

Screenshots for MLOps Assignment

1. Screenshot of MLFlow UI (Before dropping features)

a. Screenshot of all the experiments

Baseline_model_exp01

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.mean < 1 and param.model = "tree" Search Filter Clear

Showing 11 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs
<input type="checkbox"/>	10 minutes ago		Session Init...	root	ipykernel...	-	-	-	-	-	-	-1
<input type="checkbox"/>	8 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-
<input type="checkbox"/>	9 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.663	0.727	-	-
<input type="checkbox"/>	9 minutes ago		Linear Discr...	root	ipykernel...	-	sklearn	0.773	0.701	0.728	-	-
<input type="checkbox"/>	9 minutes ago		Ridge Classi...	root	ipykernel...	-	sklearn	0	0.701	0.728	-	-
<input type="checkbox"/>	9 minutes ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.784	0.71	0.74	1.0	-
<input type="checkbox"/>	9 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.817	0.735	0.761	-	-
<input type="checkbox"/>	9 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-
<input type="checkbox"/>	9 minutes ago		Extra Trees ...	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-
<input type="checkbox"/>	9 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-
<input type="checkbox"/>	9 minutes ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-

b. Screenshot of one experiments with all the artifacts visible

Baseline_model_exp01 > Light Gradient Boosting Machine

Date: 2022-10-22 11:58:50 Source: ipykernel_launcher.py User: root

Status: UNFINISHED Lifecycle Stage: active Parent Run: eea0a590ec44b00a426c2303088937d

Description Edit

Parameters (21)

Metrics (8)

Tags (5)

Artifacts

Full Path: /home/Assignment/mlruns/1/b405db01850b4b5fbffacea0cc4c6f7/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

Make Predictions

Predict on a Spark DataFrame

Import: mlflow

2. Screenshot of MLFlow UI (Before dropping features)

a. Screenshot of all the experiments

Baseline_model_exp02

Track machine learning training runs in experiments. Learn more

Experiment ID: 2

Description Edit

Refresh Compare Delete Download CSV Start Time All time

Columns Only show differences metrics.mean < 1 and param.model = "tree" Search Filter Clear

Showing 11 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs
<input type="checkbox"/>	21 minutes ago		Session Init...	root	ipykernel...	-	-	-	-	-	-	-1
<input type="checkbox"/>	19 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-
<input type="checkbox"/>	20 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.67	0.723	-	-
<input type="checkbox"/>	20 minutes ago		Linear Discr...	root	ipykernel...	-	sklearn	0.773	0.7	0.728	-	-
<input type="checkbox"/>	20 minutes ago		Ridge Classi...	root	ipykernel...	-	sklearn	0	0.7	0.728	-	-
<input type="checkbox"/>	20 minutes ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.784	0.71	0.74	1.0	-
<input type="checkbox"/>	20 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.818	0.735	0.762	-	-
<input type="checkbox"/>	20 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-
<input type="checkbox"/>	20 minutes ago		Extra Trees ...	root	ipykernel...	-	sklearn	0.817	0.737	0.758	-	-
<input type="checkbox"/>	20 minutes ago		Extreme Gra...	root	ipykernel...	-	sklearn	0.821	0.738	0.761	-	-
<input type="checkbox"/>	20 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.763	-	-

b. Screenshot of one experiments with all the artifacts visible

Light Gradient Boosting Machine

Date: 2022-10-22 12:00:47 Source: ipykernel_launcher.py User: root

Status: UNFINISHED Lifecycle Stage: active Parent Run: f54bf996382842a2811ea5f4a4e9bad7

- Description Edit
- Parameters (21)
- Metrics (8)
- 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/Assignment/mlruns/2/e0b1a02290f34006a73ef0a4612650da/artifacts/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	Make Predictions				
Input and output schema for your model. Learn more	Predict on a Spark DataFrame:				
<table border="1"> <thead> <tr> <th>Name</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Name	Type			<pre>import mlflow logged_model = 'runs:/e0b1a02290f34006a73ef0a4612650da/model'</pre>
Name	Type				

3. Create an airflow dag python file for data pipeline

a. Screenshot of successful execution Airflow DAG in graph

b. Screenshot of Airflow UI grid

DAG: Lead_Scoring_Data_Engineering_Pipeline DAG to run data pipeline for lead scoring

Schedule: @daily

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

22-10-2022 10:59:27 25 All Run Types All Run States Clear Filters

Auto-refresh

DAG Run Details Graph Mark Failed Re-run: Clear existing tasks

Status: success

Run Id: manual_2022-10-22T10:59:25.911149+00:00

Run Type: manual

Duration: 00:01:19

Last Scheduling Decision: 2022-10-22, 11:00:46 UTC

Started: 2022-10-22, 10:59:27 UTC

Ended: 2022-10-22, 11:00:46 UTC

4. Setup mlflow in utils.py

a. Screenshot of experiments with all the artifacts visible

MLflow 1.26.1 Experiments Models GitHub Docs

Experiments + < Lead_scoring_mlflow_production Share

Search Experiments: Default Lead_scoring_Train... Lead_scoring_mlflow...

Experiment ID: 2

Description Edit

Refresh Compare Delete Download CSV Start Time All time

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

Showing 2 matching runs

	Start Time	Duration	Run Name	User	Source	Version	Models	Metrics
<input type="checkbox"/>	6 minutes ago	10.2s	Lead_scoring...	root	airflow	-	LightGBM/3	AUC: 0.751, False Negative: 3891, False Positive: 8459
<input type="checkbox"/>	17 minutes ago	10.2s	Lead_scoring...	root	pykernel...	-	LightGBM/2	AUC: 0.751, False Negative: 3891, False Positive: 8459

Lead_scoring_mlflow_production > Lead_scoring_mlflow_production_22_10_2022_00_00_00

Lead_scoring_mlflow_production_22_10_2022_00_00_00

Date: 2022-10-22 17:44:42

Source: airflow

User: root

Duration: 10.2s

Status: FINISHED

Lifecycle Stage: active

Description Edit

Parameters (21)

Metrics (9)

Tags

Artifacts

models

MLmodel
conda.yaml
model.pkl
python_env.yaml
requirements.txt

Full Path: /home/Assignment/minruns/2/84cf821c0e314a5182b8c3a47381d71a/artifacts/models

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

Make Predictions

Predict on a Spark DataFrame:

```
import mlflow
logged_model = 'runs:/84cf821c0e314a5182b8c3a47381d71a/model1s'
```

b. Screenshot of model registry with model name and stage as 'production'

Registered Models > LightGBM > Version 3

Version 3

Registered At: 2022-10-22 17:44:52 Stage: Production Last Modified: 2022-10-22 17:47:55

Source Run: Lead_scoring_mlflow_production_22_10_2022_00_00_00

Description Edit

Tags

Schema

Name	Type

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

5. Create an airflow dag python file for training pipeline

a. Screenshot of successful execution Airflow DAG in graph

DAG: Lead_scoring_training_pipeline Training pipeline for Lead Scoring System success Schedule: @monthly Next Run: 2022-10-01, 00:00:00

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

2022-10-22T12:14:30Z Runs 25 Run manual_2022-10-22T12:14:29.074221+00:00 Layout Left > Right Update Find Task...

PythonOperator

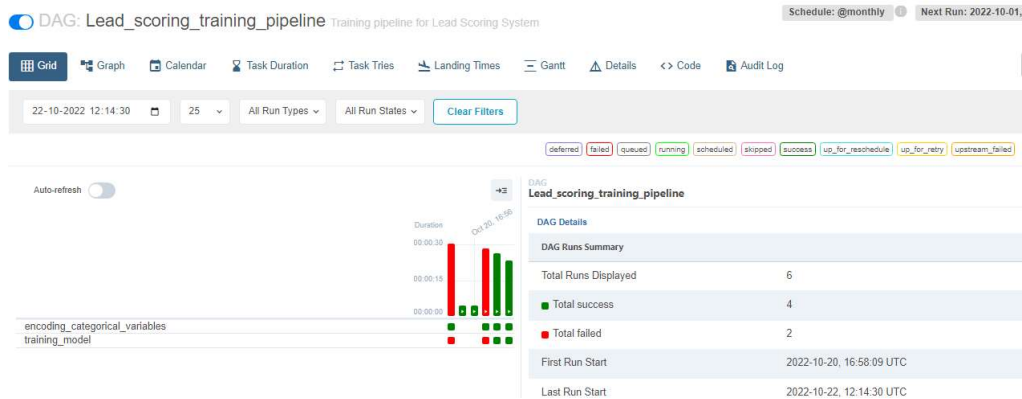
deferred failed queued running scheduled skipped success up_for_retry upstream_failed no_status

Auto-refresh

```

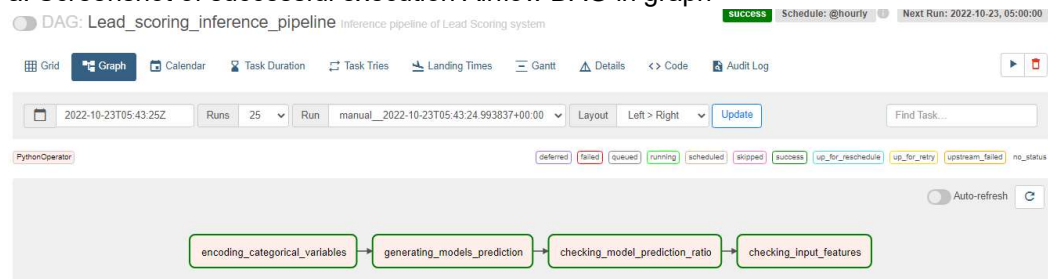
graph LR
    encoding_categorical_variables --> training_model
  
```

b. Screenshot of Airflow UI grid



6. Create an airflow dag python file for inference pipeline

a. Screenshot of successful execution Airflow DAG in graph



b. Screenshot of Airflow UI grid

