

# Screenshot of MLFlow UI

## MLFlow : All Experiments

The screenshot shows the MLFlow UI interface. At the top, there is a navigation bar with the MLflow logo, version 2.20.2, and links for Experiments and Models. On the right side of the header are GitHub and Docs links. Below the header, the main content area is titled "Experiments" and shows an experiment named "Lead\_scoring\_model\_experimentation\_01". The experiment has a status of "Active". The interface includes a search bar, filters for Runs, Evaluation, and Experimental, and a sorting section. A table below lists various runs, each with a unique ID, creation time (e.g., "2 hours ago"), duration, source (ipykernel...), and models. A "Show more columns" link is visible on the right.

Run Name	Created	Dataset	Duration	Source	Models
Session initialized 9...	2 hours ago	-	1.1min	ipykern...	-
Session initialized 5...	2 hours ago	-	56.2s	ipykern...	-
Session initialized a...	2 hours ago	-	53.3s	ipykern...	-
Session initialized 5...	2 hours ago	-	53.5s	ipykern...	-
Session initialized 0...	2 hours ago	-	55.7s	ipykern...	-
Session initialized 2...	2 hours ago	-	54.4s	ipykern...	-
Session initialized 3...	2 hours ago	-	55.5s	ipykern...	-
Session initialized 9...	2 hours ago	-	53.9s	ipykern...	-
Session initialized 9...	2 hours ago	-	56.1s	ipykern...	-
Session initialized f...	2 hours ago	-	1.9min	ipykern...	-
Session initialized 8...	2 hours ago	-	3.3min	ipykern...	-
Session initialized 0...	2 hours ago	-	3.4min	ipykern...	-

## MLFlow: Experiment with all the artifacts visible

The screenshot shows the details of a specific experiment run within the "Lead\_scoring\_model\_experimentation\_01" experiment. The run is identified by the ID "fae5". The "Overview" tab is selected, showing basic information such as creation date (2025-02-24 16:50:29), creator (rpandey1), experiment ID (1), status (Finished), run ID (ef2563e4dfe94880909257f74486c75d), duration (1.9min), and source (ipykernel\_launcher.py). Below this, sections for Parameters (28) and Metrics (0) are shown, each with a search bar and a table for parameters and values.

**mlflow** 2.20.2 Experiments Models

Lead\_scoring\_model\_experimentation\_01 >

### Session Initialized fae5

Overview Model metrics System metrics Traces Artifacts

Parameters (28)

Parameter	Value
Session id	42
Target	app_complete_flag
Target type	Binary
Original data shape	(238964, 12)
Transformed data shape	(238964, 44)
Transformed train set shape	(167274, 44)
Transformed test set shape	(71690, 44)
Numeric features	8
Categorical features	3
Preprocess	True
Imputation type	simple
Numeric imputation	mean
Categorical imputation	mode
Maximum one-hot encoding	25
Encoding method	None
Remove multicollinearity	True
Multicollinearity threshold	0.95

Metrics (0)

No metrics recorded

**mlflow** 2.20.2 Experiments Models

Lead\_scoring\_model\_experimentation\_01 >

### Session Initialized fae5

Overview Model metrics System metrics Traces Artifacts

Transformation Pipeline.pkl

Select a file to preview  
Supported formats: image, text, html, pdf, audio, geojson files

# Screenshot of MLFlow UI after dropping features

## MLFlow : All Experiments

The screenshot shows the MLFlow UI interface for managing experiments. At the top, there's a navigation bar with the MLflow logo, version 2.20.2, and links for 'Experiments' and 'Models'. On the right side of the header are icons for GitHub and Docs.

The main content area is titled 'Lead\_scoring\_model\_experiment\_02'. It includes tabs for 'Runs', 'Evaluation', and 'Experimental' (which is currently selected). Below the tabs is a search bar with the query 'metrics.rmse < 1 and params.model = "tree"'. There are also dropdown menus for 'Time created', 'State: Active', 'Datasets', and 'Sort: Created'.

The main view is a table with the following columns: Run Name, Created, Dataset, Duration, Source, and Models. The table lists 91 matching runs. Each row contains a checkbox, a run name (e.g., 'Session Initialized ...'), a timestamp ('1 day ago'), a dataset entry ('-'), a duration ('6.9min'), a source ('ipykern...'), and a model type ('-'). Some rows have colored circular icons next to them, such as green for Light Gradient ... and purple for Naive Bayes.

At the bottom left of the table, it says '91 matching runs'. On the right side of the table, there's a link 'Show more columns (93 total)'.

## MLFlow : Experiment with all the artifacts visible

mlflow 2.20.2 Experiments Models

Lead\_scoring\_model\_experiment\_02 >

### Session Initialized bb5c

Overview Model metrics System metrics Traces Artifacts

Description

No description

Details

Created at	2025-02-24 16:53:25
Created by	rpandey1
Experiment ID	2
Status	Finished
Run ID	a46067a233e3459daefeadef404aa251
Duration	4.8min
Datasets used	—
Tags	Source: setup URI: 38e480bb USI: bb5c Run Time: 1.75 Run ID: a46067a233e3459daefeadef404aa251
Source	
Logged models	—
Registered models	—

Parameters (24)

<input type="text"/> Search parameters	
Parameter	Value
Session id	42
Target	app_complete_flag
Target type	Binary
Original data shape	(238964, 7)
Transformed data shape	(238964, 41)
Transformed train set shape	(167274, 41)
Transformed test set shape	(71690, 41)
Numeric features	2
Categorical features	4
Preprocess	True
Imputation type	simple
Numeric imputation	mean
Categorical imputation	mode
Maximum one-hot encoding	25
Encoding method	None
Remove multicollinearity	True
Multicollinearity threshold	0.95

Metrics (0)

mlflow 2.20.2 Experiments Models

Lead\_scoring\_model\_experiment\_02 >

### Session Initialized bb5c

Overview Model metrics System metrics Traces Artifacts

Parameters (24)

<input type="text"/> Search parameters	
Parameter	Value
Session id	42
Target	app_complete_flag
Target type	Binary
Original data shape	(238964, 7)
Transformed data shape	(238964, 41)
Transformed train set shape	(167274, 41)
Transformed test set shape	(71690, 41)
Numeric features	2
Categorical features	4
Preprocess	True
Imputation type	simple
Numeric imputation	mean
Categorical imputation	mode
Maximum one-hot encoding	25
Encoding method	None
Remove multicollinearity	True
Multicollinearity threshold	0.95

Metrics (0)

No metrics recorded

The screenshot shows the MLFlow UI interface. At the top, there's a dark header bar with the 'mlflow' logo, version '2.20.2', and navigation links for 'Experiments' and 'Models'. On the right side of the header are icons for GitHub and Docs, along with a gear icon for settings. Below the header, the page title is 'Lead\_scoring\_model\_experiment\_02 > Session Initialized bb5c'. A horizontal navigation bar below the title includes 'Overview', 'Model metrics', 'System metrics', 'Traces', and 'Artifacts', with 'Artifacts' being the active tab. Under the 'Artifacts' tab, there's a single file listed: 'Transformation Pipeline.pkl'. To the right of the file list is a placeholder area with a document icon and the text 'Select a file to preview'. Below this placeholder, it says 'Supported formats: image, text, html, pdf, audio, geojson files'.

## MLFlow: Model registry with model name and stage as 'production'

The screenshot shows the MLFlow Model Registry UI. At the top, there's a dark header bar with the 'mlflow' logo, version '2.20.2', and navigation links for 'Experiments' and 'Models'. On the right side of the header are icons for GitHub and Docs, along with a gear icon for settings. Below the header, the page title is 'Registered Models > LightGBM'. It shows a single registered model named 'LightGBM'. Below the model name, it displays 'Created Time: 2025-02-24 19:52:45' and 'Last Modified: 2025-02-24 23:17:12'. There are three expandable sections: 'Description', 'Tags', and 'Versions'. The 'Versions' section is expanded, showing a table of registered versions. The table has columns: 'Version', 'Registered at', 'Created by', 'Stage', and 'Description'. The data in the table is as follows:

Version	Registered at	Created by	Stage	Description
Version 4	2025-02-24 23:06:40		Production	
Version 3	2025-02-24 22:48:45		Staging	
Version 2	2025-02-24 21:37:03		None	
Version 1	2025-02-24 19:52:45		None	

At the bottom right of the table, there's a link 'New model registry UI' with a circular arrow icon. A small number '1' is also visible in the bottom right corner of the table area.

## MLFlow: Lead\_scoring\_mlflow\_production

mlflow 2.20.2 Experiments Models

Lead\_scoring\_mlflow\_production > Lead\_scoring\_mlflow\_production

Model registered

Overview Model metrics System metrics Traces Artifacts

Description 

No description

Details

Created at	2025-02-24 19:52:42
Created by	rpandey1
Experiment ID	3 
Status	 Finished
Run ID	22005a59ab4f4f4586bbd4fb1418e4ea7 
Duration	2.9s
Datasets used	—
Tags	Add
Source	 airflow
Logged models	 sklearn
Registered models	 LightGBM v1

Parameters (20)

Q Search parameters	
Parameter	Value
boosting_type	gbdt

Metrics (2)

Q Search metrics	
Metric	Value
test_accuracy	0.739656855907379

mlflow 2.20.2 Experiments Models

Lead\_scoring\_mlflow\_production > Lead\_scoring\_mlflow\_production

Model registered

Overview Model metrics System metrics Traces Artifacts

Parameters (20)

Q Search parameters	
Parameter	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

Metrics (2)

Q Search metrics	
Metric	Value
test_accuracy	0.739656855907379
test_auc	0.7471604303019428

**mlflow** 2.20.2 Experiments Models

Lead\_scoring\_mlflow\_production > Model registered

## Lead\_scoring\_mlflow\_production

Overview Model metrics System metrics Traces Artifacts

**Artifacts**

Path: JUsers/rpandey1/Desktop/Upgrad/Assignment/mlruns/3/22005a59ab4f4586bbd4fbf1418e4ea7/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
No schema. See <a href="#">MLflow docs</a> for how to include input and output schema with your model.	

**Validate the model before deployment**

Run the following code to validate model inference works on the example input data and logged model dependencies, prior to deploying it to a serving endpoint

```
import mlflow

model_uri = 'runs:/22005a59ab4f4586bbd4fbf1418e4ea7/models'

# Replace INPUT_EXAMPLE with your own input example to the model
# A valid input example is a data instance suitable for pyfunc prediction
input_data = INPUT_EXAMPLE

# Verify the model with the provided input data using the logged dependencies.
# For more details, refer to:
# https://mlflow.org/docs/latest/models.html#validate-models-before-deployment
mlflow.models.predict(
    model_uri=model_uri,
    input_data=input_data,
    env_manager="uv",
)
```

**Make Predictions**

Predict on a Pandas DataFrame:



## Screenshot of AirFlow UI

## AirFlow: Airflow UI grid( All 3 Dags)



A screenshot of the Airflow web interface. The top navigation bar includes links for 'DAGs', 'Cluster Activity', 'Datasets', 'Security', 'Browse', 'Admin', and 'Docs'. The top right shows the time as '14:12 UTC' and a 'Logout' button. A yellow banner at the top states: '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.' Another yellow banner below it says: 'Do not use the SequentialExecutor in production. [Click here](#) for more information.' The main section is titled 'DAGs' and contains a table listing various DAGs. The columns include: 'Owner' (all listed as 'airflow'), 'Runs' (represented by a series of circles), 'Schedule' (e.g., '@daily', '@hourly', '@monthly'), 'Last Run' (e.g., '2025-02-23, 00:00:00'), 'Next Run' (e.g., '2025-02-24, 13:00:00'), 'Recent Tasks' (represented by a series of circles), and 'Actions' (with icons for play, stop, and more). Some DAG names have status labels like 'consumes dataset-scheduled' or 'produces dataset-scheduled'.

## AirFlow: Dags Execution

The screenshot shows the Airflow web interface under the 'DAGs' tab. At the top, there are three filter buttons: 'All 54' (selected), 'Active 3', and 'Paused 61'. Below these are buttons for 'Running 1' and 'Failed 0'. There are also 'Filter DAGs by tag' and 'Search DAGs' input fields, along with an 'Auto-refresh' button and a refresh icon.

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
Lead_Scoring_Data_Engineering_Pipeline	airflow	2 (green)	@daily	2025-02-25, 00:00:00	2025-02-26, 00:00:00	7 (green)		
Lead_scoring_inference_pipeline	airflow	2 (green)	@hourly	2025-02-26, 07:00:00	2025-02-26, 08:00:00	3 (green)		
Lead_scoring_training_pipeline	airflow	1 (green)	@monthly	2025-02-24, 17:18:32	2025-02-01, 00:00:00	8 (green)		
dataset_consumes_1	airflow	0 (white)	Dataset			On s3://dag1/output_1.txt		
dataset_consumes_1_and_2	airflow	0 (white)	Dataset			0 of 2 datasets updated		
dataset_consumes_1_never_scheduled	airflow	0 (white)	Dataset			0 of 2 datasets updated		
dataset_consumes_unknown_never_scheduled	airflow	0 (white)	Dataset			0 of 2 datasets updated		
dataset_produces_1	airflow	0 (white)	@daily		2025-02-23, 00:00:00			
dataset_produces_2	airflow	0 (white)	None					

## AirFlow: Successful execution Airflow DAG (Lead\_Scoring\_Data\_Engineering\_Pipeline)

The screenshot shows the Airflow web interface under the 'Details' tab for the 'Lead\_Scoring\_Data\_Engineering\_Pipeline' DAG. At the top, there are tabs for 'Grid', 'Graph', 'Calendar', 'Task Duration', 'Task Tries', 'Landing Times', 'Gantt', 'Details', 'Code', and 'Audit Log'. The 'Gantt' tab is selected. A status bar at the top right shows 'Schedule: @daily' and 'Next Run: 2025-02-24, 00:00:00'. Below the tabs are filters for 'All Run Types' (25), 'All Run States', and 'Clear Filters', along with an 'Auto-refresh' button.

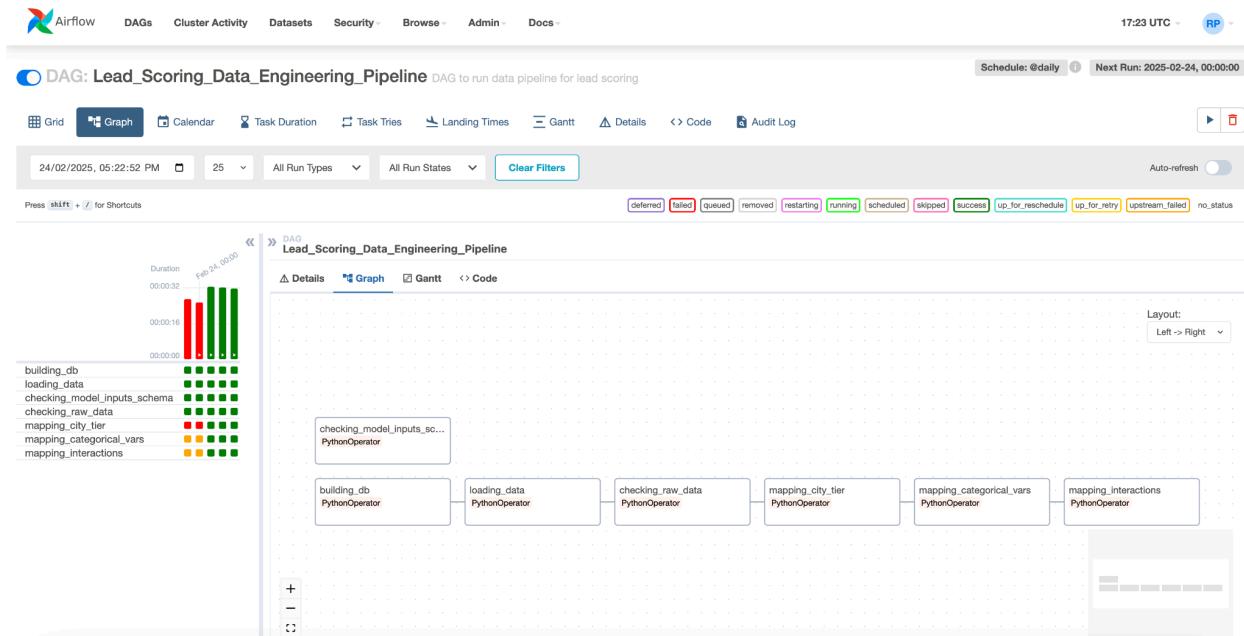
The main area displays a Gantt chart for the DAG. The tasks listed on the left are: building\_db, loading\_data, checking\_model\_inputs\_schema, checking\_raw\_data, mapping\_city\_tier, mapping\_categorical\_vars, and mapping\_interactions. The chart shows the duration of each task and their sequence. To the right of the chart is a 'DAG Runs Summary' section with the following data:

- Total Runs Displayed: 5
- Total success: 3
- Total failed: 2
- First Run Start: 2025-02-24, 14:13:44 UTC
- Last Run Start: 2025-02-24, 17:17:34 UTC
- Max Run Duration: 00:00:32
- Mean Run Duration: 00:00:29
- Min Run Duration: 00:00:25

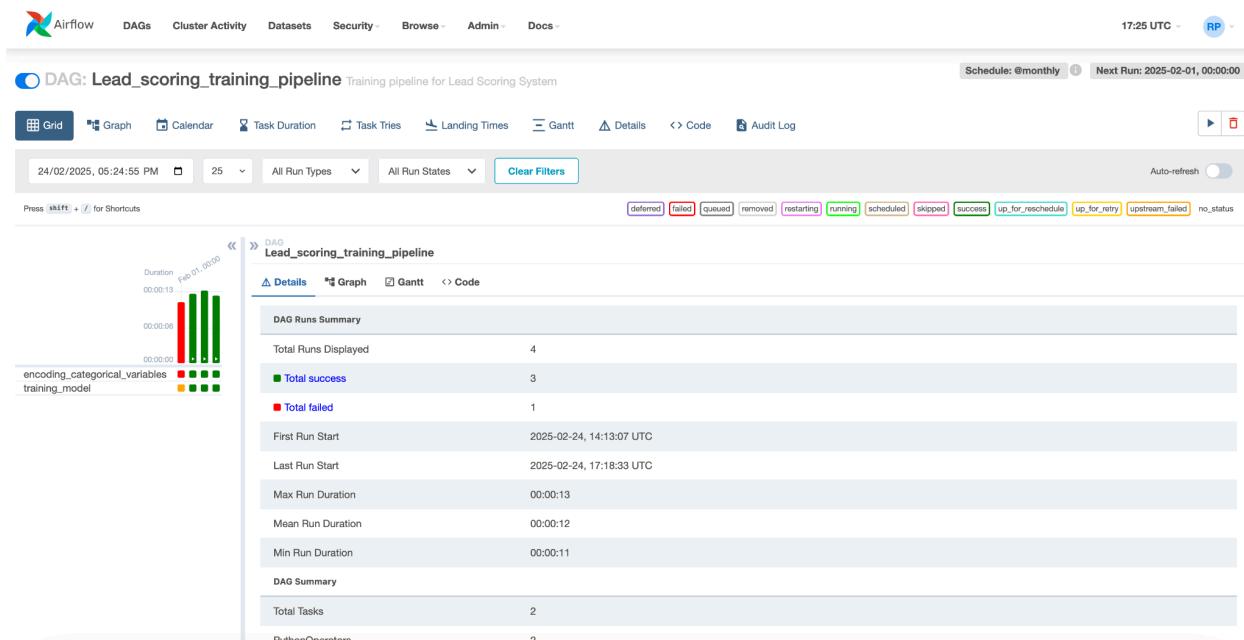
Below this is a 'DAG Summary' section with the following data:

- Total Tasks: 7
- PythonOperators: 7

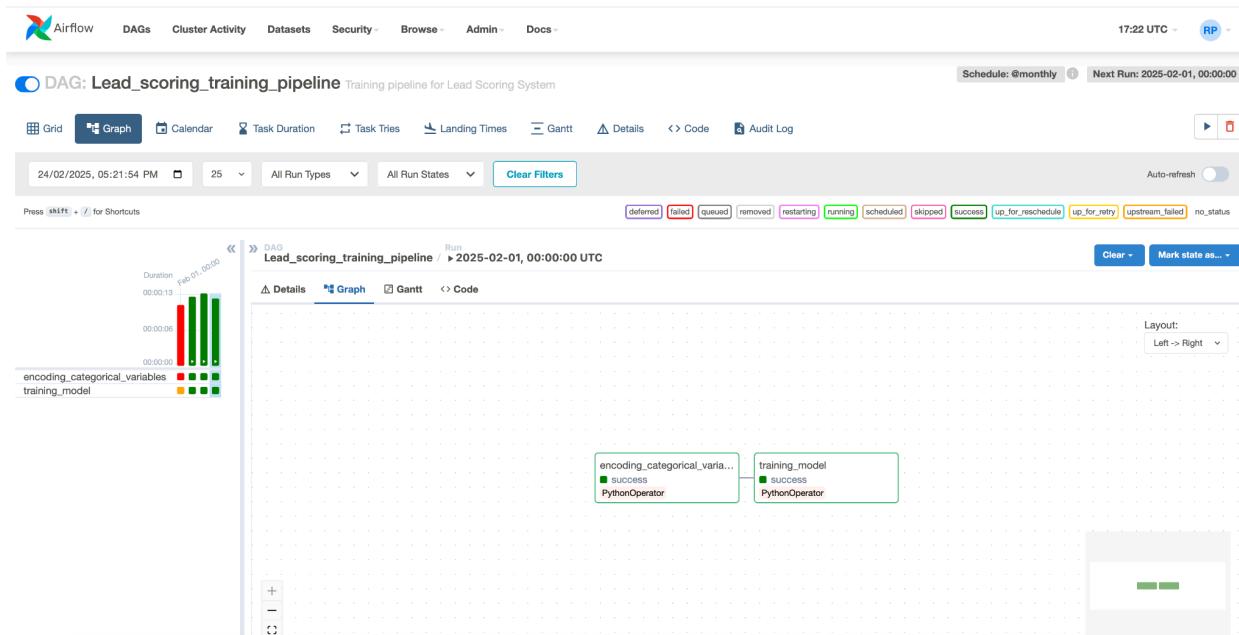
## AirFlow: Successful execution Airflow DAG in graph (Lead\_Scoring\_Data\_Engineering\_Pipeline)



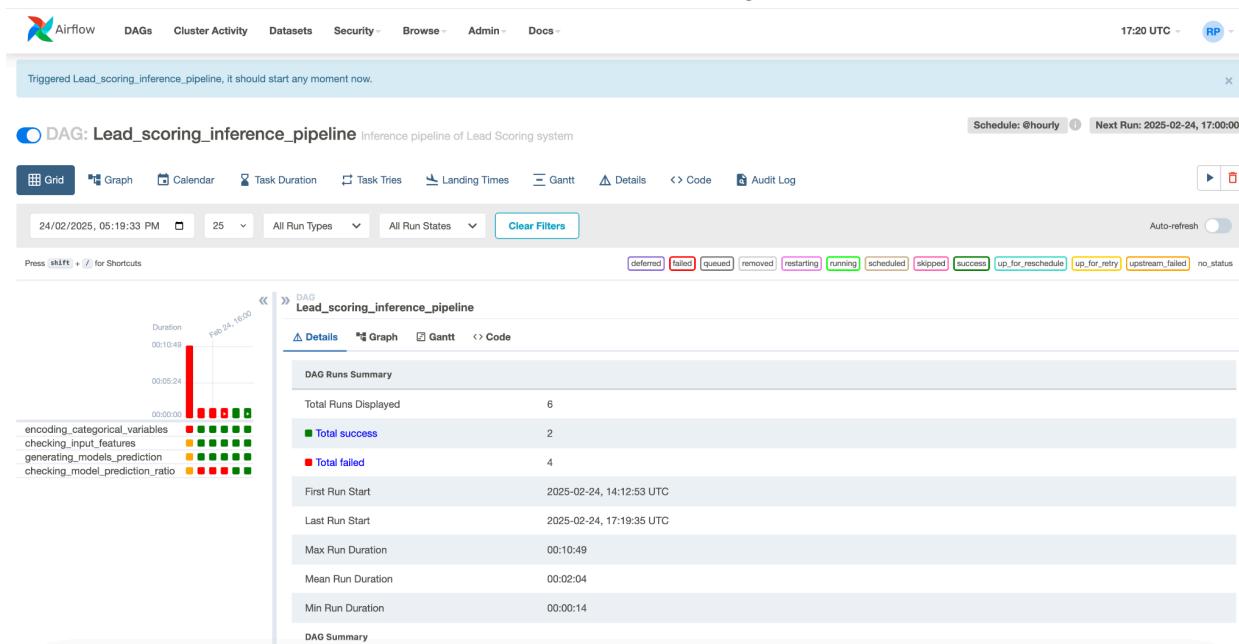
## AirFlow: Successful execution Airflow DAG (Lead\_Scoring\_Training\_Pipeline)



## AirFlow: Successful execution Airflow DAG in graph (Lead\_Scoring\_Training\_Pipeline)



## AirFlow: Successful execution Airflow DAG (Lead\_Scoring\_Inference\_Pipeline)



## AirFlow: Successful execution Airflow DAG in graph (Lead\_Scoring\_Inference\_Pipeline)

