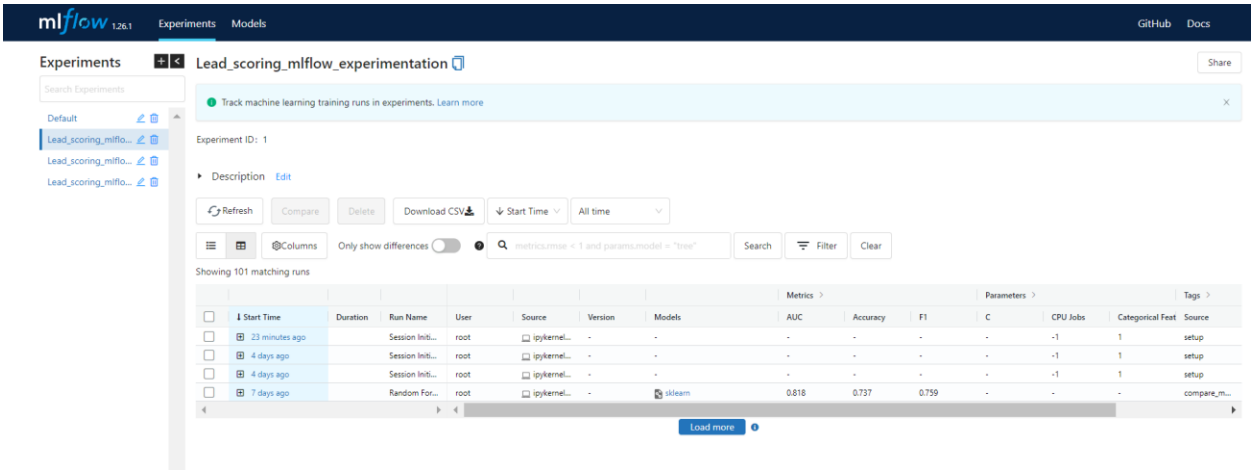
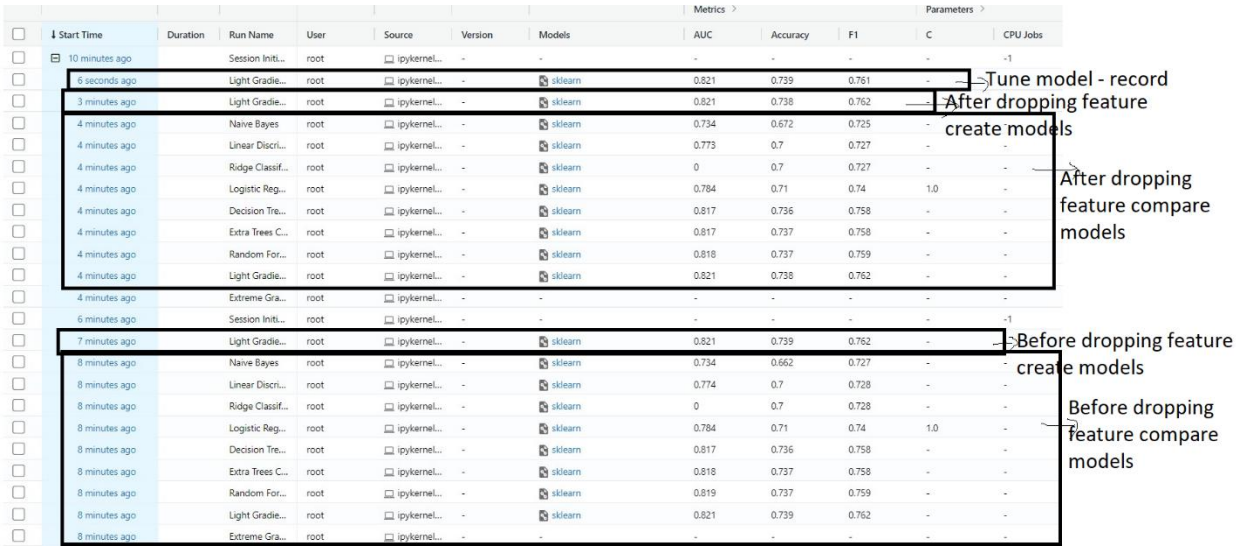
Model experimentation screenshots

Screenshot of the main page during experimentation:



Model experimentation screenshot



Before dropping features – experiment details:

[21]:
```
# setup pycaret
Baseline_model_exp01 = setup(data = encoded_df, target = 'app_complete_flag',
                  session_id = 42,fix_imbalance=False,
                  n_jobs=-1,use_gpu=False,
                  log_experiment=True,experiment_name='Lead_scoring_mlflow_experimentation',
                  log_plots=True, log_data=True,
                  silent=True, verbose=True,
                  log_profile=False)
```

	Description	Value
0	session_id	42
1	Target	app_complete_flag
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(238964, 46)
5	Missing Values	False
6	Numeric Features	44
7	Categorical Features	1
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(167274, 45)
12	Transformed Test Set	(71690, 45)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False

Lead_scoring_mlflow_experimentation > Light Gradient Boosting Machine

Light Gradient Boosting Machine

Date: 2023-01-14 16:58:42 | Source: ipykernel_launcher.py | User: root

Status: UNFINISHED | Lifecycle Stage: active | Parent Run: 8a8b5d516c7b49fb86cd6f20c2da9e46

▸ 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 (8)

Name	Value
AUC	0.821
Accuracy	0.739
F1	0.762
Kappa	0.477
MCC	0.485
Prec.	0.703
Recall	0.833
TT	1.58

▾ Tags (5)

Name	Value	Actions
Run ID	4b8f10243b784d89bad3911f34f15a8c	
Run Time	8.63	
Source	create_model	
URI	86dd81a8	

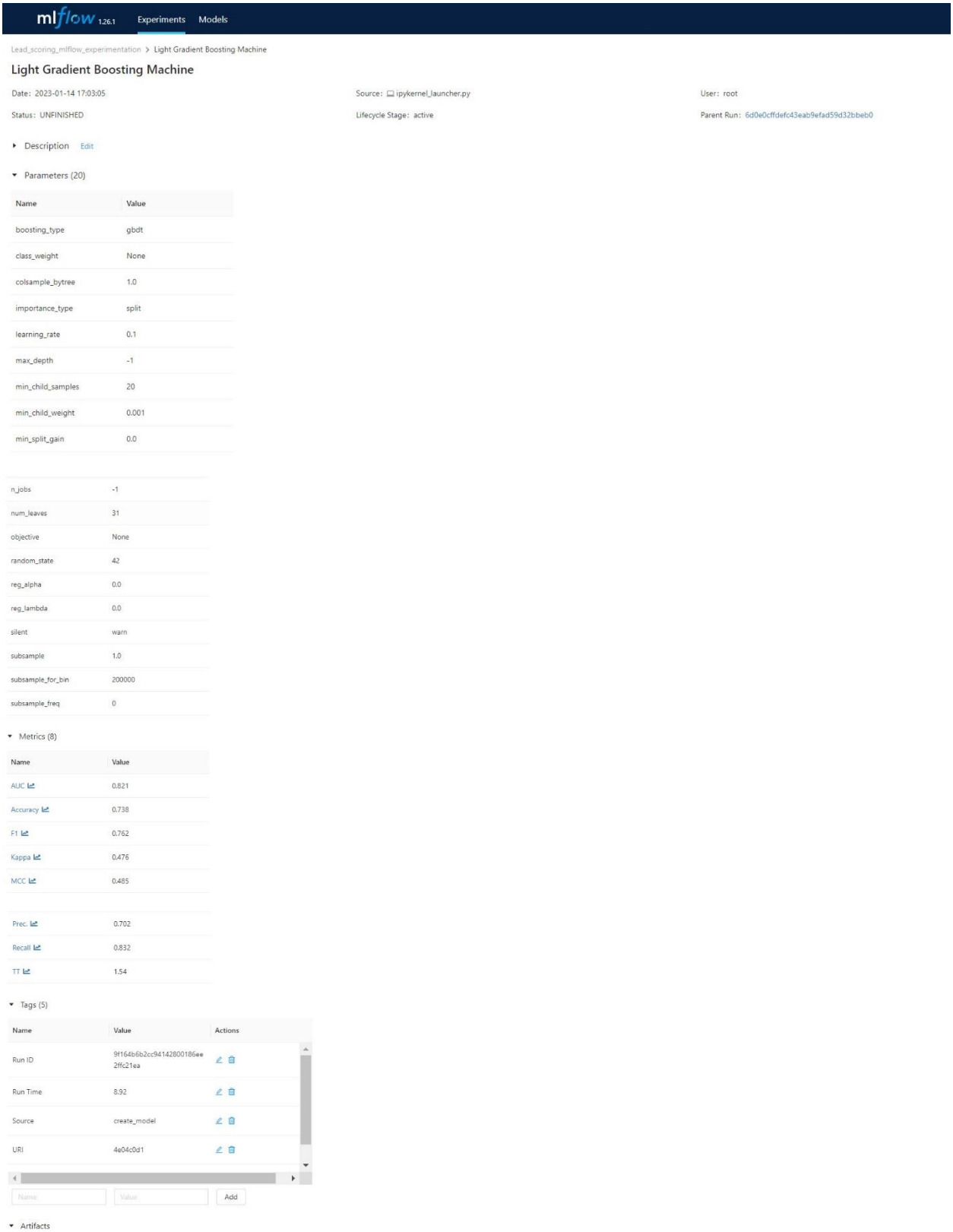
Name | Value | Add

▾ Artifacts

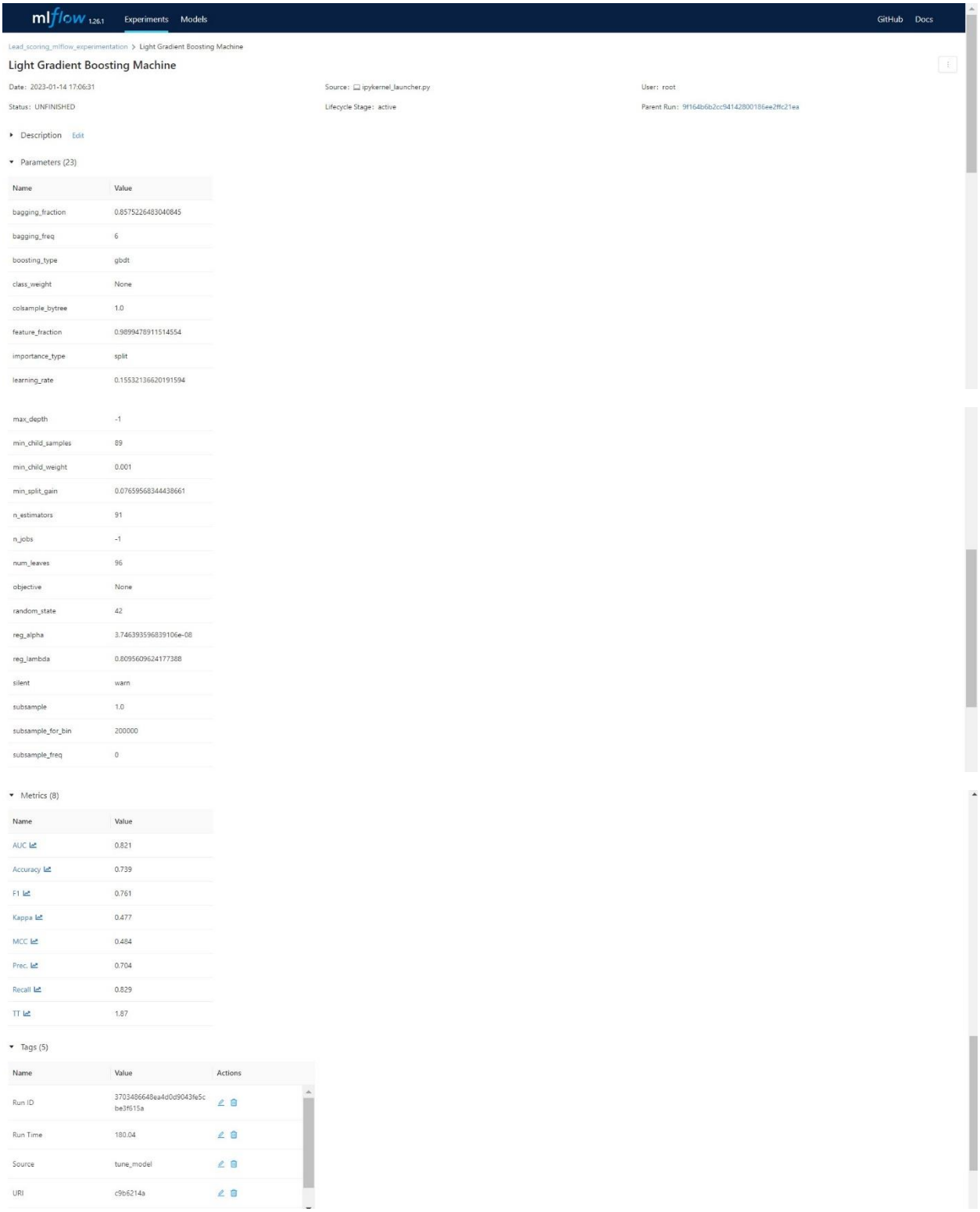
After dropping features experiment details:

```
Baseline_model_exp02 = setup(data = encoded_df_sig_features, target = 'app_complete_flag',
                  session_id = 42,fix_imbalance=False,
                  n_jobs=-1,use_gpu=False,
                  log_experiment=True,experiment_name='Lead_scoring_mlflow_experimentation',
                  log_plots=True, log_data=True,
                  silent=True, verbose=True,
                  log_profile=False, normalize=False,
                  transformation=False)
```

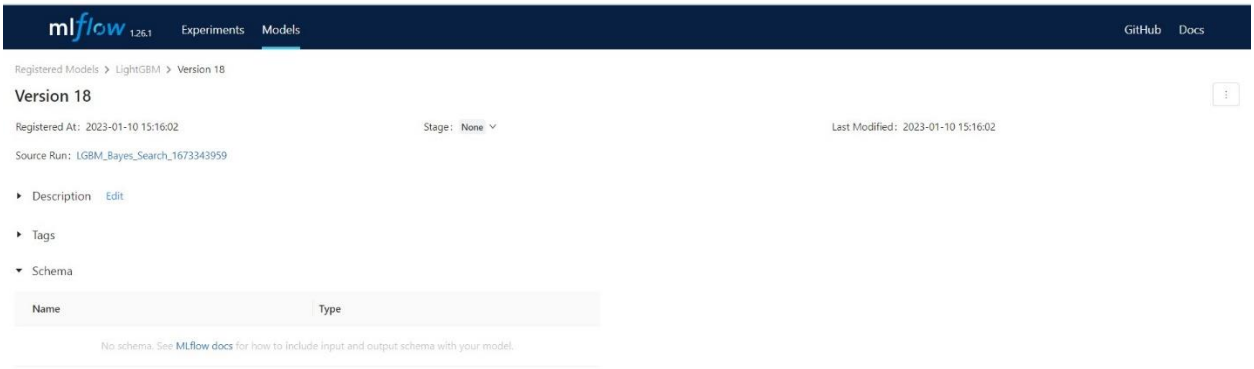
	Description	Value
0	session_id	42
1	Target	app_complete_flag
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(238964, 41)
5	Missing Values	False
6	Numeric Features	39
7	Categorical Features	1
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(167274, 40)
12	Transformed Test Set	(71690, 40)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False
15	Fold Generator	StratifiedKFold


mlflow 1.26.1
Experiments
Models
Lead_scoring_mlflow_experimentation > Light Gradient Boosting Machine
Light Gradient Boosting Machine
Date: 2023-01-14 17:03:05
Source: ipykernel_launcher.py
User: root
Status: UNFINISHED
Lifecycle Stage: active
Parent Run: 6d0e0cffdefc43eab9efad59d32bbeb0
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_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 (8)
Name
Value
AUC
0.821
Accuracy
0.738
F1
0.762
Kappa
0.476
MCC
0.485
Prec.
0.702
Recall
0.832
TT
1.54
Tags (5)
Name
Value
Actions
Run ID
9f164b6b2cc94142800186ee2ffc21ea
Run Time
8.92
Source
create_model
URI
4e04c0d1
Name
Value
Add
Artifacts

Tuned model experimentation details:

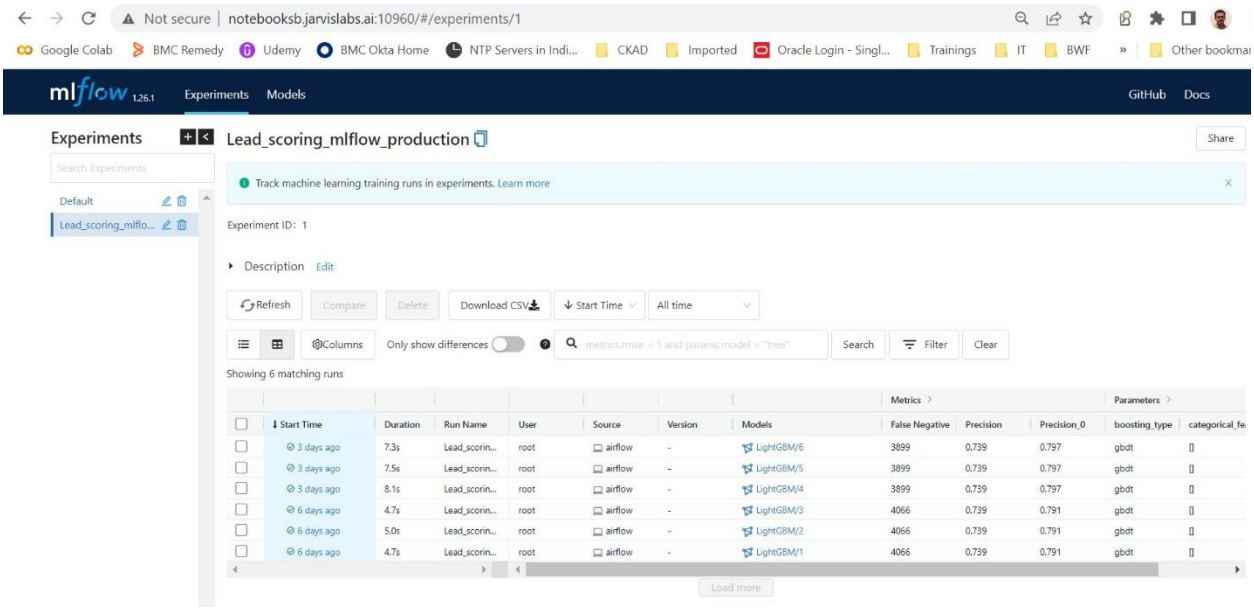


One of the model details:

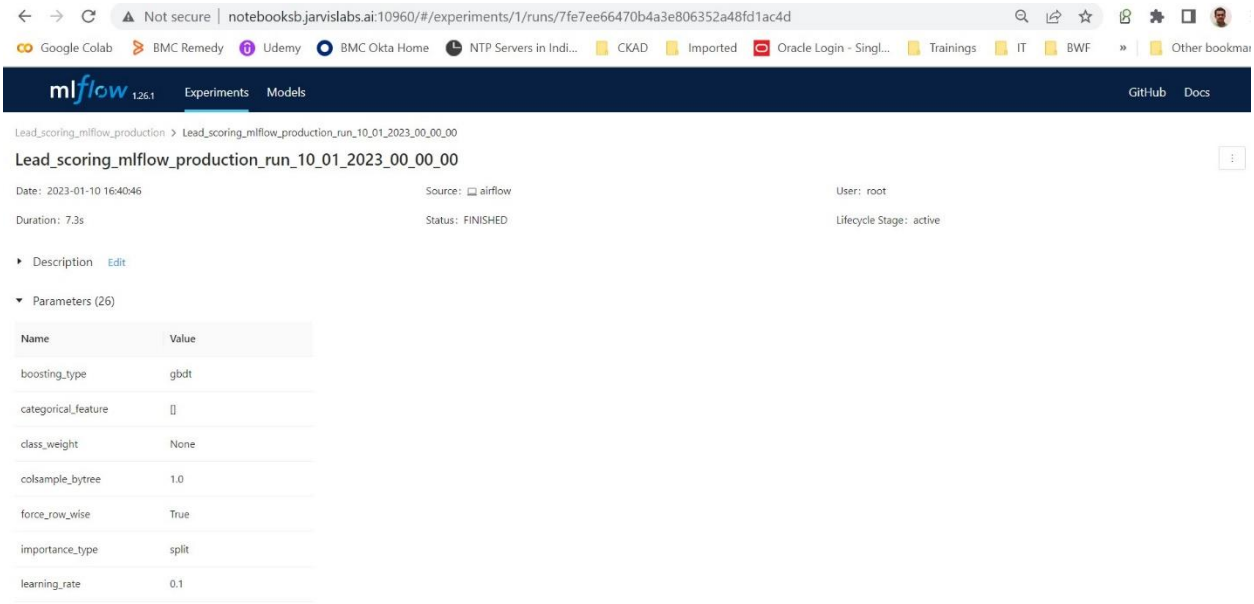

mlflow 1.26.1
Experiments
Models
GitHub
Docs
Registered Models > LightGBM > Version 18
Version 18
Registered At: 2023-01-10 15:16:02
Stage: None
Last Modified: 2023-01-10 15:16:02
Source Run: LGBM_Bayes_Search_1673343959
Description
Edit
Tags
Schema
Name
Type
No schema. See MLflow docs for how to include input and output schema with your model.

Model training screenshots

Run pipeline multiple times



Following 5 screenshot highlight the details of the LightGBM/6 model from above screenshot:

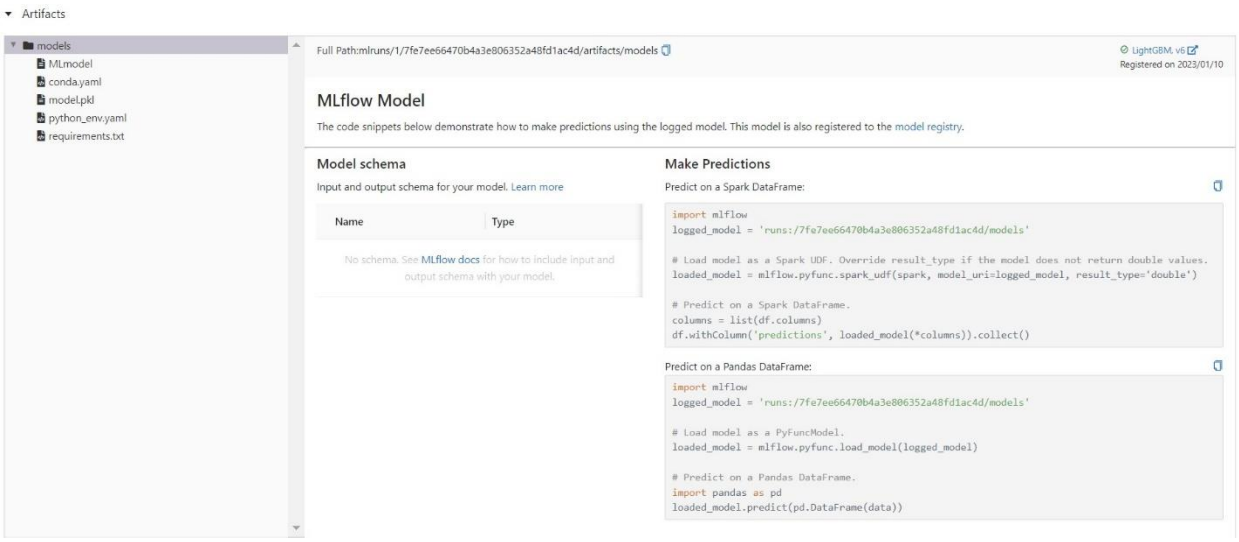


max_bin	255
max_depth	-1
metric	roc_auc
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	100
n_jobs	-1
num_boost_round	200
num_leaves	31
objective	binary
random_state	42
reg_alpha	0.0
reg_lambda	0.0

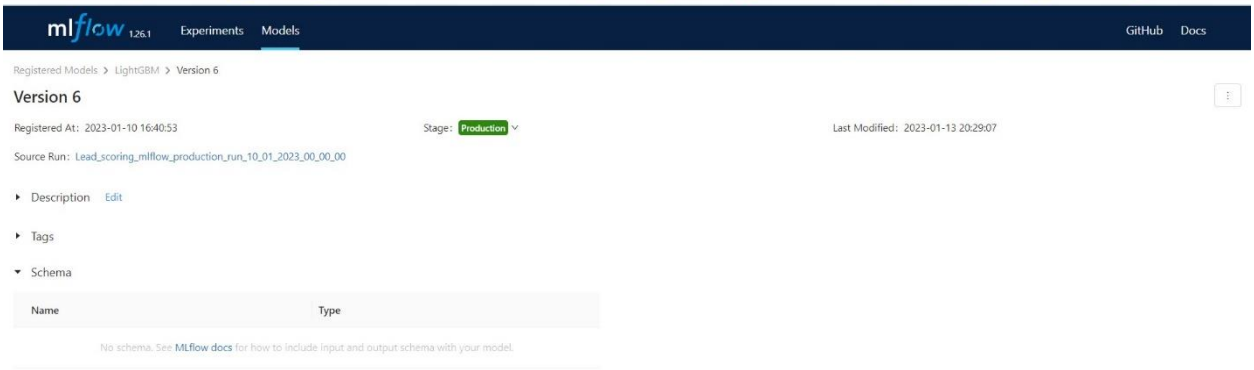
reg_lambda	0.0
silent	warn
subsample	1.0
subsample_for_bin	200000
subsample_freq	0
verbose	-1

▼ Metrics (13)

Name	Value
False Negative	3899
Precision	0.739
Precision_0	0.797
Precision_1	0.701
Recall	0.749
Recall_0	0.641
Recall_1	0.837
True Negative	15286
f1 score	0.737
f1_0	0.711
f1_1	0.763
roc_auc	0.749
test_accuracy	0.739

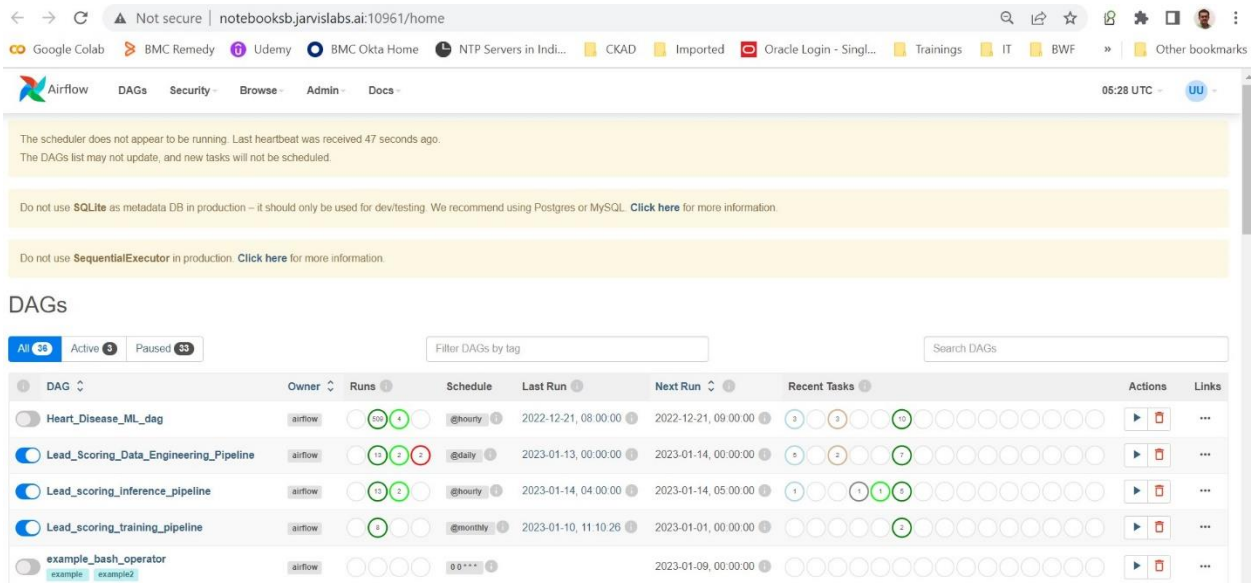


Model moved to Production stage

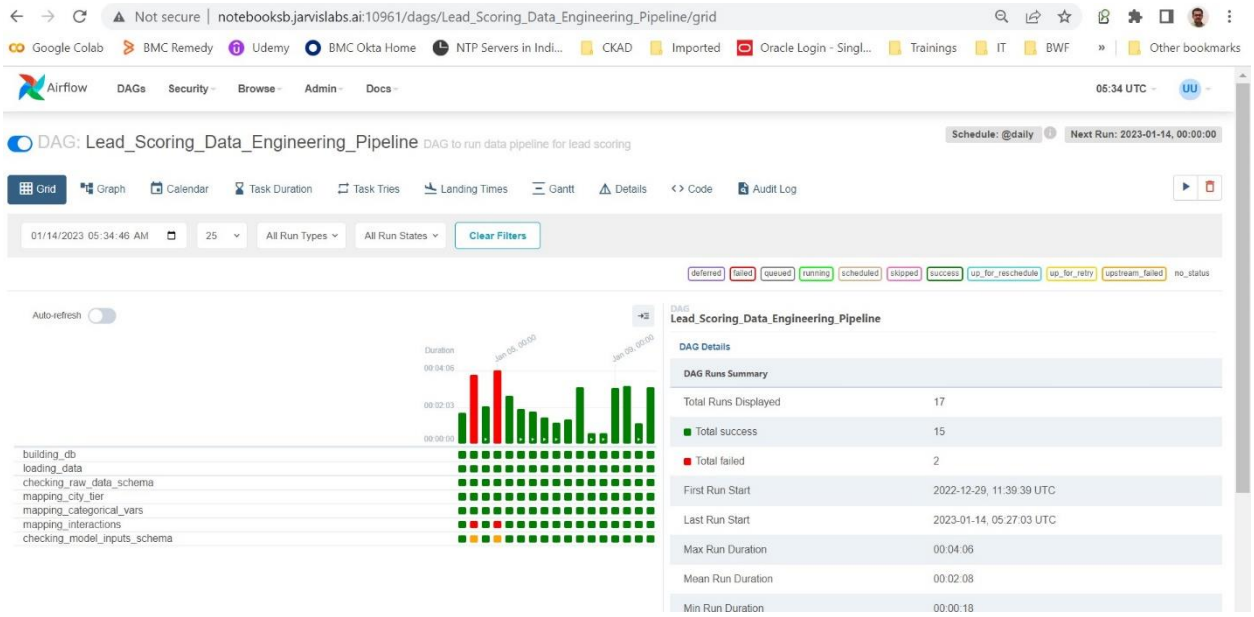


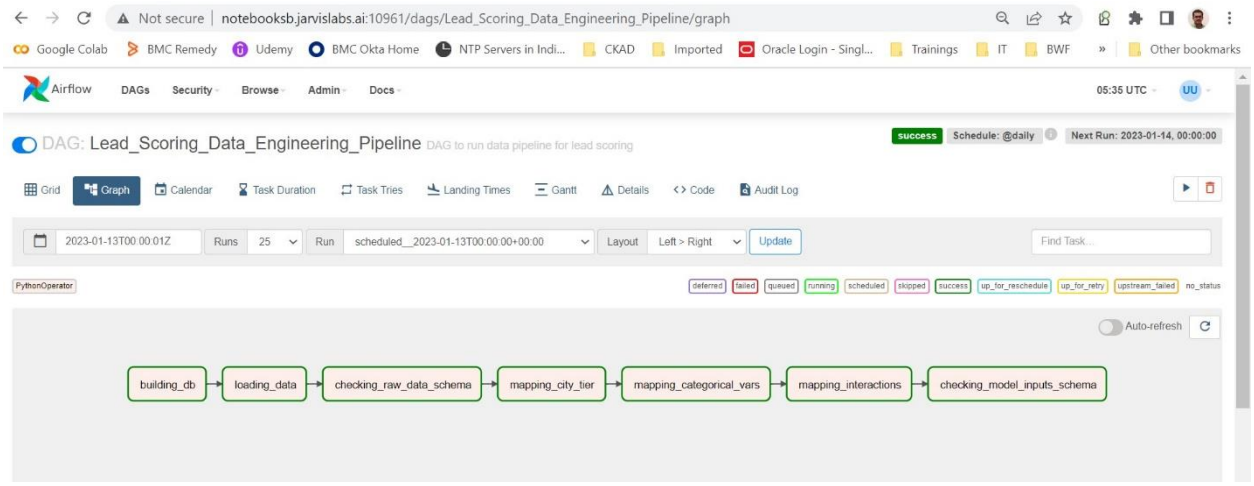
Airflow Screenshots

Airflow main page:

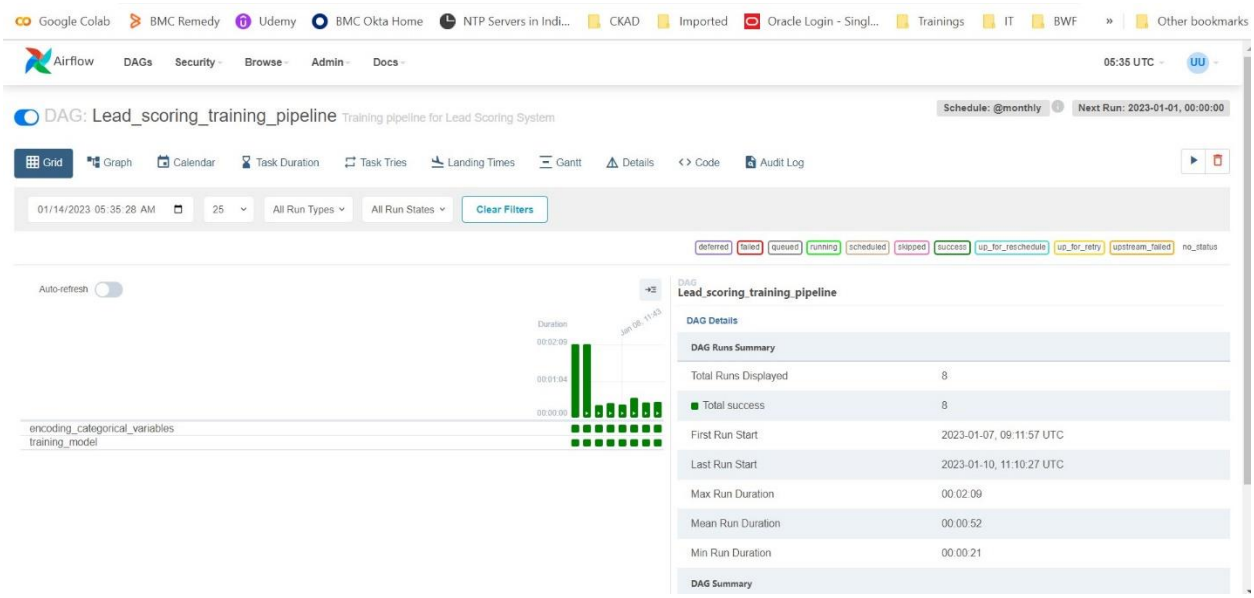


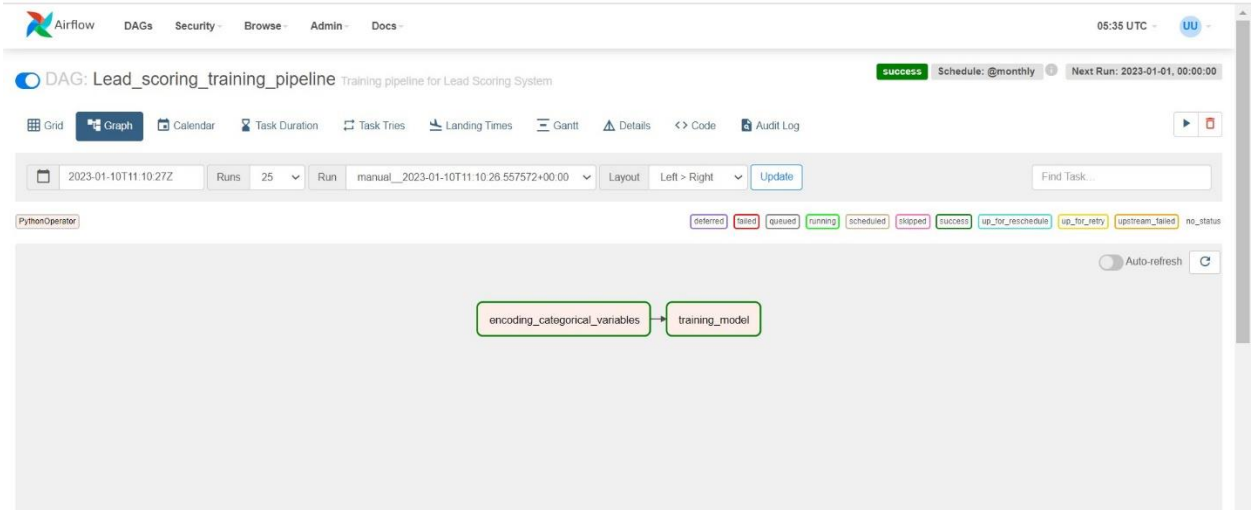
Data engineering pipeline



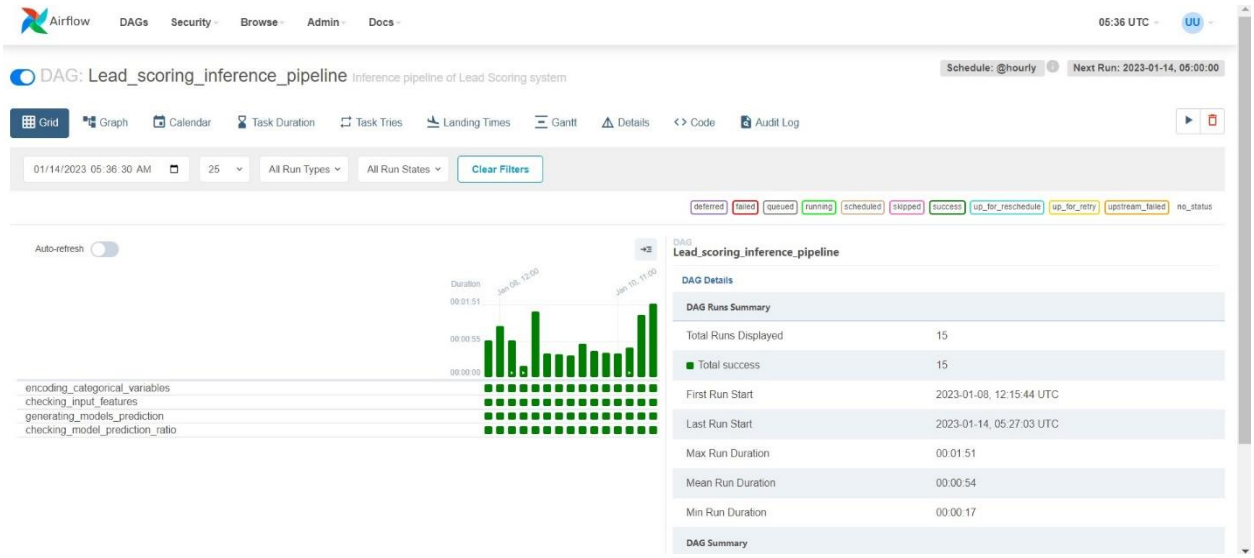


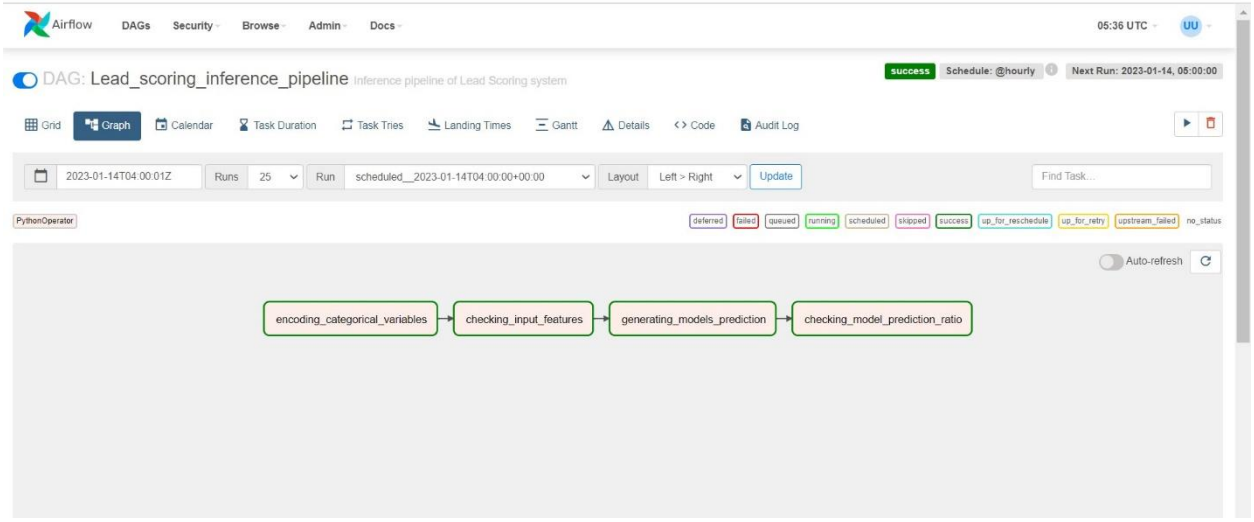
Model training pipeline





Inference pipeline




Airflow
DAGs
Security
Browse
Admin
Docs
05:36 UTC
UU
DAG: Lead_scoring_inference_pipeline
Inference pipeline of Lead Scoring system
success
Schedule: @hourly
Next Run: 2023-01-14, 05:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
2023-01-14T04:00:01Z
Runs
25
Run
scheduled__2023-01-14T04:00:00+00:00
Layout
Left > Right
Update
Find Task...
PythonOperator
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
encoding_categorical_variables
checking_input_features
generating_models_prediction
checking_model_prediction_ratio