

## Screenshot of ML Flow UI

The screenshot shows the MLflow web interface. The top navigation bar includes 'mlflow 1.26.1', 'Experiments', and 'Models'. The breadcrumb trail is 'Registered Models > LightGBM > Version 5'. The main content area displays 'Version 5' with a 'Production' stage. It shows the registration date (2023-04-23 19:28:35) and the last modified date (2023-04-23 20:25:57). The source run is 'Lead\_Scoring\_Training\_Pipeline'. There are links for 'Description' and 'Edit'. Under the 'Schema' section, there is a table with columns 'Name' and 'Type', but it is empty with a message: 'No schema. See [MLflow docs](#) for how to include input and output schema with your model.'

The screenshot shows the MLflow 'Artifacts' page for a specific model. The left sidebar shows a file tree with 'models' containing 'MLmodel', 'conda.yaml', 'model.pkl', 'python\_env.yaml', and 'requirements.txt'. The main content area shows the 'Full Path' and 'LightGBM, v5' model details. Below this, there are two sections: 'Model schema' and 'Make Predictions'. The 'Model schema' section has a table with columns 'Name' and 'Type', but it is empty with a message: 'No schema. See [MLflow docs](#) for how to include input and output schema with your model.' The 'Make Predictions' section shows a code snippet for loading the model and making predictions on a Spark DataFrame.

```
import mlflow
logged_model = 'runs:/5bd33c4640ec48e983c69dc4f2a6cca3/model1s'

# 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()
```

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

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 5	-	Version 5	2023-04-23 20:25:57	-

< 1 > 10 / page

Type here to search 29°C Partly cloudy 19:54 26-04-2023

## Screenshot of Airflow UI:

## Screenshot of Lead Scoring Data Pipeline:

The top screenshot displays the Apache Airflow web interface for the DAG `Lead_Scoring_Data_Engineering_Pipeline`. The interface shows the DAG's graph, which consists of the following tasks: `building_db`, `checking_raw_data_schema`, `load_data_db`, `map_city_tier`, `map_categorical_vars`, `interactions_mapping`, and `model_input_schema_check`. The DAG is currently in a `success` state, with a schedule of `@daily` and a next run time of `2023-04-22, 00:00:00`. The interface also includes a search bar, a filter dropdown, and a list of task instances.

The bottom screenshot shows the same DAG in a task instance view. The interface displays a bar chart of task durations and a summary table of run statistics. The summary table is as follows:

DAG Runs Summary	
Total Runs Displayed	4
Total success	1
Total failed	3
First Run Start	2023-04-22, 14:41:46 UTC
Last Run Start	2023-04-22, 15:15:58 UTC
Max Run Duration	00:01:31

## Screenshot of Lead Scoring training Pipeline:

The first screenshot shows the Airflow DAG view for 'Lead\_scoring\_training\_pipeline/graph'. It displays a simple DAG with two tasks: 'encode\_cat\_features' and 'training\_model'. The interface includes a top navigation bar with 'Airflow', 'DAGs', 'Security', 'Browse', 'Admin', and 'Docs'. Below the navigation bar are tabs for 'Grid', 'Graph', 'Calendar', 'Task Duration', 'Task Tries', 'Landing Times', 'Gantt', and 'Details'. A filter bar at the top shows a date range from '2023-04-23T13:58:20Z' to '2023-04-23T13:58:19.831161+00:00' with 25 runs. A legend below the filter bar shows various task states: deferred, failed, queued, running, scheduled, skipped, success, up\_for\_reschedule, up\_for\_retry, upstream\_failed, and no\_status. The DAG graph shows 'encode\_cat\_features' as a parent task and 'training\_model' as a child task.

The second screenshot shows the Airflow DAG view for 'Lead\_scoring\_training\_pipeline/grid?root='. It displays a more detailed view of the DAG, including a 'DAG Details' section on the right. The 'DAG Details' section includes a 'DAG Runs Summary' table with the following data:

DAG Runs Summary	
Total Runs Displayed	13
Total success	1
Total failed	12
First Run Start	2023-04-23, 12:59:55 UTC

The 'DAG Details' section also includes a 'DAG Runs Summary' table with the following data:

DAG Runs Summary	
Total Runs Displayed	13
Total success	1
Total failed	12
First Run Start	2023-04-23, 12:59:55 UTC

The 'DAG Details' section also includes a 'DAG Runs Summary' table with the following data:

DAG Runs Summary	
Total Runs Displayed	13
Total success	1
Total failed	12
First Run Start	2023-04-23, 12:59:55 UTC

## Screenshot of Lead Scoring Inference Pipeline:

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

status: success Schedule: @hourly Next Run: 2023-04-23, 16:00:00

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

2023-04-23T16:18:11Z Runs: 25 Run: manual\_2023-04-23T16:18:10.113702+00:00 Layout: Left > Right Update Find Task...

PythonOperator

Auto-refresh

```
graph LR; encoding_cat_var --> get_models_prediction; get_models_prediction --> prediction_ratio_check; prediction_ratio_check --> checking_input_features;
```

MLflow Grid: Lead\_scoring\_inference\_pipeline Inference pipeline of Lead Scoring system

Triggered Lead\_scoring\_inference\_pipeline, it should start any moment now.

status: success Schedule: @hourly Next Run: 2023-04-23, 16:00:00

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

23-04-2023 16:18:12 25 All Run Types All Run States Clear Filters

Auto-refresh

Task	Duration	Status
encoding_cat_var	00:00:25	Failed
get_models_prediction	00:00:12	Failed
prediction_ratio_check	00:00:00	Failed
checking_input_features	00:00:00	Failed

DAG Details

Lead\_scoring\_inference\_pipeline

DAG Runs Summary	
Total Runs Displayed	5
Total success	1
Total failed	4
First Run Start	2023-04-23, 16:09:51 UTC
Last Run Start	2023-04-23, 16:18:11 UTC
Max Run Duration	00:00:25
Mean Run Duration	00:00:23

Initially lot of errors were observed while running all the respective pipelines shown in red above. Accordingly as per log, code in scripts was debugged and Dags were re-run and finally for all the pipelines were implemented.