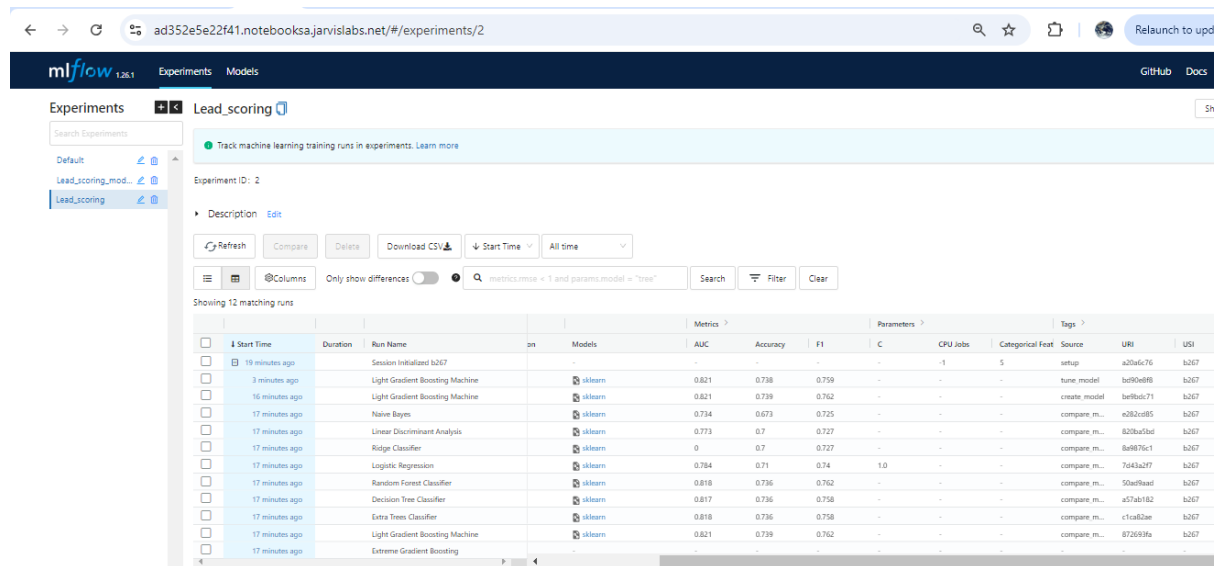


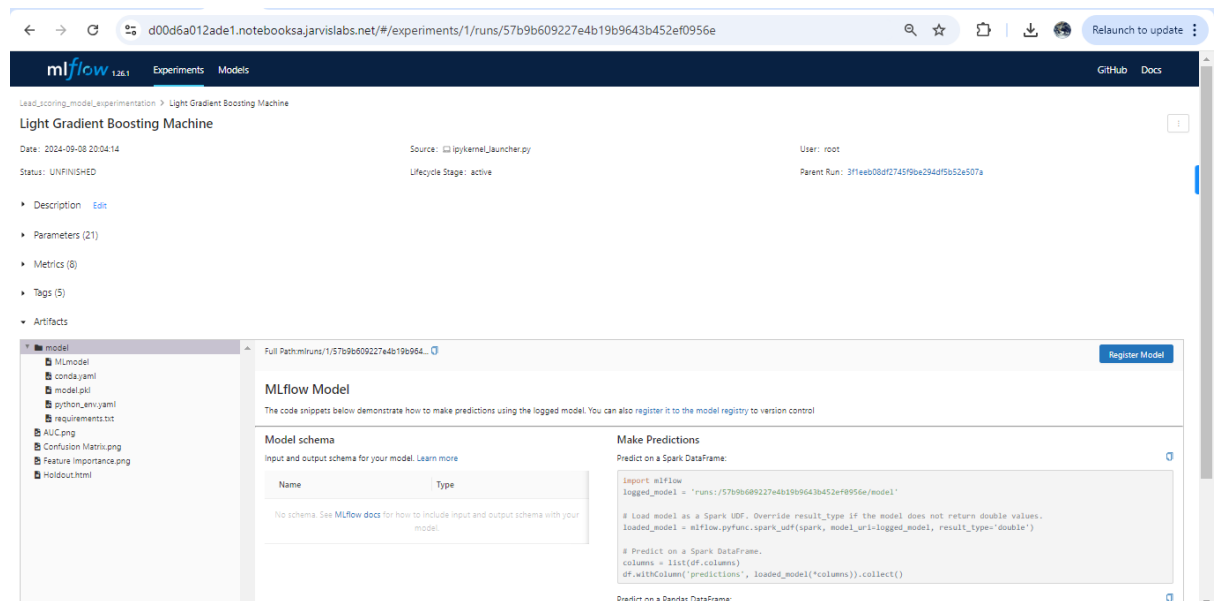
MLFlow Screenshots

Model Experimentation:



The screenshot shows the MLFlow Experiments page for the 'Lead_scoring' experiment. The interface includes a search bar, a list of experiments, and a table of runs. The table displays various metrics such as AUC, Accuracy, F1, and CPU Jobs for different models like Light Gradient Boosting Machine, Naive Bayes, and Ridge Classifier.

Start Time	Duration	Run Name	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source	URI	URI
19 minutes ago		Session Initialized b267	-	-	-	-	-	-	5	setup	a204dc76	b267
3 minutes ago		Light Gradient Boosting Machine	sklearn	0.821	0.738	0.759	-	-	-	tune.model	b950d8f8	b267
16 minutes ago		Light Gradient Boosting Machine	sklearn	0.821	0.739	0.762	-	-	-	create.model	bef8bd71	b267
17 minutes ago		Naive Bayes	sklearn	0.734	0.673	0.725	-	-	-	compare.m...	e282c885	b267
17 minutes ago		Linear Discriminant Analysis	sklearn	0.773	0.7	0.727	-	-	-	compare.m...	820ba5bd	b267
17 minutes ago		Ridge Classifier	sklearn	0	0.7	0.727	-	-	-	compare.m...	8a8876c1	b267
17 minutes ago		Logistic Regression	sklearn	0.784	0.71	0.74	1.0	-	-	compare.m...	7d43a297	b267
17 minutes ago		Random Forest Classifier	sklearn	0.818	0.736	0.762	-	-	-	compare.m...	50a9f8ad	b267
17 minutes ago		Decision Tree Classifier	sklearn	0.817	0.736	0.758	-	-	-	compare.m...	a57ab182	b267
17 minutes ago		Extra Trees Classifier	sklearn	0.818	0.736	0.758	-	-	-	compare.m...	c1a8d3ae	b267
17 minutes ago		Light Gradient Boosting Machine	sklearn	0.821	0.739	0.762	-	-	-	compare.m...	872693fa	b267
17 minutes ago		Extreme Gradient Boosting	-	-	-	-	-	-	-	-	-	-



The screenshot shows the MLFlow Model page for the 'Light Gradient Boosting Machine' model. The page displays the model's schema, metrics, and tags. It also includes a section for 'Make Predictions' with code snippets for using the model in a Spark DataFrame and a Pandas DataFrame.

```
import mlflow
logged_model = 'runs:/57b9b609227e4b19b9643b452ef0956e/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, logged_model, result_type='double')

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

Lead_Scoring_Training_Pipeline > Lead_Scoring_Training_Pipeline0809_2024_00_00_00

Lead_Scoring_Training_Pipeline0809_2024_00_00_00

Date: 2024-08-08 06:03:06

Source: airflow

User: root

Duration: 4.9s

Status: FINISHED

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

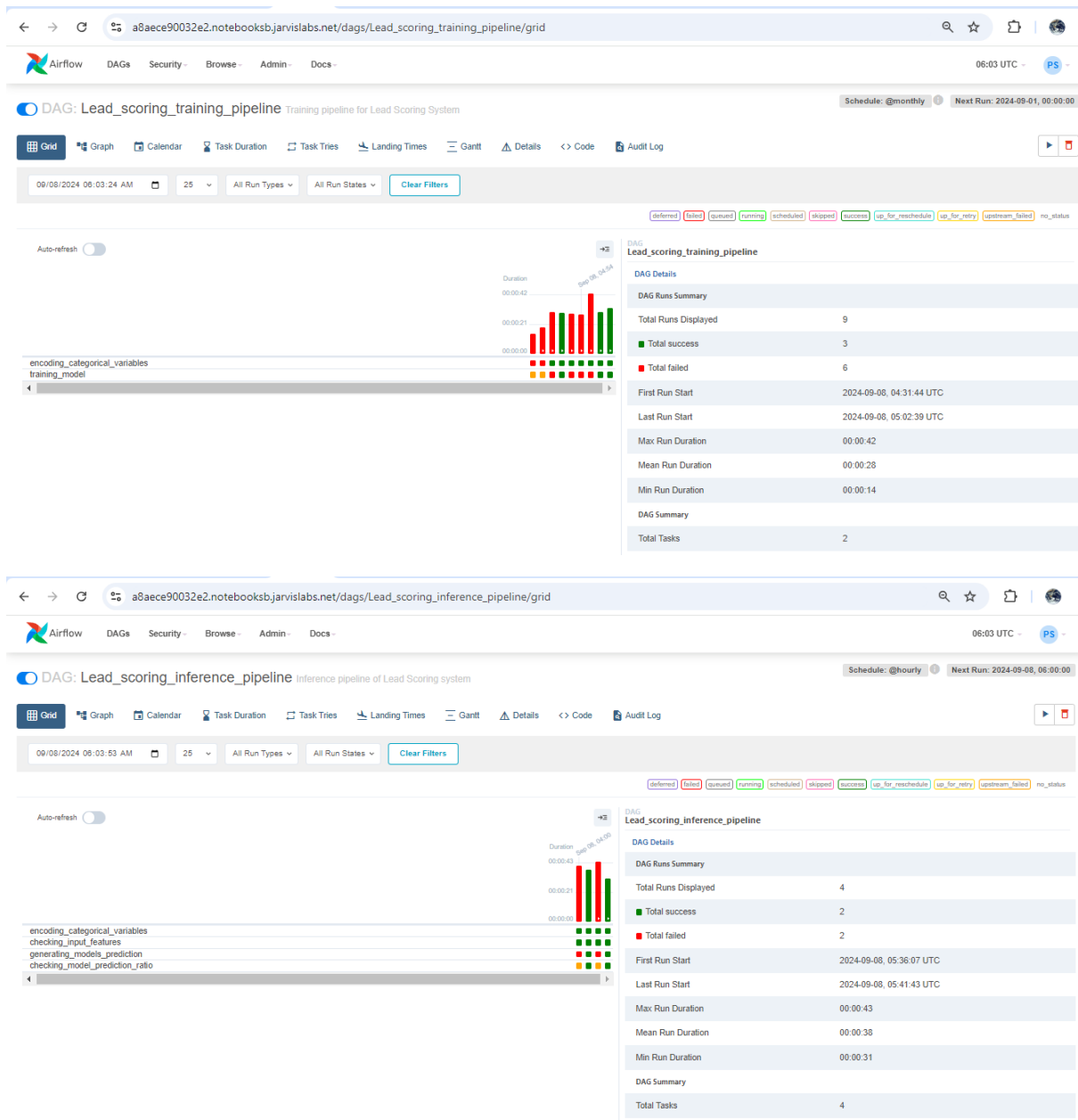
subsample_freq	0
----------------	---

▼ Metrics (9)

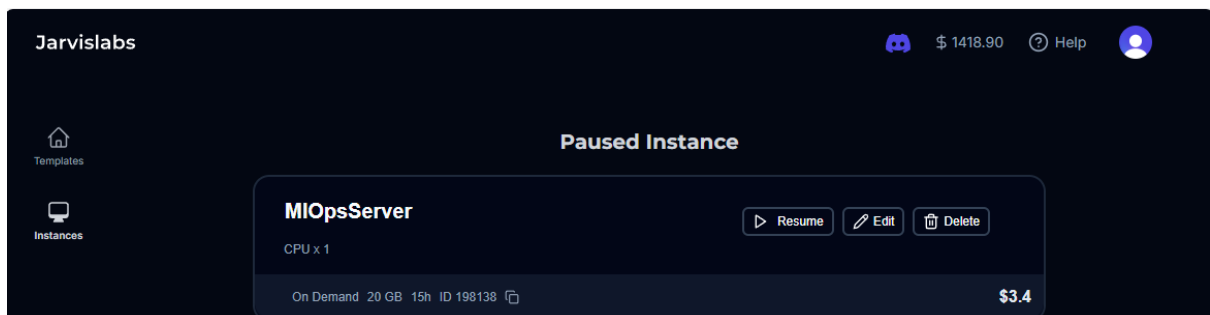
Name	Value
AUC	0.751
False Negative	3915
False Positive	8436
Precision	0.741
Recall	0.751
True Negative	15199
True Positive	20243
f1	0.739
test_accuracy	0.742

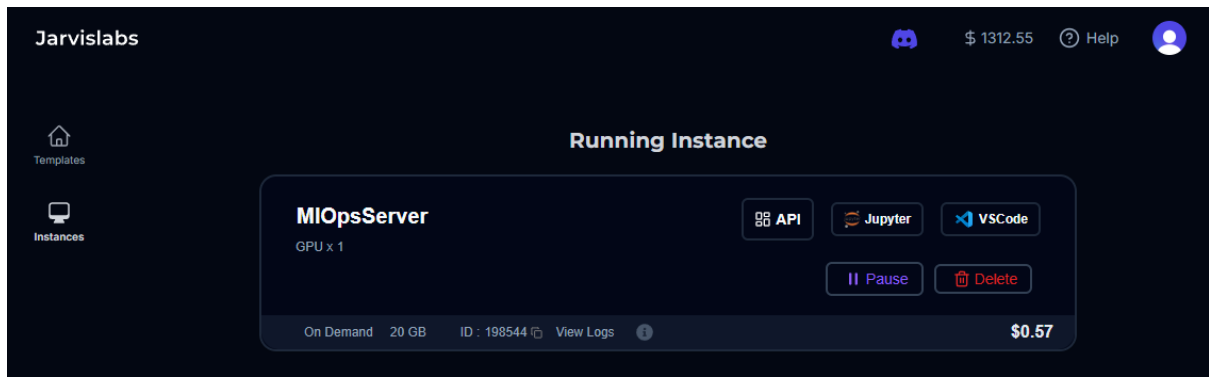
▸ Tags

▼ Artifacts



Jarvis cost:





27-Aug-2024:12.43 bst

