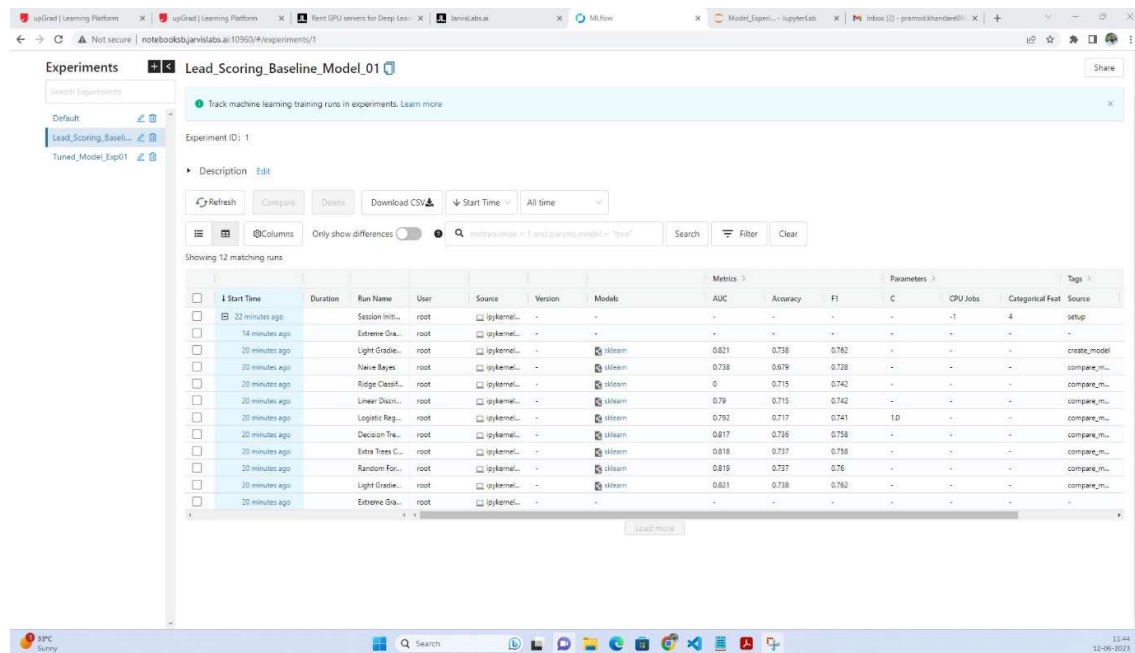


# MODEL EXPERIMENTATION:

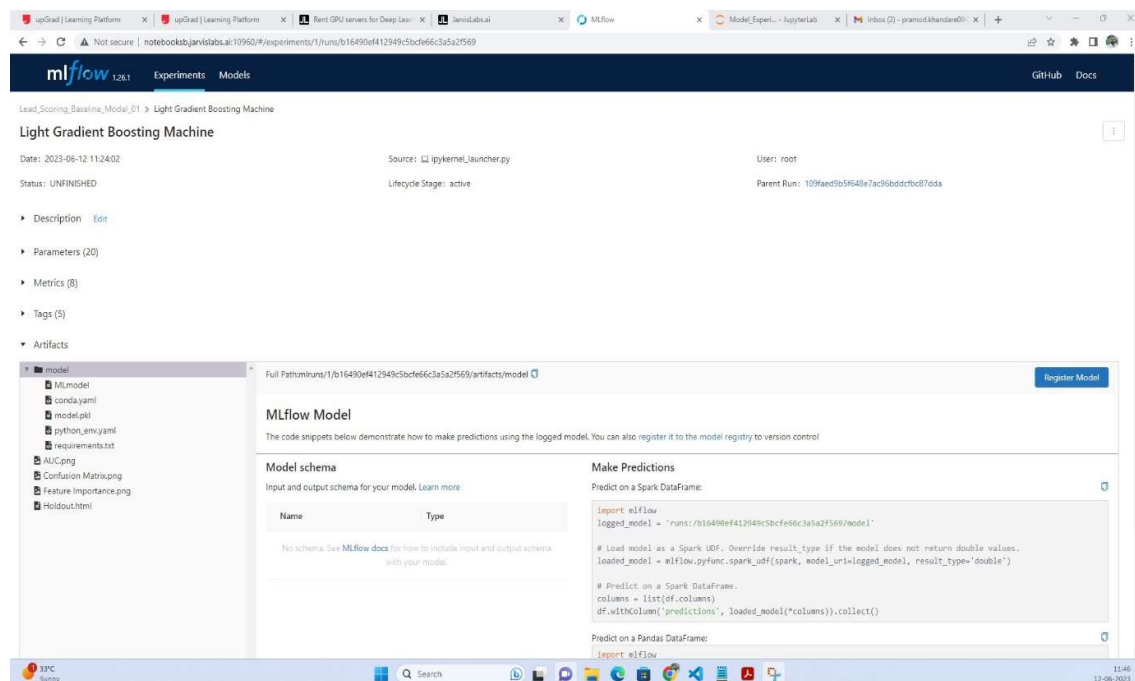
## Screenshot of mlflow ui- all experiments



The screenshot shows the MLflow Experiments page for the experiment 'Lead\_Scoring\_Baseline\_Model\_01'. The page displays a table of 12 runs, sorted by start time. The table includes columns for Start Time, Duration, Run Name, User, Source, Version, Models, Metrics (AUC, Accuracy, F1, C), Parameters (CPU Jobs, Categorical Features), and Tags. The runs are listed in descending order of start time, with the most recent run at the top.

Start Time	Duration	Run Name	User	Source	Version	Models	Metrics	Parameters	Tags
32 minutes ago		Session Init...	root	ipykernel...	-	-	-	-	-
14 minutes ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-
20 minutes ago		Light Gradie...	root	ipykernel...	-	libsvm	0.821, 0.738, 0.762	-	create_model
20 minutes ago		Naive Bayes	root	ipykernel...	-	libsvm	0.738, 0.679, 0.728	-	compare_m...
20 minutes ago		Ridge Classi...	root	ipykernel...	-	libsvm	0, 0.715, 0.742	-	compare_m...
20 minutes ago		Linear Discr...	root	ipykernel...	-	libsvm	0.79, 0.715, 0.742	-	compare_m...
20 minutes ago		Logistic Reg...	root	ipykernel...	-	libsvm	0.782, 0.717, 0.741	1.0	compare_m...
20 minutes ago		Decision Tre...	root	ipykernel...	-	libsvm	0.817, 0.736, 0.758	-	compare_m...
20 minutes ago		Extra Trees C...	root	ipykernel...	-	libsvm	0.818, 0.737, 0.758	-	compare_m...
20 minutes ago		Random For...	root	ipykernel...	-	libsvm	0.819, 0.737, 0.76	-	compare_m...
20 minutes ago		Light Gradie...	root	ipykernel...	-	libsvm	0.821, 0.738, 0.762	-	compare_m...
20 minutes ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-

## Screenshot of mlflow ui-- screenshot of one experiment with all the artifacts visible



The screenshot shows the MLflow UI for a specific experiment, 'Lead\_Scoring\_Baseline\_Model\_01'. The page displays the experiment's details, including its name, date, status, source, user, and parent run. It also shows a list of artifacts, including the ML model, training data, and evaluation results. The 'Artifacts' section is expanded, showing a list of files and their details.

**Experiment Details:**

- Name: Lead\_Scoring\_Baseline\_Model\_01
- Date: 2023-06-12 11:24:02
- Status: UNFINISHED
- Source: ipykernel\_launcher.py
- User: root
- Parent Run: 109faed9b5f548e7ac3b0bdcdbd7dda

**Artifacts:**

- MLmodel
- conda.yaml
- model.pkl
- python\_env.yaml
- requirements.txt
- AUC.png
- Confusion Matrix.png
- Feature Importance.png
- Holdout.html

**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:/109faed9b5f548e7ac3b0bdcdbd7dda/model'

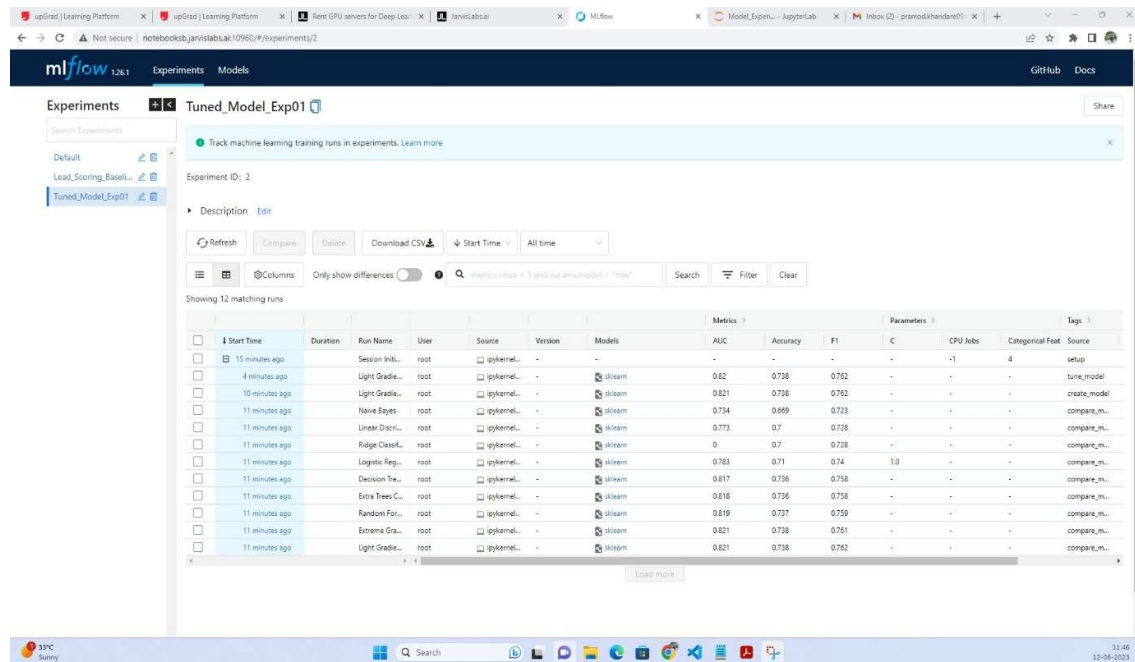
# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=logged_model, result_type='double')

# Predict on a Spark DataFrame.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
```

Predict on a Pandas DataFrame:

```
import mlflow
```

## Screenshot of mlflow ui after dropping features—all experiments

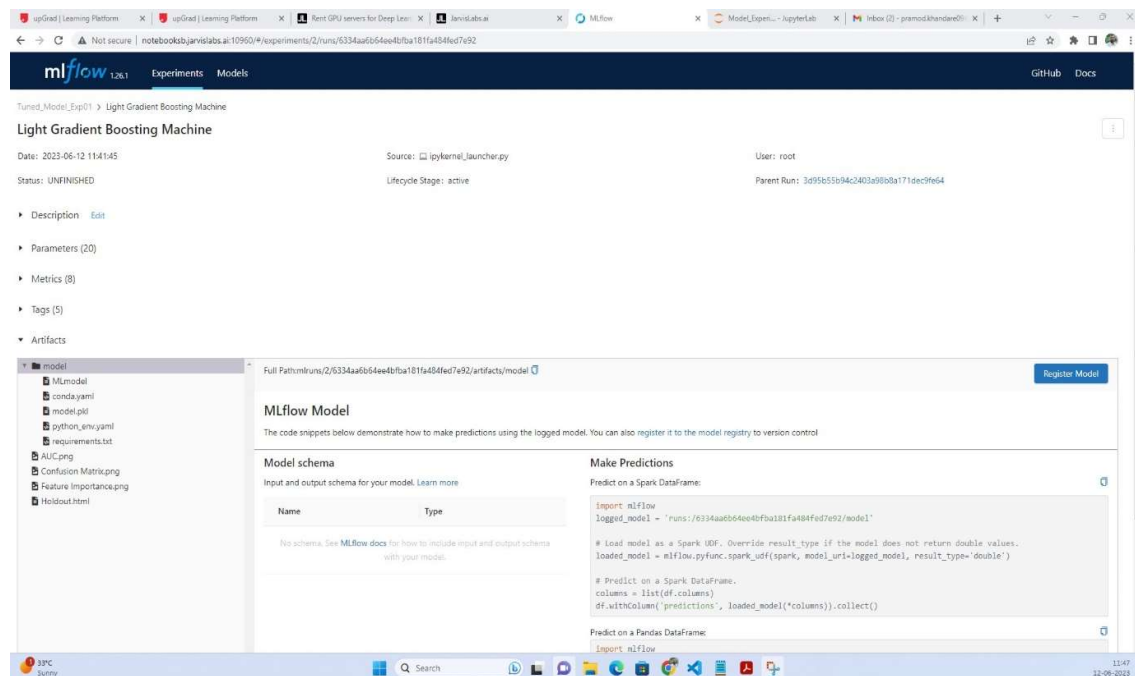


The screenshot shows the MLflow Experiments page for 'Tuned\_Model\_Exp01'. It displays a table of 12 runs, sorted by start time. The table includes columns for Start Time, Duration, Run Name, User, Source, Version, Models, Metrics (AUC, Accuracy, F1), Parameters (C, CPU Jobs, Categorical Feat), and Tags. The runs are listed in descending order of start time, with the most recent run at the top.

Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Tags
11 minutes ago	Session Init...	root	ipykernel...	-	-	-	-	-	-	-	-1	4	setup
4 minutes ago	Light Gradie...	root	ipykernel...	-	-	sklearn	0.82	0.738	0.762	-	-	-	tune_model
10 minutes ago	Light Gradie...	root	ipykernel...	-	-	sklearn	0.821	0.738	0.762	-	-	-	create_model
11 minutes ago	Naive Bayes	root	ipykernel...	-	-	sklearn	0.734	0.669	0.723	-	-	-	compare_m...
11 minutes ago	Linear Discrim...	root	ipykernel...	-	-	sklearn	0.773	0.7	0.728	-	-	-	compare_m...
11 minutes ago	Ridge Classi...	root	ipykernel...	-	-	sklearn	0	0.7	0.728	-	-	-	compare_m...
11 minutes ago	Logistic Reg...	root	ipykernel...	-	-	sklearn	0.783	0.71	0.74	1.0	-	-	compare_m...
11 minutes ago	Decision Tree...	root	ipykernel...	-	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...
11 minutes ago	Extra Trees C...	root	ipykernel...	-	-	sklearn	0.816	0.736	0.758	-	-	-	compare_m...
11 minutes ago	Random For...	root	ipykernel...	-	-	sklearn	0.819	0.737	0.759	-	-	-	compare_m...
11 minutes ago	Extrema Gra...	root	ipykernel...	-	-	sklearn	0.821	0.738	0.761	-	-	-	compare_m...
11 minutes ago	Light Gradie...	root	ipykernel...	-	-	sklearn	0.821	0.738	0.762	-	-	-	compare_m...

## Screenshot of mlflow ui after dropping features- one experiment with all artifacts

(Model, AUC, Confusion Matrix, Feature Importance, Holdout.html)



The screenshot shows the MLflow UI for a specific experiment run, 'Tuned\_Model\_Exp01 > Light Gradient Boosting Machine'. The page displays the run's date (2023-06-12 11:41:45), source (ipykernel\_launcher.py), user (root), and lifecycle stage (active). It also shows the parent run ID. The 'Artifacts' section is expanded, showing a list of artifacts including 'MLmodel', 'conda.yaml', 'model.pkl', 'python\_env.yaml', 'requirements.txt', 'AUC.png', 'Confusion Matrix.png', 'Feature Importance.png', and 'Holdout.html'. The 'Model' artifact is selected, showing the 'MLflow Model' details, including the model schema and code snippets for making predictions using the logged 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 <a href="#">MLflow docs</a> for how to include input and output schema with your model.	

**Make Predictions**

Predict on a Spark DataFrame:

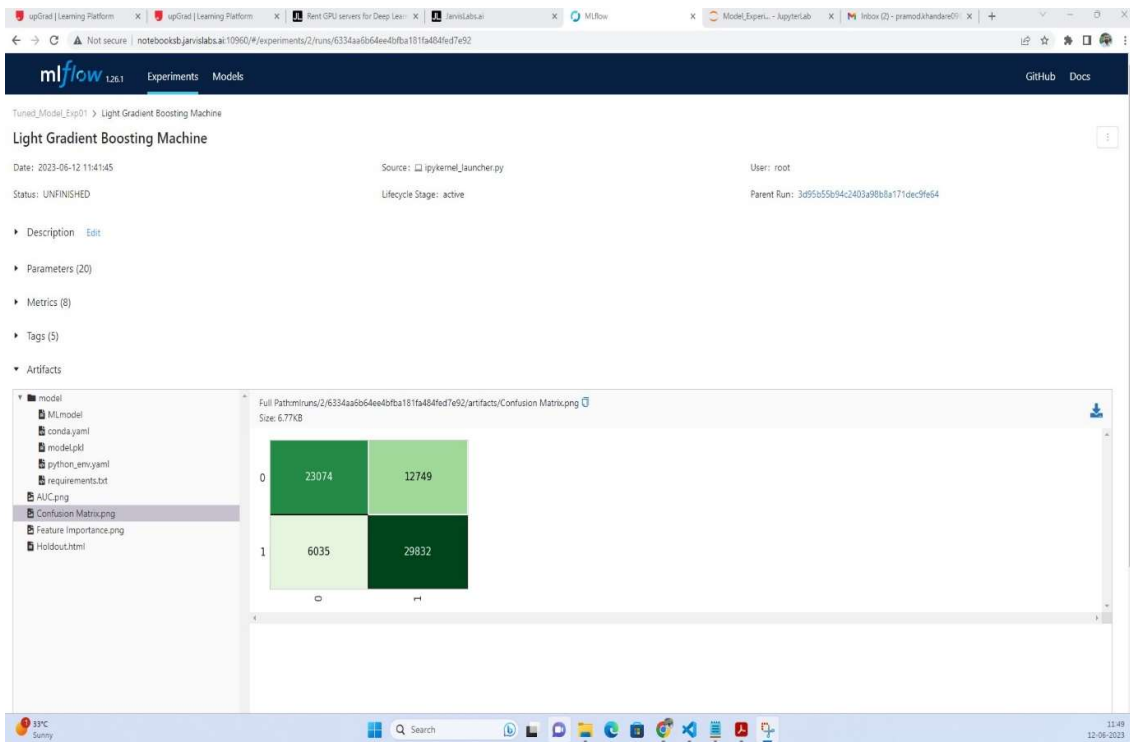
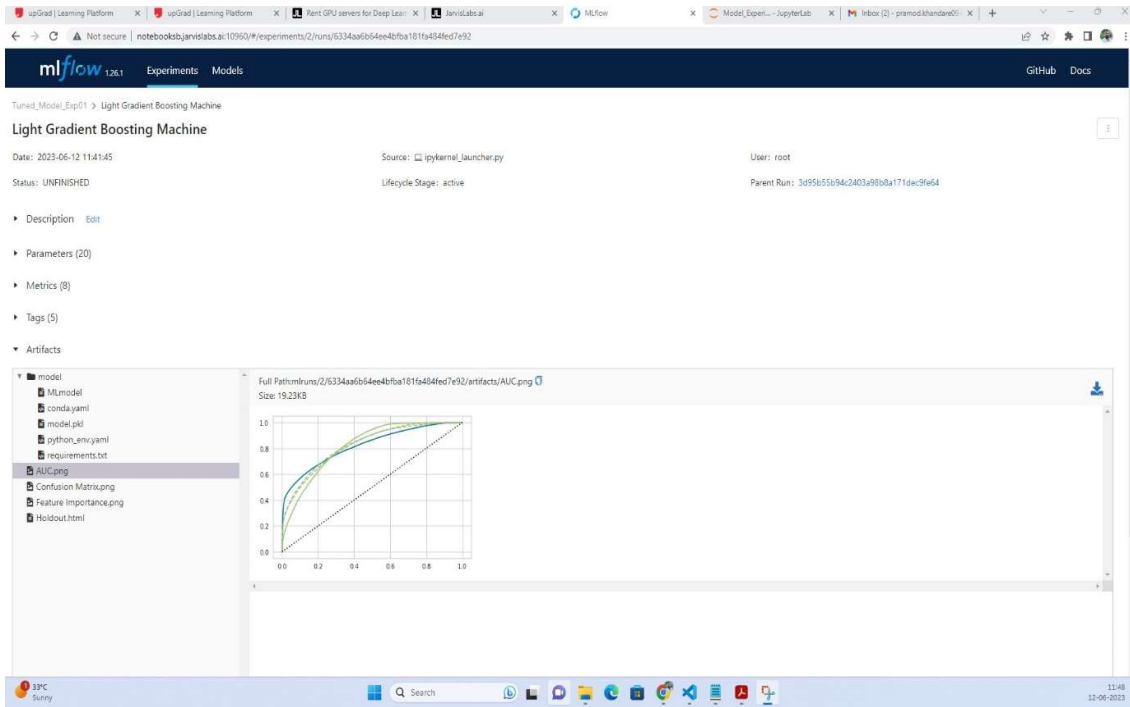
```
import mlflow
logged_model = 'runs:/6334a6b64ee4bfa181fa484fed7e92/model'

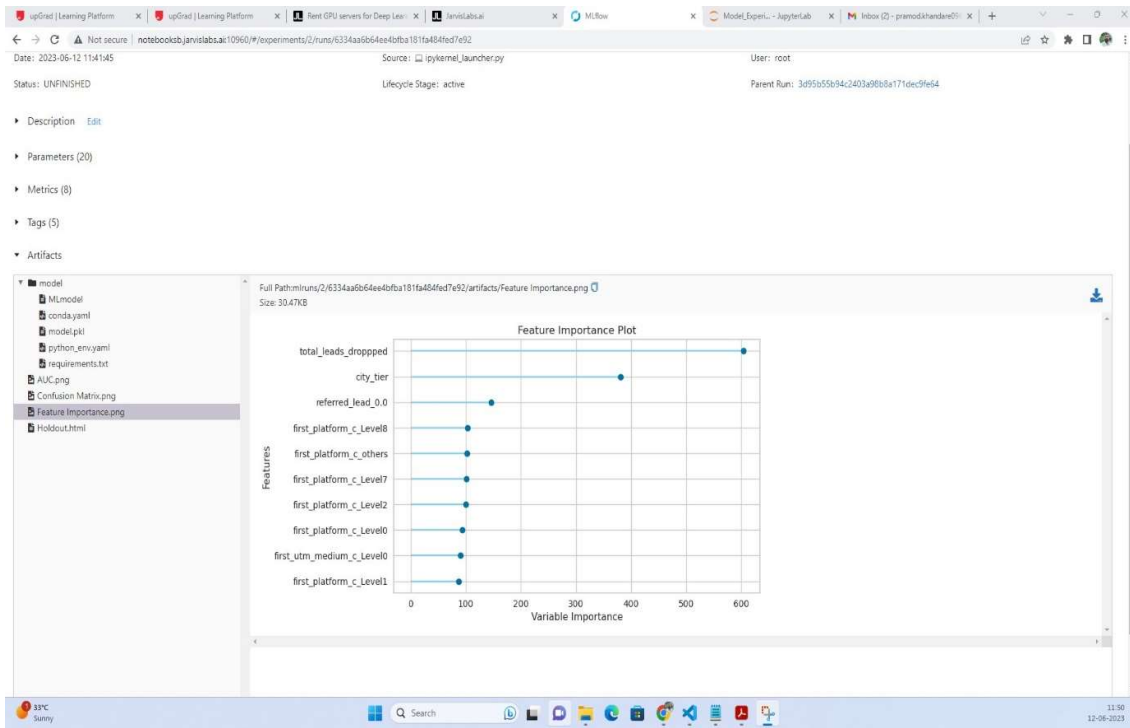
# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=logged_model, result_type='double')

# Predict on a Spark DataFrame.
columns = list(df.columns)
df.withColumn('predictions', loaded_model(*columns)).collect()
```

Predict on a Pandas DataFrame:

```
import mlflow
```



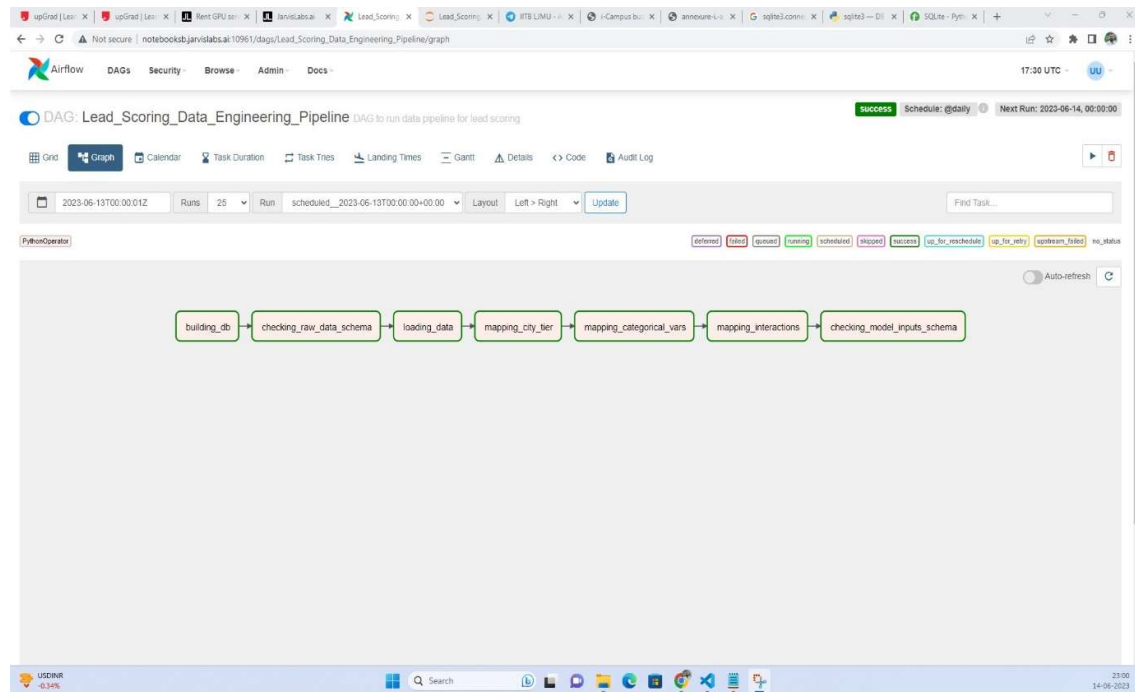


Full Path: minisuns/2/6334aa6b64ee4bfb181fa484ed7e92/artifacts/Holdout.html  
Size: 774B

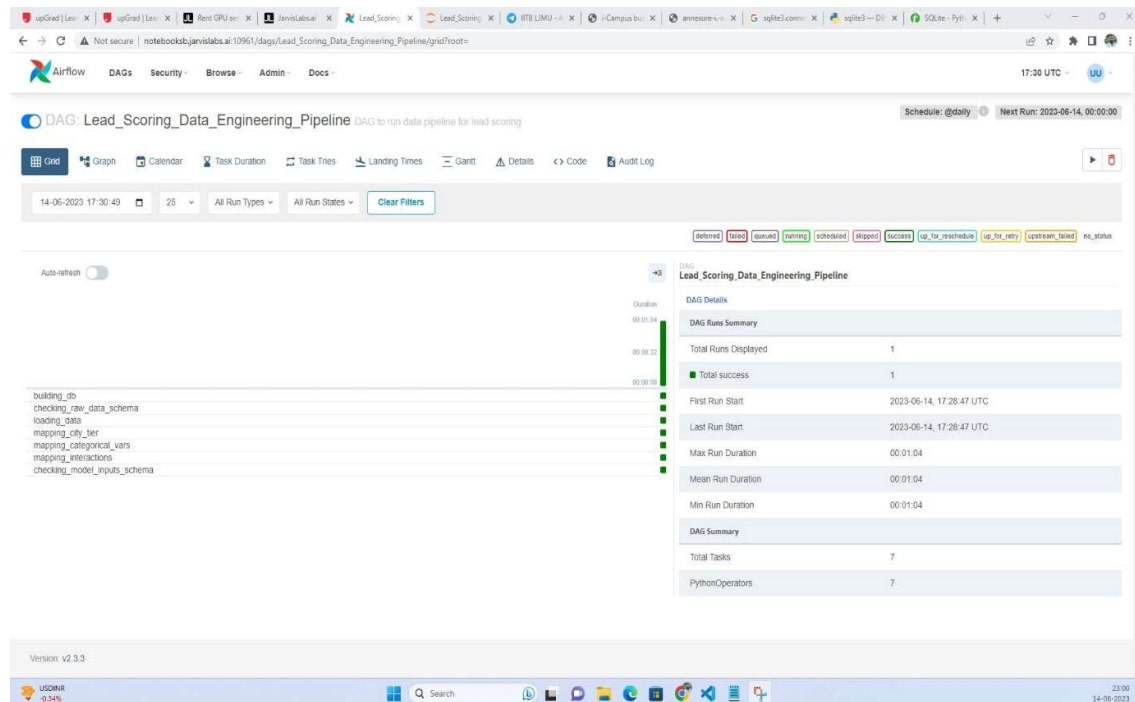
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Light Gradient Boosting Machine	0.738	0.8199	0.8317	0.7006	0.7606	0.4759	0.4845

# DATA PIPELINE

## Screenshot of successful execution Airflow DAG in graph



## Screenshot of Airflow UI grid



# TRAINING PIPELINE

screenshot of experiments with all the artifacts visible and also metric AUC

The screenshot shows the mlflow Experiments page for the experiment 'Lead\_scoring\_mlflow\_production'. The page displays a table of 15 matching runs. The table columns include Start Time, Duration, Run Name, User, Source, Version, Models, Metrics (False Negative, Precision, Precision@0), and Parameters (boosting\_type, class\_weight, colsample\_bytree). The runs are sorted by Start Time, showing runs from 5 hours ago to 14 hours ago. The metrics for all runs are consistent: False Negative is 6190, Precision is 0.739, and Precision@0 is 0.789. The parameters are also consistent: boosting\_type is gbm, class\_weight is None, and colsample\_bytree is 1.0.

Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision@0	boosting_type	class_weight	colsample_bytree
5 hours ago	6.7s	run_LightGBM	root	airflow	+	LightGBM/20	6190	0.739	0.789	gbm	None	1.0
5 hours ago	3.6s	run_LightGBM	root	airflow	+	LightGBM/19	-	-	-	gbm	None	1.0
5 hours ago	3.6s	run_LightGBM	root	airflow	+	LightGBM/18	-	-	-	gbm	None	1.0
5 hours ago	3.6s	run_LightGBM	root	airflow	+	LightGBM/17	-	-	-	gbm	None	1.0
5 hours ago	4.1s	run_LightGBM	root	airflow	+	LightGBM/16	-	-	-	gbm	None	1.0
5 hours ago	4.2s	run_LightGBM	root	airflow	+	LightGBM/15	-	-	-	gbm	None	1.0
5 hours ago	3.7s	run_LightGBM	root	airflow	+	LightGBM/14	-	-	-	gbm	None	1.0
5 hours ago	3.6s	run_LightGBM	root	airflow	+	LightGBM/13	-	-	-	gbm	None	1.0
5 hours ago	3.6s	run_LightGBM	root	airflow	+	LightGBM/12	-	-	-	gbm	None	1.0
5 hours ago	3.8s	run_LightGBM	root	airflow	+	LightGBM/11	-	-	-	gbm	None	1.0
7 hours ago	5.2s	run_LightGBM	root	airflow	+	LightGBM/9	6190	0.739	0.789	gbm	None	1.0
7 hours ago	5.3s	run_LightGBM	root	airflow	+	LightGBM/7	6190	0.739	0.789	gbm	None	1.0
7 hours ago	5.3s	run_LightGBM	root	airflow	+	LightGBM/5	6190	0.739	0.789	gbm	None	1.0
7 hours ago	6.1s	run_LightGBM	root	airflow	+	LightGBM/3	6190	0.739	0.789	gbm	None	1.0
14 hours ago	4.9s	run_LightGBM	root	airflow	+	LightGBM/1	6190	0.739	0.789	gbm	None	1.0

The screenshot shows the mlflow Models page for the model 'run\_LightGBM'. The page displays the model's metadata, including the date (2023-06-18 15:09:27), source (airflow), user (root), duration (5.2s), status (FINISHED), and lifecycle stage (active). The page also shows the model's artifacts, including the MLmodel file, conda.yaml, model.pkl, python\_env.yaml, and requirements.txt. The page provides a description of the model and instructions on how to make predictions using the logged model. The instructions include a code snippet for loading the model and making predictions on a Spark DataFrame, and another code snippet for making predictions on a Pandas DataFrame.

**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:/fa9b79e88e58419d8e7faac825c6dc20/models"

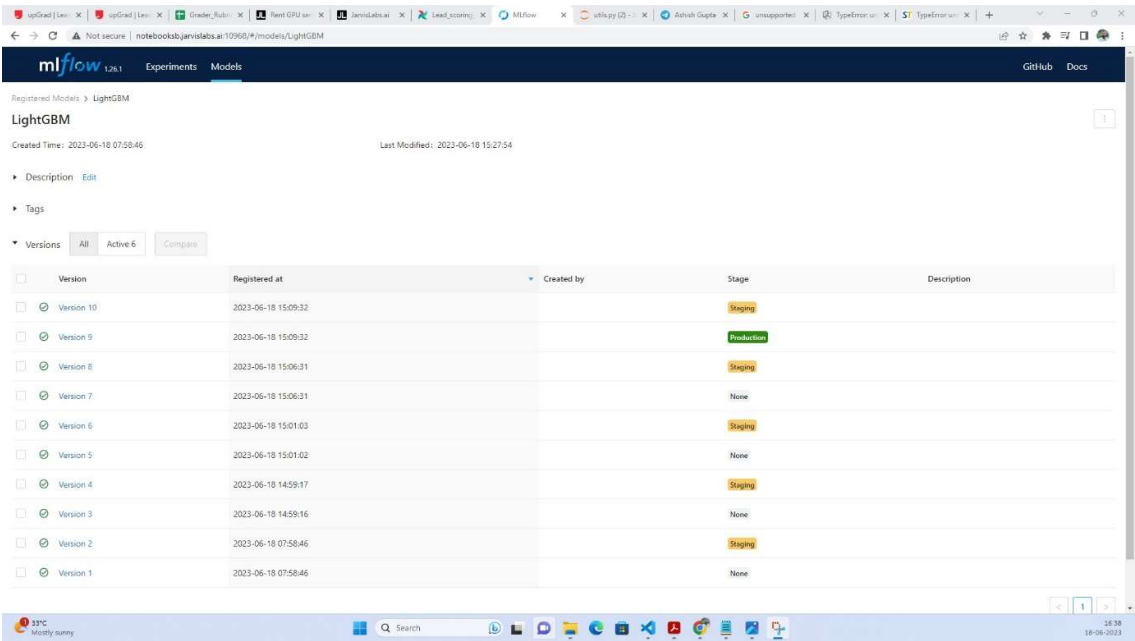
# Load model as a Spark UDF. Override result_type if the model does not return double values.
loaded_model = mlflow.pyfunc.spark_udf(spark, model_uri=logged_model, result_type='double')

# Predict on a Spark DataFrame.
columns = list(df.columns)
df.withColumn("predictions", loaded_model(*columns)).collect()
```

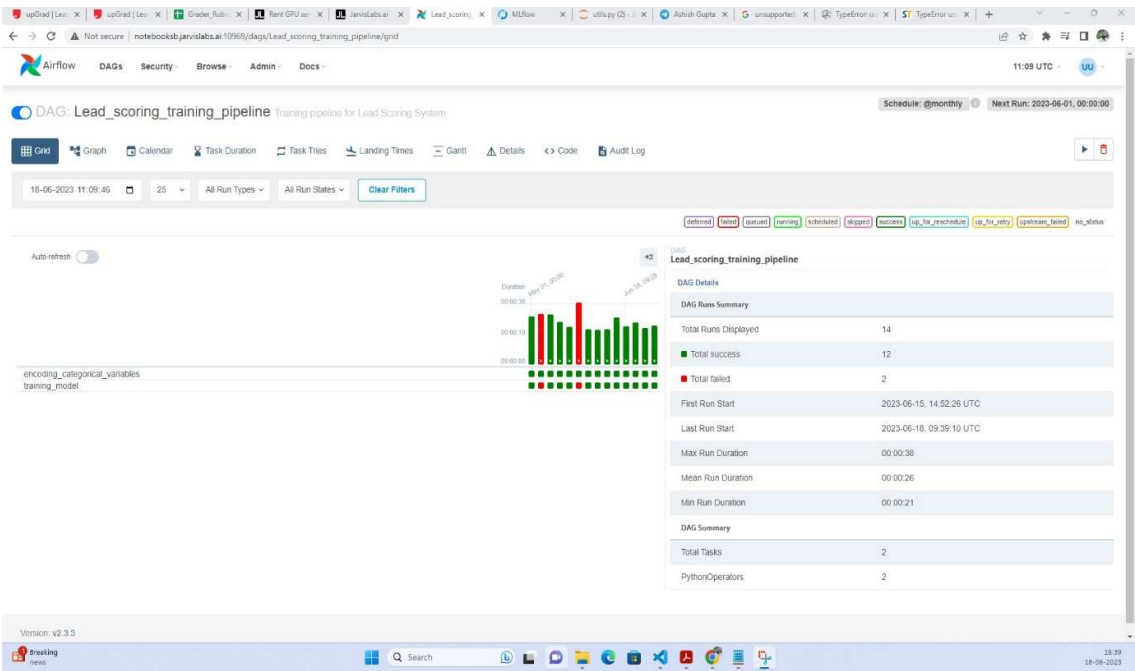
Predict on a Pandas DataFrame

```
import mlflow
```

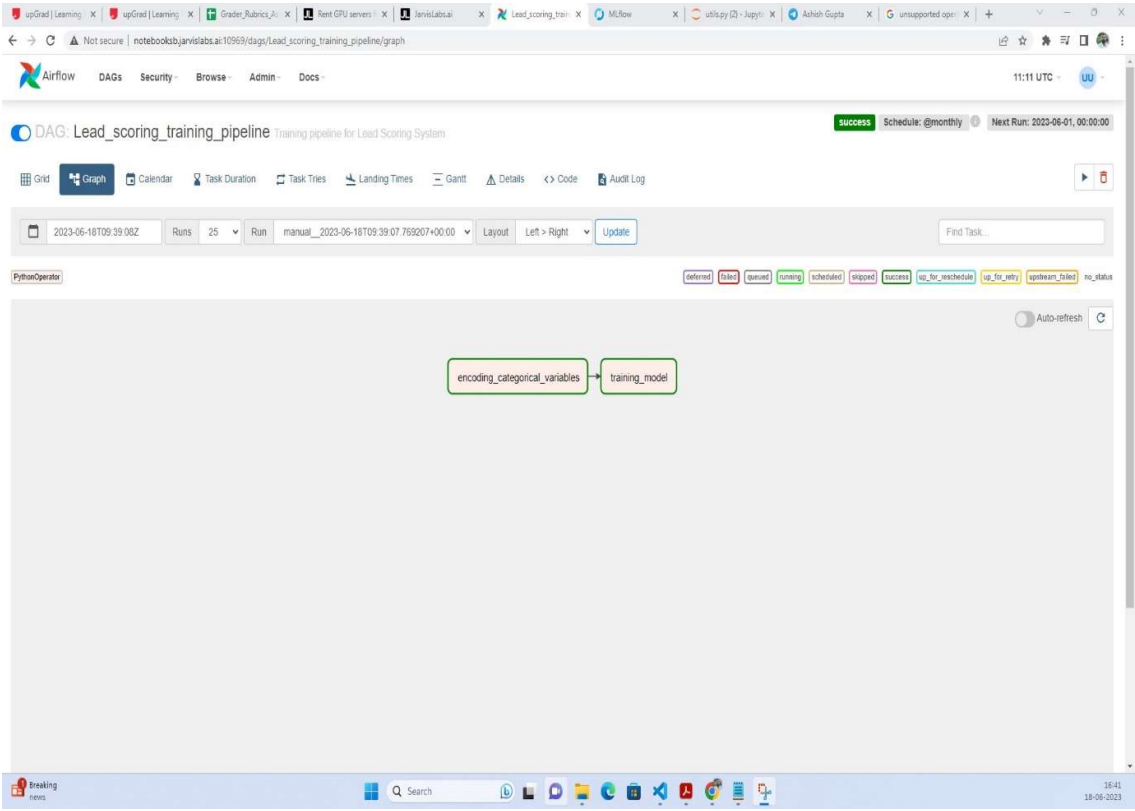
# Screenshot of model registry with model name and stage as 'Production'



# Screenshot of Airflow UI grid



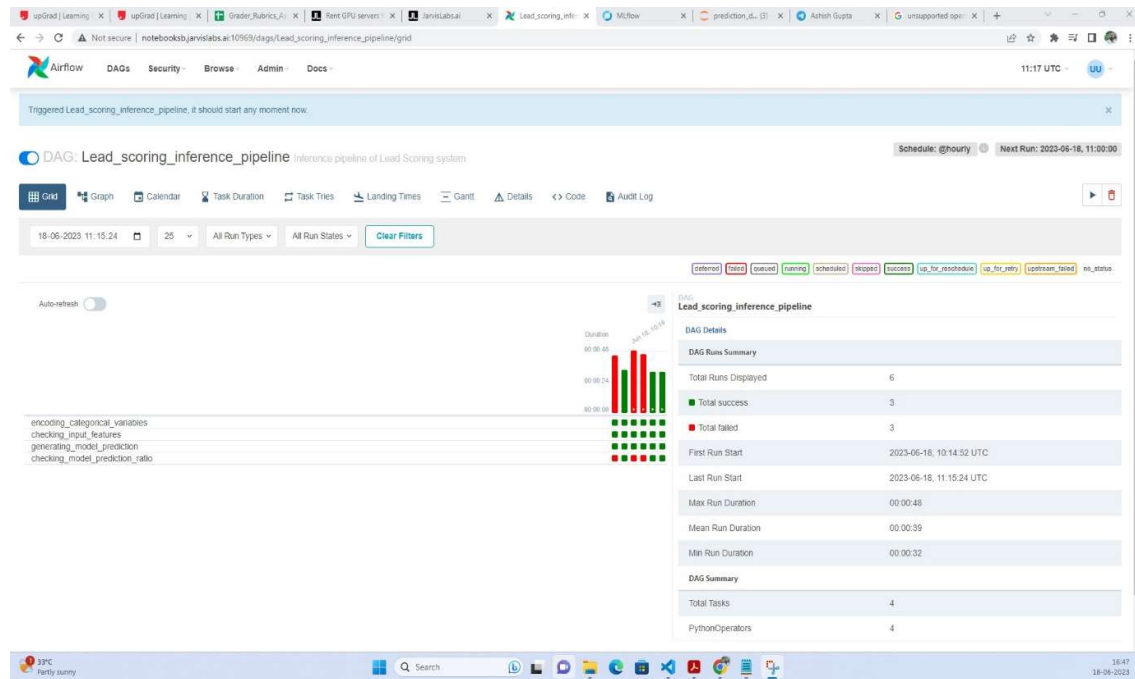
# Screenshot of successful execution Airflow DAG in graph



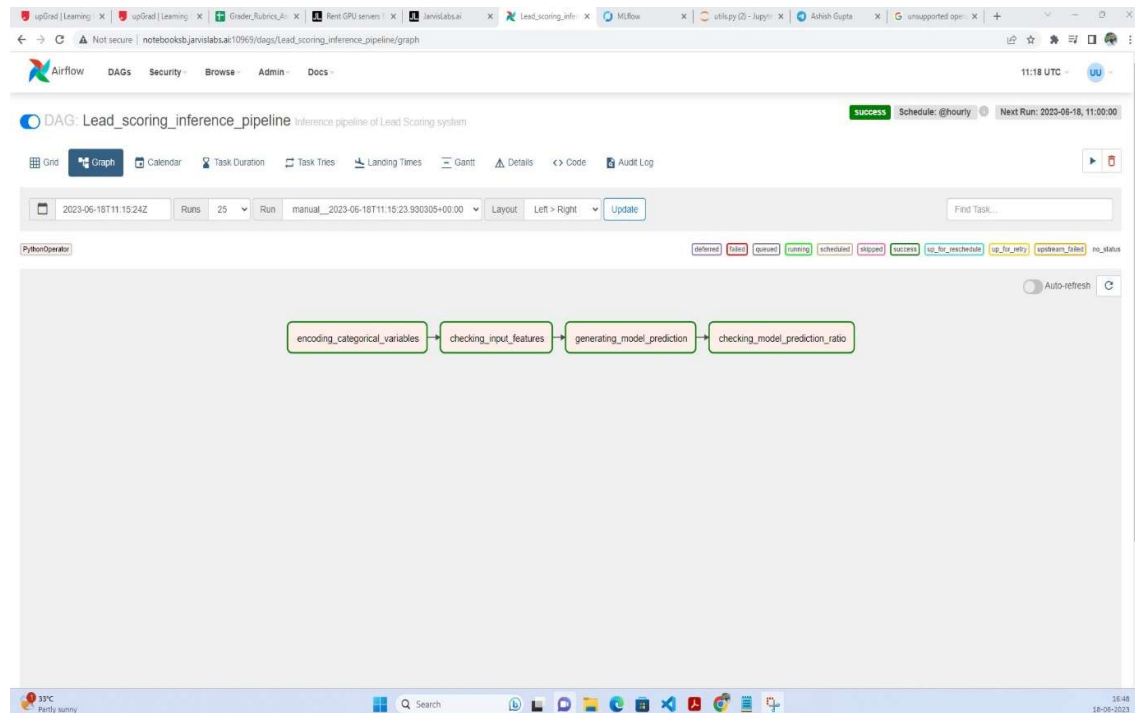


# INFERENCE PIPELINE

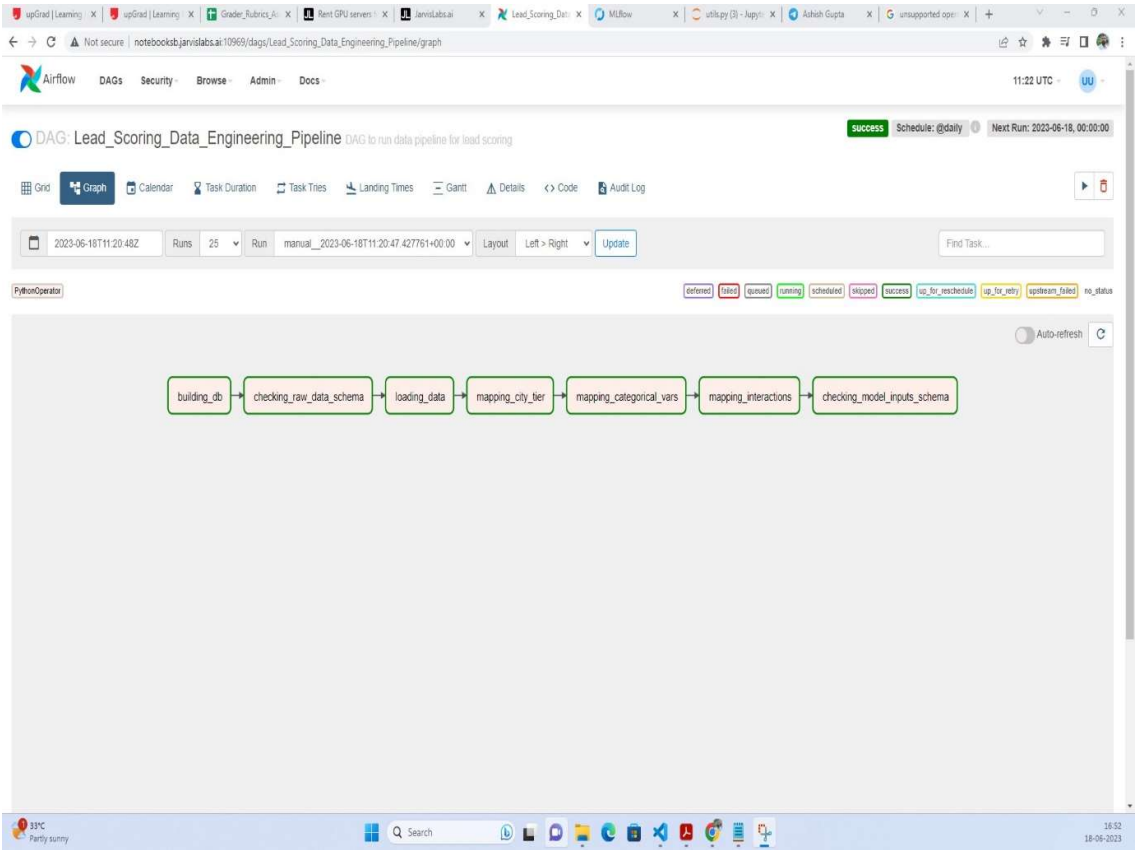
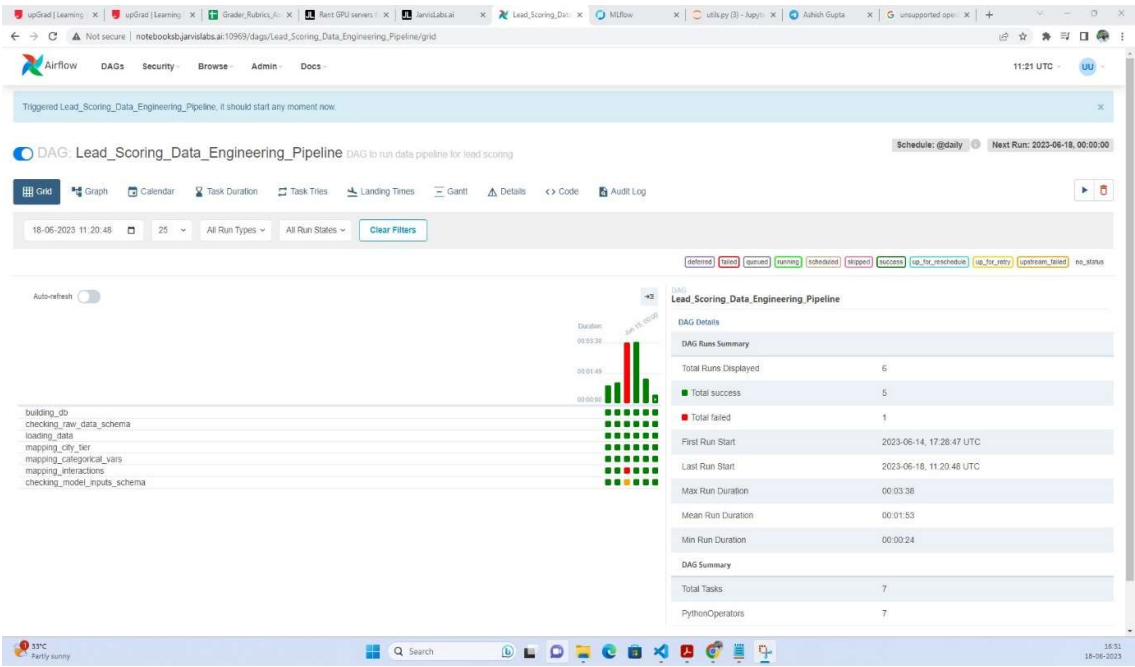
## Screenshot of Airflow UI grid



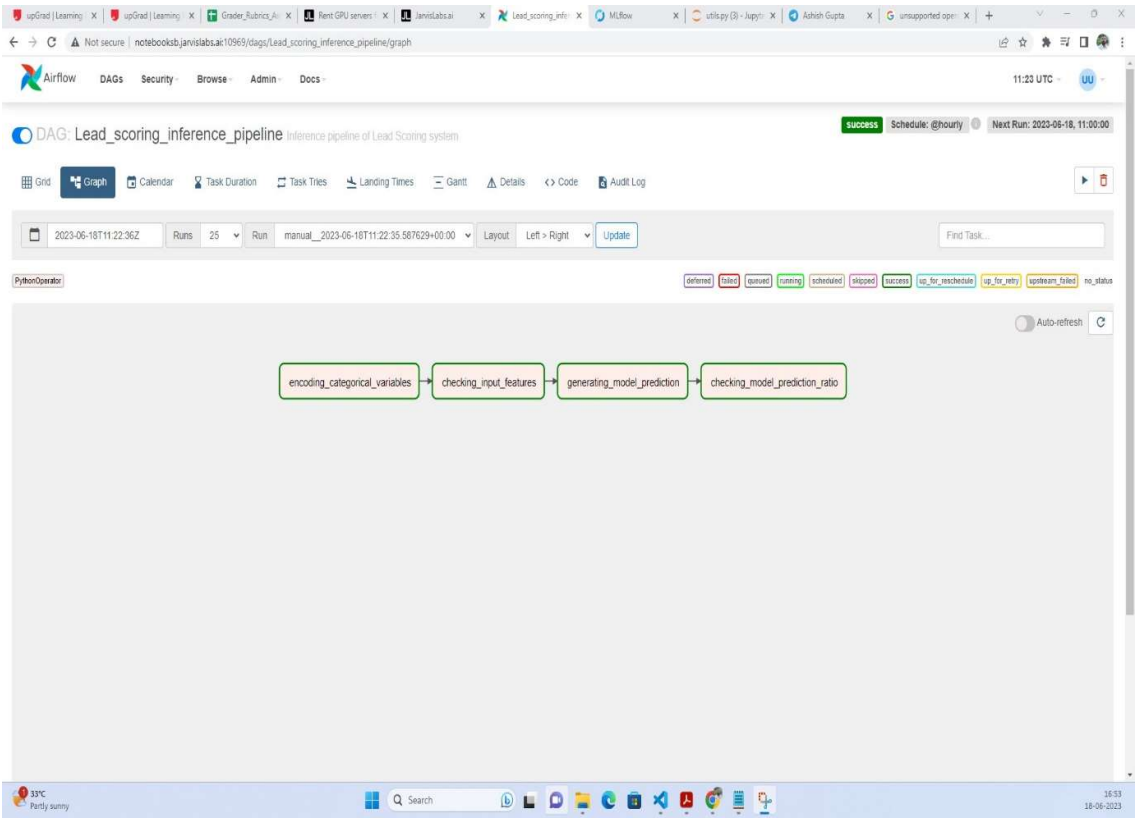
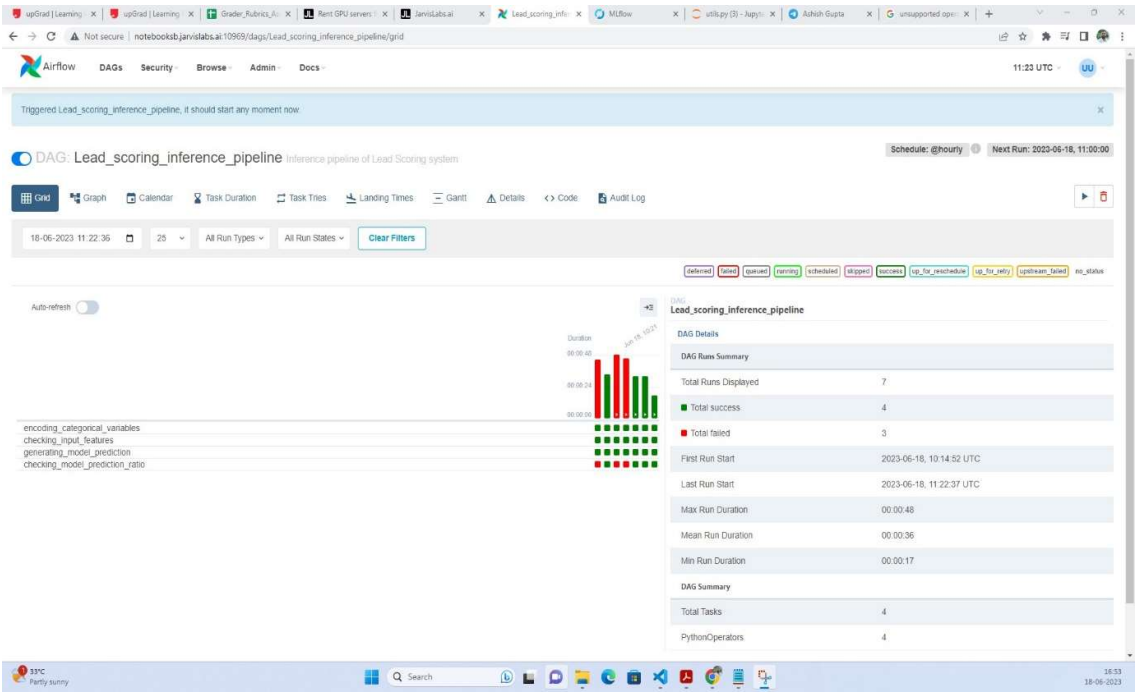
## Screenshot of successful execution Airflow DAG in graph



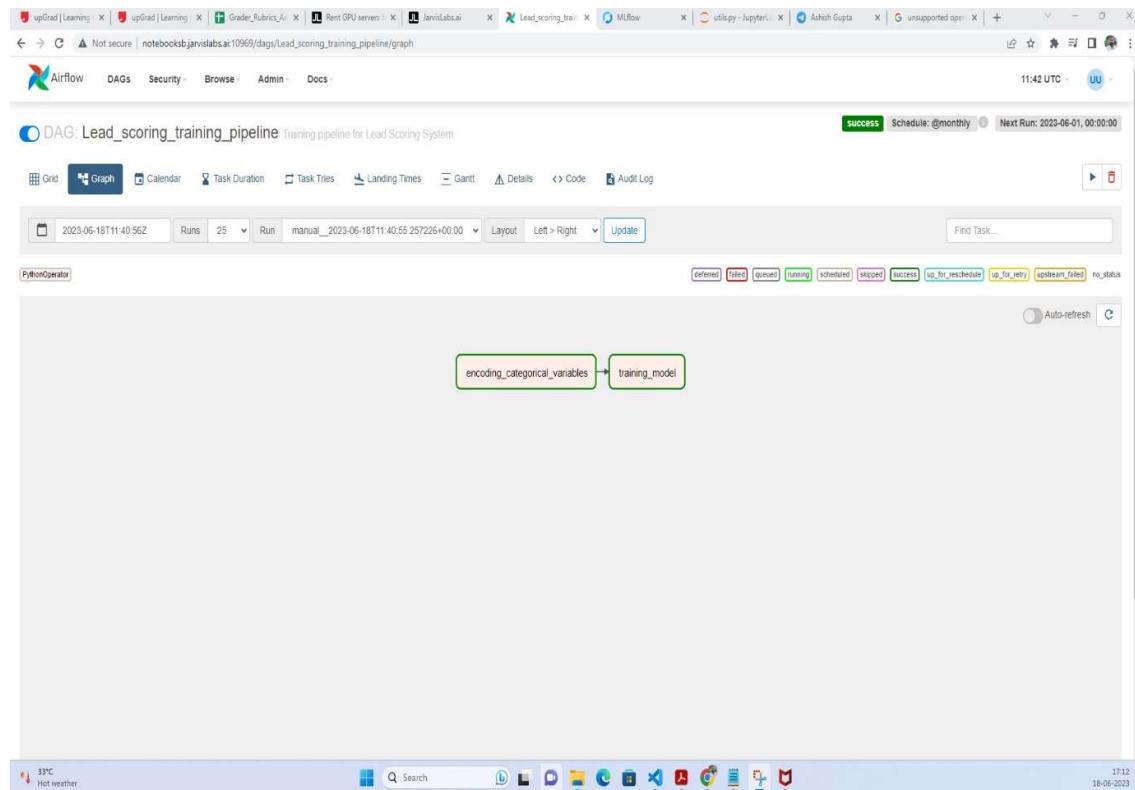
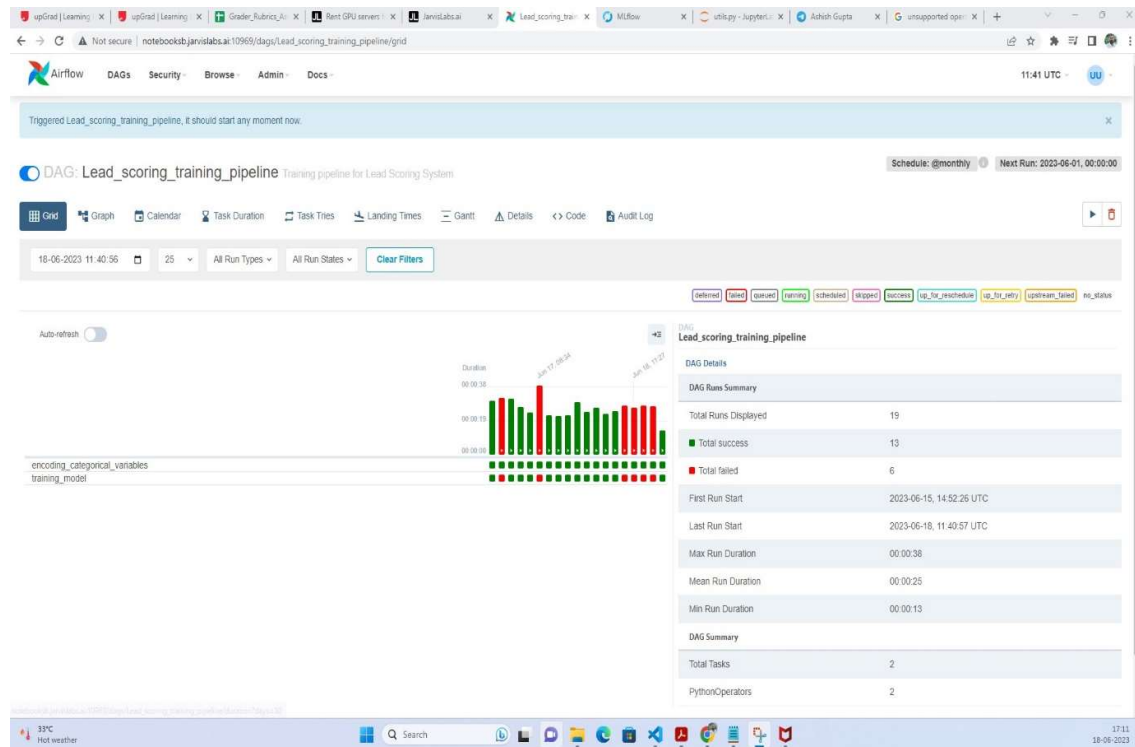
# Running data cleaning pipeline on leadscoring\_inference.csv



# Running data inference pipeline on cleaned leadscoring\_inference.csv



## Running training pipeline on cleaned leadscoring\_inference.csv



upGrad | Learning | upGrad | Learning | Grader\_Rubrics\_Air | Rent GPU servers | Jarvislabs.ai | Lead\_scoring\_train | MLflow | utility - JupyterLab | Ashish Gupta | unsupported open |

Not secure | notebooks.jarvislabs.ai/10968/#/experiments/1

mlflow 1.26.1 Experiments Models GitHub Docs

Experiments **Lead\_scoring\_mlflow\_production** Share

Search Experiments

Default **Lead\_scoring\_mlflow...**

Track machine learning training runs in experiments. Learn more

Experiment ID: 1

Description Edit

Refresh Compare Delete Download CSV Start Time All time

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

Showing 14 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	False Negative	Precision	Precision@0	boosting.type	class_weight	colsample_bytree
<input type="checkbox"/>	1 minute ago	3.6s	run_LightGBM	root	airflow	-	LightGBM/19	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	6 minutes ago	3.6s	run_LightGBM	root	airflow	-	LightGBM/18	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	6 minutes ago	3.6s	run_LightGBM	root	airflow	-	LightGBM/17	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	8 minutes ago	4.1s	run_LightGBM	root	airflow	-	LightGBM/16	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	9 minutes ago	4.2s	run_LightGBM	root	airflow	-	LightGBM/15	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	14 minutes ago	3.7s	run_LightGBM	root	airflow	-	LightGBM/14	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	15 minutes ago	3.6s	run_LightGBM	root	airflow	-	LightGBM/13	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	17 minutes ago	3.6s	run_LightGBM	root	airflow	-	LightGBM/12	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	17 minutes ago	3.8s	run_LightGBM	root	airflow	-	LightGBM/11	-	-	-	gbdt	None	1.0
<input type="checkbox"/>	2 hours ago	5.2s	run_LightGBM	root	airflow	-	LightGBM/9	6190	0.739	0.789	gbdt	None	1.0
<input type="checkbox"/>	2 hours ago	5.3s	run_LightGBM	root	airflow	-	LightGBM/7	6190	0.739	0.789	gbdt	None	1.0
<input type="checkbox"/>	2 hours ago	5.3s	run_LightGBM	root	airflow	-	LightGBM/5	6190	0.739	0.789	gbdt	None	1.0
<input type="checkbox"/>	2 hours ago	6.1s	run_LightGBM	root	airflow	-	LightGBM/3	6190	0.739	0.789	gbdt	None	1.0
<input type="checkbox"/>	9 hours ago	4.9s	run_LightGBM	root	airflow	-	LightGBM/1	6190	0.739	0.789	gbdt	None	1.0

Load more

33°C High weather 17:13 18-06-2023

upGrad | Learning | upGrad | Learning | Grader\_Rubrics\_Air | Rent GPU servers | Jarvislabs.ai | Lead\_scoring\_train | MLflow | utility - JupyterLab | Ashish Gupta | unsupported open |

Not secure | notebooks.jarvislabs.ai/10968/#/runs/6d0be147999340658584f92ef941c905

mlflow 1.26.1 Experiments Models GitHub Docs

Lead\_scoring\_mlflow\_production > run\_LightGBM

run\_LightGBM

Date: 2023-06-18 17:11:05 Source: airflow User: root

Duration: 3.6s Status: FINISHED Lifecycle Stage: active

Description Edit

Parameters (20)

Metrics

Tags

Artifacts

models

- MLmodel
- conda.yaml
- model.pkl
- python\_env.yaml
- requirements.txt

Full Path: /home/airflow/dags/Lead\_Scoring\_Training\_Pipeline/miruna/1/6d0be147999340658584f92ef941c905/artifacts/models

LightGBM v19 Registered on 2023-06-18

### 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:/6d0be147999340658584f92ef941c906/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
```

17:13 18-06-2023