Spark DataFrame:
import mlflow
logged_model = 'runs:/f7246d345a094bfdab7981d7ad5f2275/model'


Not secure | notebooksa.jarvislabs.ai:10984/#/experiments/2
Lead_scoring
Description Edit
Refresh
Compare
Delete
Download CSV
Start Time
All time
Columns
Only show differences
Search
Filter
Clear
Showing 24 matching runs
Start Time
Duration
Run Name
User
Source
Version
Models
Metrics
AUC
Accuracy
F1
16 minutes ago Session Initi... root ipykernel... - - - - -
3 minutes ago Light Gradie... root ipykernel... - sklearn 0.821 0.738 0.761 -
7 minutes ago Light Gradie... root ipykernel... - sklearn 0.821 0.739 0.762 -
9 minutes ago Naive Bayes root ipykernel... - sklearn 0.734 0.673 0.725 -
9 minutes ago Linear Discri... root ipykernel... - sklearn 0.773 0.7 0.727 -
9 minutes ago Ridge Classif... root ipykernel... - sklearn 0 0.7 0.727 -
9 minutes ago Logistic Reg... root ipykernel... - sklearn 0.784 0.71 0.74 -
9 minutes ago Decision Tre... root ipykernel... - sklearn 0.817 0.736 0.758 -
9 minutes ago Extra Trees C... root ipykernel... - sklearn 0.818 0.736 0.758 -
9 minutes ago Random For... root ipykernel... - sklearn 0.819 0.737 0.759 -
9 minutes ago Extreme Gra... root ipykernel... - sklearn 0.821 0.738 0.761 -
9 minutes ago Light Gradie... root ipykernel... - sklearn 0.821 0.739 0.762 -

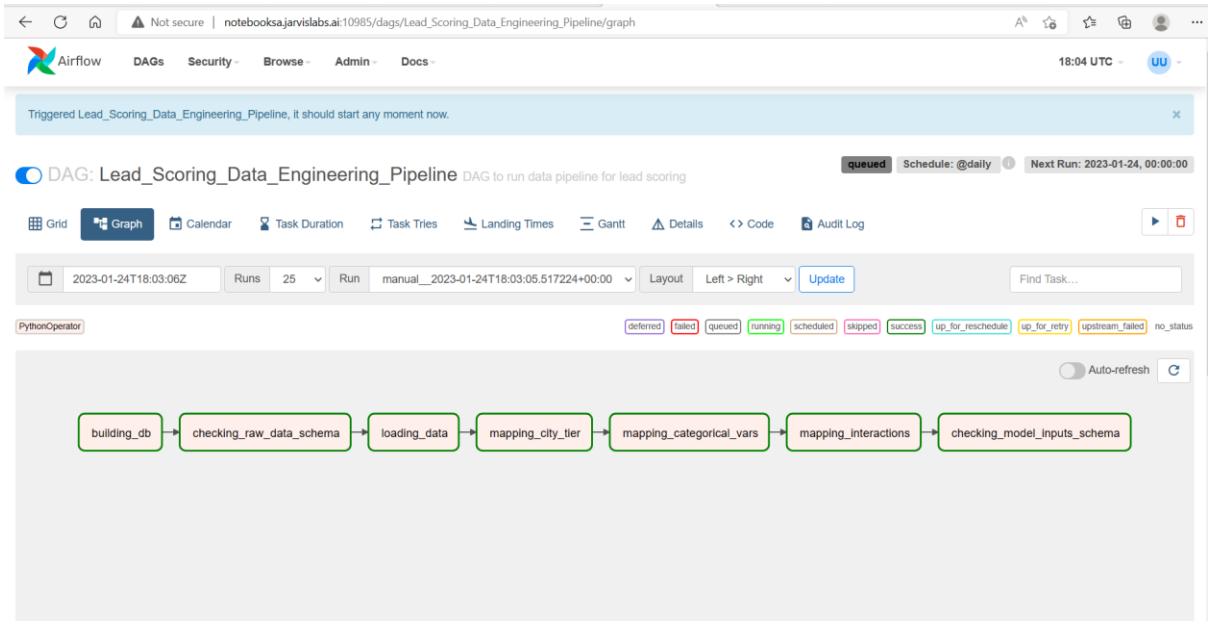
2) **Lead_scoring_data_pipeline**



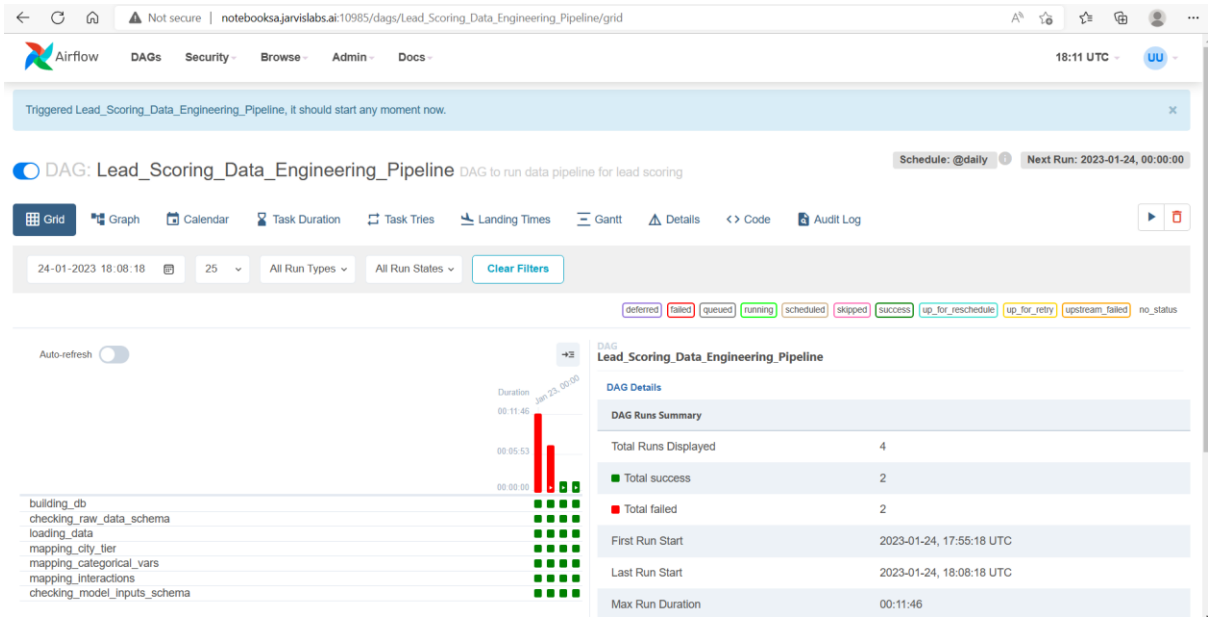
```
root@b3de51368704:~/airflow/dags/Lead_scoring_data_pipeline# tree
.
├── Testing.ipynb
├── __pycache__
│   ├── constants.cpython-38.pyc
│   ├── data_validation_checks.cpython-38.pyc
│   ├── lead_scoring_data_pipeline.cpython-38.pyc
│   ├── schema.cpython-38.pyc
│   └── utils.cpython-38.pyc
├── constants.py
├── data
│   ├── cleaned_data.csv
│   └── leadscoring.csv
├── data_cleaning_template.ipynb
├── data_validation_checks.py
├── database
│   └── lead_scoring_data_cleaning.db
├── lead_scoring_data_cleaning.db
├── lead_scoring_data_pipeline.py
├── mapping
│   ├── __pycache__
│   │   ├── city_tier.cpython-38.pyc
│   │   └── significant_categorical_level.cpython-38.pyc
│   ├── city_tier.py
│   ├── interaction_mapping.csv
│   └── significant_categorical_level.py
├── profile_report
│   ├── cleaned_data_report.html
│   └── raw_data_report.html
├── schema.py
└── utils.py

6 directories, 23 files
```

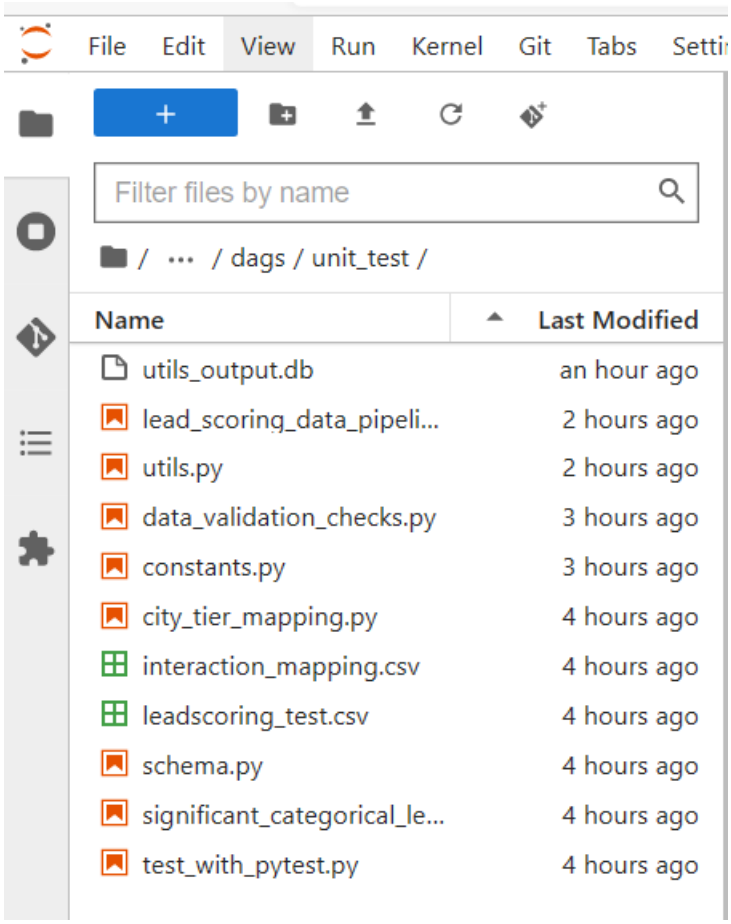
Screenshot of successful execution Airflow DAG in graph :



Screenshot of Airflow UI grid



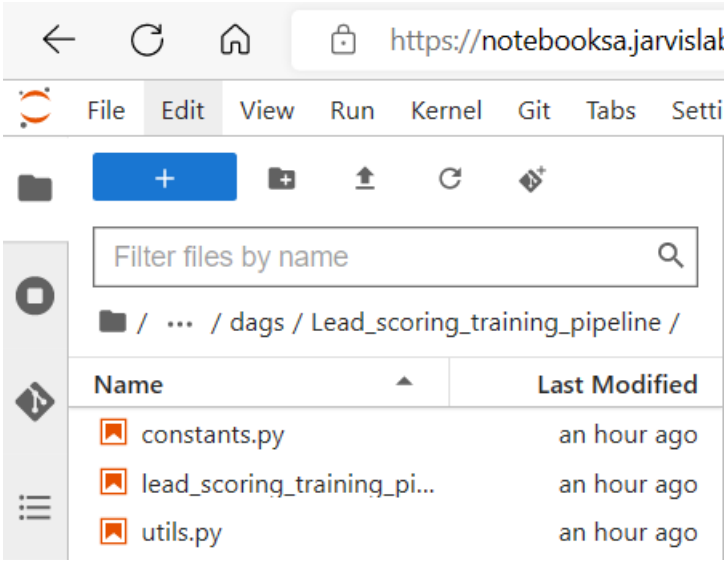
3) unit_test



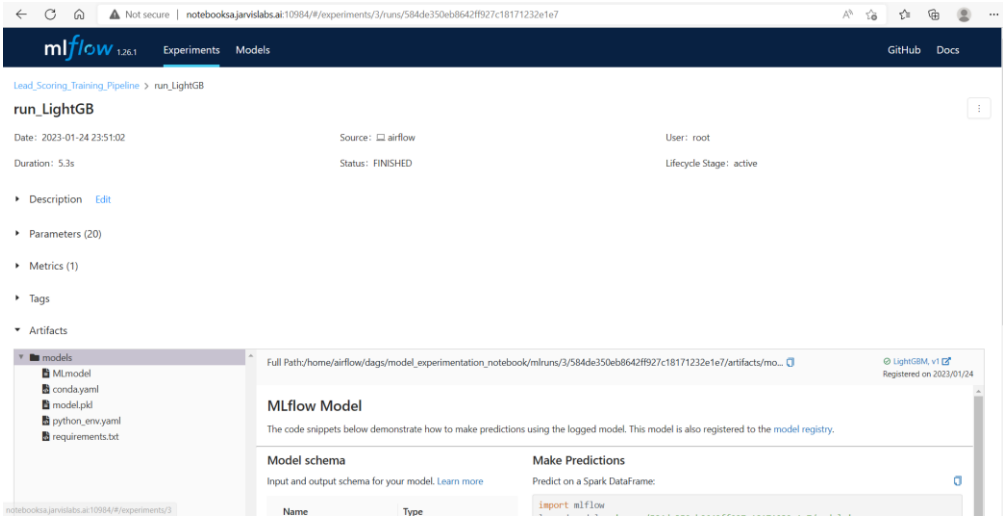

```
root@b3de51368704:~/airflow/dags/unit_test# tree
.
├── __pycache__
│   ├── city_tier_mapping.cpython-38.pyc
│   ├── constants.cpython-38.pyc
│   ├── data_validation_checks.cpython-38.pyc
│   ├── lead_scoring_data_pipeline.cpython-38.pyc
│   ├── significant_categorical_level.cpython-38.pyc
│   └── utils.cpython-38.pyc
├── city_tier_mapping.py
├── constants.py
├── data_validation_checks.py
├── interaction_mapping.csv
├── lead_scoring_data_pipeline.py
├── leadscoring_test.csv
├── schema.py
├── significant_categorical_level.py
├── test_with_pytest.py
├── utils.py
└── utils_output.db

1 directory, 17 files
```

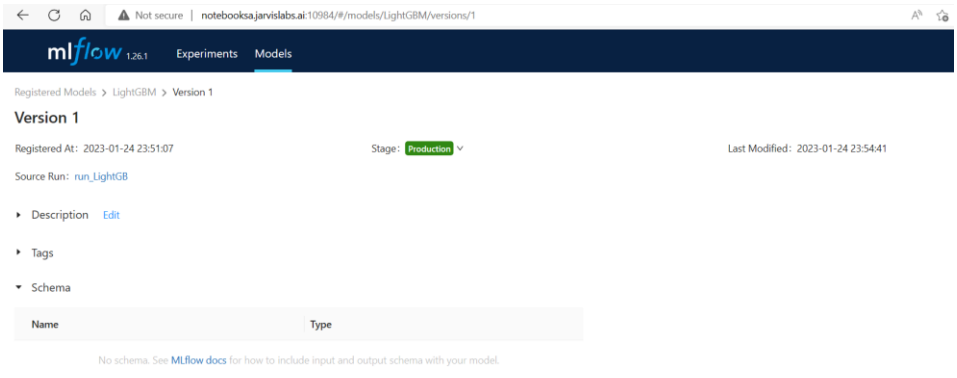
4) **Lead_scoring_training_pipeline**

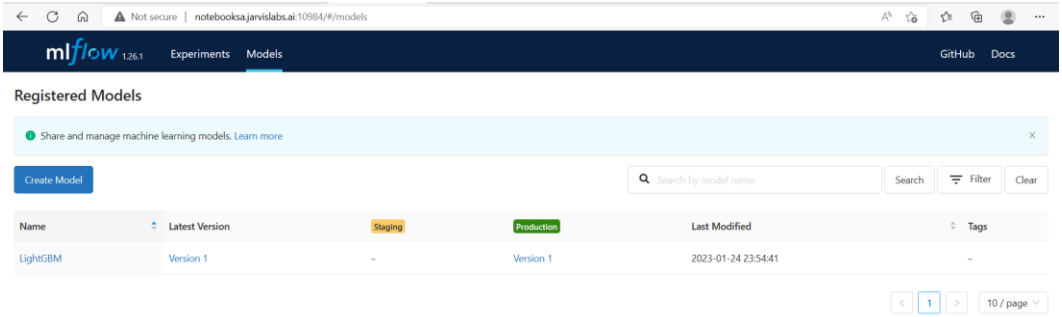


screenshot of experiments with all the artifacts visible :-

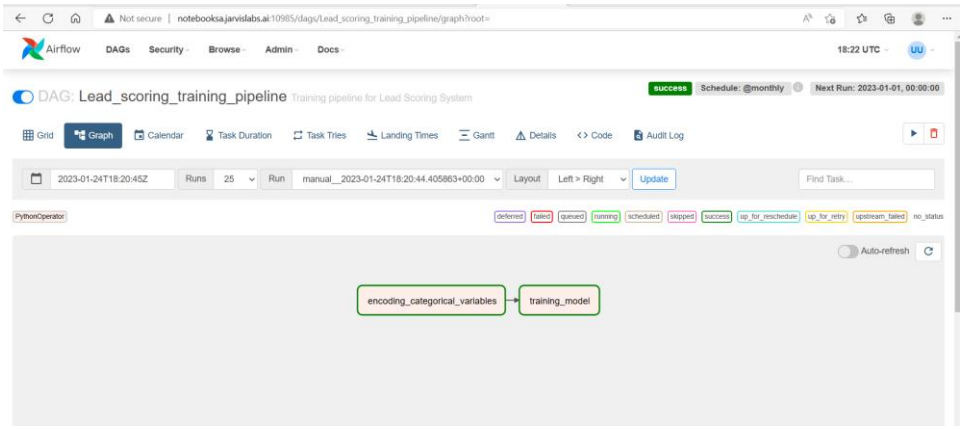


Screenshot of model registry with model name and stage as 'production'

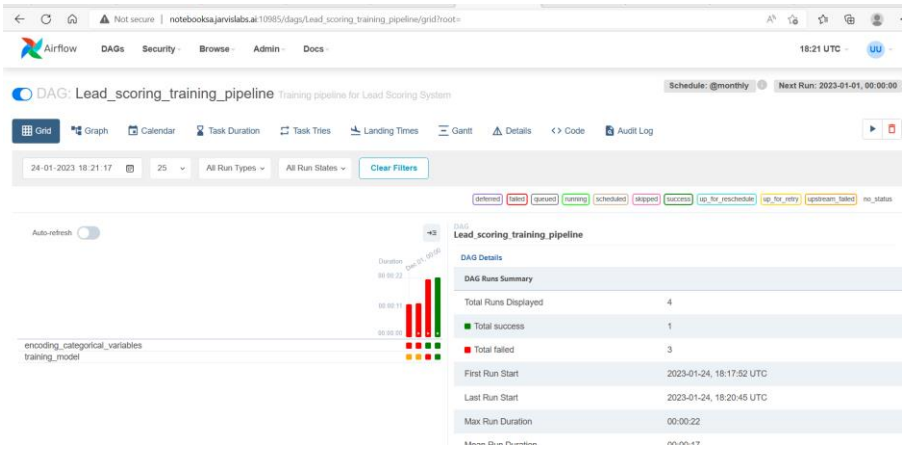




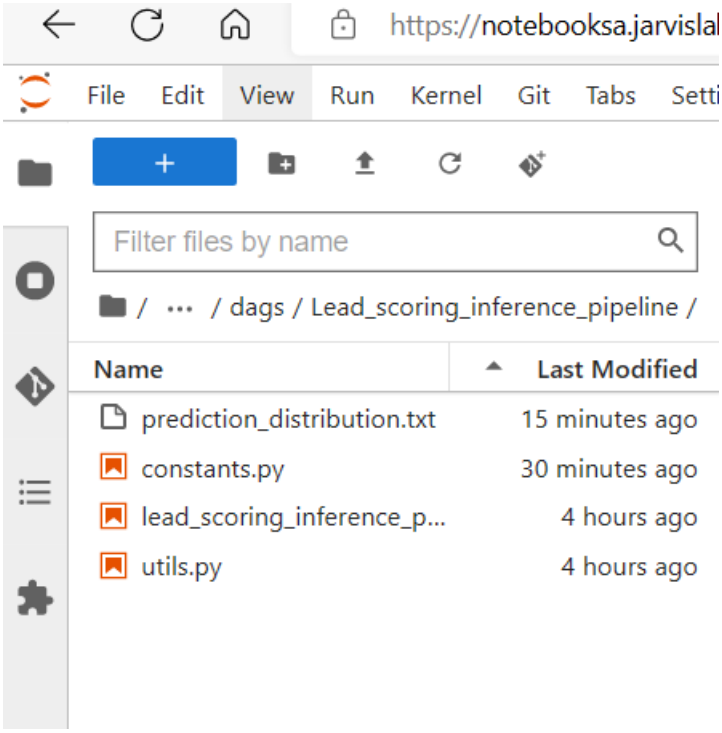
Screenshot of sucessful execution Airflow DAG in graph



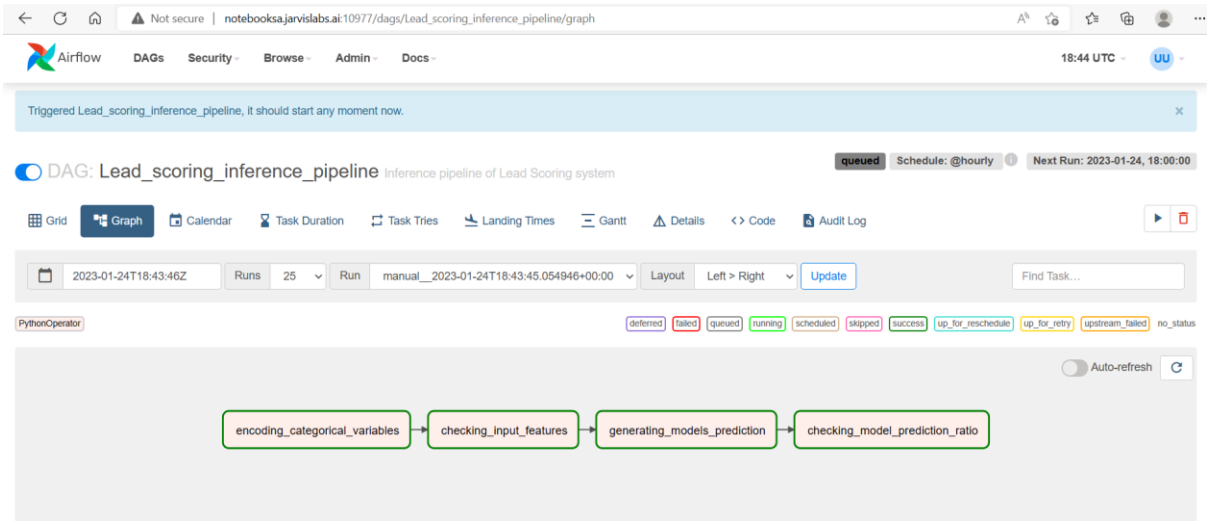
Screenshot of Airflow UI grid



5) Lead_scoring_inference_pipeline



Screenshot of successful execution Airflow DAG in graph



Screenshot of Airflow UI grid