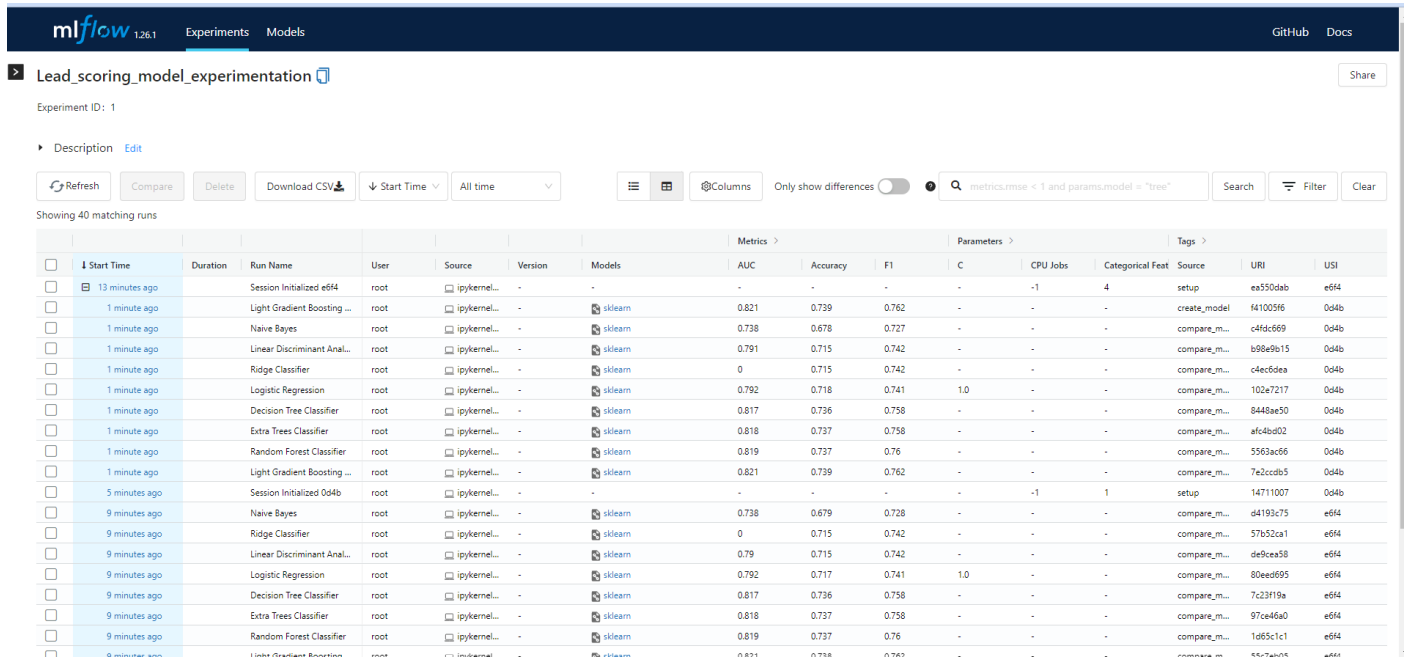


SCREENSHOTS OF ML Ops ASSIGNMENT :

Model Experimentation

Before dropping features:

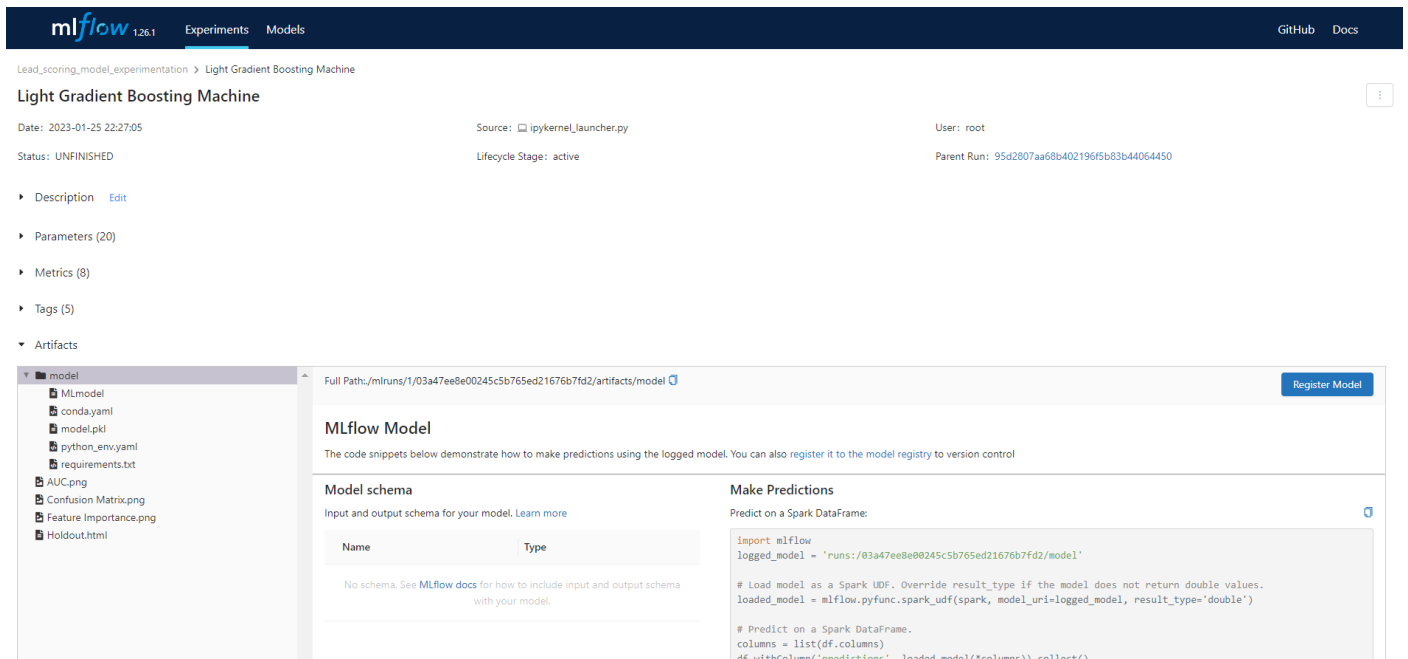
- screenshot of all the experiments



The screenshot shows the mlflow Experiments page for an experiment named 'Lead_scoring_model_experimentation'. The table lists 20 runs, each with a checkbox, a timestamp (e.g., '13 minutes ago'), a duration, a run name, a user, a source, a version, a list of models, and various metrics (AUC, Accuracy, F1, C, CPU Jobs, Categorical Feat). The runs are sorted by start time. The table also includes columns for Source, URI, and USI.

	Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Source	URI	USI
<input type="checkbox"/>	13 minutes ago		Session Initialized e6f4	root	ipykernel...	-	-	-	-	-	-	-1	4	setup	ea5504ab	e6f4
<input type="checkbox"/>	1 minute ago		Light Gradient Boosting ...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	create_model	641005f6	0d4b
<input type="checkbox"/>	1 minute ago		Naive Bayes	root	ipykernel...	-	sklearn	0.738	0.678	0.727	-	-	-	compare_m...	c4fd669	0d4b
<input type="checkbox"/>	1 minute ago		Linear Discriminant Anal...	root	ipykernel...	-	sklearn	0.791	0.715	0.742	-	-	-	compare_m...	b98e9b15	0d4b
<input type="checkbox"/>	1 minute ago		Ridge Classifier	root	ipykernel...	-	sklearn	0	0.715	0.742	-	-	-	compare_m...	c4ec6dea	0d4b
<input type="checkbox"/>	1 minute ago		Logistic Regression	root	ipykernel...	-	sklearn	0.792	0.718	0.741	1.0	-	-	compare_m...	102e7217	0d4b
<input type="checkbox"/>	1 minute ago		Decision Tree Classifier	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	8448ae50	0d4b
<input type="checkbox"/>	1 minute ago		Extra Trees Classifier	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	afcd0d02	0d4b
<input type="checkbox"/>	1 minute ago		Random Forest Classifier	root	ipykernel...	-	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	5563ac66	0d4b
<input type="checkbox"/>	1 minute ago		Light Gradient Boosting ...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	compare_m...	7e2ccdb5	0d4b
<input type="checkbox"/>	5 minutes ago		Session Initialized 0d4b	root	ipykernel...	-	-	-	-	-	-	-1	1	setup	14711007	0d4b
<input type="checkbox"/>	9 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.738	0.679	0.728	-	-	-	compare_m...	d4193c75	e6f4
<input type="checkbox"/>	9 minutes ago		Ridge Classifier	root	ipykernel...	-	sklearn	0	0.715	0.742	-	-	-	compare_m...	57b52ca1	e6f4
<input type="checkbox"/>	9 minutes ago		Linear Discriminant Anal...	root	ipykernel...	-	sklearn	0.79	0.715	0.742	-	-	-	compare_m...	de9cea58	e6f4
<input type="checkbox"/>	9 minutes ago		Logistic Regression	root	ipykernel...	-	sklearn	0.792	0.717	0.741	1.0	-	-	compare_m...	80eed695	e6f4
<input type="checkbox"/>	9 minutes ago		Decision Tree Classifier	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...	7c23f19a	e6f4
<input type="checkbox"/>	9 minutes ago		Extra Trees Classifier	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	-	compare_m...	97ce46a0	e6f4
<input type="checkbox"/>	9 minutes ago		Random Forest Classifier	root	ipykernel...	-	sklearn	0.819	0.737	0.76	-	-	-	compare_m...	1d65c1c1	e6f4
<input type="checkbox"/>	9 minutes ago		Light Gradient Boosting ...	root	ipykernel...	-	sklearn	0.821	0.738	0.762	-	-	-	compare_m...	5574b095	e6f4

- screenshot of one experiment with all the artifacts visible



The screenshot shows the mlflow Model page for a model named 'Light Gradient Boosting Machine'. The page displays the model's metadata, including its date (2023-01-25 22:27:05), status (UNFINISHED), source (ipykernel_launcher.py), user (root), lifecycle stage (active), and parent run (95d2807aa68b402196f5b83b44064450). The page also shows the model's description, parameters (20), metrics (8), tags (5), and artifacts (5). The artifacts section lists various files, including MLmodel, conda.yaml, model.pkl, python_env.yaml, requirements.txt, AUC.png, Confusion Matrix.png, Feature Importance.png, and Holdout.html. The page also includes a section for the MLflow Model schema, which shows the input and output schema for the model. The schema is defined by the following code snippet:

```
import mlflow
logged_model = 'runs:/03a47ee8e00245c5b765ed21676b7fd2/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()
```

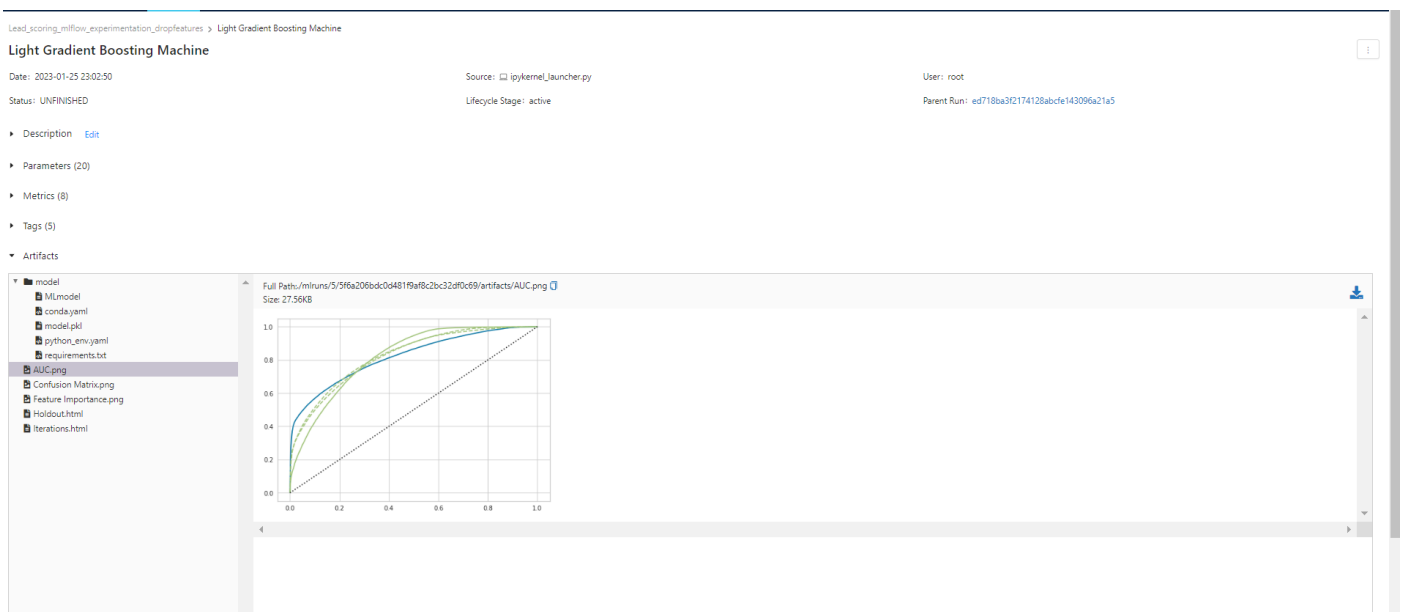
After dropping features:

- screenshot of all the experiments

The screenshot shows the mlflow Experiments page for the experiment 'Lead_scoring_mlflow_experimentation_dropfeatures'. The interface includes a search bar, a list of experiments on the left, and a table of runs. The table columns are: Start Time, Duration, Run Name, User, Source, Version, Models, Metrics (AUC, Accuracy, F1), Parameters (C, CPU Jobs, Categorical Feat), and Tags. The runs are sorted by start time, with the most recent run at the top.

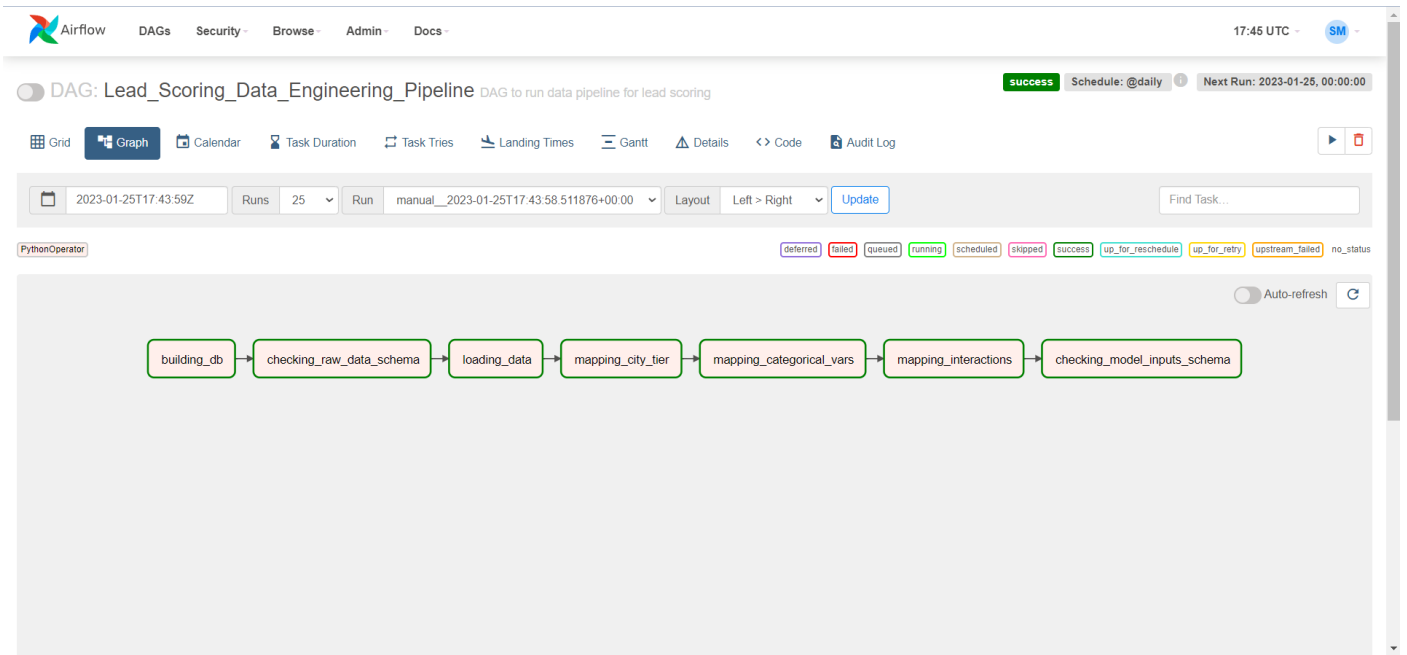
Start Time	Duration	Run Name	User	Source	Version	Models	AUC	Accuracy	F1	C	CPU Jobs	Categorical Feat	Tags
22 minutes ago		Session Init...	root	ipykernel...	-	-	-	-	-	-	-1	1	setup
1 minute ago	3.2s	run_LightGB...	root	ipykernel...	-	LightGBM/5	-	-	-	-	-	-	-
1 minute ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.738	0.76	-	-	-	tune_model
9 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	create_model
10 minutes ago		Naive Bayes	root	ipykernel...	-	sklearn	0.734	0.674	0.725	-	-	-	compare_m...
10 minutes ago		Linear Discr...	root	ipykernel...	-	sklearn	0.773	0.7	0.727	-	-	-	compare_m...
10 minutes ago		Ridge Classif...	root	ipykernel...	-	sklearn	0	0.7	0.727	-	-	-	compare_m...
10 minutes ago		Logistic Reg...	root	ipykernel...	-	sklearn	0.784	0.71	0.74	1.0	-	-	compare_m...
10 minutes ago		Decision Tre...	root	ipykernel...	-	sklearn	0.817	0.736	0.758	-	-	-	compare_m...
10 minutes ago		Extra Trees C...	root	ipykernel...	-	sklearn	0.818	0.737	0.758	-	-	-	compare_m...
10 minutes ago		Random For...	root	ipykernel...	-	sklearn	0.819	0.737	0.759	-	-	-	compare_m...
10 minutes ago		Light Gradie...	root	ipykernel...	-	sklearn	0.821	0.739	0.762	-	-	-	compare_m...
10 minutes ago		Extreme Gra...	root	ipykernel...	-	-	-	-	-	-	-	-	-

- screenshot of one experiment with all the artifacts visible

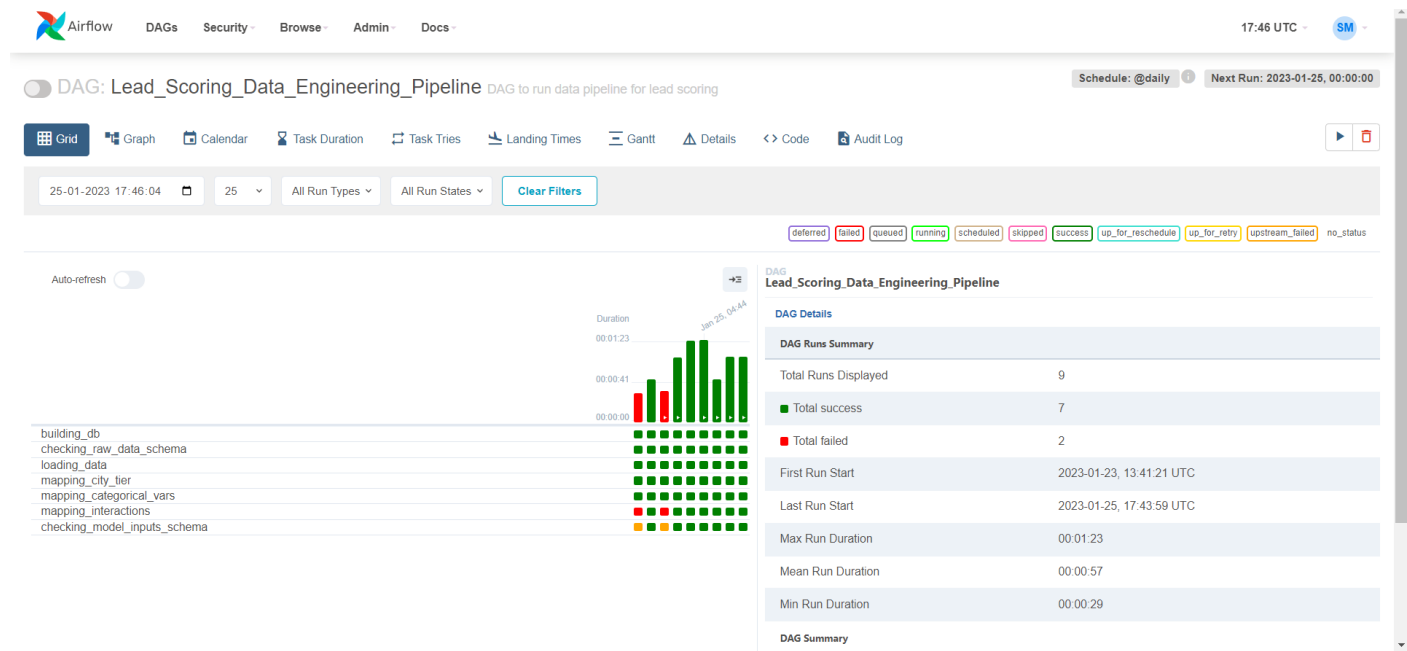


Data Pipeline

- Screenshot of successful execution Airflow DAG in graph



- Screenshot of Airflow UI grid



Training Pipeline

ML Flow setup:

- screenshot of experiments with all the artifacts visible

The screenshot shows the MLflow Experiments page. The breadcrumb navigation is 'Lead_scoring_mlflow_production > Lead_scoring_mlflow_production_25_01_2023_00_00_00'. The experiment name is 'Lead_scoring_mlflow_production_25_01_2023_00_00_00'. The date is '2023-01-25 23:17:32', the source is 'airflow', the user is 'root', the status is 'FINISHED', and the lifecycle stage is 'active'. The duration is '5.1s'. On the left, there are tabs for 'Description', 'Parameters (26)', 'Metrics (9)', 'Tags', and 'Artifacts'. The 'Artifacts' tab is selected, showing a tree view of artifacts under 'models'. The main content area displays the 'MLflow Model' details, including the full path, a description, the model schema (input and output), and code snippets for making predictions on a Spark DataFrame and a Pandas DataFrame.

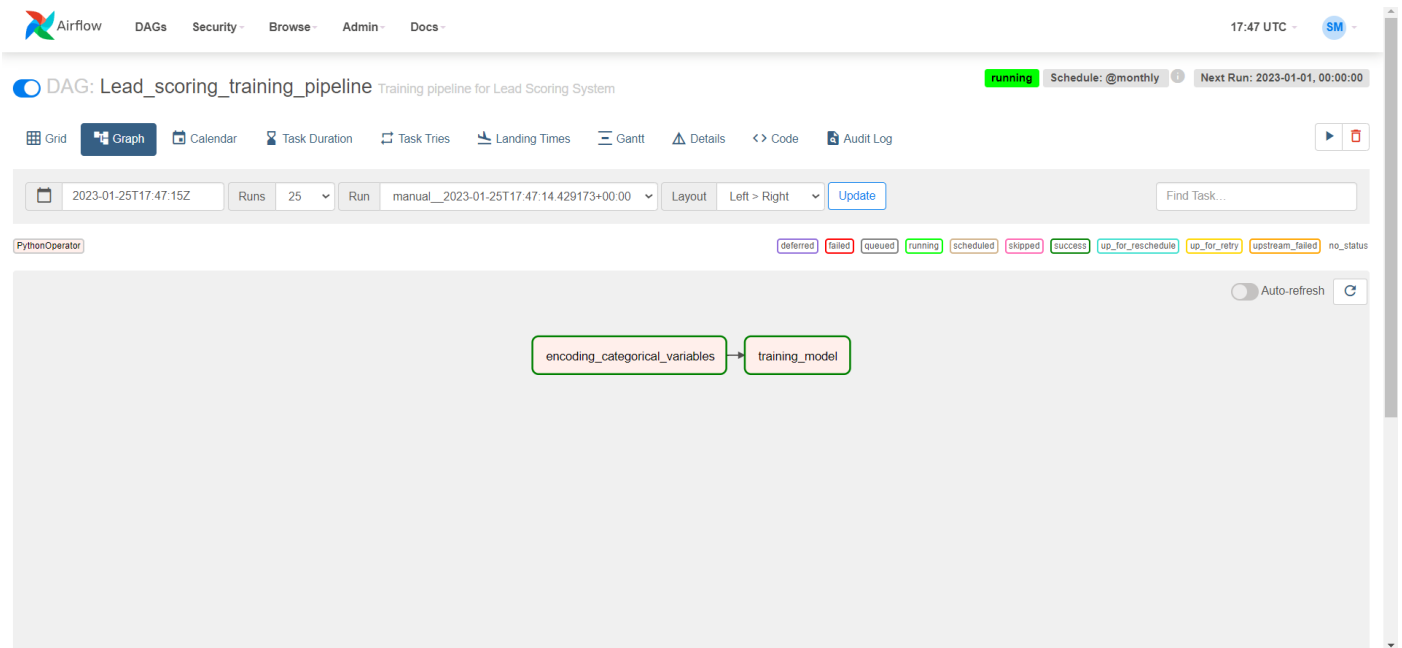
- screenshot of model registry with model name and stage as 'production'

The screenshot shows the MLflow Models page. The breadcrumb navigation is 'Registered Models > LightGBM'. The model name is 'LightGBM'. The created time is '2023-01-24 23:44:13' and the last modified time is '2023-01-25 23:19:42'. On the left, there are tabs for 'Description', 'Tags', and 'Versions'. The 'Versions' tab is selected, showing a table of model versions. The table has columns for 'Version', 'Registered at', 'Created by', 'Stage', and 'Description'. The 'Stage' column shows 'Production' for Version 7 and 'None' for all other versions.

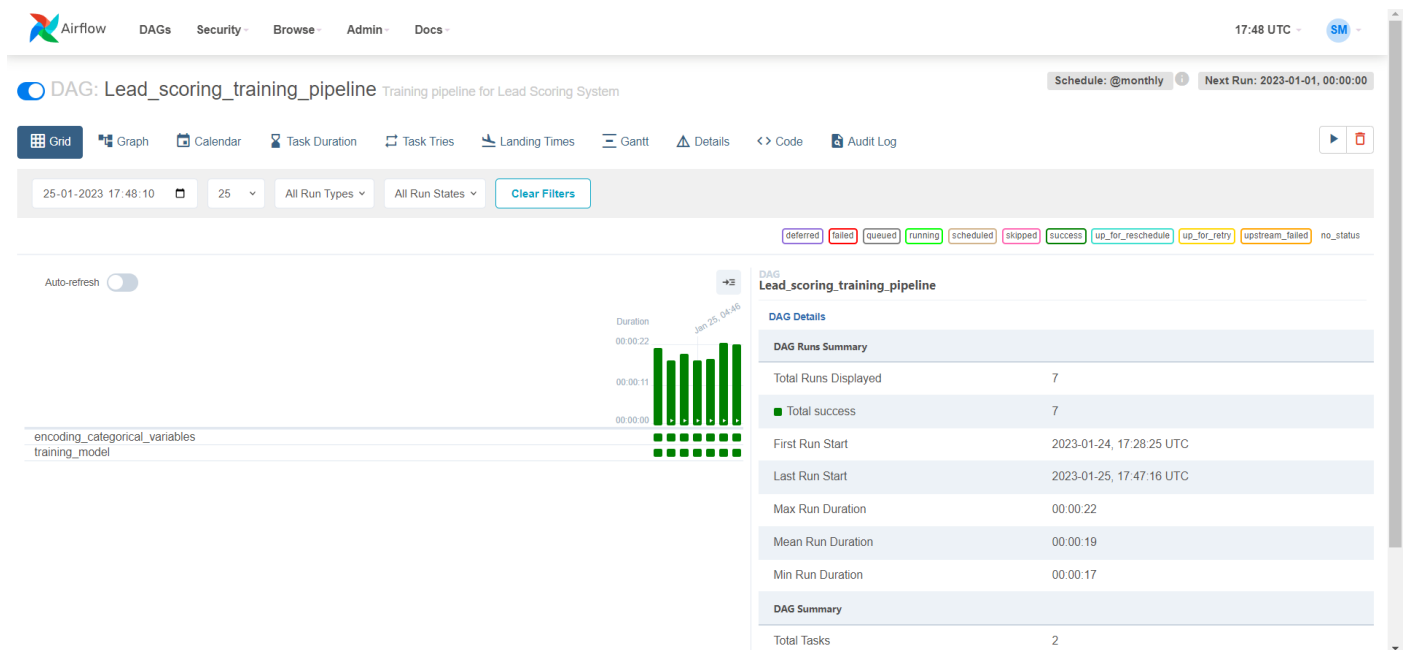
Version	Registered at	Created by	Stage	Description
Version 7	2023-01-25 23:17:36		Production	
Version 6	2023-01-25 21:01:15		None	
Version 5	2023-01-25 10:30:14		None	
Version 4	2023-01-25 10:16:40		None	
Version 3	2023-01-24 23:53:21		None	
Version 2	2023-01-24 23:48:21		None	
Version 1	2023-01-24 23:44:13		None	

Air Flow setup:

- Screenshot of successful execution Airflow DAG in graph

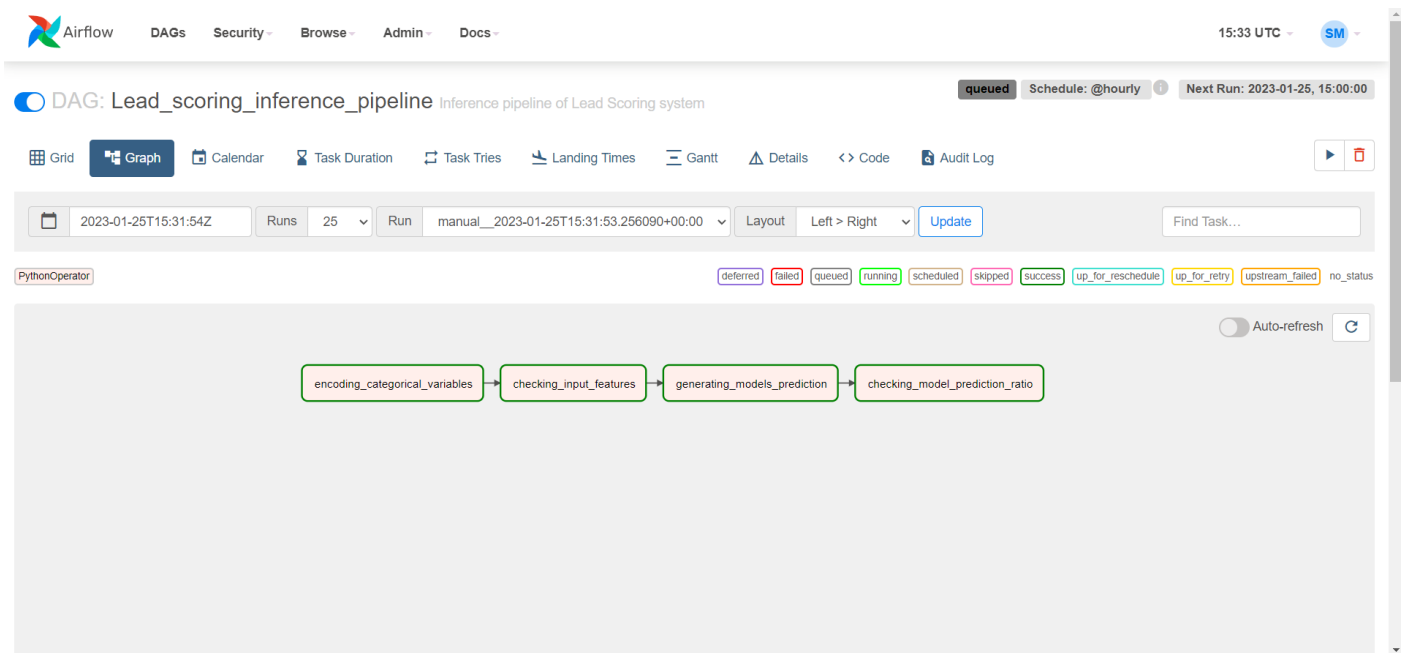


- Screenshot of Airflow UI grid

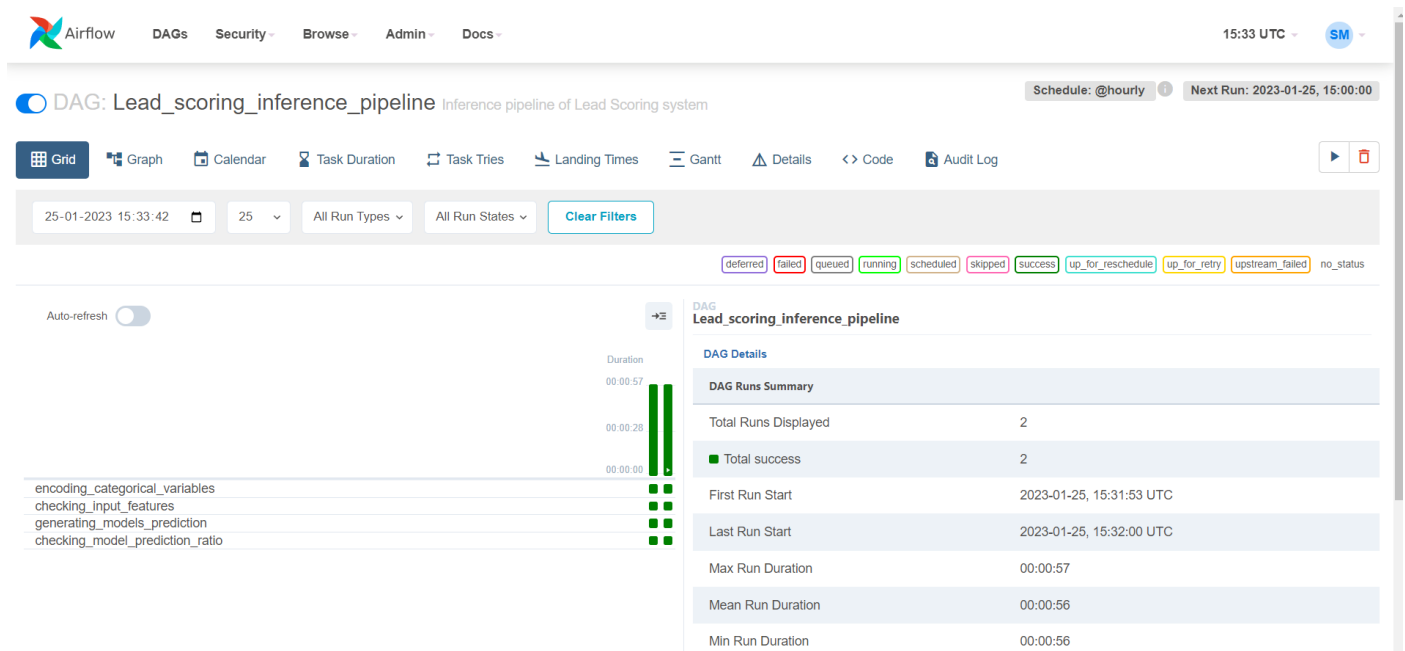


Inference Pipeline


- Screenshot of successful execution Airflow DAG in graph



- Screenshot of Airflow UI grid



ALL DAGS :

 Airflow

DAGs

Security

Browse

Admin

Docs

15:34 UTC

SM

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

DAGs

All 35Active 0Paused 35

Filter DAGs by tag

Search DAGs