

MLOps Assignment

- Screenshots at different levels from feature engineering to training and production.

Submitted by:

Name: Khalid Mohamad Aladin Hafez Omar

Email: Khalidomarasu@gmail.com

Airflow Main UI Page

scripts (7)JupyterLabRent GPUJarvisLabsDAGs - Airdag_id%3MLflowMLflowAirflow isDatabasepython

Not secure | notebooksbjarvislabs.ai:10961/home?status=active

23:38 (+01:00)

Airflow

DAGsSecurityBrowseAdminDocs

Do not use SQLite as metadata DB in production – it should only be used for dev/testing. We recommend using Postgres or MySQL. [Click here](#) for more information.

Do not use SequentialExecutor in production. [Click here](#) for more information.

DAGs

All35Active3Paused32

Filter DAGs by tag

Search DAGs

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
<input checked="" type="checkbox"/> Lead_Scoring_Data_Engineering_Pipeline	airflow	8	@daily	2023-06-26, 23:14:03	2023-06-26, 01:00:00	7
<input checked="" type="checkbox"/> Lead_scoring_inference_pipeline	airflow	10	@hourly	2023-06-26, 23:27:48	2023-06-26, 23:00:00	4
<input checked="" type="checkbox"/> Lead_scoring_training_pipeline	airflow	61	@monthly	2023-06-26, 23:15:56	2023-06-01, 01:00:00	11

« < 1 > »

Showing 1-3 of 3 DAGs

- 3 Active DAGs from Airflow Main Page
1. Lead_Scoring_Data_Engineering_Pipeline

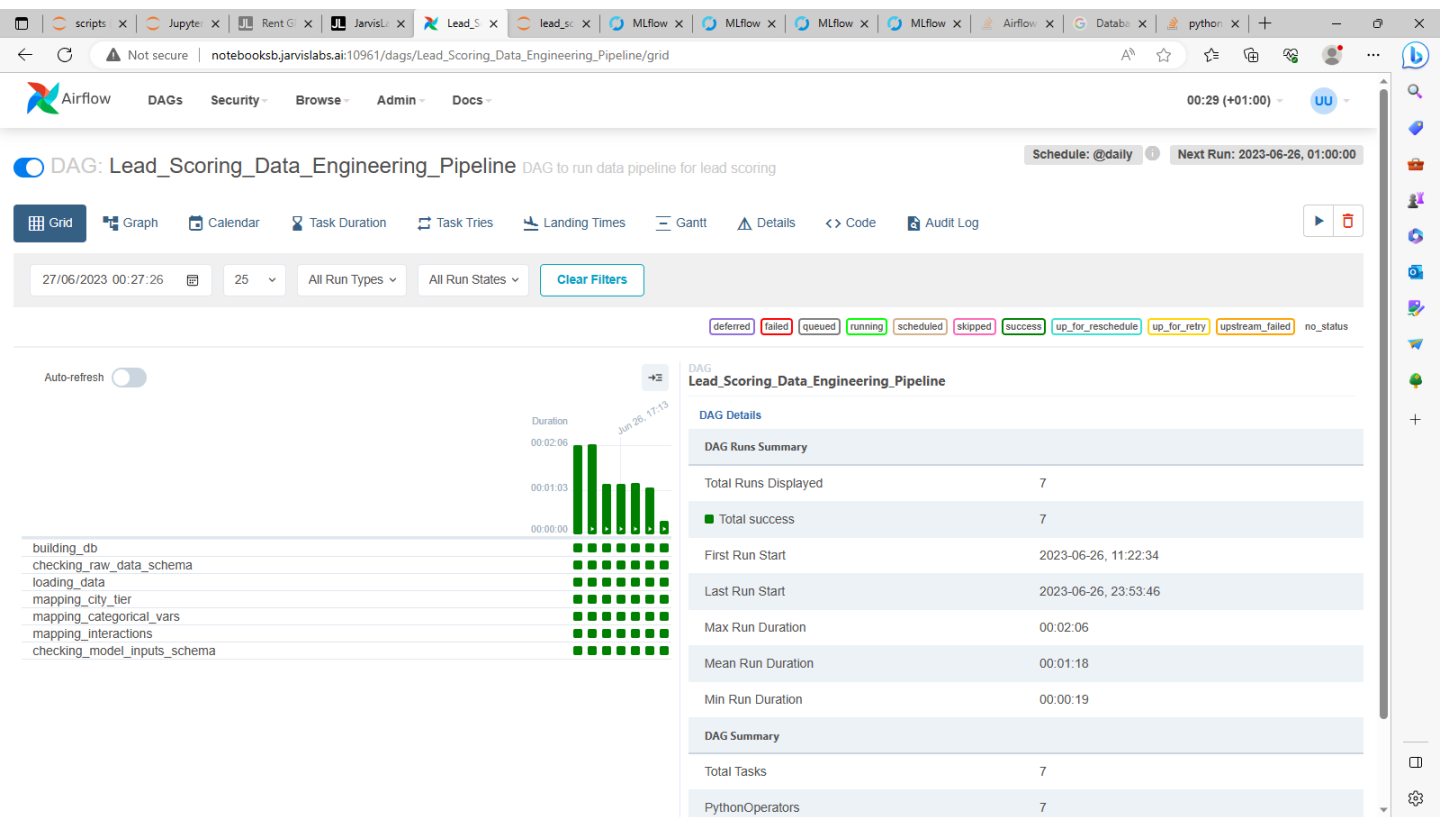
2. Lead_scoring_training_pipeline

3. Lead_scoring_inference_pipeline

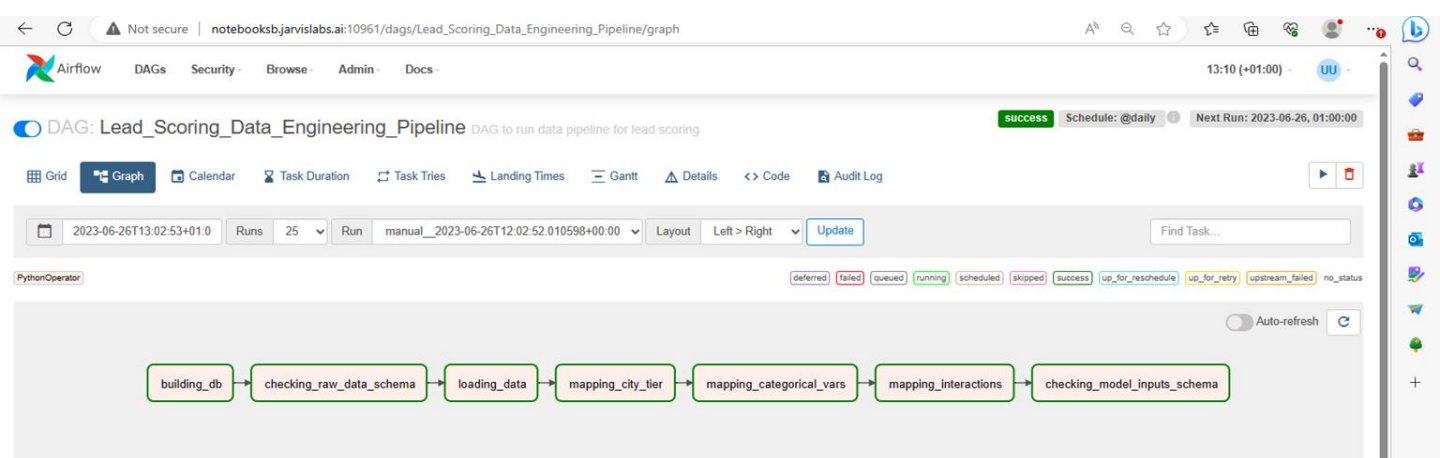
Airflow - Data Pipeline

Dag of Lead_Scoring_Data_Engineering_Pipeline

- Screenshot of Airflow UI grid



- Screenshot of successful execution Airflow DAG in graph

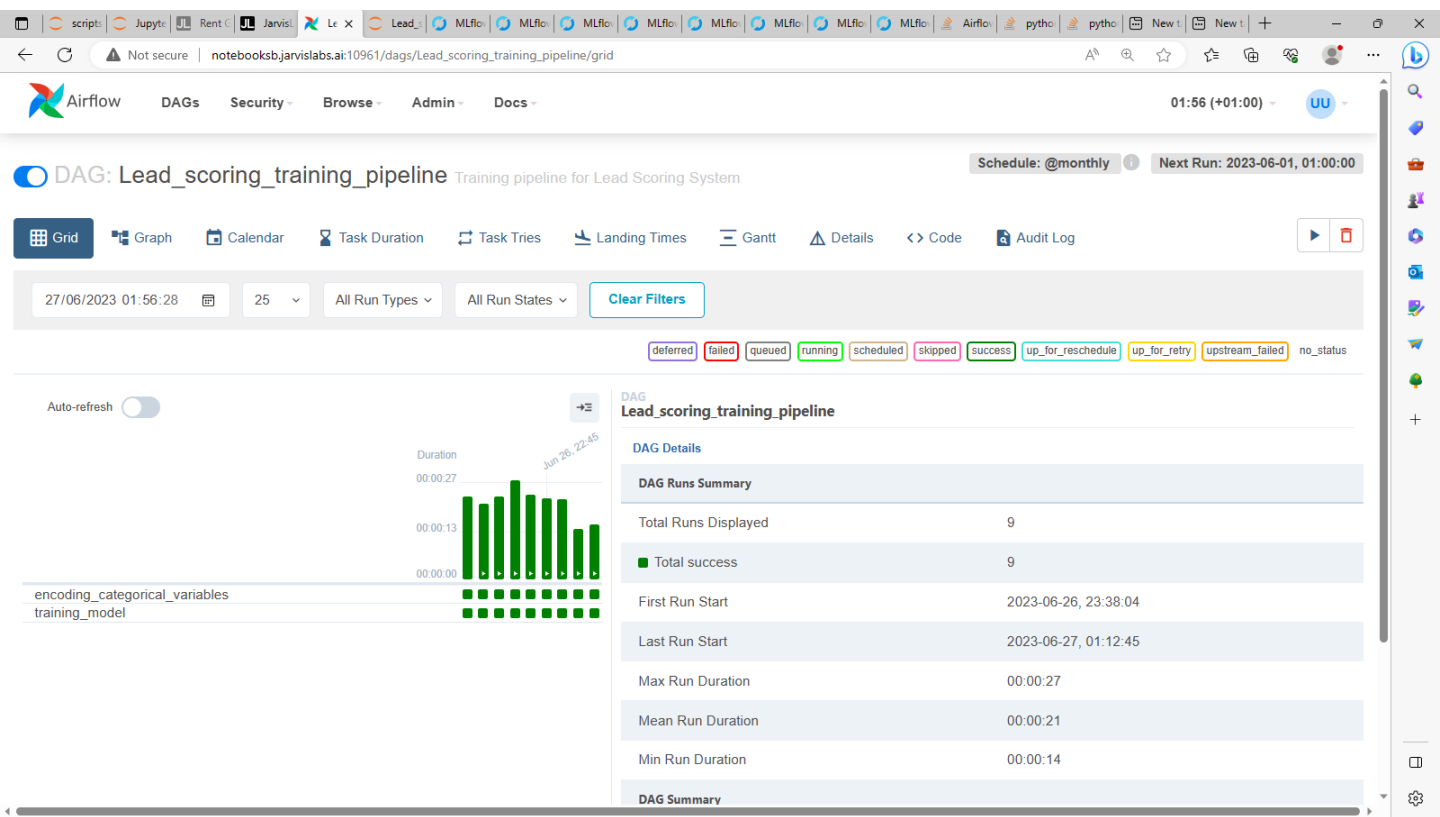


Manual & Scheduled Run with Successful Runs and 7 tasks done

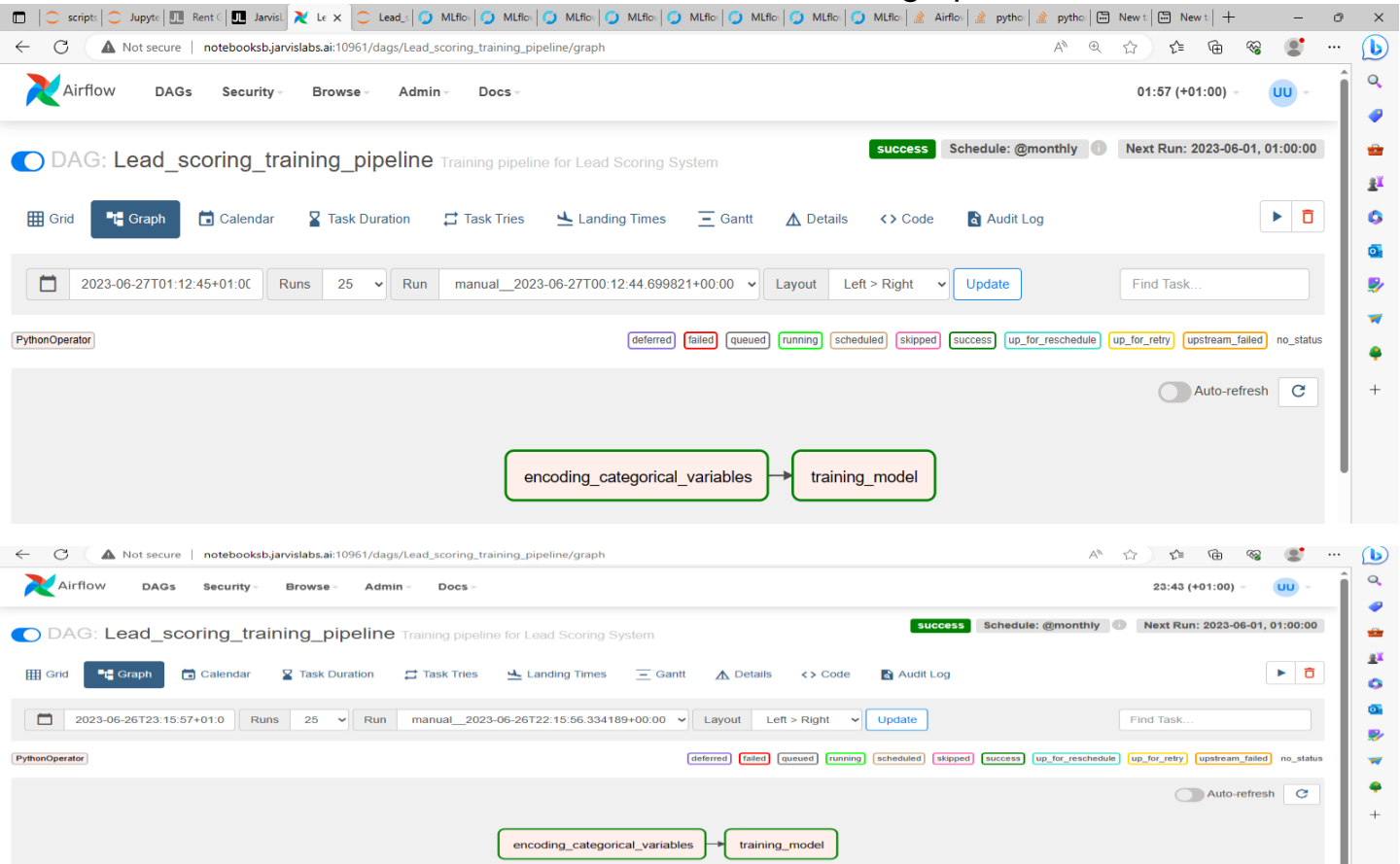
Airflow - Training Pipeline

Dag of Lead_scoring_training_pipeline

- Screenshot of Airflow UI grid



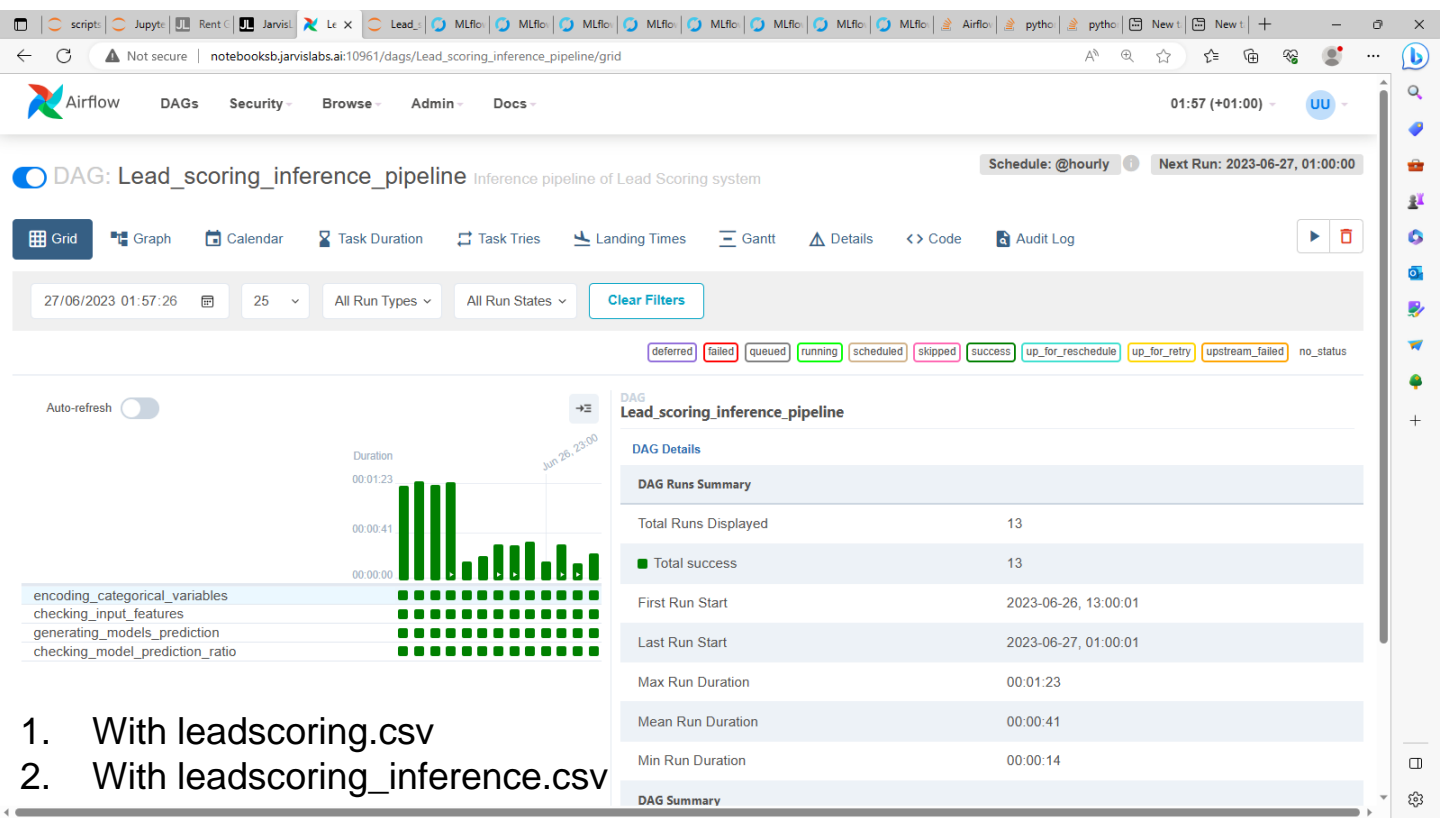
- Screenshot of successful execution Airflow DAG in graph



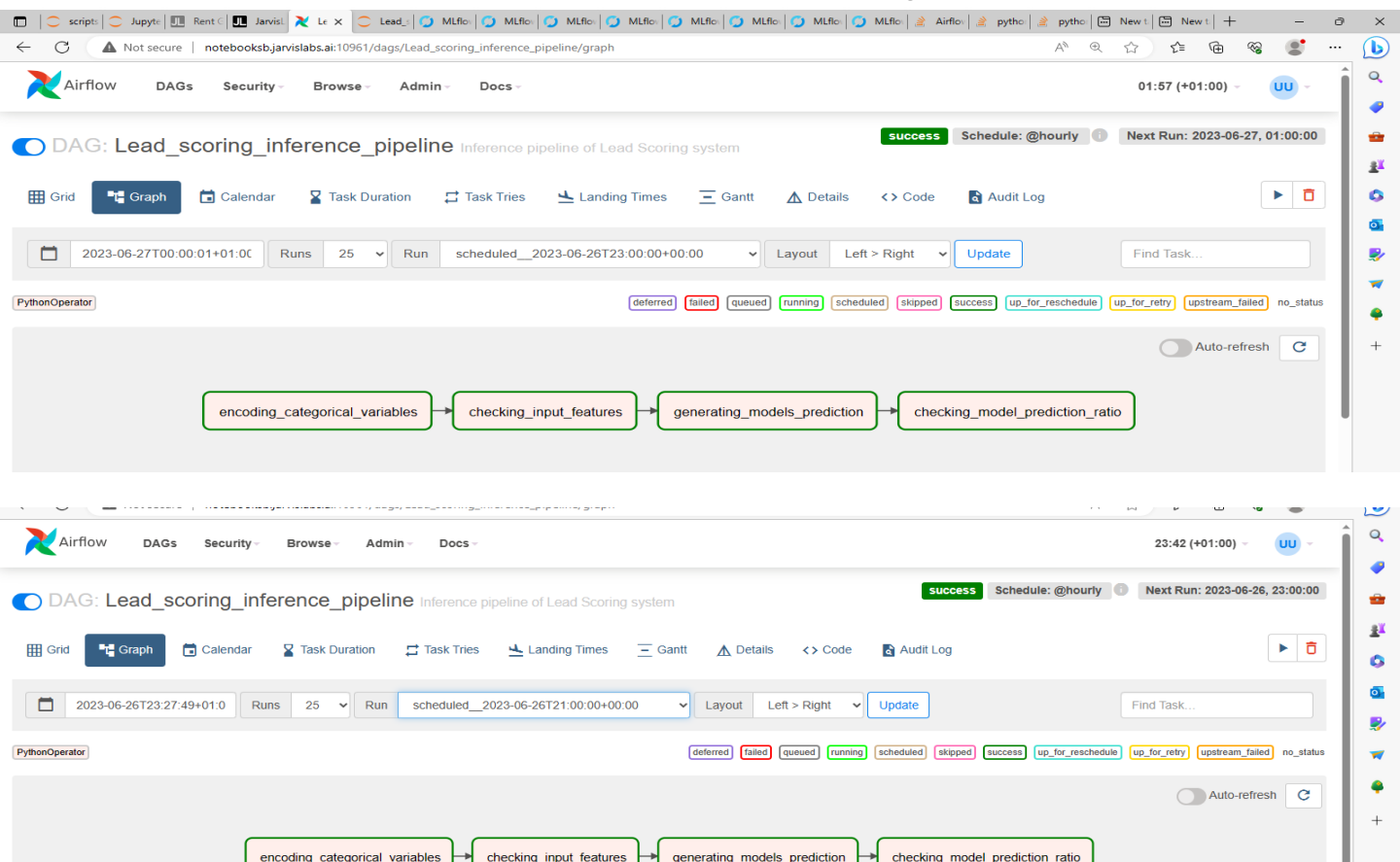
Airflow - Inference Pipeline

Dag of Lead_scoring_inference_pipeline

- Screenshot of Airflow UI grid



- Screenshot of successful execution Airflow DAG in graph



ML FLOW

ML FLOW – Random Screenshots

Registry View

mlflow 1.26.1ExperimentsModels

Lead_scoring_model_experimentation > Light Gradient Boosting Machine

Date: 2023-06-26 15:29:29

Status: UNFINISHED

Description Edit

None

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

Duration: 4.8s

Description Edit

Parameters (26)

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

Tags

Artifacts

models

Full Path/home/mlruns/3/b0c3aa4715d645

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

mlflow 1.26.1ExperimentsModels

Lead_scoring_mlflow_production > Lead_scoring_mlflow_production_run_26_06_2023_00_00_00

Date: 2023-06-26 23:38:22

Source: airflow

Duration: 4.8s

Status: FINISHED

Description Edit

Parameters (26)

Metrics (13)

Tags

Artifacts

models

Full Path/home/mlruns/3/b0c3aa4715d645

MLmodel

conda.yaml

model.pkl

python_env.yaml

requirements.txt

MLflow Model

The code snippets below demonstrate how

Model schema

Input and output schema for your model. [Learn more](#)

Name	Type
------	------

No schema. See [MLflow docs](#) for how to define input and output schema with your model.

mlflow 1.26.1ExperimentsModels

Experiments

Lead_scoring_model_experimentation

Search Experiments

Default

Lead_scoring_model_experimentation

Lead_scoring

Lead_scoring_mlflow_production

Lead_scoring_Training

Experiment ID: 1

Description Edit

Refresh

Compare

Delete

Download CSV

Start Time

All time

Columns

Only show differences

metricsname < 1 and paramsmodel = "tree"

Search

Filter

Clear

Showing 22 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	CPU Jobs	Categorical Feat	Source	URI	URI
	8 hours ago		Session Init...	root	ipykernel...	-	-	-	-	-	-1	4	setup	fb2c5333	61e6
	8 hours ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	-	-	create_model	c4b36639	61e6
	8 hours ago		Naive Bayes	root	ipykernel...	-	sklearn	0.738	0.679	0.728	-	-	compare_m...	01a020e1	61e6
	8 hours ago		Ridge Classi...	root	ipykernel...	-	sklearn	0	0.715	0.742	-	-	compare_m...	9d9d9448	61e6
	8 hours ago		Linear Discri...	root	ipykernel...	-	sklearn	0.79	0.715	0.742	-	-	compare_m...	12649216	61e6
	8 hours ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.792	0.717	0.741	1.0	-	compare_m...	3369210a	61e6
	8 hours ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	compare_m...	820ee16	61e6
	8 hours ago		Extra Trees...	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	compare_m...	194b0a35	61e6
	8 hours ago		Random For...	root	ipykernel...	-	sklearn	0.819	0.737	0.76	-	-	compare_m...	134597f1	61e6
	8 hours ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	-	-	compare_m...	9858db6d	61e6
	8 hours ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-
	8 hours ago		Session Init...	root	ipykernel...	-	-	-	-	-	-1	4	setup	8a9e31d5	ac17
	8 hours ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	-	-	create_model	88e4e135	ac17
	8 hours ago		Naive Bayes	root	ipykernel...	-	sklearn	0.738	0.679	0.728	-	-	compare_m...	d374d183	ac17
	8 hours ago		Ridge Classi...	root	ipykernel...	-	sklearn	0	0.715	0.742	-	-	compare_m...	d3715377	ac17
	8 hours ago		Linear Discri...	root	ipykernel...	-	sklearn	0.79	0.715	0.742	-	-	compare_m...	cd194e29	ac17
	8 hours ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.792	0.717	0.741	1.0	-	compare_m...	6ee70c53	ac17
	8 hours ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	compare_m...	d3802897	ac17
	8 hours ago		Extra Trees...	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	compare_m...	f15cd136	ac17
	8 hours ago		Random For...	root	ipykernel...	-	sklearn	0.819	0.737	0.76	-	-	compare_m...	b89090fa	ac17
	8 hours ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	-	-	compare_m...	53040ac4	ac17
	8 hours ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-

mlflow 1.26.1ExperimentsModels

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 8	-	Version 6	2023-06-26 23:38:26	-

ML FLOW - Training Pipeline (Model Training

Registry View

scriptsJupyterRent GJarvisLaDAGsLead_scMLxMLflowMLflowMLflowMLflowMLflowAirflowpythonpythonNew taNew ta+

←↻⚠ Not secure | notebooks.jarvislabs.ai:10960/#/experiments/3

⌕🌟🔖🗑🔌👤⋮

mlflow1.26.1ExperimentsModels

GitHubDocs

Experiments

+

<

Lead_Scoring_Training_Pipeline📄

Search Experiments

Default🔗🗑

Lead_scoring_model_e...🔗🗑

Lead_scoring🔗🗑

Lead_Scoring_Training...🔗🗑

Lead_Scoring_Training...🔗🗑

📌 Track machine learning training runs in experiments. [Learn more](#)

Experiment ID: 3

▶ Description [Edit](#)

↻ Refresh

Compare

Delete

Download CSV📄

↓ Start Time ▾

All time ▾

☰

⌵ Columns

Only show differences ☐

🔍 metrics.rmse < 1 and params.model = "tree" S

Showing 8 matching runs

	↓ Start Time	Duration	Run Name	User	Source	Version	Model
<input type="checkbox"/>	2 hours ago	4.8s	Lead_scorin...	root	airflow	-	LightGBM/8
<input type="checkbox"/>	2 hours ago	5.5s	Lead_scorin...	root	airflow	-	LightGBM/7
<input type="checkbox"/>	3 hours ago	6.0s	Lead_scorin...	root	airflow	-	LightGBM/6
<input type="checkbox"/>	7 hours ago	5.1s	Lead_scorin...	root	airflow	-	LightGBM/5
<input type="checkbox"/>	8 hours ago	6.3s	Lead_scorin...	root	airflow	-	LightGBM/4
<input type="checkbox"/>	9 hours ago	7.1s	Lead_scorin...	root	ipykernel...	-	LightGBM/3
<input type="checkbox"/>	9 hours ago	7.1s	Lead_scorin...	root	airflow	-	LightGBM/2
<input type="checkbox"/>	10 hours ago	6.2s	Lead_scorin...	root	airflow	-	LightGBM/1

Multiple Pipeline Runs

scriptsJupyterRent GJarvisLaDAGsLead_scMLxMLflowMLflowMLflowMLflowMLflowAirflowpythonpythonNew taNew ta+

←↻⚠ Not secure | notebooks.jarvislabs.ai:10960/#/experiments/3

⌕🌟🔖🗑🔌👤⋮

mlflow1.26.1ExperimentsModels

GitHubDocs

Experiments

+

<

Lead_scoring_mlflow_production📄

Search Experiments

Default🔗🗑

Lead_scoring_model_...🔗🗑

Lead_scoring🔗🗑

Lead_scoring_mlflow_...🔗🗑

Lead_Scoring_Trainin...🔗🗑

📌 Track machine learning training runs in experiments. [Learn more](#)

Experiment ID: 3

▶ 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

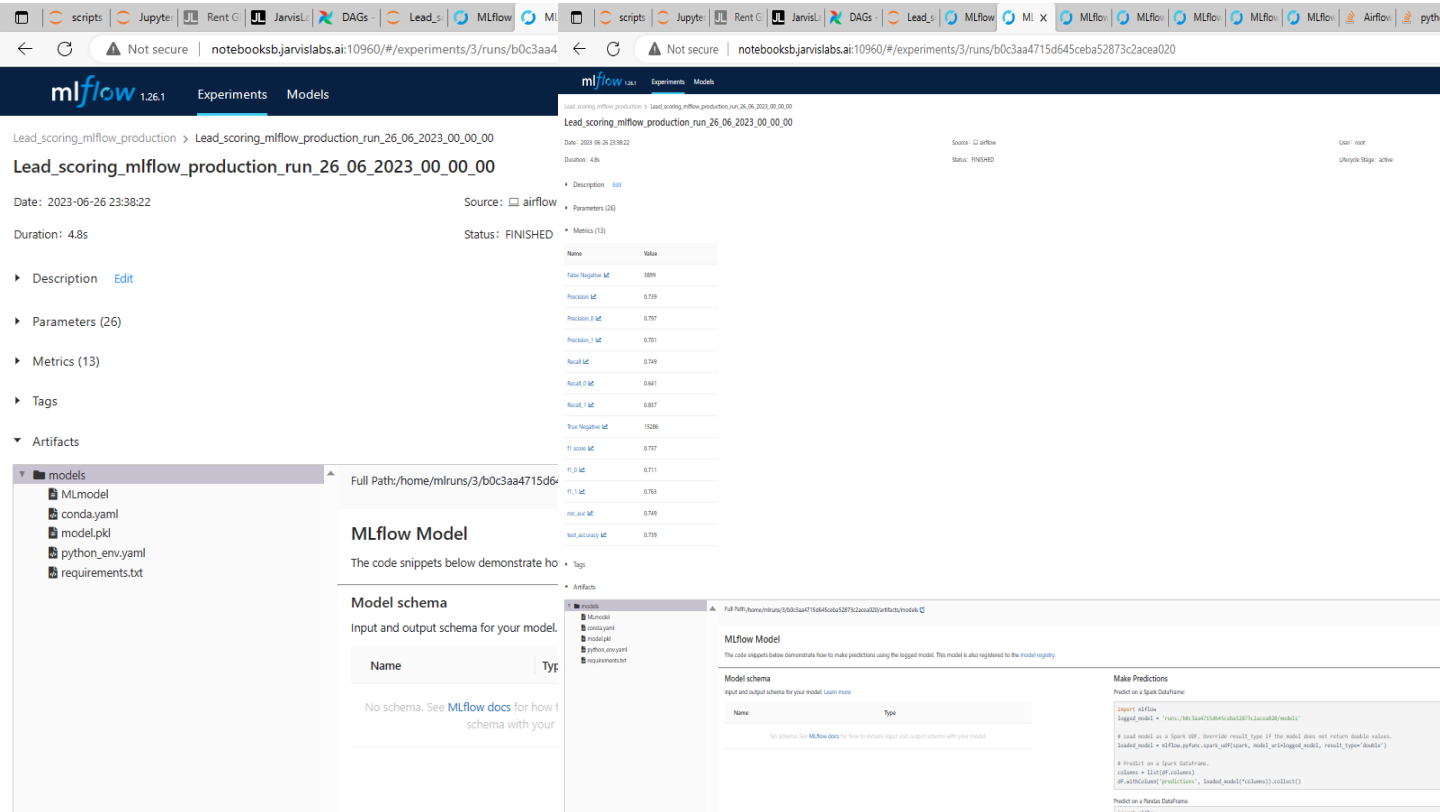
Showing 8 matching runs

	↓ Start Time	Duration	Run Name	User	Source	Version	Models	Metrics >	
								AUC	False Negative
<input type="checkbox"/>	2 hours ago	4.8s	Lead_scorin...	root	airflow	-	LightGBM/8	-	3899
<input type="checkbox"/>	2 hours ago	5.5s	Lead_scorin...	root	airflow	-	LightGBM/7	-	3899
<input type="checkbox"/>	3 hours ago	6.0s	Lead_scorin...	root	airflow	-	LightGBM/6	-	3899
<input type="checkbox"/>	7 hours ago	5.1s	Lead_scorin...	root	airflow	-	LightGBM/5	-	3899
<input type="checkbox"/>	8 hours ago	6.3s	Lead_scorin...	root	airflow	-	LightGBM/4	-	3899
<input type="checkbox"/>	9 hours ago	7.1s	Lead_scorin...	root	ipykernel...	-	LightGBM/3	-	3899
<input type="checkbox"/>	9 hours ago	7.1s	Lead_scorin...	root	airflow	-	LightGBM/2	0.751	3915
<input type="checkbox"/>	10 hours ago	6.2s	Lead_scorin...	root	airflow	-	LightGBM/1	0.751	3915

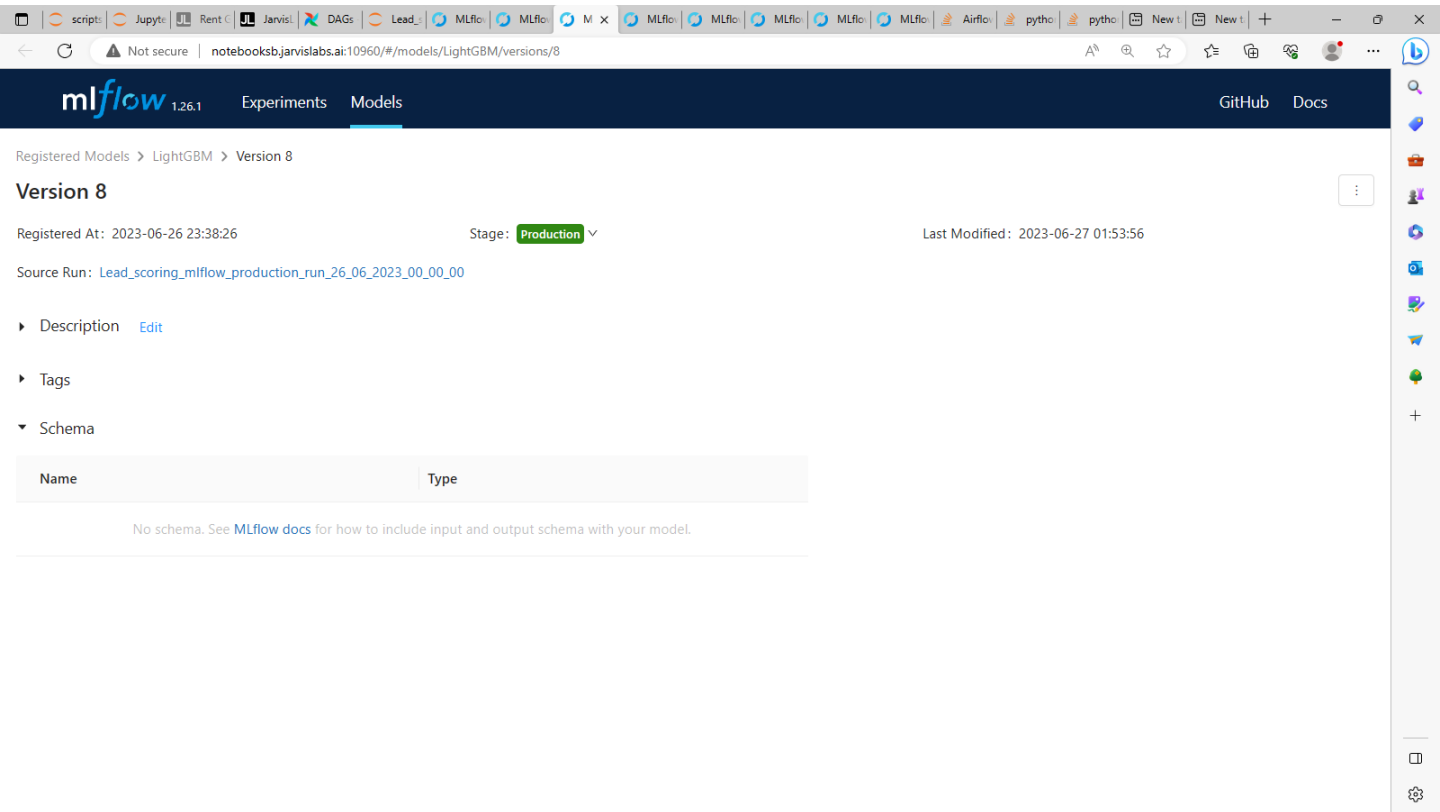
Load more

ML FLOW - Training Pipeline (Model Training)

- Details of model at Production Stage
- Screenshot of experiments with all the artifacts visible



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



ML FLOW - Model Experimentation

● Screenshot of all the experiments at Model Experimentation

Experiments

Search Experiments

Default

Lead_scoring_model...

Lead_scoring

Lead_scoring_miflow...

Lead_scoring_Traini...

Lead_scoring_model_experimentation

Track machine learning training runs in experiments. Learn more

Experiment ID: 1

Description

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

	Duration	Run Name	User	Source	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source	URI	USI
<input type="checkbox"/>	11 hours ago	Session Init...	root	ipykernel...	-	-	-	-	-	-1	4	setup	8b25333	61e0
<input type="checkbox"/>	11 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	create_model	c4b36039	61e0
<input type="checkbox"/>	11 hours ago	Naive Bayes	root	ipykernel...	sklearn	0.738	0.679	0.728	-	-	-	compare_m...	01a0026f	61e0
<input type="checkbox"/>	11 hours ago	Ridge Classi...	root	ipykernel...	sklearn	0	0.715	0.742	-	-	-	compare_m...	9d49d448	61e0
<input type="checkbox"/>	11 hours ago	Linear Discr...	root	ipykernel...	sklearn	0.79	0.715	0.742	-	-	-	compare_m...	12d49716	61e0
<input type="checkbox"/>	11 hours ago	Logistic Reg...	root	ipykernel...	sklearn	0.792	0.717	0.741	1.0	-	-	compare_m...	33d9210a	61e0
<input type="checkbox"/>	11 hours ago	Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	82ddee16	61e0
<input type="checkbox"/>	11 hours ago	Extra Trees ...	root	ipykernel...	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	1f1d8b35	61e0
<input type="checkbox"/>	11 hours ago	Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	134597f1	61e0
<input type="checkbox"/>	11 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	compare_m...	9658dbdd	61e0
<input type="checkbox"/>	12 hours ago	Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-
<input type="checkbox"/>	12 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	create_model	884e4135	ac17
<input type="checkbox"/>	12 hours ago	Naive Bayes	root	ipykernel...	sklearn	0.738	0.679	0.728	-	-	-	compare_m...	d374d183	ac17
<input type="checkbox"/>	12 hours ago	Ridge Classi...	root	ipykernel...	sklearn	0	0.715	0.742	-	-	-	compare_m...	d5715377	ac17
<input type="checkbox"/>	12 hours ago	Linear Discr...	root	ipykernel...	sklearn	0.79	0.715	0.742	-	-	-	compare_m...	cd194a29	ac17
<input type="checkbox"/>	12 hours ago	Logistic Reg...	root	ipykernel...	sklearn	0.792	0.717	0.741	1.0	-	-	compare_m...	6ee78c53	ac17
<input type="checkbox"/>	12 hours ago	Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	d3802897	ac17
<input type="checkbox"/>	12 hours ago	Extra Trees ...	root	ipykernel...	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	f15cd136	ac17
<input type="checkbox"/>	12 hours ago	Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	b890a0fa	ac17
<input type="checkbox"/>	12 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	compare_m...	53040ac4	ac17
<input type="checkbox"/>	12 hours ago	Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-

Load more

After Dropping Feature

Before Dropping Feature

mlflow 1.26.1

Experiments

Models

GitHub

Docs

Lead_scoring_model_experimentation

Track machine learning training runs in experiments. Learn more

Experiment ID: 1

Description

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

	Duration	Run Name	User	Source	Models	AUC	Accuracy	F1	C	CPU Jobs	Categor
<input type="checkbox"/>	8 hours ago	Session Init...	root	ipykernel...	-	-	-	-	-	-1	4
<input type="checkbox"/>	8 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-
<input type="checkbox"/>	8 hours ago	Naive Bayes	root	ipykernel...	sklearn	0.738	0.679	0.728	-	-	-
<input type="checkbox"/>	8 hours ago	Ridge Classi...	root	ipykernel...	sklearn	0	0.715	0.742	-	-	-
<input type="checkbox"/>	8 hours ago	Linear Discr...	root	ipykernel...	sklearn	0.79	0.715	0.742	-	-	-
<input type="checkbox"/>	8 hours ago	Logistic Reg...	root	ipykernel...	sklearn	0.792	0.717	0.741	1.0	-	-
<input type="checkbox"/>	8 hours ago	Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	-	-	-
<input type="checkbox"/>	8 hours ago	Extra Trees C...	root	ipykernel...	sklearn	0.818	0.737	0.758	-	-	-
<input type="checkbox"/>	8 hours ago	Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	-	-	-
<input type="checkbox"/>	8 hours ago	Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-

ML FLOW – Jupyter Lab

● Screenshot of the details of the model experiment before dropping Features

Untitled1.ipynb (7) - JupyterLab JarvisLabs.ai lead_scoring... - JupyterLab DAGs - Airflow MLflow New tab

https://notebooksbjarvislabs.ai/065/lab/tree/Assignment/02_training_pipeline/notebooks/lead_scoring_model_experimentation.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

02_training_pipeline / notebooks /

Name

Last Modified

Data

10 days ago

Lead_scoring_mlflow_production.db

a day ago

lead_scoring_model_experimentation.db

7 hours ago

lead_scoring_model_experimentation.ipynb

a day ago

logslog

a day ago

Terminal 1 Terminal 2 Terminal 3 lead_scoring_model_experim X

Python 3 (ipykernel)

```
mlflow.set_tracking_uri("http://0.0.0.0:5006")

[8]: # setup pycaret
exp_lead_scoring = setup(data=dataset, target = 'app_complete_flag',
                        fold_shuffle=True,
                        session_id = 42,
                        normalize = True,
                        transformation = True,
                        remove_multicollinearity = True, multicollinearity_threshold = 0.95,
                        n_jobs=-1,use_gpu=False,
                        log_experiment=True,experiment_name='Lead_scoring_model_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, 12)
5	Missing Values	False
6	Numeric Features	7
7	Categorical Features	4
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(167274, 43)
12	Transformed Test Set	(71690, 43)
13	Shuffle Train-Test	True

Simple 3 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 lead_scoring_model_experimentation.ipynb

Untitled1.ipynb (7) - JupyterLab JarvisLabs.ai lead_scoring... - JupyterLab DAGs - Airflow

https://notebooksbjarvislabs.ai/065/lab/tree/Assignment/02_training_pipeline/notebooks/lead_scoring_model_experimentation.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

02_training_pipeline / notebooks /

Name

Last Modified

Data

10 days ago

Lead_scoring_mlflow_production.db

a day ago

lead_scoring_model_experimentation.db

7 hours ago

lead_scoring_model_experimentation.ipynb

a day ago

logslog

a day ago

Terminal 1 Terminal 2 Terminal 3 lead_scoring_model_experim X

Python 3 (ipykernel)

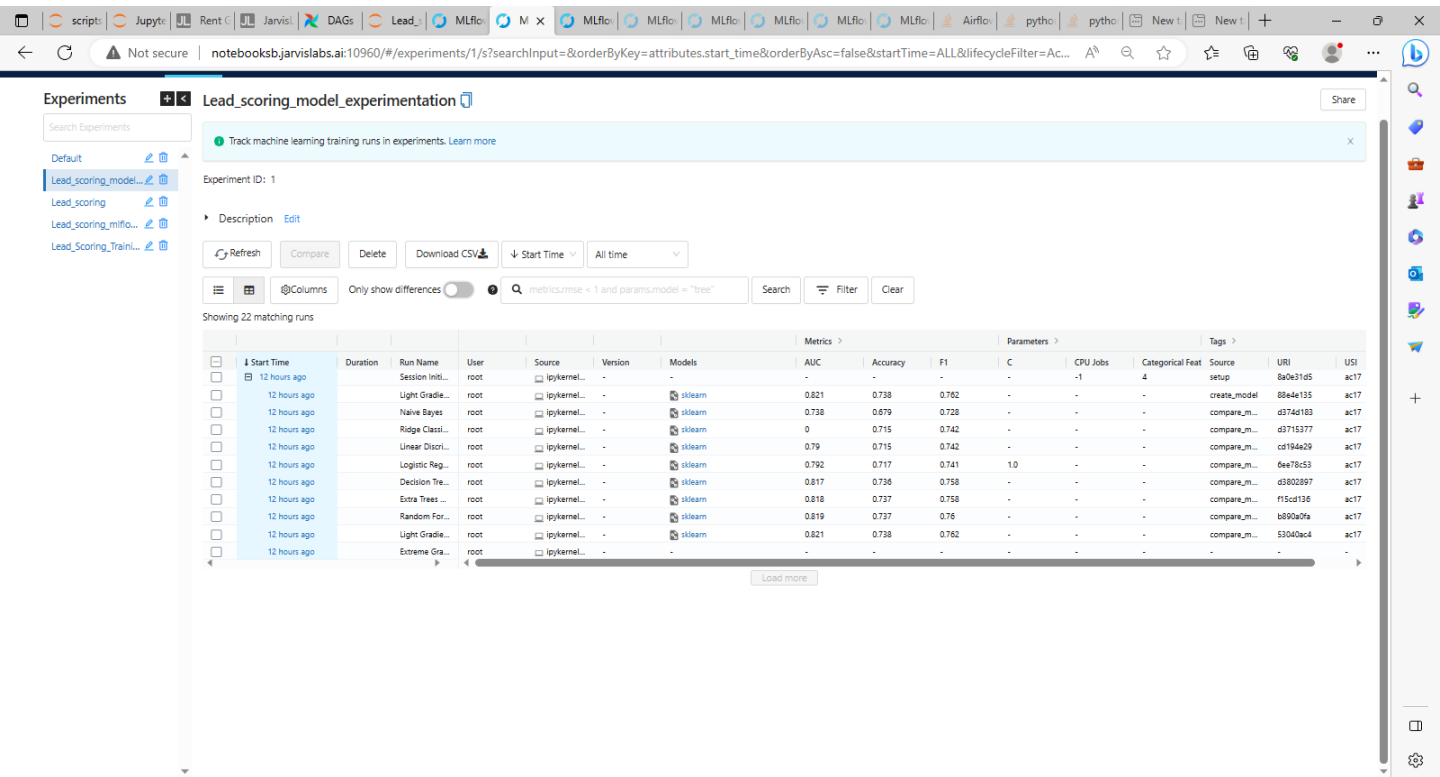
```
silent=True, verbose=True,
log_profile=False)
```

	Description	Value
0	session_id	42
1	Target	app_complete_flag
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(238964, 12)
5	Missing Values	False
6	Numeric Features	7
7	Categorical Features	4
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(167274, 43)
12	Transformed Test Set	(71690, 43)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	True
20	Experiment Name	Lead_scoring_model_experimentation
21	USI	61e6
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	mean
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent
29	Normalize	True
30	Normalize Method	zscore
31	Transformation	True
32	Transformation Method	yeo-johnson
33	PCA	False
34	PCA Method	None

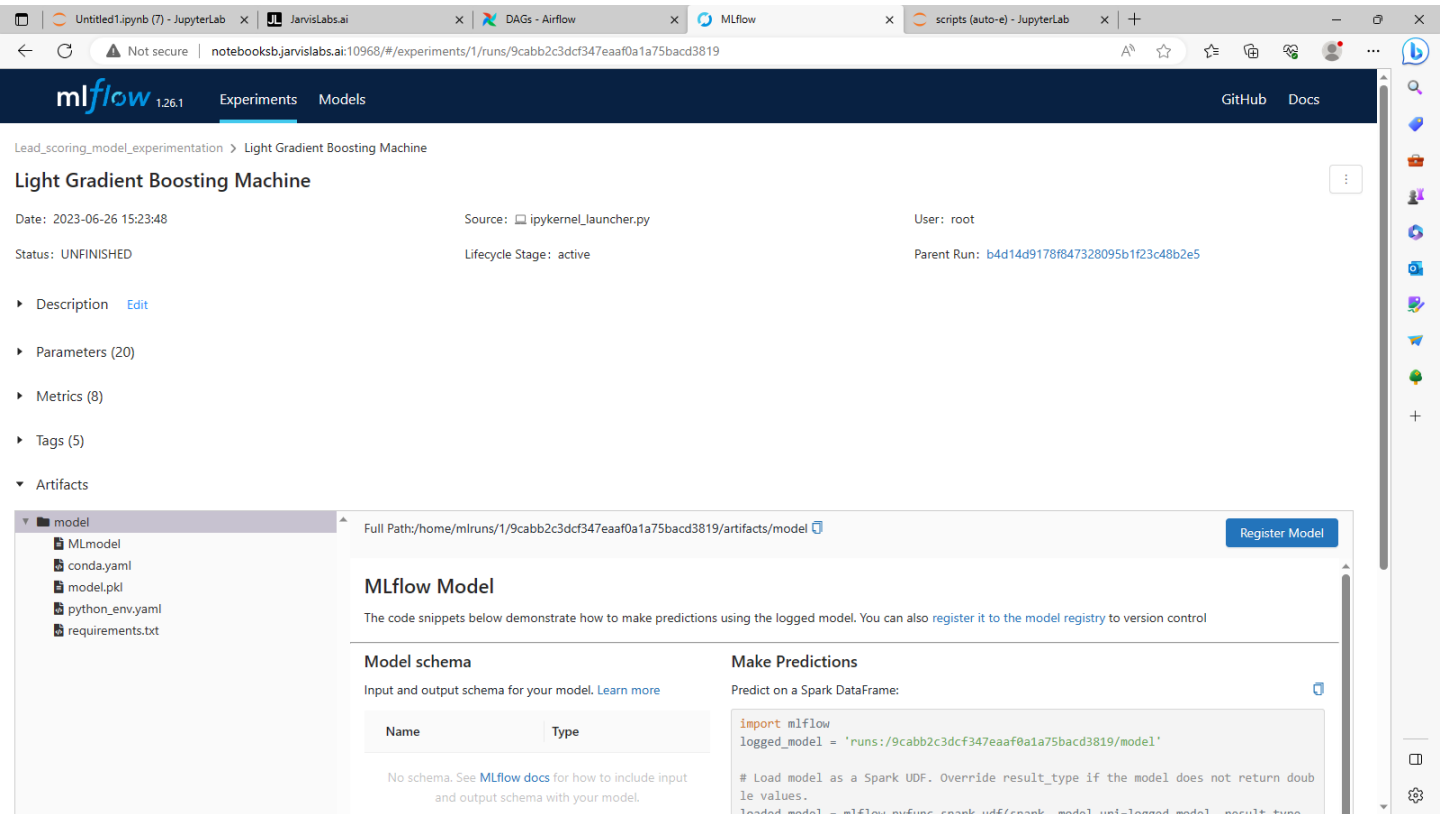
Simple 3 1 Python 3 (ipykernel) | Idle

ML FLOW - Model Experimentation

- Screenshot of all the experiments before dropping Features



- Screenshot of one experiments with all the artifacts visible



ML FLOW – Jupyter Lab

● Screenshot of the model Experimentation Notebook after dropping features

5 : Model Experimentation after dropping features

From the above feature tests we can clearly see that some of the features are not significant. We will now drop all the insignificant features and select only the significant ones. The list of the significant features is ['total_leads_dropped', 'city_tier', 'referred_lead', 'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c']. So now you will train your model with only these features.

Also note that in our previous experiments we saw that tree based models are our top performers. In case of tree based models we do not require transformations normalization, scaling etc. So make sure that you use setup pycaret in the proper way, i.e. make sure that you use normalize = False and transformation = False.

```
[12]: # ['total_leads_dropped', 'city_tier', 'referred_lead', 'app_complete_flag', 'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c']
#
# Train the model using the features listed above. Since we are using tree models we do not require any transformations
# such as normalization, scaling etc. So make sure that you use setup pycaret in the proper way, i.e. make sure that you use
# normalize = False and transformation = False.
cols = ['total_leads_dropped', 'city_tier', 'referred_lead', 'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c', 'app_complete_flag']
dataset_cols = dataset[cols]
```

```
[13]: from pycaret.classification import *

exp_lead_scoring = setup(data=dataset_cols, target = 'app_complete_flag',
                        remove_multicollinearity = True, multicollinearity_threshold = 0.95,
                        categorical_features = ['city_tier', 'first_platform_c', 'first_utm_medium_c', 'first_utm_source_c'],
                        fold_shuffle=True,
                        session_id = 42,
                        n_jobs=-1, use_gpu=False,
                        log_experiment=True, experiment_name='Lead_scoring',
                        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, 7)
5	Missing Values	False
6	Numeric Features	1
7	Categorical Features	5
8	Ordinal Features	False
9	High Cardinality Features	False

```
log_pycaret=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, 7)
5	Missing Values	False
6	Numeric Features	1
7	Categorical Features	5
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
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	True
20	Experiment Name	Lead_scoring
21	URI	IDS
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	mean
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent

ML FLOW

● Screenshot of all the experiments after dropping features

scriptsJupyterRent JarvisLeadMLflowMLflowMLflowMLflowAirflowDatabapython+

Not secure | notebooks.jarvislabs.ai:10960/#/experiments/1/s?searchInput=&orderByKey=attributes.start_time&orderByAsc=false&startTime=ALL&lifecycleFilter=Ac...

mlflow1.26.1ExperimentsModelsGitHubDocs

Lead_scoring_model_experimentation

Share

Track machine learning training runs in experiments. [Learn more](#)

Experiment ID: 1

DescriptionEdit

RefreshCompareDeleteDownload CSVStart TimeAll Time

ColumnsOnly show differencesmetrics.mse < 1 and params.model = "tree"SearchFilterClear

Showing 22 matching runs

	Start Time	Duration	Run Name	User	Source	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source	URI	USI
	8 hours ago		Session Init...	root	ipykernel...	-	-	-	-	-	-1	4	setup	fb2c25333	61e6
	8 hours ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	create_model	c4b36639	61e6
	8 hours ago		Naive Bayes	root	ipykernel...	sklearn	0.738	0.679	0.728	-	-	-	compare_m...	01a0026f	61e6
	8 hours ago		Ridge Classi...	root	ipykernel...	sklearn	0	0.715	0.742	-	-	-	compare_m...	9d49d448	61e6
	8 hours ago		Linear Discr...	root	ipykernel...	sklearn	0.79	0.715	0.742	-	-	-	compare_m...	126d2716	61e6
	8 hours ago		Logistic Reg...	root	ipykernel...	sklearn	0.792	0.717	0.741	1.0	-	-	compare_m...	33d9210a	61e6
	8 hours ago		Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	820ee16	61e6
	8 hours ago		Extra Trees ...	root	ipykernel...	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	19f48435	61e6
	8 hours ago		Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	13459771	61e6
	8 hours ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	compare_m...	9658dbd6	61e6
	8 hours ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-
	9 hours ago		Session Init...	root	ipykernel...	-	-	-	-	-	-1	4	setup	8a0e31d5	ac17
	9 hours ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	create_model	884e135	ac17
	9 hours ago		Naive Bayes	root	ipykernel...	sklearn	0.738	0.679	0.728	-	-	-	compare_m...	d3746163	ac17
	9 hours ago		Ridge Classi...	root	ipykernel...	sklearn	0	0.715	0.742	-	-	-	compare_m...	d3715377	ac17
	9 hours ago		Linear Discr...	root	ipykernel...	sklearn	0.79	0.715	0.742	-	-	-	compare_m...	cd194a29	ac17
	9 hours ago		Logistic Reg...	root	ipykernel...	sklearn	0.792	0.717	0.741	1.0	-	-	compare_m...	6ee78c53	ac17
	9 hours ago		Decision Tre...	root	ipykernel...	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	d3802897	ac17
	9 hours ago		Extra Trees ...	root	ipykernel...	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	f15cd136	ac17
	9 hours ago		Random For...	root	ipykernel...	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	b890a09a	ac17
	9 hours ago		Light Gradie...	root	ipykernel...	sklearn	0.821	0.738	0.762	-	-	-	compare_m...	53040ac4	ac17
	9 hours ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-	-

Load more

● Screenshot of one experiments with all the artifacts visible

scriptsJupyterRent JarvisDAGsLeadM xMLflowMLflowMLflowMLflowMLflowMLflowMLflowAirflowpythopythonNewNew+

Not secure | notebooks.jarvislabs.ai:10960/#/experiments/1/runs/286af6552ccf4c92a7db533d99e54680

mlflow1.26.1ExperimentsModelsGitHubDocs

Lead_scoring_model_experimentation > Light Gradient Boosting Machine

...

Date: 2023-06-26 15:29:46

Source: ipykernel_launcher.py

User: root

Status: UNFINISHED

Lifecycle Stage: active

Parent Run: e058498e724241e59fe4c800fc3c6cdf

DescriptionEdit

Parameters (20)

Metrics (8)

Tags (5)

Artifacts

model

Full Path: /home/mlruns/1/286af6552ccf4c92a7db533d99e54680/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:/286af6552ccf4c92a7db533d99e54680/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:/286af6552ccf4c92a7db533d99e54680/model'

# Load model as a PyFuncModel.
loaded_model = mlflow.pyfunc.load_model(logged_model)

# Predict on a Pandas DataFrame.
import pandas as pd
```