

## Data Pipeline:

Airflow DAG: Lead\_Scoring\_Data\_Engineering\_Pipeline DAG to run data pipeline for lead scoring

success Schedule: @daily Next Run: 2023-04-15, 00:00:00

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

Audit Log

2023-04-15T10:04:53Z Runs 25 Run manual\_2023-04-15T10:04:52.396864+00:00 Find Task...

Layout Left > Right Update

PythonOperator deferred failed queued running scheduled skipped success up\_for\_reschedule up\_for\_retry upstream\_failed no\_status

Auto-refresh

```
graph LR; building_db --> checking_raw_data_schema; checking_raw_data_schema --> loading_data; loading_data --> mapping_city_tier; mapping_city_tier --> mapping_categorical_vars; mapping_categorical_vars --> mapping_interactions; mapping_interactions --> checking_model_inputs_schema;
```

Airflow DAG: Lead\_Scoring\_Data\_Engineering\_Pipeline DAG to run data pipeline for lead scoring

Schedule: @daily Next Run: 2023-04-15, 00:00:00

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

Audit Log

15-04-2023 10:06:17 25 All Run Types All Run States Clear Filters

Auto-refresh

Duration

00:01:23

00:00:41

00:00:00

Apr 15 09:55

building\_db

checking\_raw\_data\_schema

loading\_data

mapping\_city\_tier

mapping\_categorical\_vars

mapping\_interactions

checking\_model\_inputs\_schema

DAG Lead\_Scoring\_Data\_Engineering\_Pipeline

DAG Details

DAG Runs Summary

|                      |                          |
|----------------------|--------------------------|
| Total Runs Displayed | 8                        |
| Total success        | 3                        |
| Total failed         | 5                        |
| First Run Start      | 2023-04-15, 09:43:07 UTC |
| Last Run Start       | 2023-04-15, 10:04:53 UTC |
| Max Run Duration     | 00:01:23                 |

Model experimentation:

Search Experiments

Default

model\_exp02

lead\_scoring\_model\_...

lead\_scoring\_model\_...

Experiment ID: 2

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 56 matching runs

|  |                |          |                  |      |              |         |         | Metrics |
|--|----------------|----------|------------------|------|--------------|---------|---------|---------|
|  | Start Time     | Duration | Run Name         | User | Source       | Version | Models  | AUC     |
|  | 11 minutes ago |          | Session Initi... | root | ipykernel... | -       | -       | -       |
|  | 9 minutes ago  |          | Light Gradie...  | root | ipykernel... | -       | sklearn | 0.814   |
|  | 9 minutes ago  |          | Naive Bayes      | root | ipykernel... | -       | sklearn | 0.723   |
|  | 9 minutes ago  |          | Linear Discri... | root | ipykernel... | -       | sklearn | 0.764   |
|  | 9 minutes ago  |          | Ridge Classif... | root | ipykernel... | -       | sklearn | 0       |
|  | 9 minutes ago  |          | Logistic Reg...  | root | ipykernel... | -       | sklearn | 0.777   |
|  | 9 minutes ago  |          | Decision Tre...  | root | ipykernel... | -       | sklearn | 0.81    |
|  | 9 minutes ago  |          | Extra Trees C... | root | ipykernel... | -       | sklearn | 0.811   |
|  | 9 minutes ago  |          | Random For...    | root | ipykernel... | -       | sklearn | 0.812   |
|  | 9 minutes ago  |          | Extreme Gra...   | root | ipykernel... | -       | sklearn | 0.814   |
|  | 9 minutes ago  |          | Light Gradie...  | root | ipykernel... | -       | sklearn | 0.814   |

mlflow1.26.1ExperimentsModels

lead\_scoring\_model\_experimentation > Light Gradient Boosting Machine

Light Gradient Boosting Machine

Date: 2023-04-19 20:38:17

Source: ipykernel\_launcher.py

User: root

Status: UNFINISHED

Lifecycle Stage: active

Parent Run: b05330a9b64f4fe1ba3e2fed51ac44fc

Description Edit

Parameters (20)

Metrics (8)

| Name     | Value |
|----------|-------|
| AUC      | 0.814 |
| Accuracy | 0.736 |
| F1       | 0.761 |
| Kappa    | 0.472 |
| MCC      | 0.481 |
| Prec.    | 0.7   |
| Recall   | 0.833 |

lead\_scoring\_model\_experimentation\_significant > Light Gradient Boosting Machine

## Light Gradient Boosting Machine



Date: 2023-04-19 20:42:35

Source: ipykernel\_launcher.py

User: root

Status: UNFINISHED

Lifecycle Stage: active

Parent Run: 606cc1c40a194eda97c4c1b94d63b1b8

▸ Description [Edit](#)

▸ Parameters (20)

▼ Metrics (8)

| Name                     | Value |
|--------------------------|-------|
| <a href="#">AUC</a>      | 0.814 |
| <a href="#">Accuracy</a> | 0.736 |
| <a href="#">F1</a>       | 0.76  |
| <a href="#">Kappa</a>    | 0.471 |
| <a href="#">MCC</a>      | 0.48  |
| <a href="#">Prec.</a>    | 0.7   |

▸ Description [Edit](#)

Refresh Compare Delete Download CSV Start Time All time



Columns Only show differences  Search Filter Clear

Showing 62 matching runs


|                          |                 |          |                                 |         |         | Metrics > |          |
|--------------------------|-----------------|----------|---------------------------------|---------|---------|-----------|----------|
| <input type="checkbox"/> | ⌄ Start Time    | Duration | Run Name                        | Version | Models  | AUC       | Accuracy |
| <input type="checkbox"/> | ⌄ 5 minutes ago |          | Session Initialized d1dc        |         | -       | -         | -        |
| <input type="checkbox"/> | 1 minute ago    |          | Light Gradient Boosting Machine |         | sklearn | 0.814     | 0.736    |
| <input type="checkbox"/> | 1 minute ago    |          | Naive Bayes                     |         | sklearn | 0.723     | 0.66     |
| <input type="checkbox"/> | 1 minute ago    |          | Linear Discriminant Analysis    |         | sklearn | 0.764     | 0.699    |
| <input type="checkbox"/> | 1 minute ago    |          | Ridge Classifier                |         | sklearn | 0         | 0.699    |
| <input type="checkbox"/> | 2 minutes ago   |          | Logistic Regression             |         | sklearn | 0.777     | 0.709    |
| <input type="checkbox"/> | 2 minutes ago   |          | Decision Tree Classifier        |         | sklearn | 0.81      | 0.734    |
| <input type="checkbox"/> | 2 minutes ago   |          | Extra Trees Classifier          |         | sklearn | 0.811     | 0.734    |
| <input type="checkbox"/> | 2 minutes ago   |          | Random Forest Classifier        |         | sklearn | 0.812     | 0.735    |
| <input type="checkbox"/> | 2 minutes ago   |          | Extreme Gradient Boosting       |         | sklearn | 0.814     | 0.736    |
| <input type="checkbox"/> | 2 minutes ago   |          | Light Gradient Boosting Machine |         | sklearn | 0.814     | 0.736    |
| <input type="checkbox"/> | ⌄ 23 hours ago  |          | Session Initialized eca6        |         | -       | -         | -        |
| <input type="checkbox"/> | ⌄ 23 hours ago  |          | Session Initialized 7efd        |         | -       | -         | -        |
| <input type="checkbox"/> | ⌄ 3 days ago    |          | Session Initialized 365a        |         | -       | -         | -        |
| <input type="checkbox"/> | ⌄ 3 days ago    |          | Session Initialized 8903        |         | -       | -         | -        |

## Training pipeline:












← ↻ 🏠 ⚠️ Not secure | notebookse.jarvislabs.ai:10977/dags/Lead\_scoring\_training\_pipeline/graph 📄 🗑️ ⚙️ ⌘ ⌕ 🌐


 Airflow DAGs Security ▾ Browse ▾ Admin ▾ Docs ▾ 16:07 UTC 



---

 **DAG: Lead\_scoring\_training\_pipeline** Training pipeline for Lead Scoring System



**success** Schedule: @monthly ⓘ Next Run: 2023-04-01, 00:00:00

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

 Audit Log

 2023-04-16T15:11:34Z Runs 25 ▾ Run manual\_\_2023-04-16T15:11:33.057542+00:00 ▾ Layout Left > Right ▾ Find Task... 

PythonOperator deferred failed queued running scheduled skipped success up\_for\_reschedule up\_for\_retry upstream\_failed no\_status

 Auto-refresh 

```
graph LR; encoding_categorical_variables --> training_model
```

mlflow1.26.1ExperimentsModelsGitHubDocs

Default > test

test

Date: 2023-04-16 20:41:52

Source: airflow

User: root

Duration: 6.9s

Status: FINISHED

Lifecycle Stage: active

DescriptionEdit

Parameters (20)

Metrics (13)

Tags

Artifacts

models

MLmodelconda.yamlmodel.pklpython\_env.yamlrequirements.txt

Full Path:mlruns/0/2e71da59796046b88259acac550f46d1/artifacts/models

LightGBM, v1Registered on 2023/04/16

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](#)

Make Predictions

Predict on a Spark DataFrame:

mlflow1.26.1ExperimentsModelsGitHubDocs

ExperimentsDefault

Search Experiments

Default

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

Experiment ID: 0

DescriptionEdit

RefreshCompareDeleteDownload CSVStart TimeAll time

ColumnsOnly show differences

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

SearchFilterClear

Showing 1 matching run

|  | Start Time | Duration | Run Name | User | Source  | Version | Models     | Metric   |
|--|------------|----------|----------|------|---------|---------|------------|----------|
|  | 3 days ago | 6.9s     | test     | root | airflow | -       | LightGBM/1 | AUC0.747 |

Load more

## Inference pipeline:

Airflow DAG: Lead\_scoring\_inference\_pipeline Inference pipeline of Lead Scoring system

success Schedule: @hourly Next Run: 2023-04-19, 15:00:00

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

Audit Log

2023-04-19T15:38:29Z Runs 25 Run manual\_\_2023-04-19T15:38:28.568028+00:00 Layout Left > Right Find Task...

Update

PythonOperator deferred failed queued running scheduled skipped success up\_for\_reschedule up\_for\_retry upstream\_failed no\_status

Auto-refresh

```
graph LR; encoding_categorical_variables --> checking_input_features; checking_input_features --> generating_models_prediction; generating_models_prediction --> checking_model_prediction_ratio;
```

19-04-2023 15:39:18 25 All Run Types All Run States Clear Filters

Auto-refresh

Duration: 13:00 Apr 17, 14:55

encoding\_categorical\_variables  
checking\_input\_features  
generating\_models\_prediction  
checking\_model\_prediction\_ratio

DAG Lead\_scoring\_inference\_pipeline

DAG Details

DAG Runs Summary

|                      |                          |
|----------------------|--------------------------|
| Total Runs Displayed | 14                       |
| Total success        | 3                        |
| Total failed         | 11                       |
| First Run Start      | 2023-04-17, 14:28:52 UTC |
| Last Run Start       | 2023-04-19, 15:38:29 UTC |
| Max Run Duration     | 00:03:15                 |