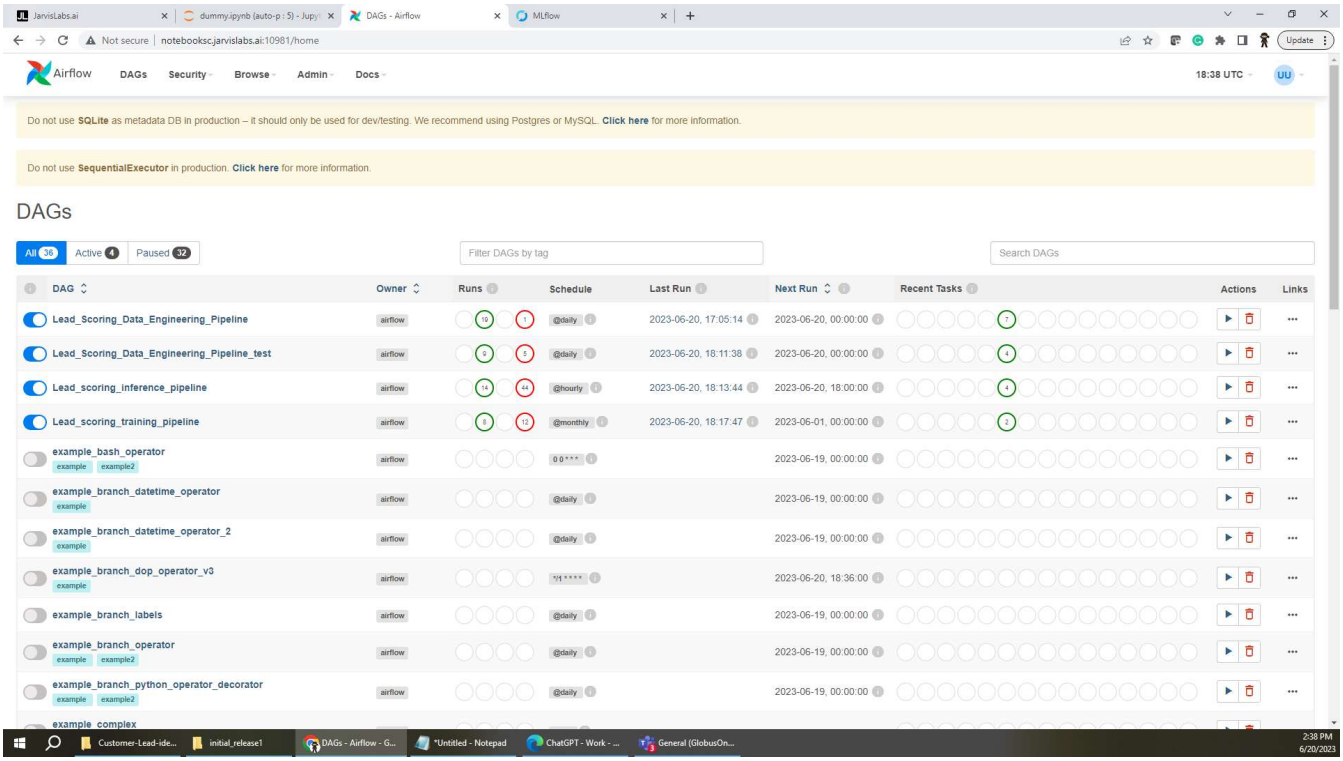
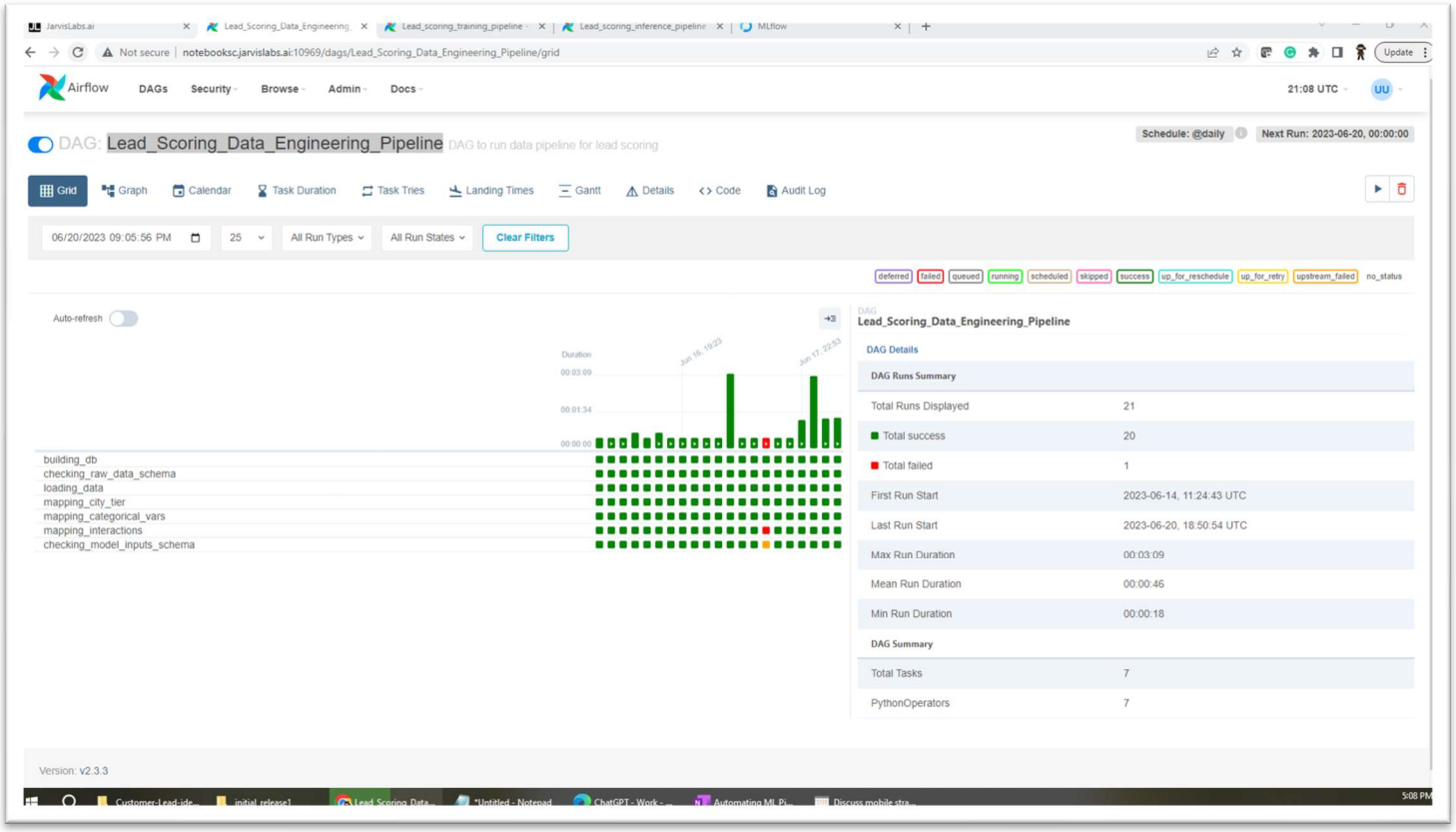
MLOps Assignment

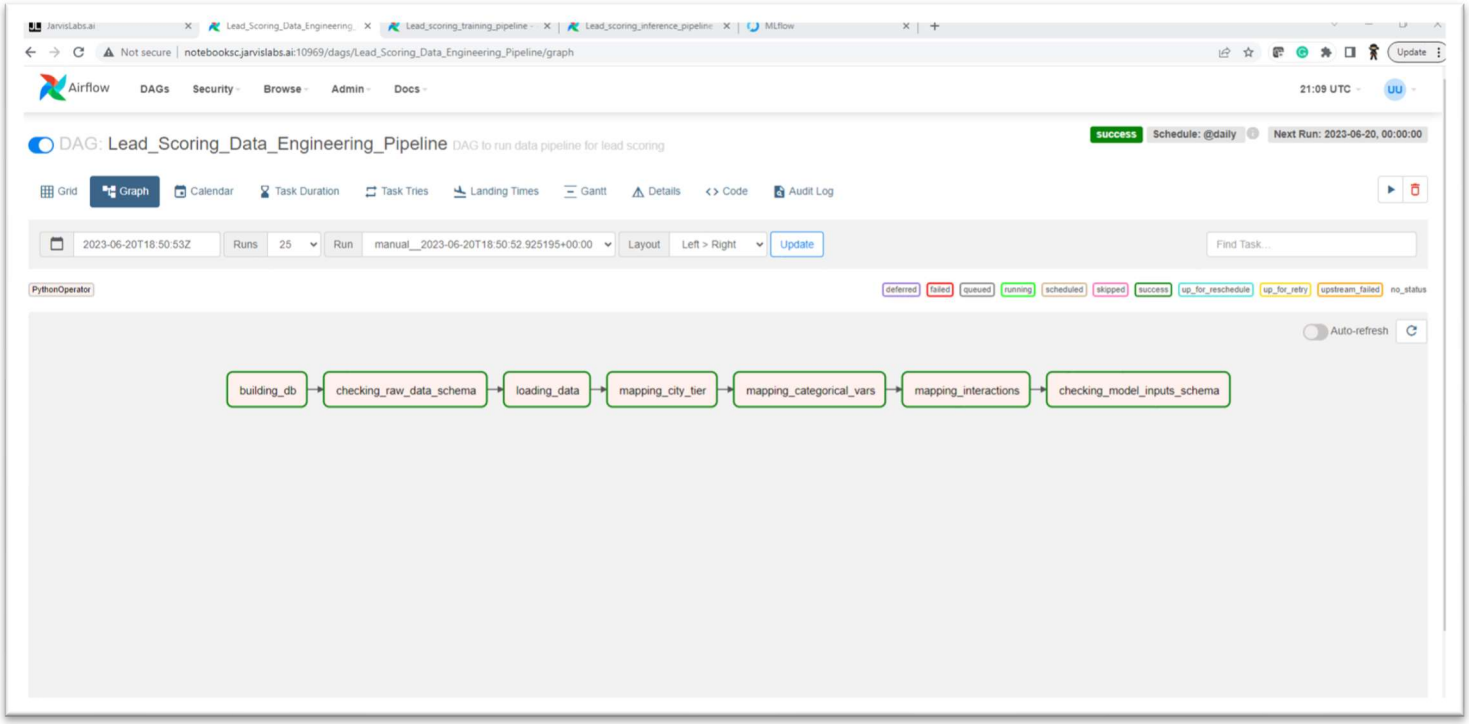
Contents

1. Initial DAG Landing Page

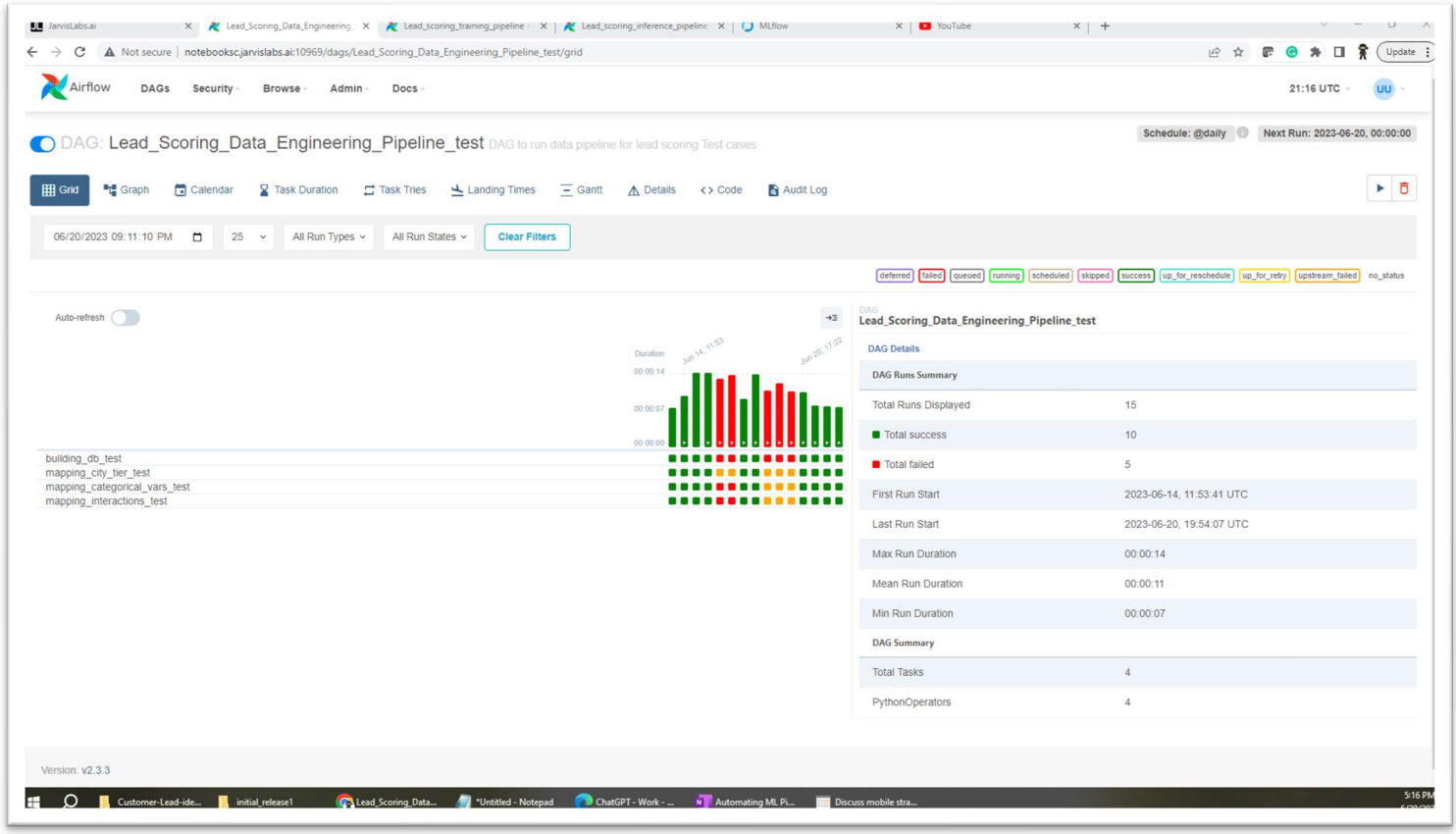


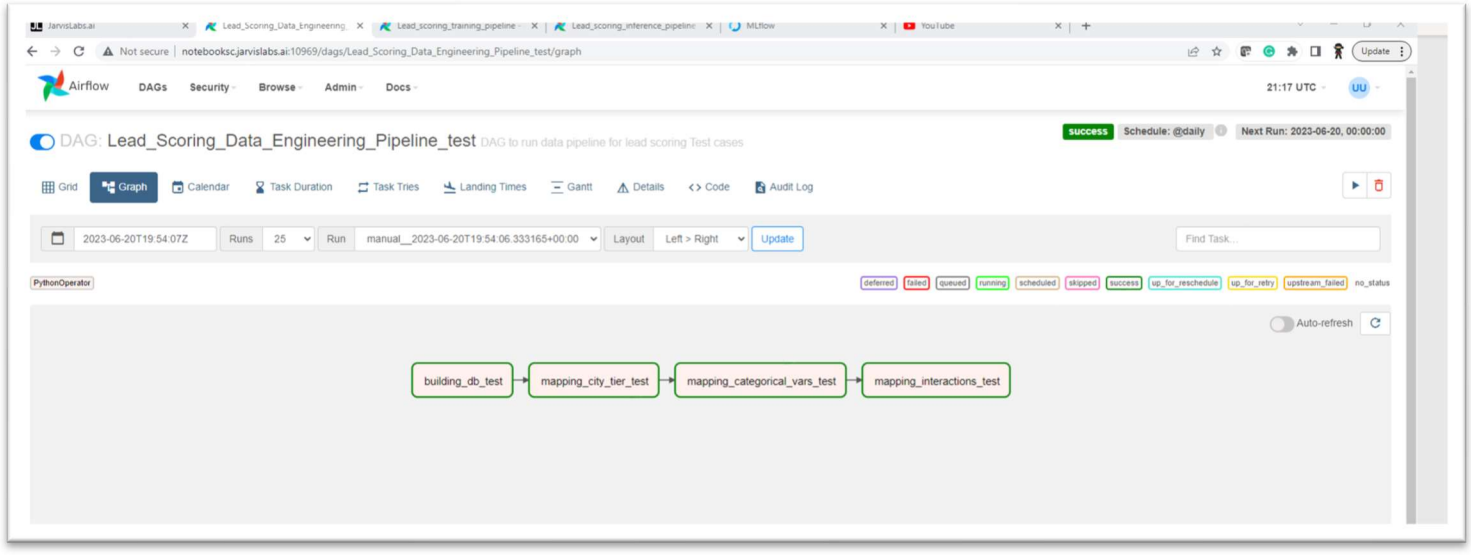
2. Lead_Scoring_Data_Engineering_Pipeline



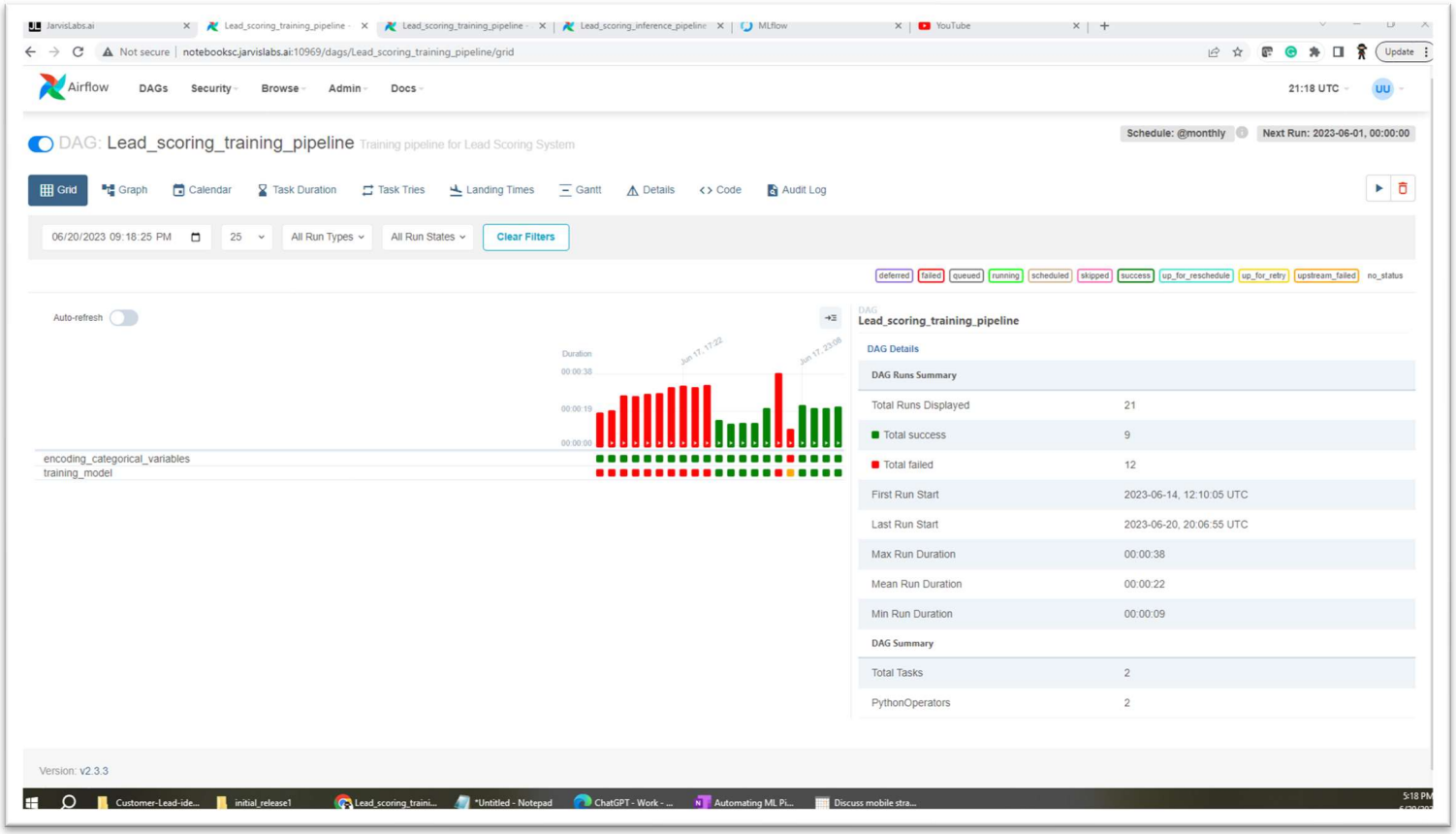


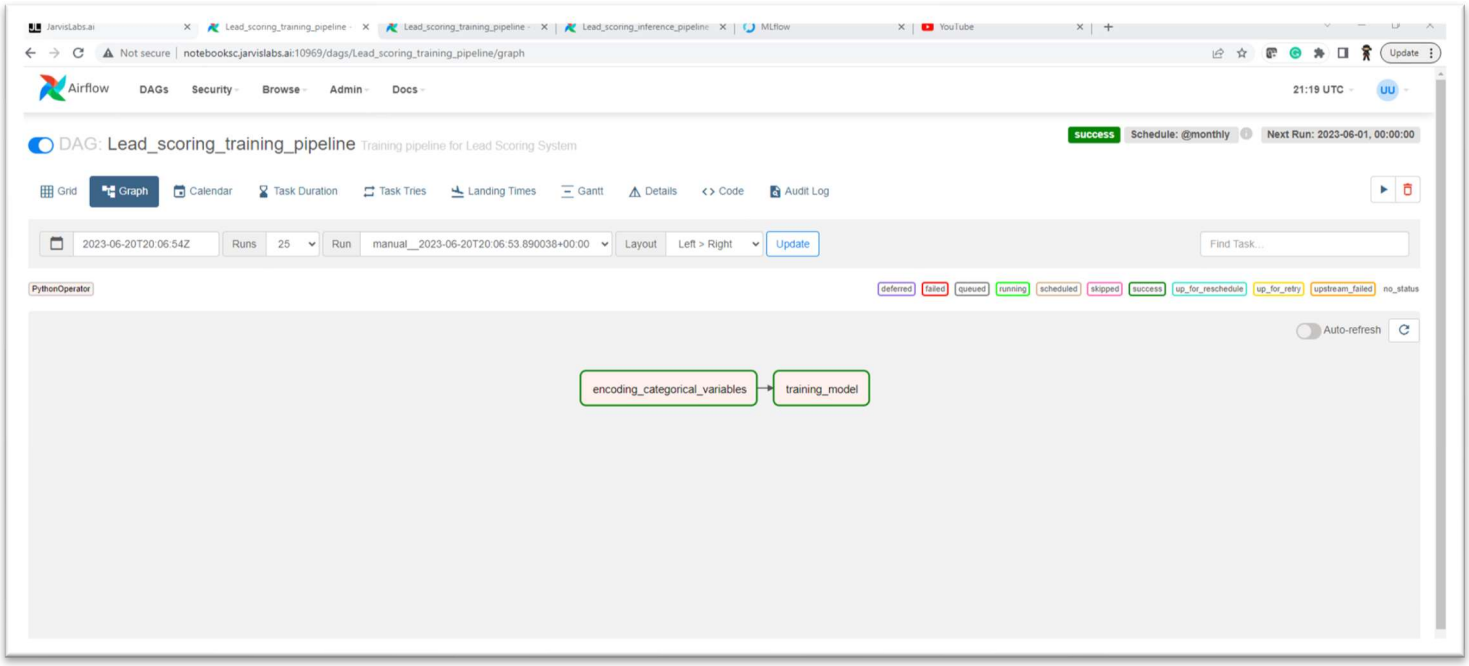
3. Lead_Scoring_Data_Engineering_Pipeline



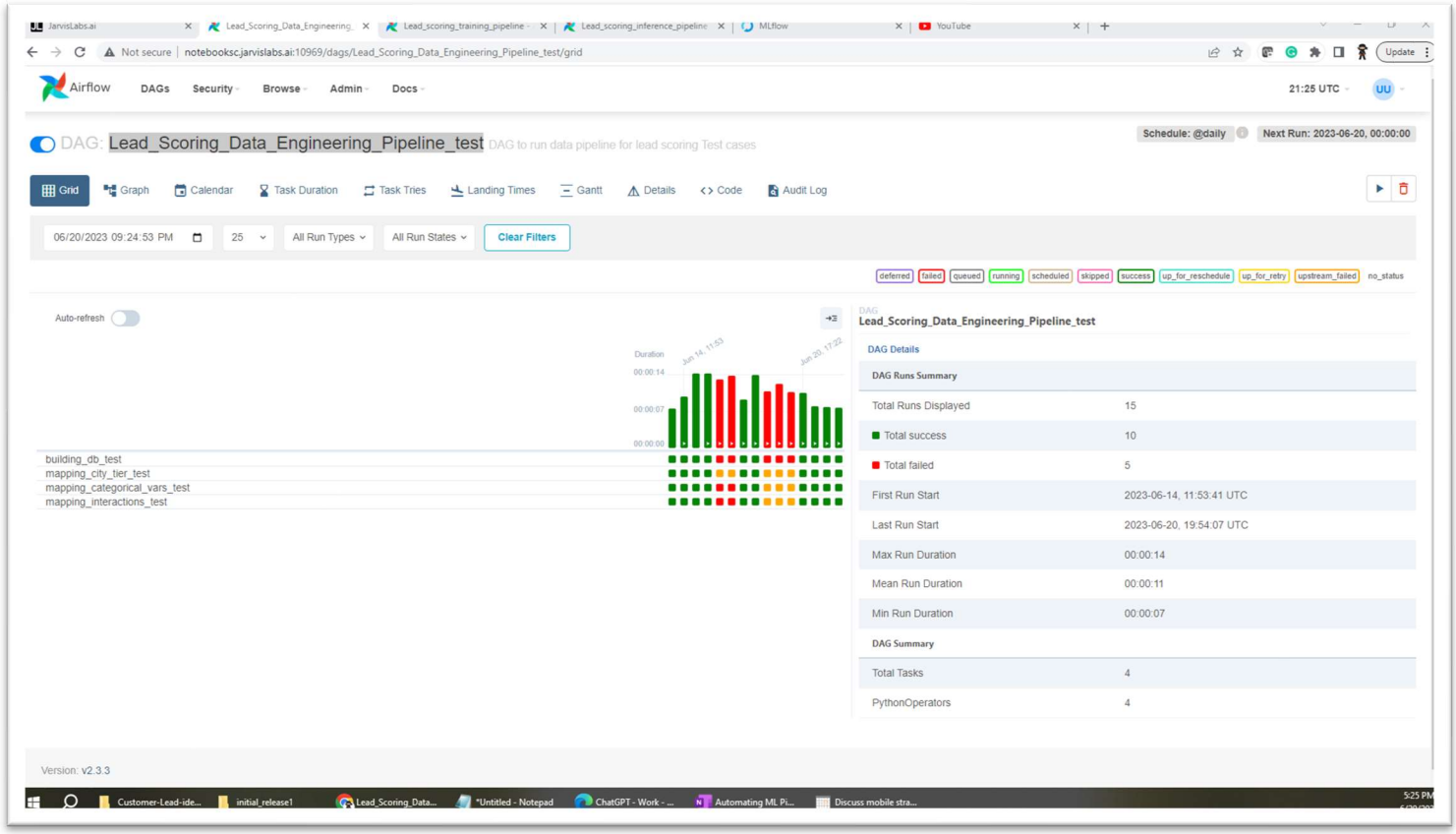


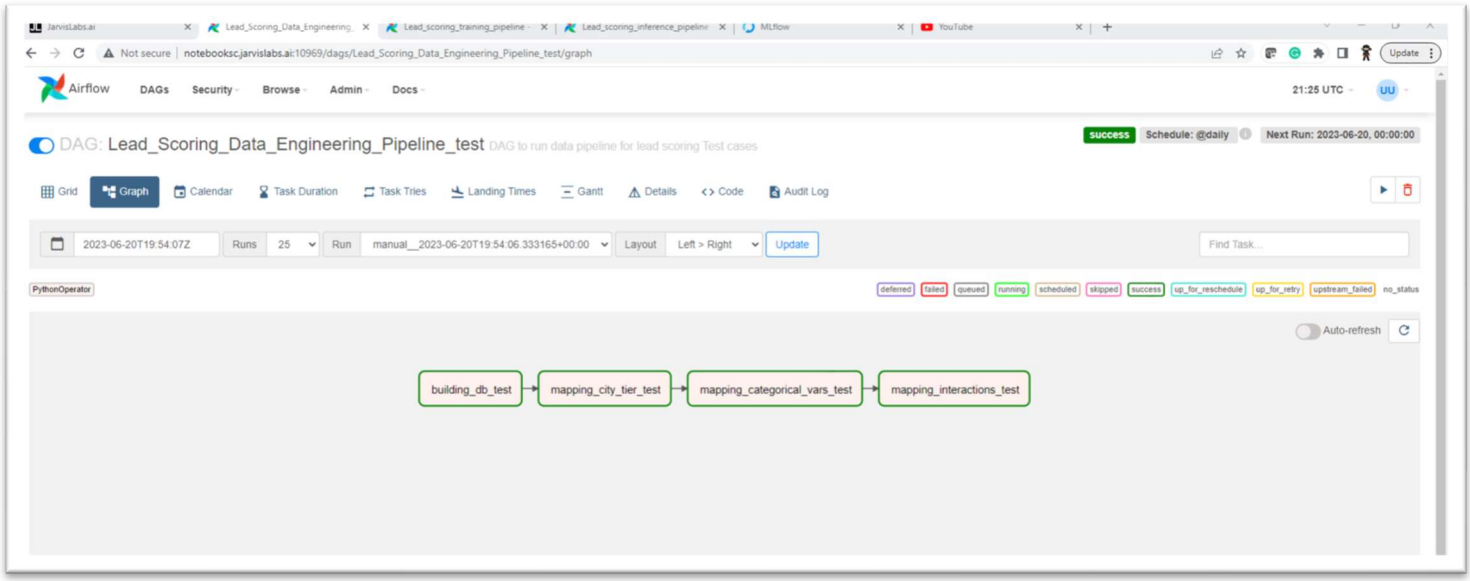
4. Lead_Scoring_Data_Engineering_Pipeline



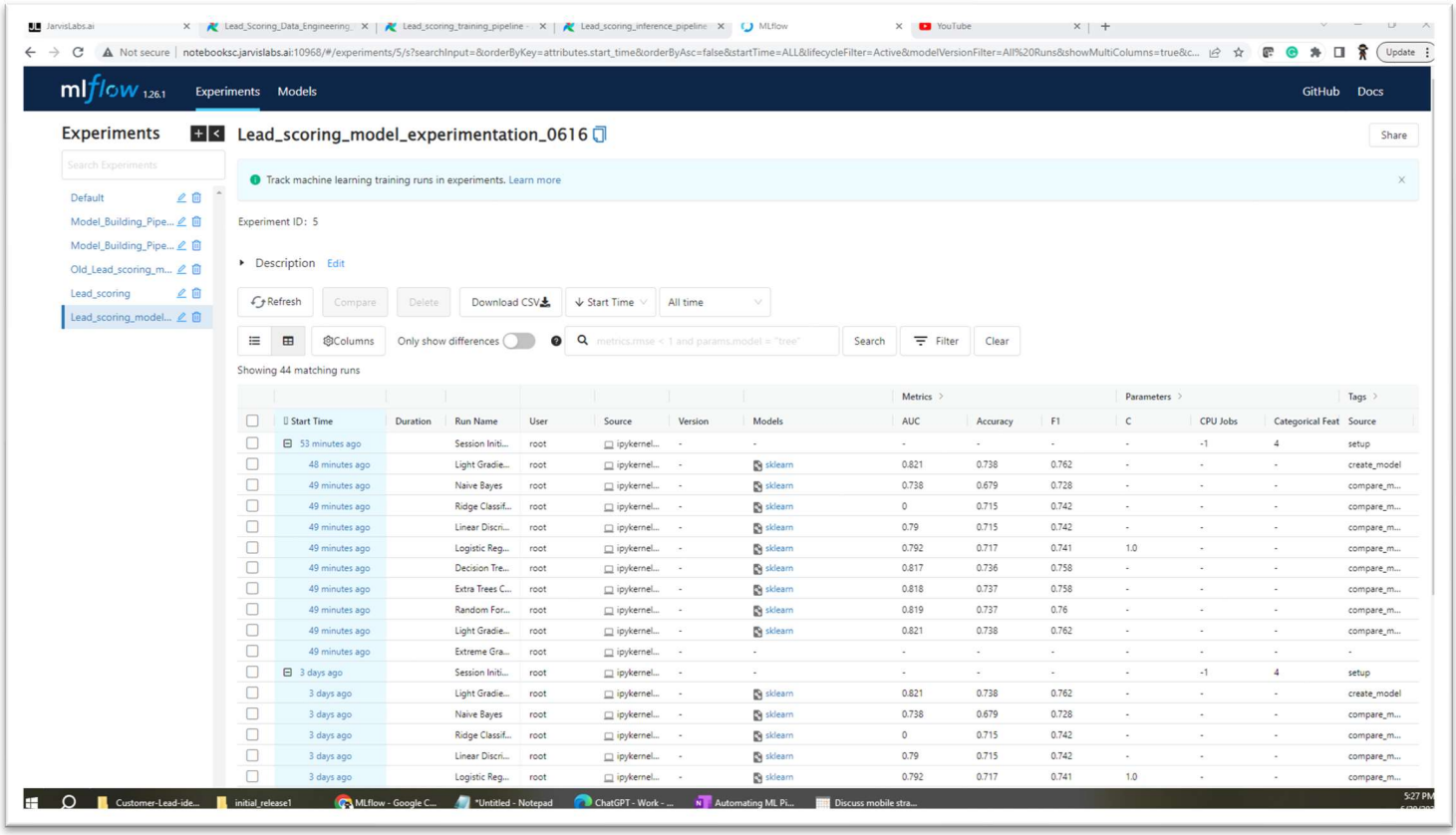


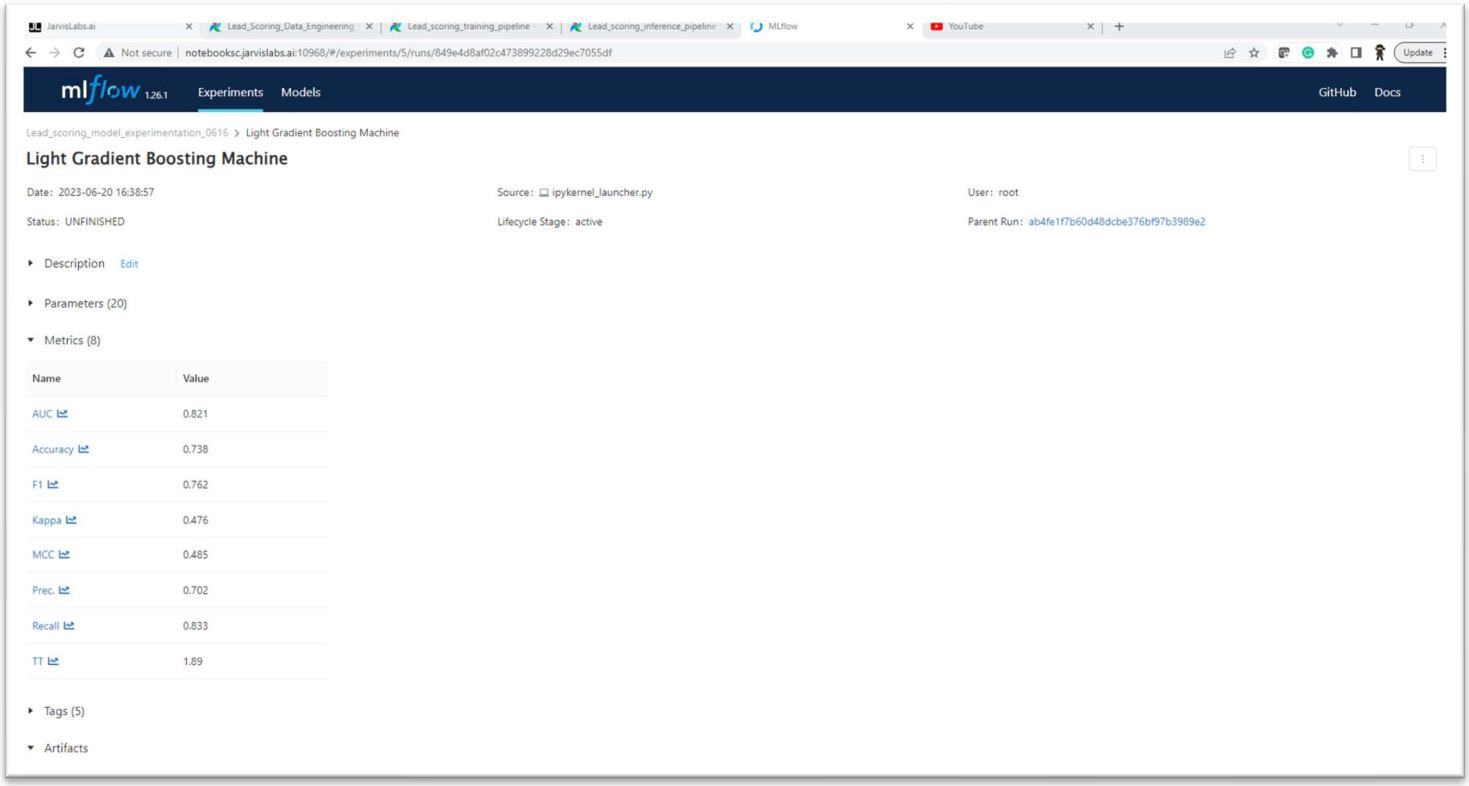
5. Lead_Scoring_Data_Engineering_Pipeline_test

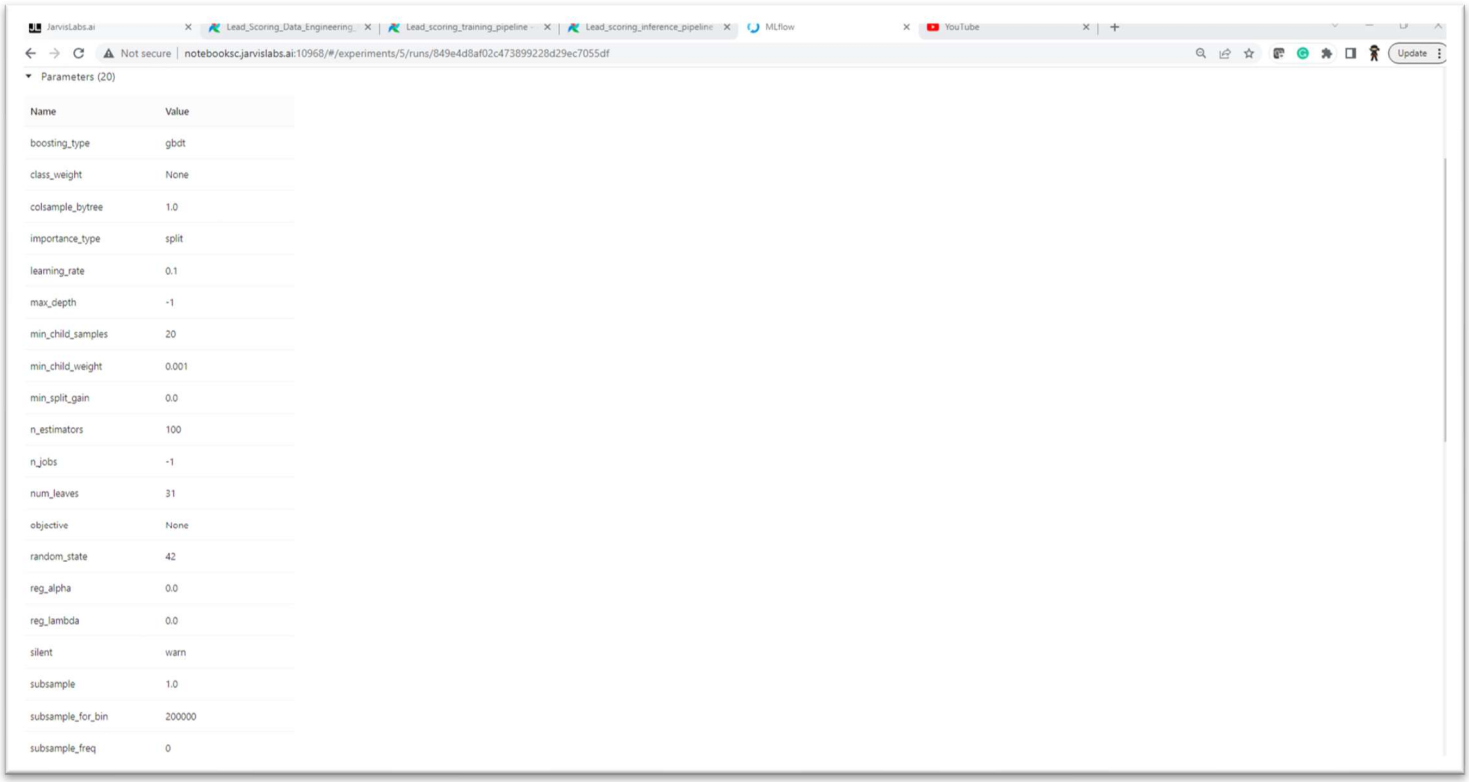


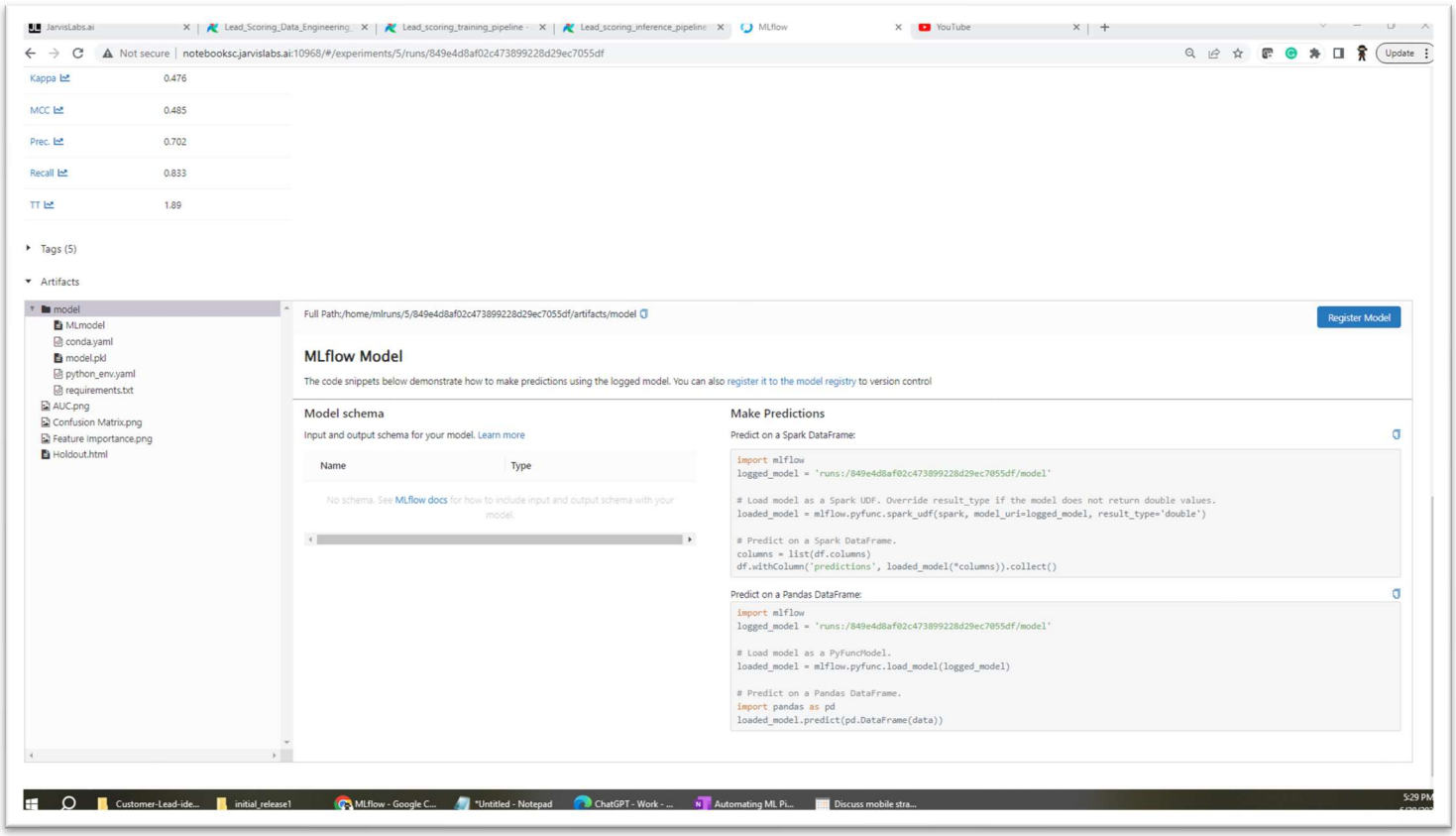


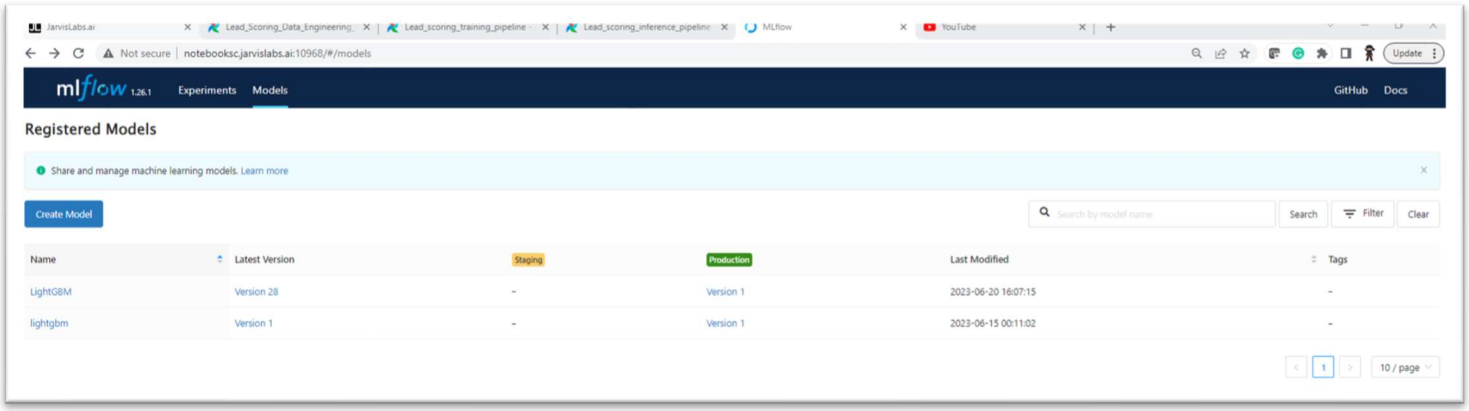
6. mlFlow Lead Scoring Model Experimentation

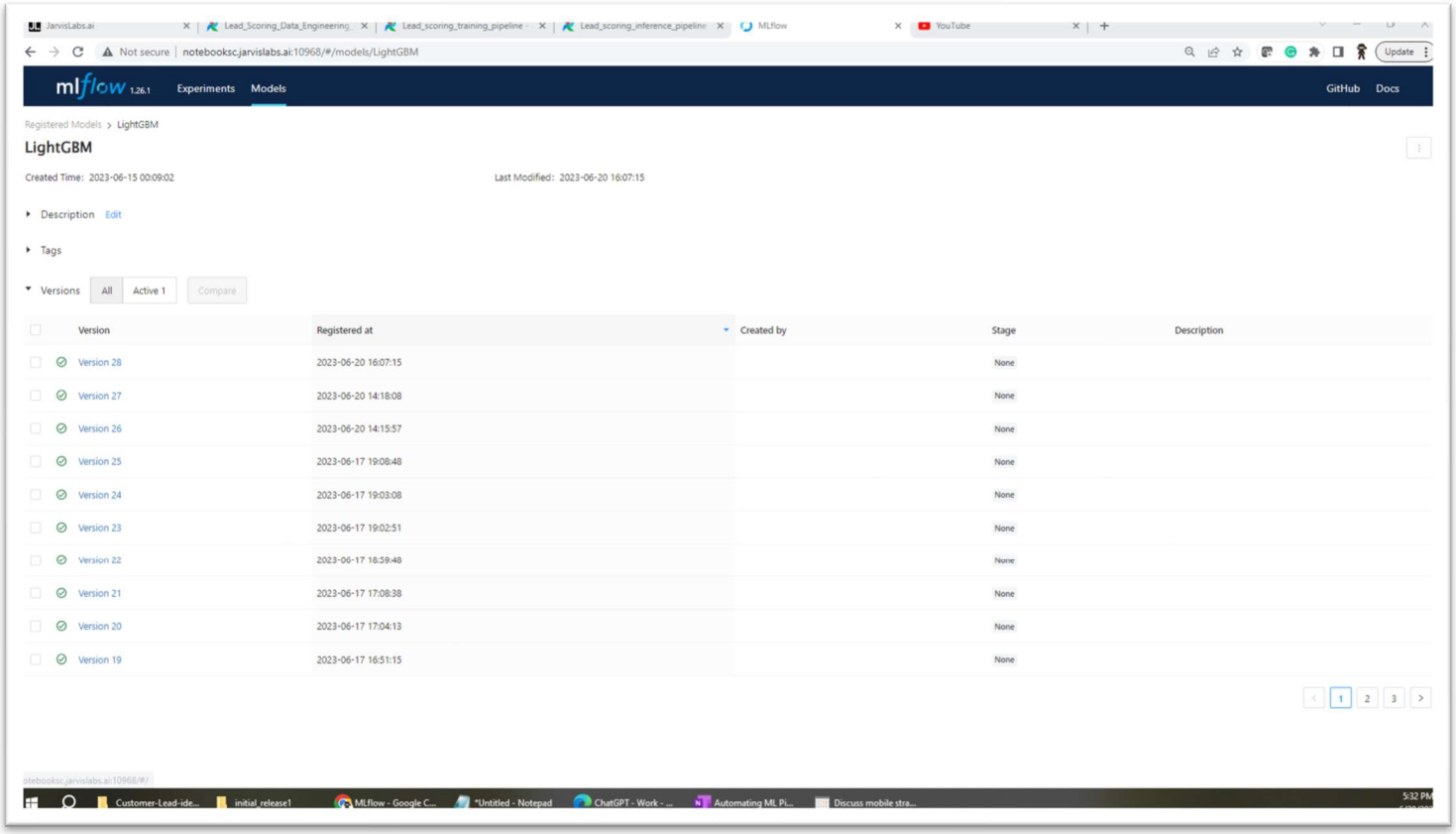


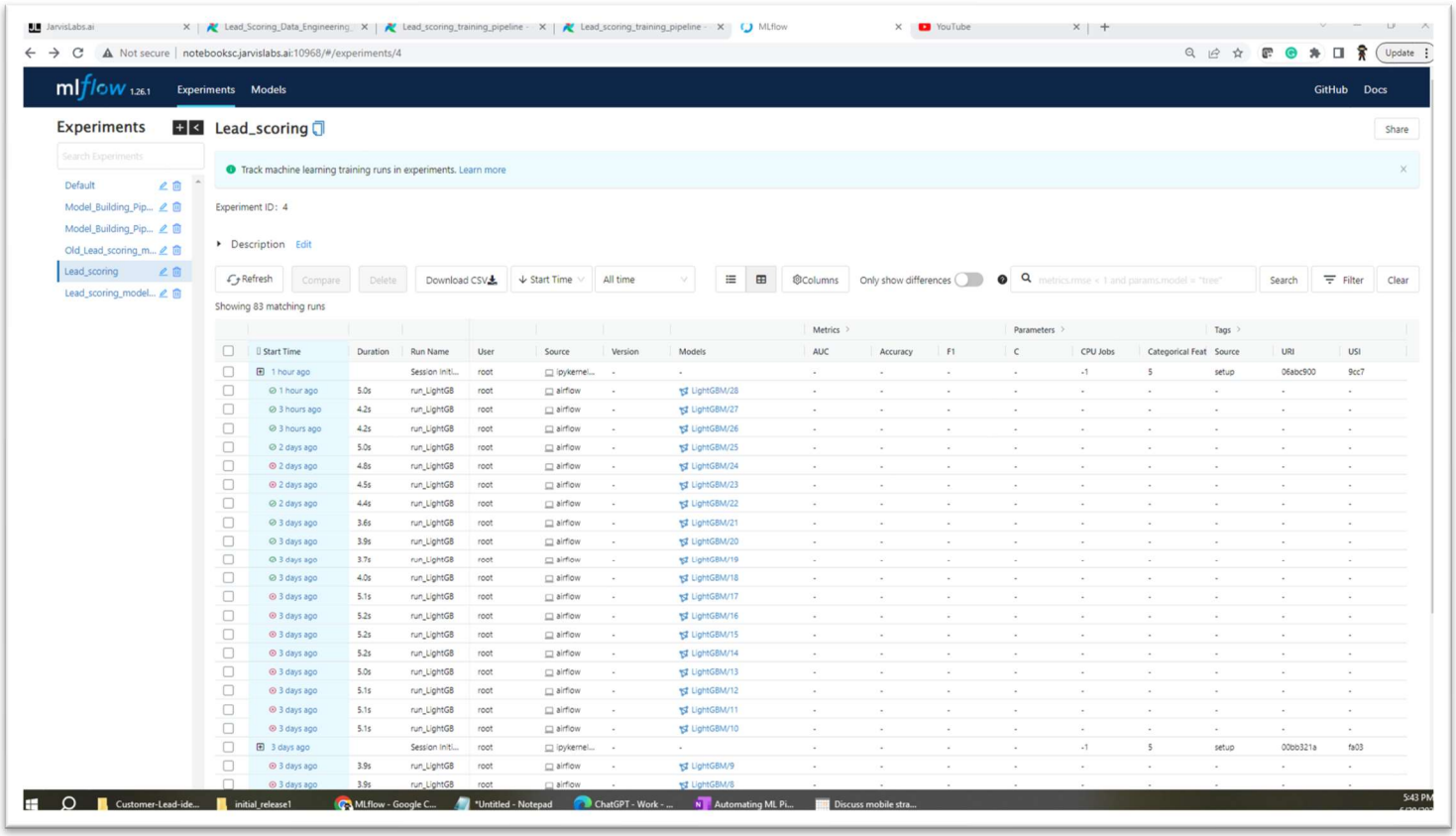


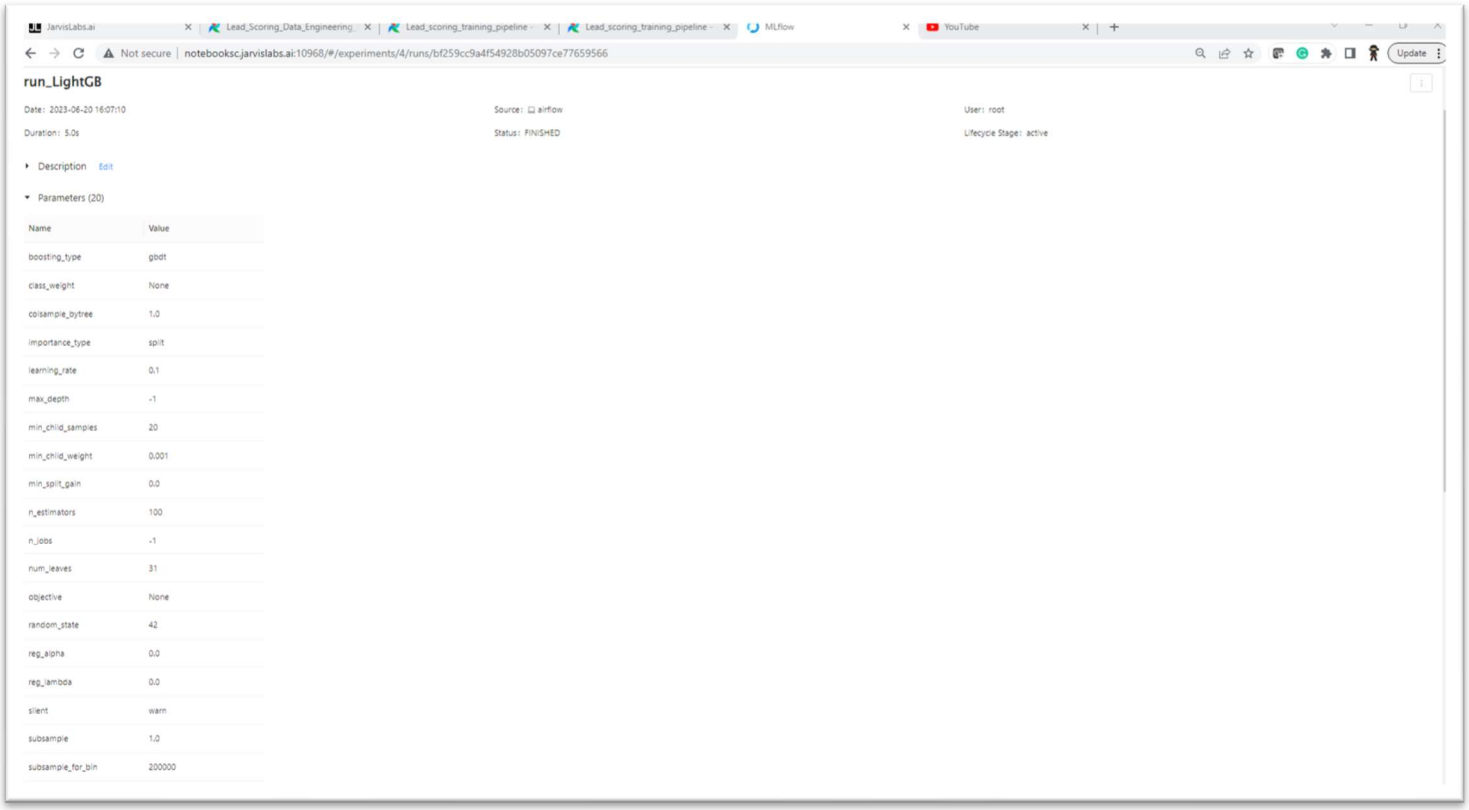

JarvisLabs.ai
Lead_Scoring_Data_Engineering
Lead_scoring_training_pipeline
Lead_scoring_inference_pipeline
MLflow
YouTube
Not secure | notebooksc.jarvislabs.ai:10968/#/experiments/5/runs/849e4d8af02c473899228d29ec7055df
Update
Parameters (20)
Name
Value
boosting_type
gbdt
class_weight
None
colsample_bytree
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

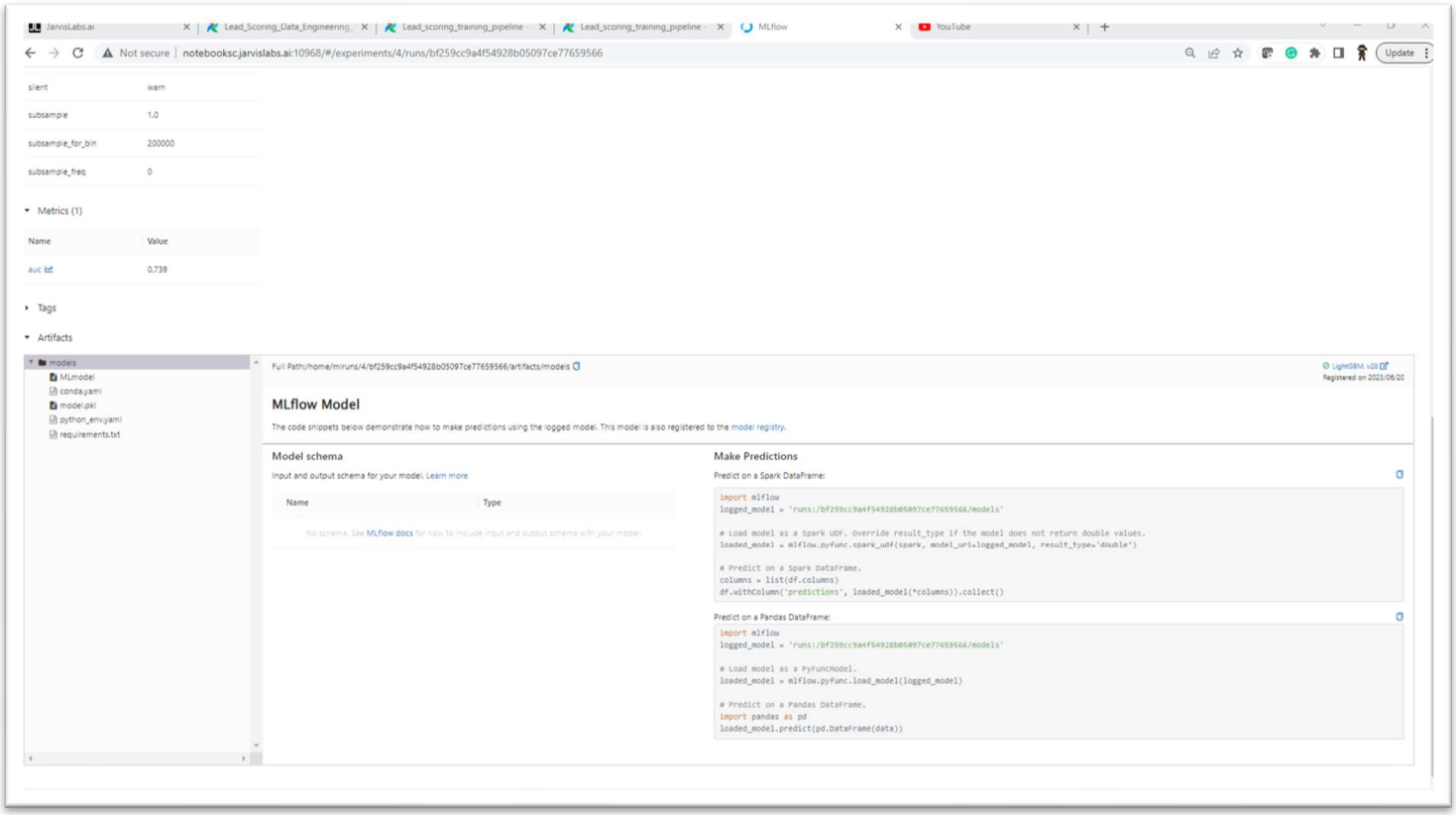

JarvisLabs.ai
Lead_Scoring_Data_Engineering
Lead_scoring_training_pipeline
Lead_scoring_inference_pipeline
MLflow
YouTube
Not secure | notebooksc.jarvislabs.ai:10968/#/experiments/5/runs/849e4d8af02c473899228d29ec7055df
Update
Kappa
0.476
MCC
0.485
Prec.
0.702
Recall
0.833
TT
1.89
Tags (5)
Artifacts
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/5/849e4d8af02c473899228d29ec7055df/artifacts/model
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
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:/849e4d8af02c473899228d29ec7055df/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
logged_model = 'runs:/849e4d8af02c473899228d29ec7055df/model'
# 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))
Customer-Lead-ide...
initial_release1
MLflow - Google C...
*Untitled - Notepad
ChatGPT - Work - ...
Automating ML Pi...
Discuss mobile stra...
5:29 PM


Not secure | notebooksc.jarvislabs.ai:10968/#/models
Update
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 28
-
Version 1
2023-06-20 16:07:15
-
lightgbm
Version 1
-
Version 1
2023-06-15 00:11:02
-
1
10 / page


Not secure | notebooksc.jarvislabs.ai:10968/#/models/LightGBM
Update
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM
LightGBM
Created Time: 2023-06-15 00:09:02
Last Modified: 2023-06-20 16:07:15
Description
Edit
Tags
Versions
All
Active 1
Compare
Version
Registered at
Created by
Stage
Description
Version 28
2023-06-20 16:07:15
None
Version 27
2023-06-20 14:18:08
None
Version 26
2023-06-20 14:15:57
None
Version 25
2023-06-17 19:08:48
None
Version 24
2023-06-17 19:03:08
None
Version 23
2023-06-17 19:02:51
None
Version 22
2023-06-17 18:59:48
None
Version 21
2023-06-17 17:08:38
None
Version 20
2023-06-17 17:04:13
None
Version 19
2023-06-17 16:51:15
None
1
2
3
Customer-Lead-ide...
initial_release1
MLflow - Google C...
*Untitled - Notepad
ChatGPT - Work - ...
Automating ML Pi...
Discuss mobile stra...
5:32 PM


JarvisLabs.ai
Lead_Scoring_Data_Engineering
Lead_scoring_training_pipeline
Lead_scoring_training_pipeline
MLflow
YouTube
Not secure | notebooksc.jarvislabs.ai:10968/#/experiments/4
Update
mlflow 1.26.1
Experiments
Models
GitHub
Docs
Experiments
Lead_scoring
Share
Search Experiments
Track machine learning training runs in experiments. Learn more
Default
Model_Building_Pip...
Model_Building_Pip...
Old_Lead_scoring_m...
Lead_scoring
Lead_scoring_model...
Experiment ID: 4
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 83 matching runs
Metrics
Parameters
Tags
Start Time
Duration
Run Name
User
Source
Version
Models
AUC
Accuracy
F1
C
CPU Jobs
Categorical Feat
Source
URI
USI
1 hour ago
Session Initi...
root
ipykernel...
-
-
-1
5
setup
06abc900
9cc7
1 hour ago
5.0s
run_LightGB
root
airflow
LightGBM/28
3 hours ago
4.2s
run_LightGB
root
airflow
LightGBM/27
3 hours ago
4.2s
run_LightGB
root
airflow
LightGBM/26
2 days ago
5.0s
run_LightGB
root
airflow
LightGBM/25
2 days ago
4.8s
run_LightGB
root
airflow
LightGBM/24
2 days ago
4.5s
run_LightGB
root
airflow
LightGBM/23
2 days ago
4.4s
run_LightGB
root
airflow
LightGBM/22
3 days ago
3.6s
run_LightGB
root
airflow
LightGBM/21
3 days ago
3.9s
run_LightGB
root
airflow
LightGBM/20
3 days ago
3.7s
run_LightGB
root
airflow
LightGBM/19
3 days ago
4.0s
run_LightGB
root
airflow
LightGBM/18
3 days ago
5.1s
run_LightGB
root
airflow
LightGBM/17
3 days ago
5.2s
run_LightGB
root
airflow
LightGBM/16
3 days ago
5.2s
run_LightGB
root
airflow
LightGBM/15
3 days ago
5.2s
run_LightGB
root
airflow
LightGBM/14
3 days ago
5.0s
run_LightGB
root
airflow
LightGBM/13
3 days ago
5.1s
run_LightGB
root
airflow
LightGBM/12
3 days ago
5.1s
run_LightGB
root
airflow
LightGBM/11
3 days ago
5.1s
run_LightGB
root
airflow
LightGBM/10
3 days ago
Session Initi...
root
ipykernel...
-
-
-1
5
setup
00bb321a
fa03
3 days ago
3.9s
run_LightGB
root
airflow
LightGBM/9
3 days ago
3.9s
run_LightGB
root
airflow
LightGBM/8
Customer-Lead-ide...
initial_release1
MLflow - Google C...
*Untitled - Notepad
ChatGPT - Work - ...
Automating ML Pi...
Discuss mobile stra...
5:43 PM


JarvisLabs.ai
Lead_Scoring_Data_Engineering
Lead_scoring_training_pipeline
Lead_scoring_training_pipeline
MLflow
YouTube
Not secure | notebooksc.jarvislabs.ai:10968/#/experiments/4/runs/bf259cc9a4f54928b05097ce77659566
Update
run_LightGB
Date: 2023-06-20 16:07:10
Source: airflow
User: root
Duration: 5.0s
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
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


JarvisLabs.ai
Lead_Scoring_Data_Engineering
Lead_scoring_training_pipeline
Lead_scoring_training_pipeline
MLflow
YouTube
Not secure | notebooksc.jarvislabs.ai:10968/#/experiments/4/runs/bf259cc9a4f54928b05097ce77659566
Update
silent warn
subsample 1.0
subsample_for_bin 200000
subsample_freq 0
Metrics (1)
Name Value
auc 0.739
Tags
Artifacts
models
MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt
Full Path/home/miruns/4/bf259cc9a4f54928b05097ce77659566/artifacts/models
LightGBM, v28
Registered on 2023/06/20
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:/bf259cc9a4f54928b05097ce77659566/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:/bf259cc9a4f54928b05097ce77659566/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))