

AIRFLOW HOMEPAGE

The screenshot shows the Airflow homepage with the title "AIRFLOW HOMEPAGE". At the top, there is a navigation bar with links for "DAGs", "Security", "Browse", "Admin", and "Docs". The time "06:18 UTC" is displayed on the right. Below the navigation bar, the title "DAGs" is centered. A search bar for "Search DAGs" is present. Below the search bar, there are three tabs: "All 35", "Active 3", and "Paused 32". A "Filter DAGs by tag" input field is also available. The main area displays a table of DAGs with columns: DAG, Owner, Runs, Schedule, Last Run, Next Run, and Recent Tasks. Three DAGs are listed:

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks
Lead_Scoring_Data_Engineering_Pipeline	airflow	4 (green), 2 (red)	@daily	2022-09-20, 06:02:24	2022-09-20, 00:00:00	7 (green)
Lead_scoring_inference_pipeline	airflow	3 (green), 8 (red)	@hourly	2022-09-20, 06:05:41	2022-09-20, 06:00:00	4 (green)
Lead_scoring_training_pipeline	airflow	8 (green)	@monthly	2022-09-20, 06:04:14	2022-09-01, 00:00:00	2 (green)

Pagination controls at the bottom show page 1 of 3, with a total of 3 DAGs.

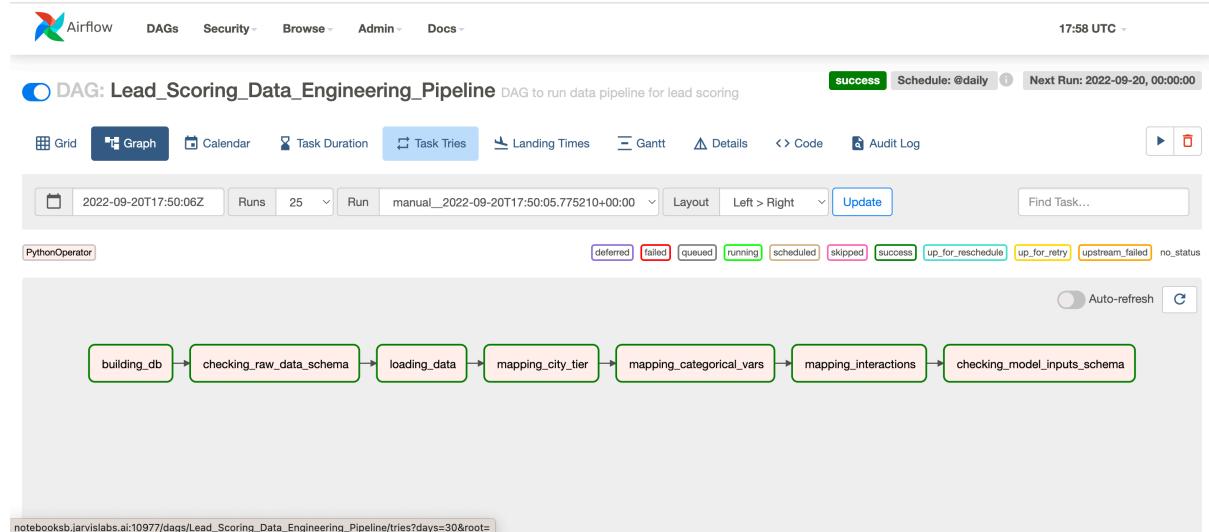
DATA PIPELINE

The screenshot shows the details page for the "Lead_Scoring_Data_Engineering_Pipeline". The top header includes the DAG name, schedule (@daily), next run (2022-09-20, 00:00:00), and refresh buttons. Below the header are navigation tabs: Grid, Graph, Calendar, Task Duration, Task Tries, Landing Times, Gantt, Details, Code, and Audit Log. A "Clear Filters" button is also present. The main area features a histogram of task durations for runs on September 19 and 20. A legend below the histogram defines run states: deferred (purple), failed (red), queued (yellow), running (green), scheduled (orange), skipped (pink), success (blue), up_for_reschedule (cyan), up_for_retry (light blue), upstream_failed (yellow-orange), and no_status (grey). A legend also indicates success (green) and failure (red). On the left, a list of tasks is shown: building_db, checking_raw_data_schema, loading_data, mapping_city_tier, and mapping_categorical_vars. The right side displays "DAG Details" and "DAG Runs Summary" sections. The "DAG Details" section shows the DAG name and a summary table:

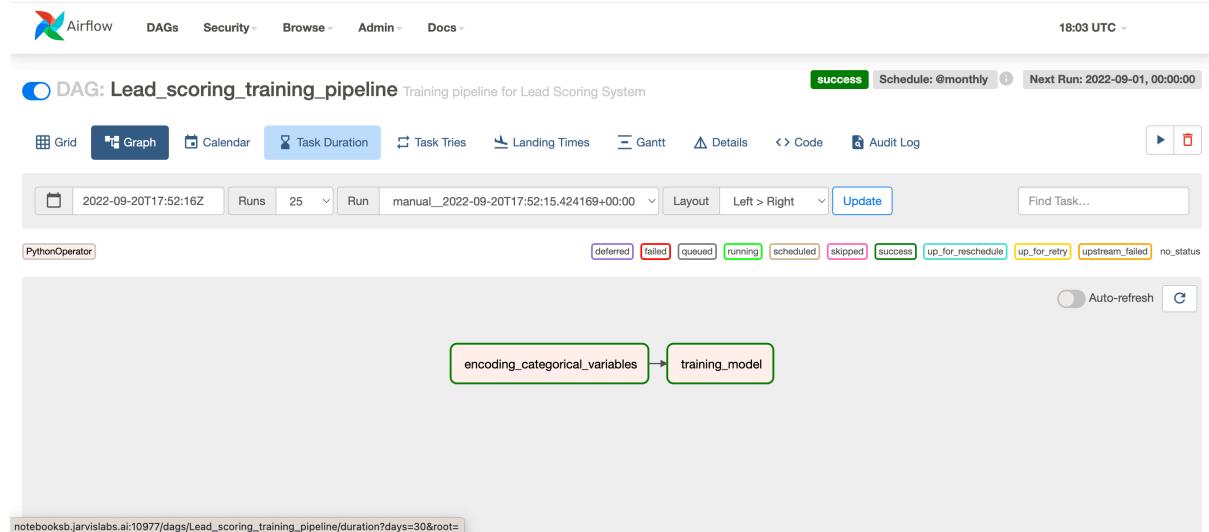
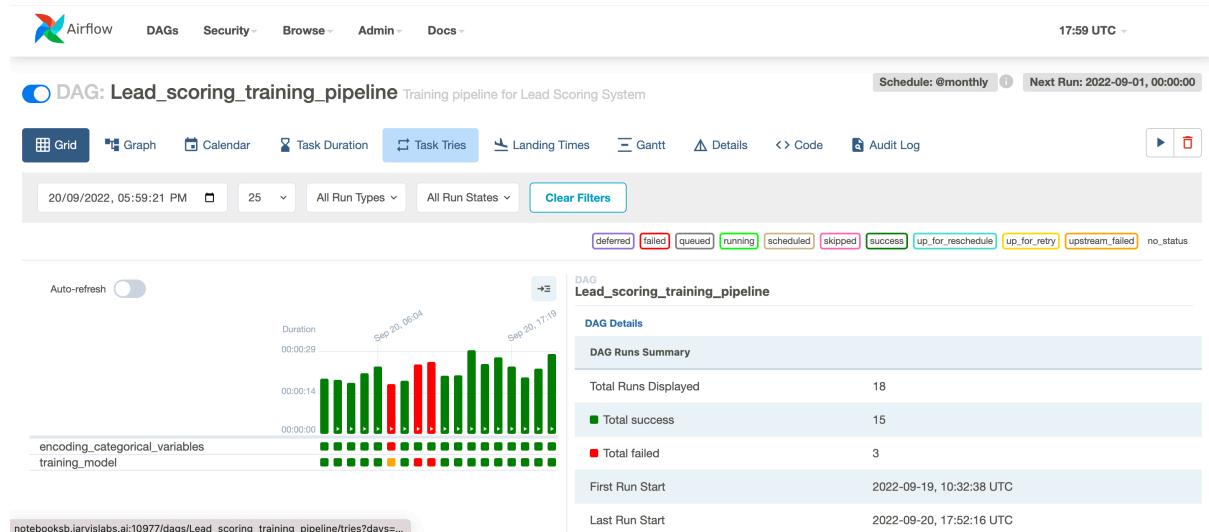
Total Runs Displayed	18
Total success	16
Total failed	2

The "DAG Runs Summary" section provides specific run details:

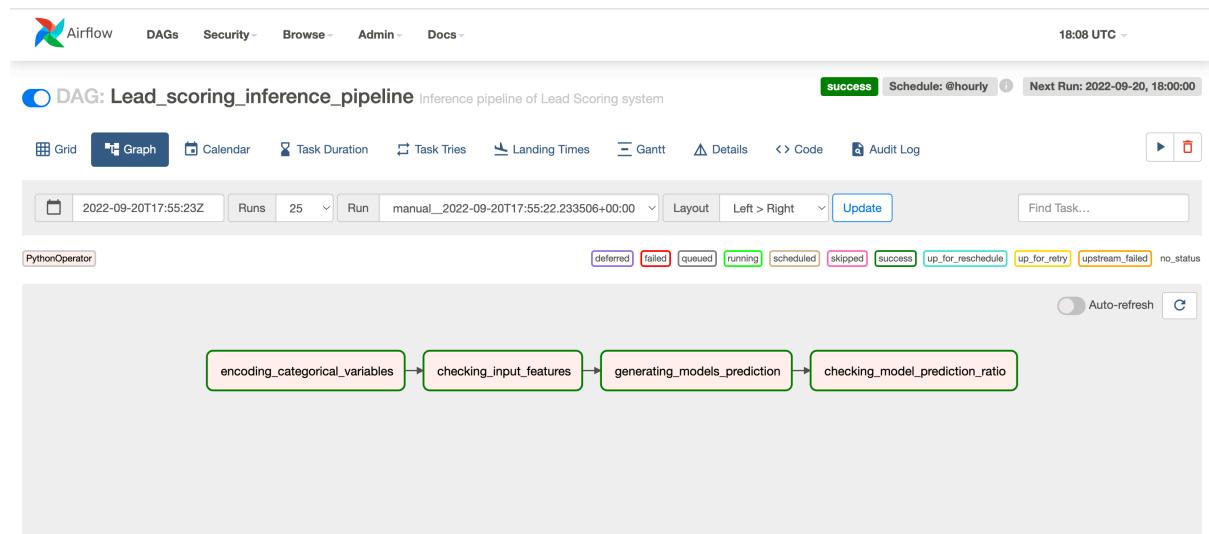
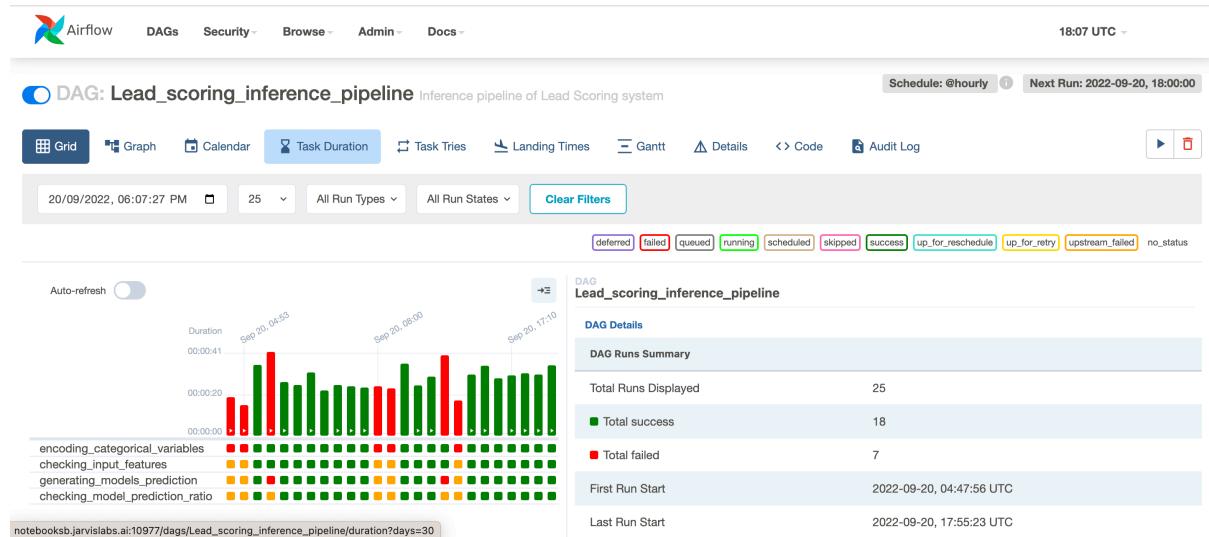
First Run Start	2022-09-18, 10:42:09 UTC
Last Run Start	2022-09-20, 17:50:08 UTC



TRAINING PIPELINE



INFERENCE PIPELINE



After making changes in Data Pipeline for Inference file Data

The screenshot shows a Jupyter Notebook interface. On the left, a file browser displays the contents of the `/dags/Lead_scoring_data_pipeline` directory. The `constants.py` file is selected and its content is shown on the right:

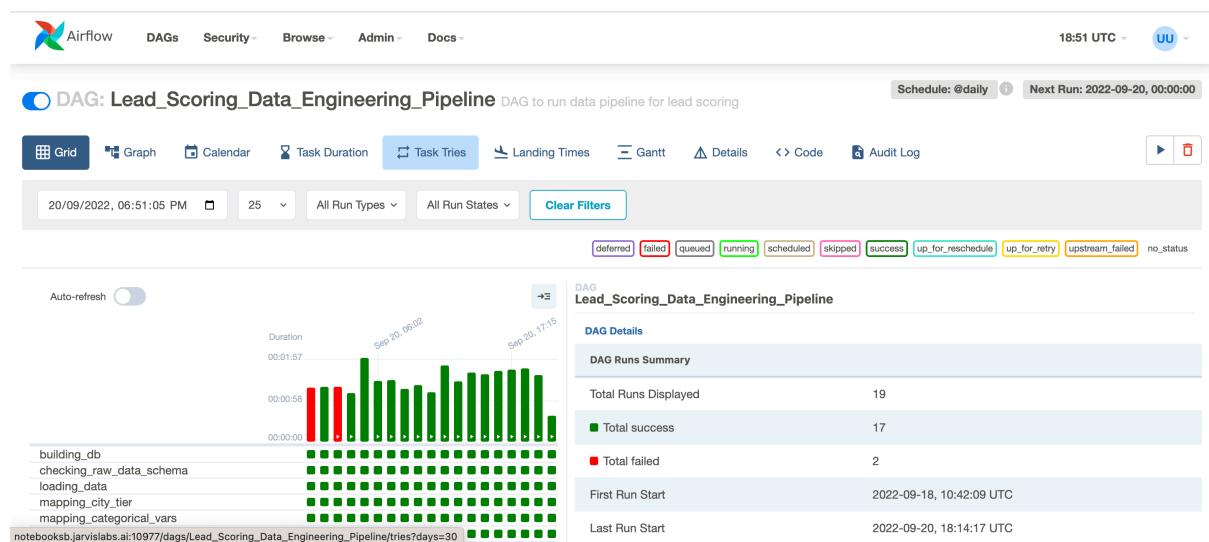
```

1 # You can create more variables according to your project. The following are the basic
2 DB_PATH = '/home/airflow/dags/Lead_scoring_data_pipeline/'
3 DB_FILE_NAME = 'lead_scoring_data_cleaning.db'
4 DATA_DIRECTORY = '/home/airflow/dags/Lead_scoring_data_pipeline/data/'
5 INTERACTION_MAPPING = '/home/airflow/dags/Lead_scoring_data_pipeline/mapping/'
6 INDEX_COLUMNS = ['created_date', 'city_tier', 'first_platform_c', 'first_utm_medium_c',
7 '#NOT_FEATURES = '
8 LEAD_SCORING_FILE='leadscore_inference_final_v2'

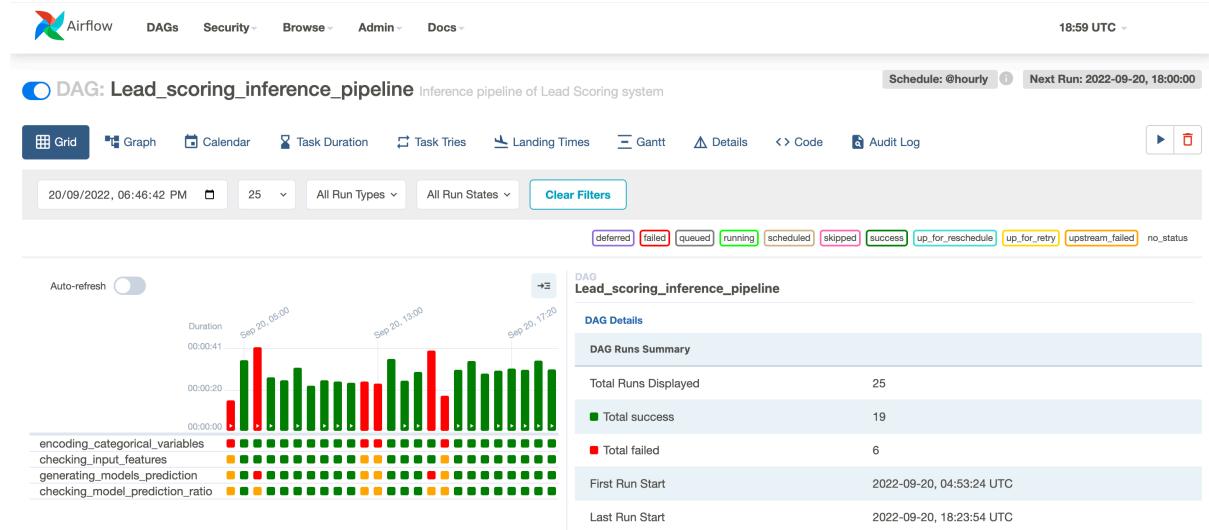
```

The notebook interface includes tabs for `Untitled.ipynb`, `prediction_distribution_.py`, and `constants.py`. The status bar at the bottom indicates "Ln 8, Col 51 Spaces: 4 constants.py".

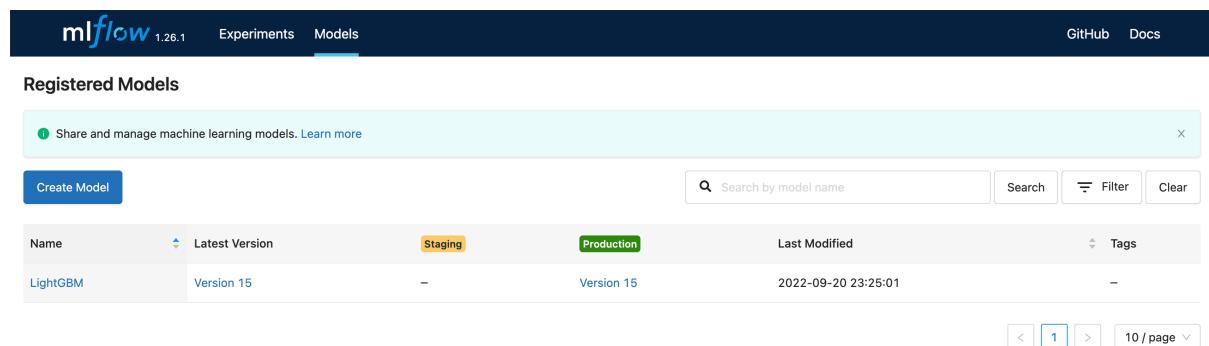
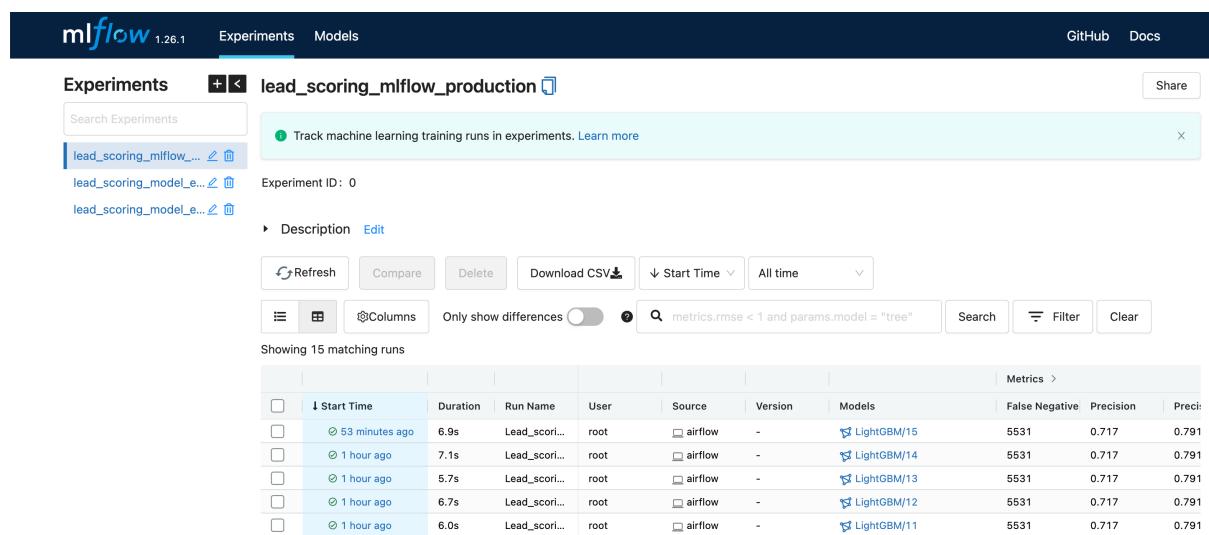
Data Pipeline



Data Inference



ML Flow Server



Models

The screenshot shows the MLflow UI for a registered model named "LightGBM, v15". The model was registered on 2022/09/20. It includes a list of artifacts such as MLmodel, conda.yaml, model.pkl, python_env.yaml, and requirements.txt. Below this, there's a section titled "MLflow Model" with code snippets for making predictions using a Spark DataFrame and a Pandas DataFrame.

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 MLflow docs for how to include input and output schema with your model.	

Make Predictions

Predict on a Spark DataFrame:

```
import mlflow
logged_model = 'runs:/2fc45d56b03244d1882cd9035b8c737f/models'
```

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()

Predict on a Pandas DataFrame:

```
import mlflow
logged_model = 'runs:/2fc45d56b03244d1882cd9035b8c737f/models'
```

Metrics

The screenshot shows the MLflow UI for an experiment named "Lead_scoring_mlflow_production". The experiment was run on 2022-09-20 23:22:33 with a duration of 6.9s. The status is FINISHED. The user was root and the lifecycle stage is active. The metrics page displays various performance metrics:

Name	Value
False Negative %	5531
Precision %	0.717
Precision_0 %	0.791
Precision_1 %	0.674
Recall %	0.732
Recall_0 %	0.588
Recall_1 %	0.846
True Negative %	20946
F1_0 %	0.675
F1_1 %	0.75
roc_auc %	0.717
test_accuracy %	0.717

Tags

Artifacts

The artifacts section lists the same files as the "Models" screen: MLmodel, conda.yaml, model.pkl, python_env.yaml, and requirements.txt.

Parameters

MLC31 - Google D... The Super Duper... Computer Vision I... abhishek (Abhish... speech_recogniti... AWS Management... Encrypting passw... How To Add Goog... Adobe Creative Cl... »

► Description [Edit](#)

► Parameters (20)

Name	Value
boosting_type	gbdt
class_weight	None
criterion	gini
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
subsample	1.0
subsample_for_bin	200000
subsample_freq	0

► Metrics (12)

► Tags

▼ Artifacts