SCREENSHOT Document for UpGrad's Assignment Submission

Prepared By,

Sanjeev Surendran

Contents

Table of Figures

Airflow Main UI Page

This is the main Airflow UI page that contains all the registered DAGS. We can see our 3 DAGS:

- Lead_Scoring_Data_Engineering_Pipeline
- Lead_scoring_training_pipeline
- Lead_scoring_inference_pipeline

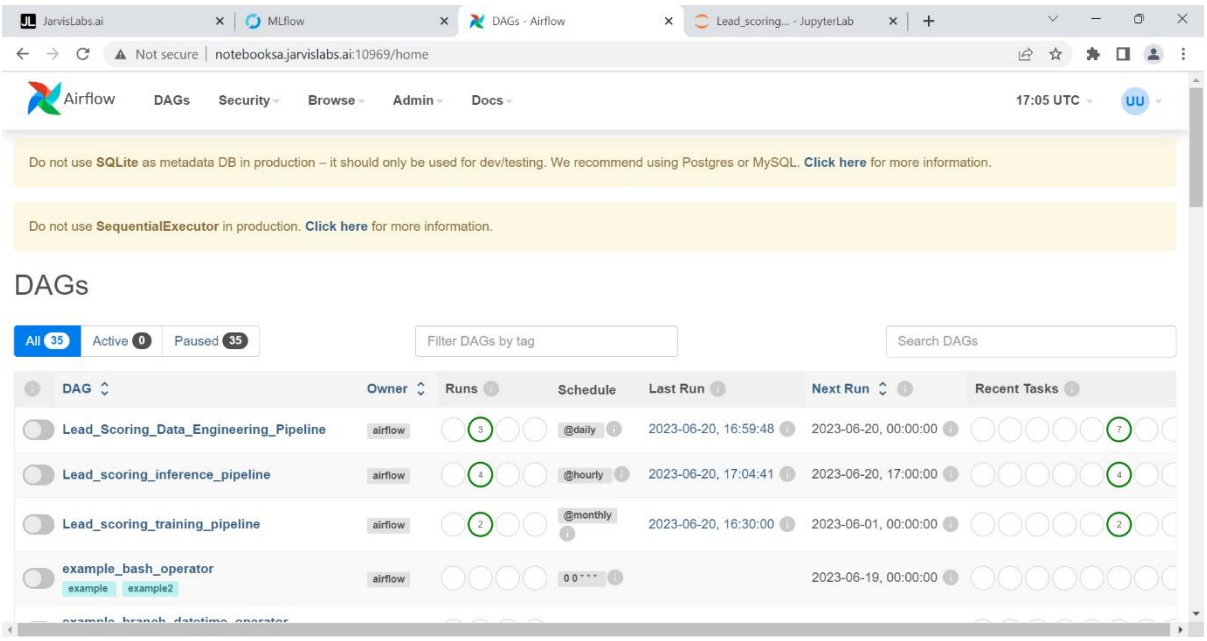

Figure 1: Airflow Main UI Page

Initial run of Lead_Scoring_Data_Engineering_Pipeline

We have triggered manual run of Lead_Scoring_Data_Engineering_Pipeline for data engineering and cleaning data.

Both manual and scheduled runs can be seen in picture. Manual runs have white coloured arrow on top of green bars in Grid view.

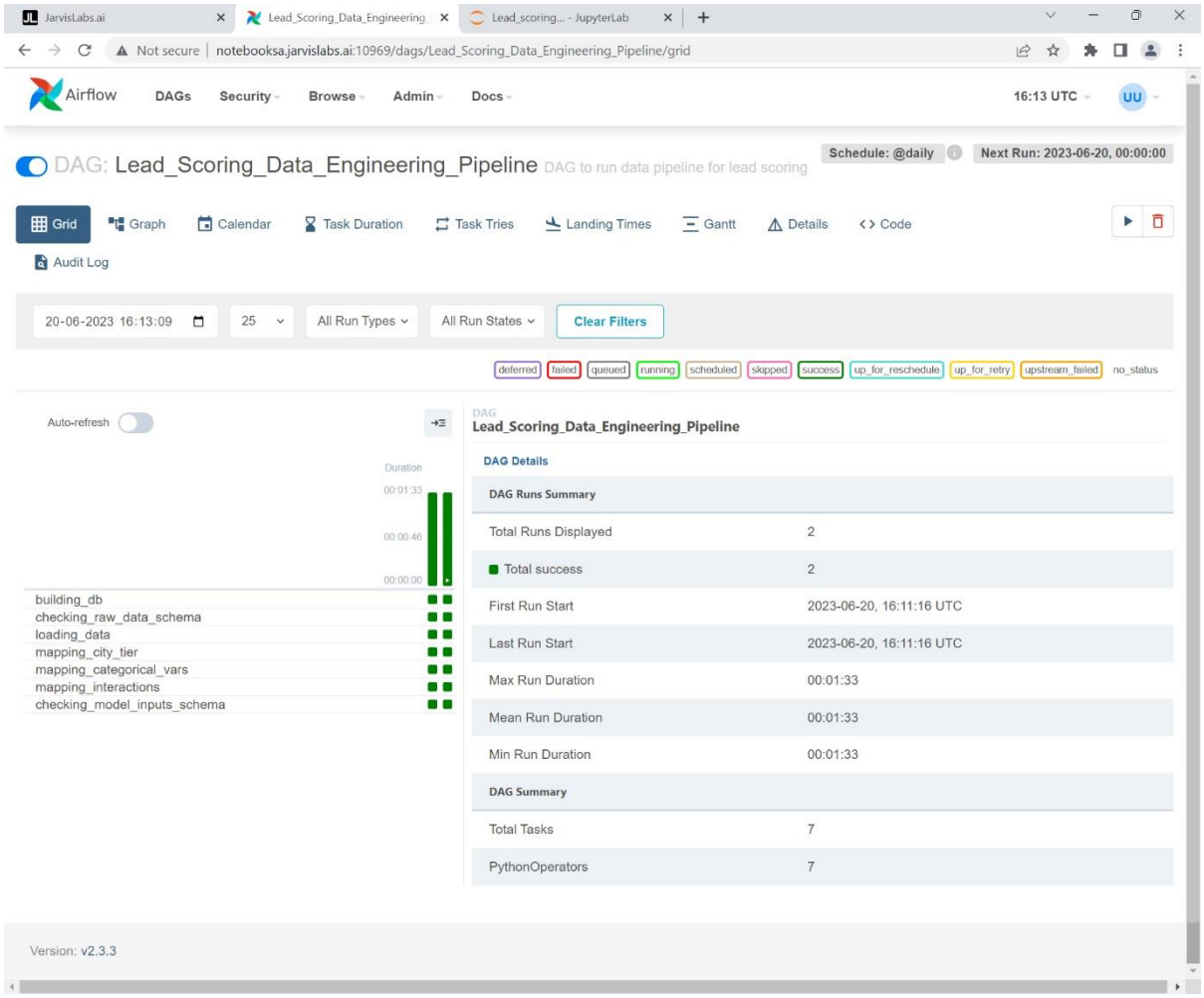

Figure 2: Grid view of Lead_Scoring_Data_Engineering_Pipeline

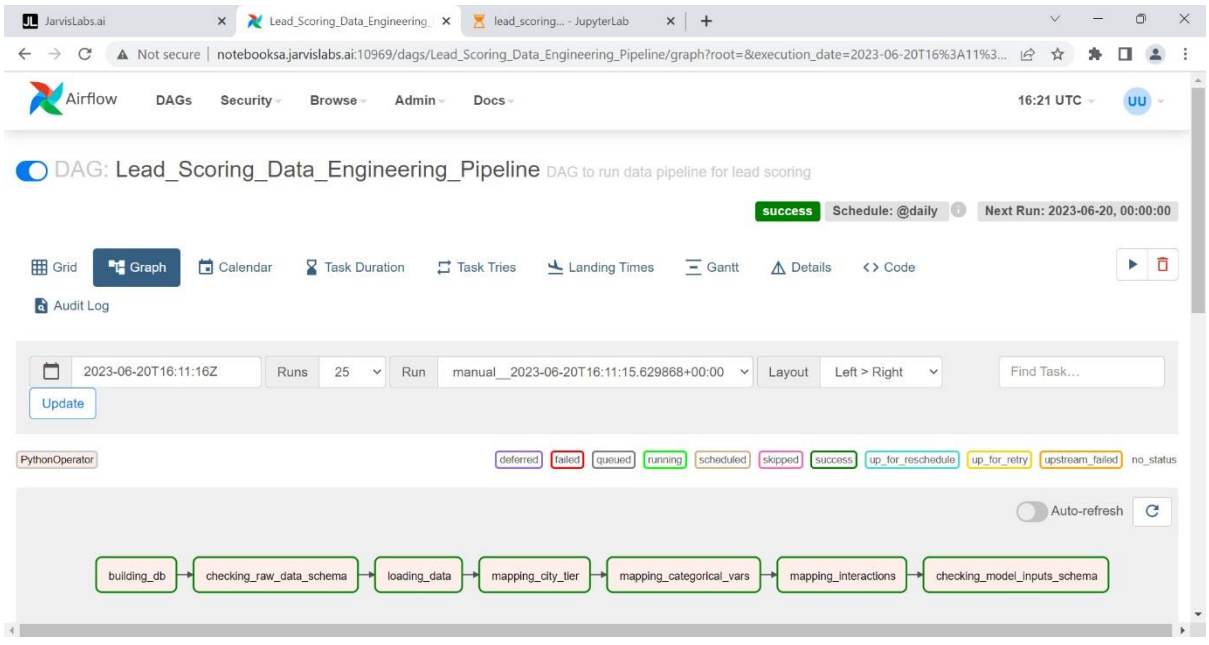

Figure 3: Graph view of Lead_Scoring_Data_Engineering_Pipeline

Initial run of Lead_scoring_training_pipeline

We have triggered manual run of Lead_scoring_training_pipeline for training cleaned data.

Both manual and scheduled runs can be seen in picture. Manual runs have white coloured arrow on top of green bars in Grid view.

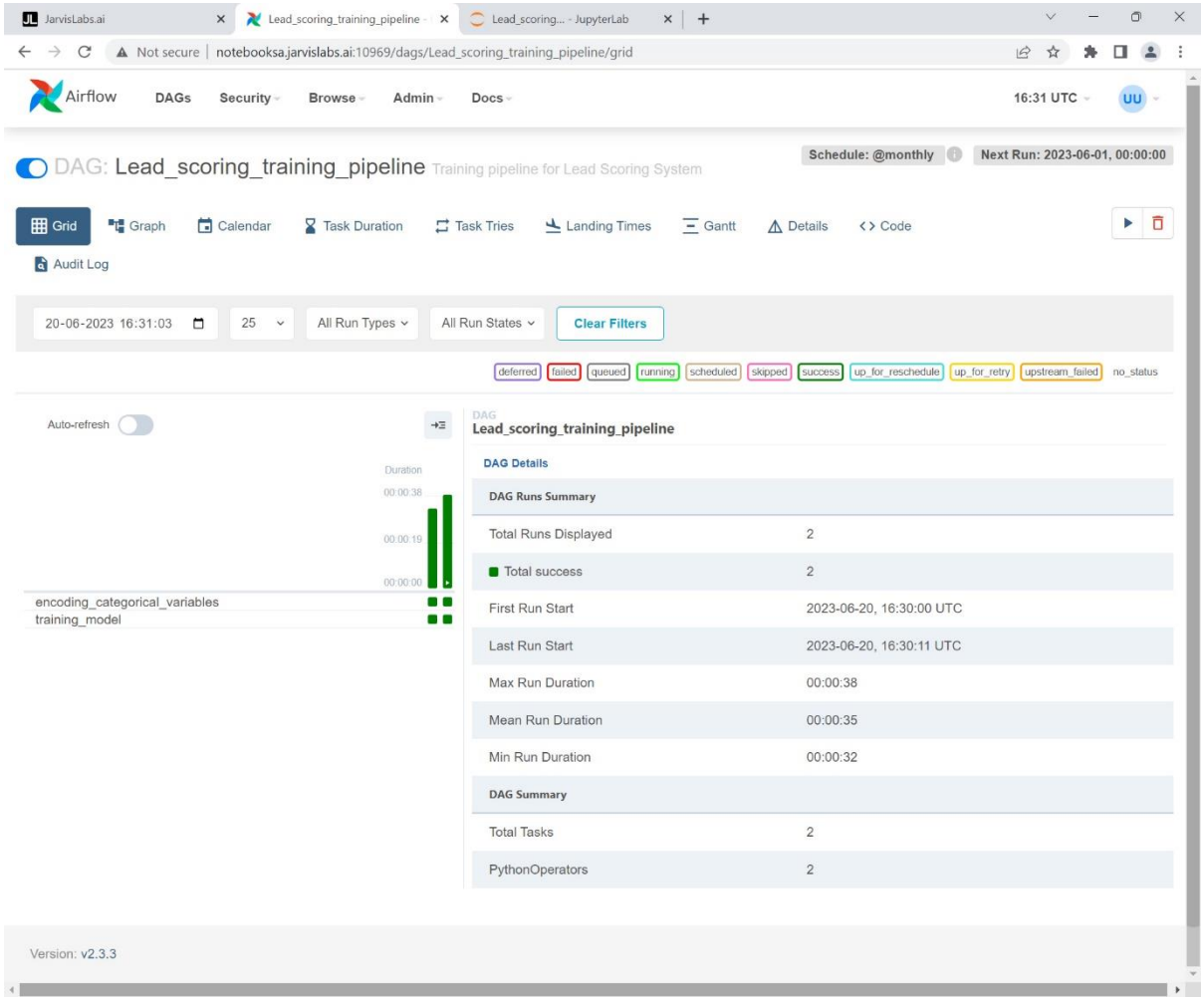

Figure 4: Grid view of Lead_scoring_training_pipeline

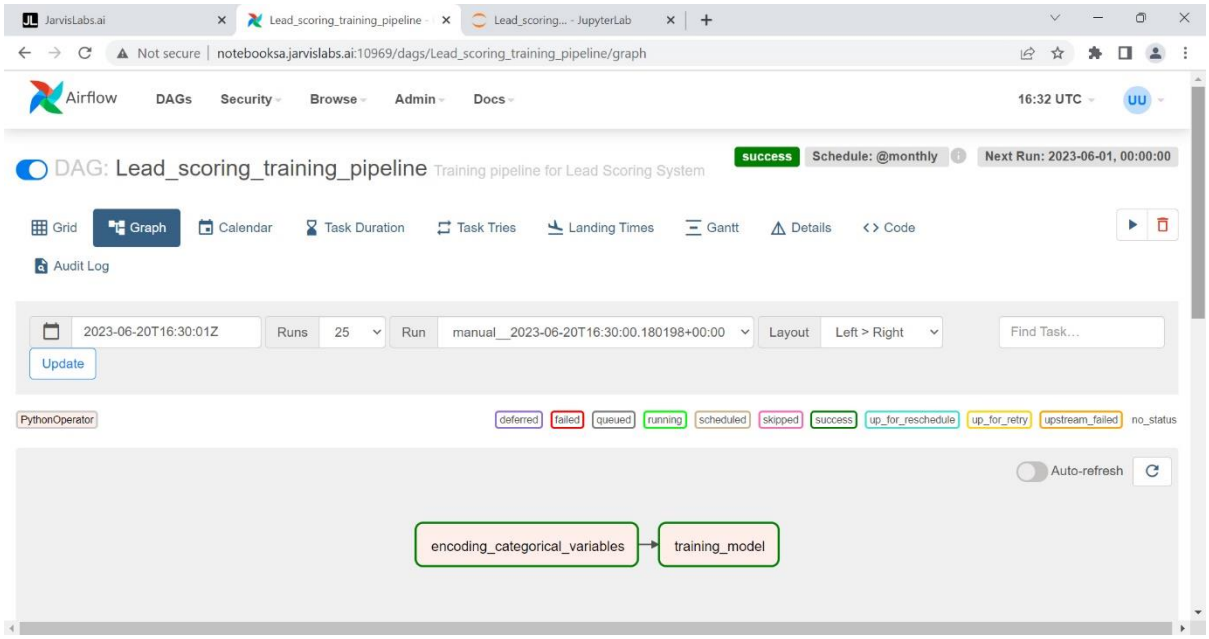

Figure 5: Graph view of Lead_scoring_training_pipeline

Initial run of Lead_scoring_inference_pipeline

We have triggered manual run of Lead_scoring_inference_pipeline for inferencing incoming data and predict outcome.

Both manual and scheduled runs can be seen in picture. Manual runs have white coloured arrow on top of green bars in Grid view.

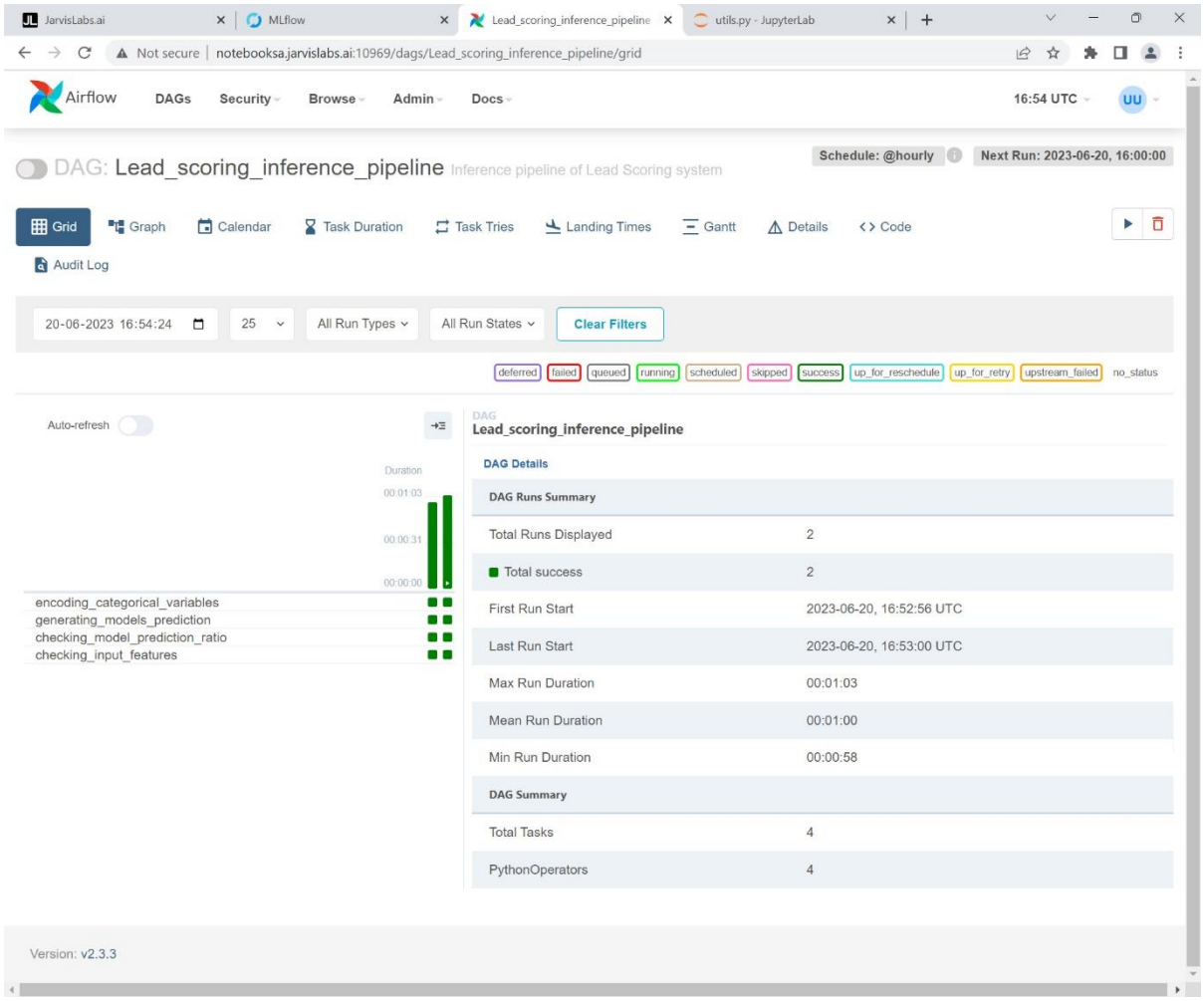

Figure 6: Grid view of Lead_scoring_inference_pipeline

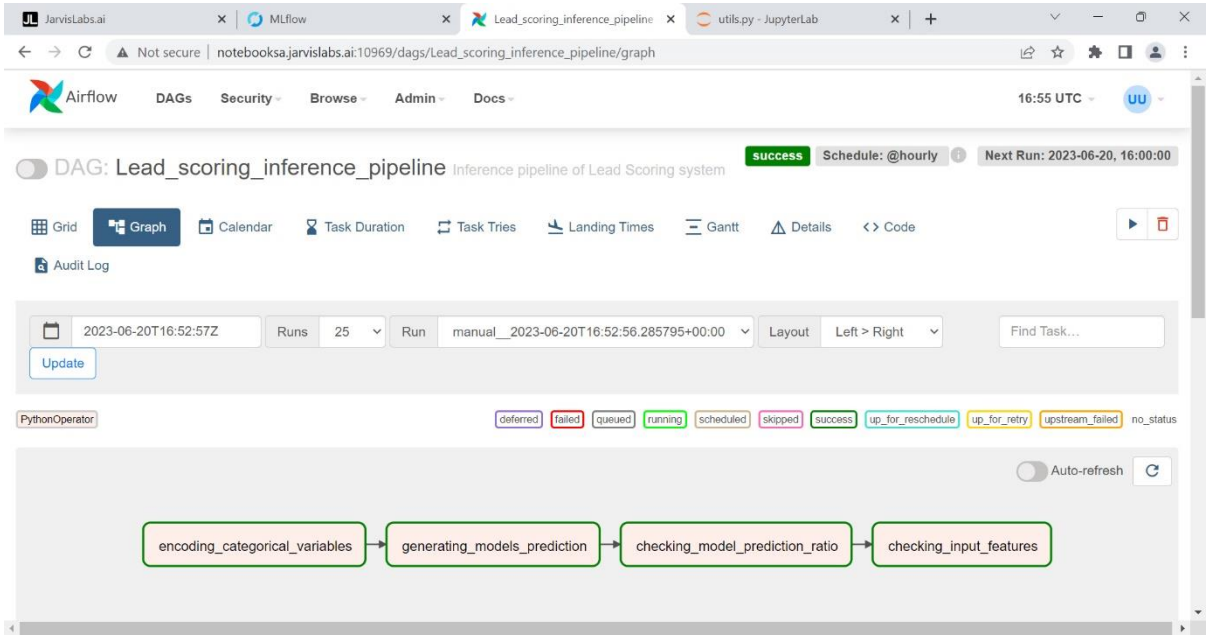

Figure 7: Graph view of Lead_scoring_inference_pipeline

Next run of Lead_Scoring_Data_Engineering_Pipeline

We have triggered manual run of Lead_Scoring_Data_Engineering_Pipeline for cleaning **leadscoring_inference.csv**.

Both manual and scheduled runs can be seen in picture. Manual runs have white coloured arrow on top of green bars in Grid view.

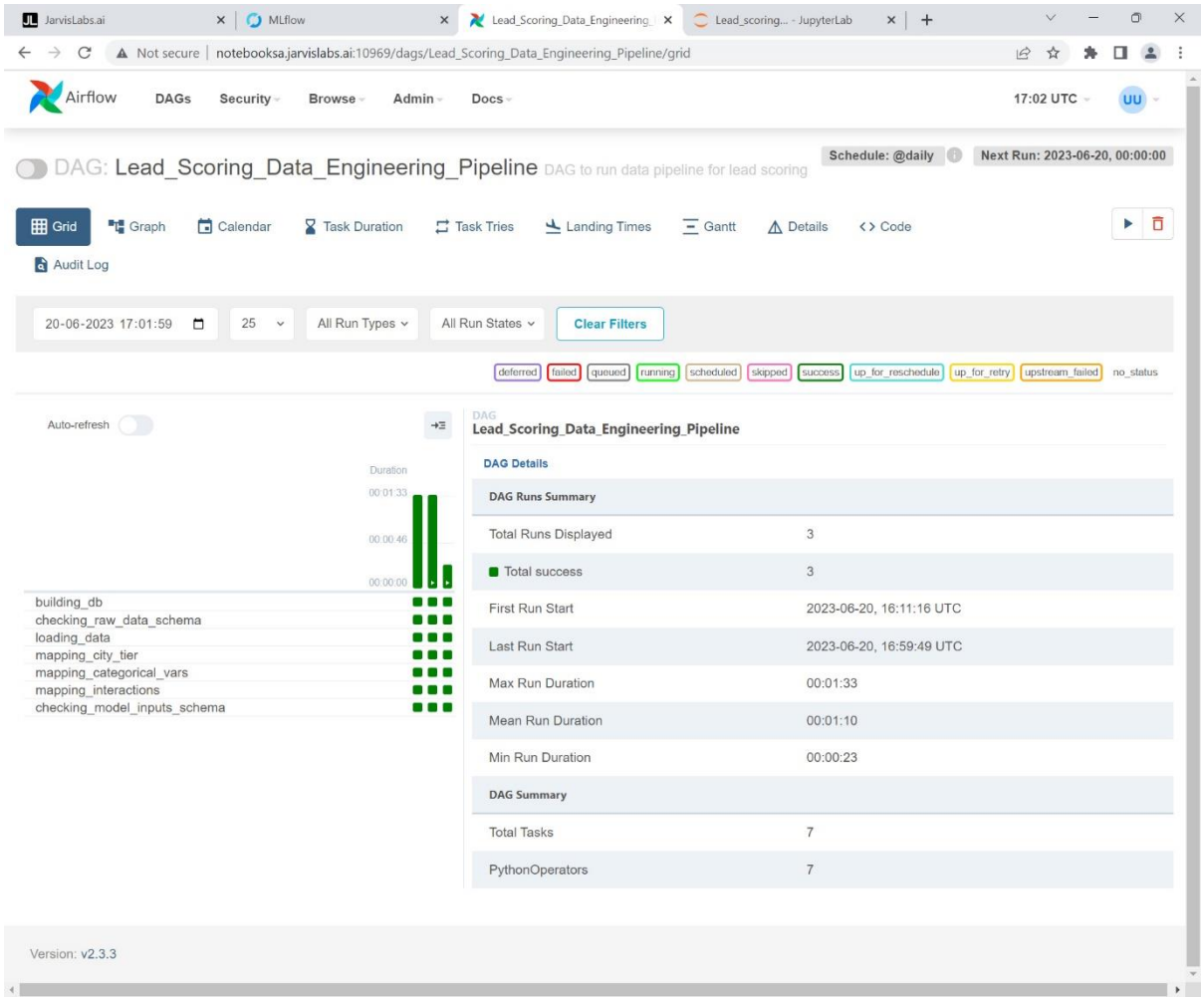

Figure 8: Grid view of Lead_Scoring_Data_Engineering_Pipeline

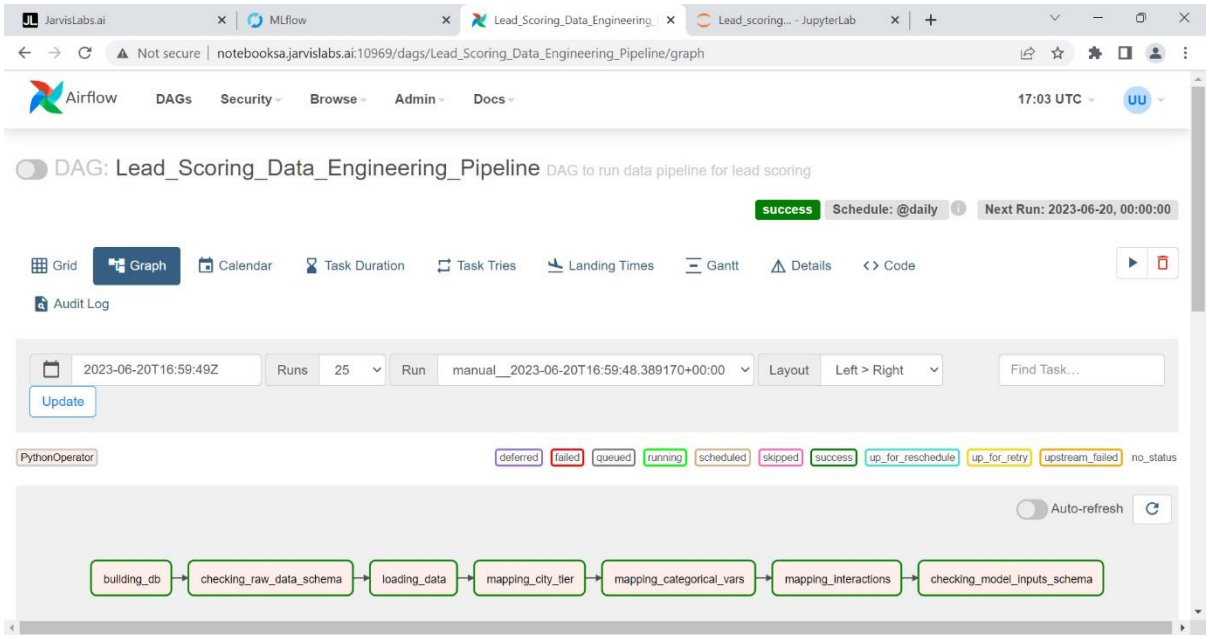

Figure 9: Graph view of Lead_Scoring_Data_Engineering_Pipeline

Next run of Lead_scoring_inference_pipeline

We have triggered manual run of Lead_scoring_inference_pipeline for inferencing incoming data and predict outcome.

Both manual and scheduled runs can be seen in picture. Manual runs have white coloured arrow on top of green bars in Grid view.

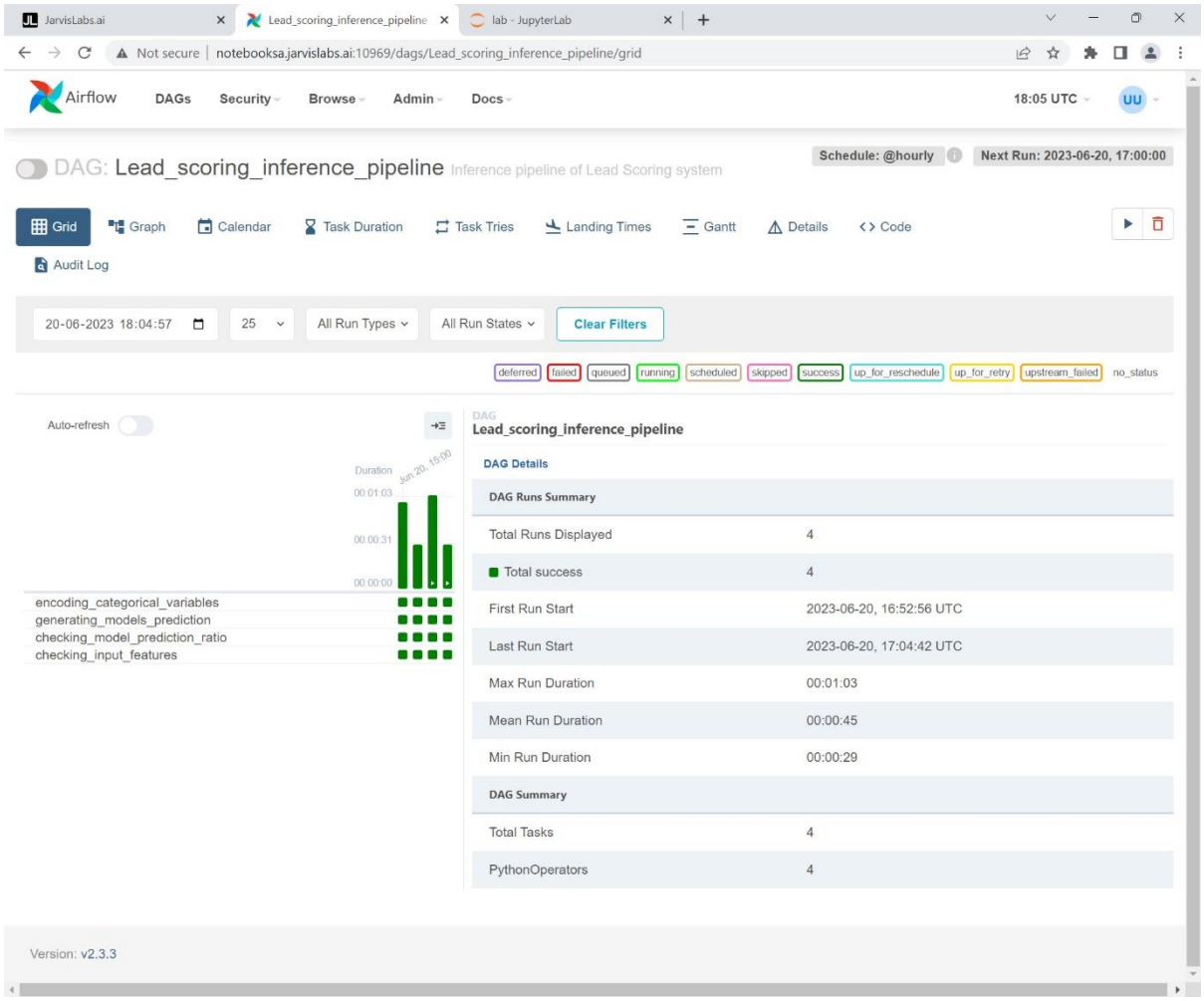

Figure 10: Grid view of Lead_scoring_inference_pipeline

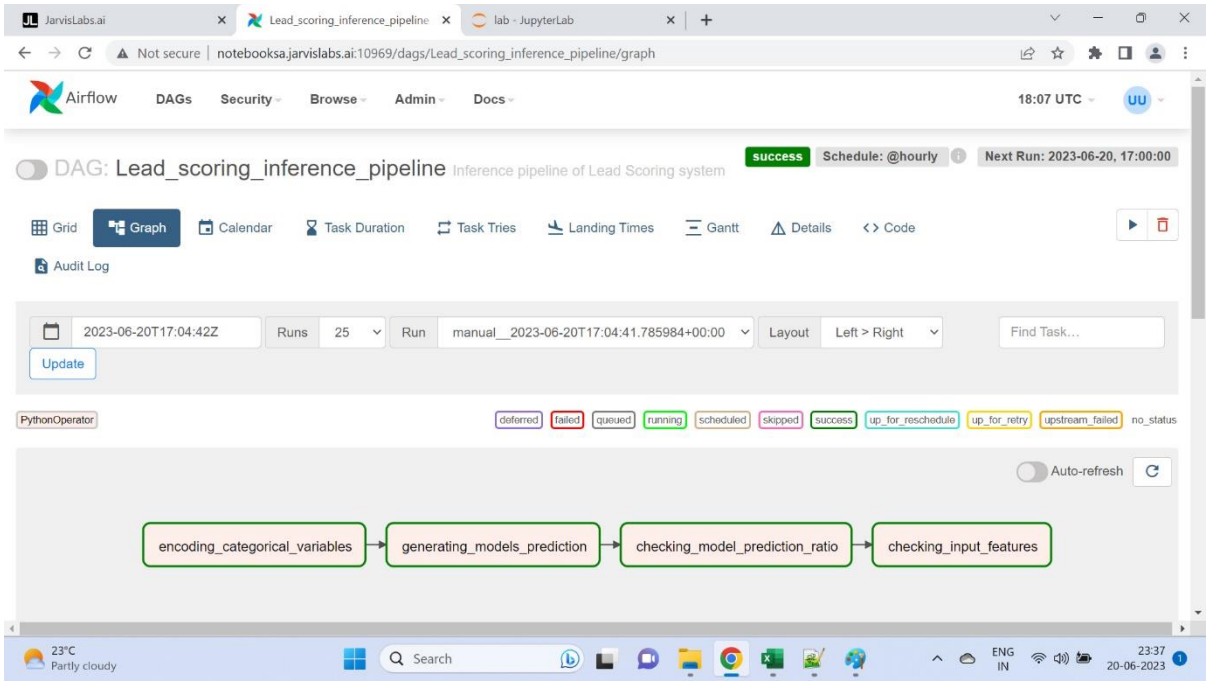

Figure 11: Graph view of Lead_scoring_inference_pipeline

MLFlow of Baseline_model_exp01

Baseline_model_exp01 is the initial pycaret setup experiment done for training model.

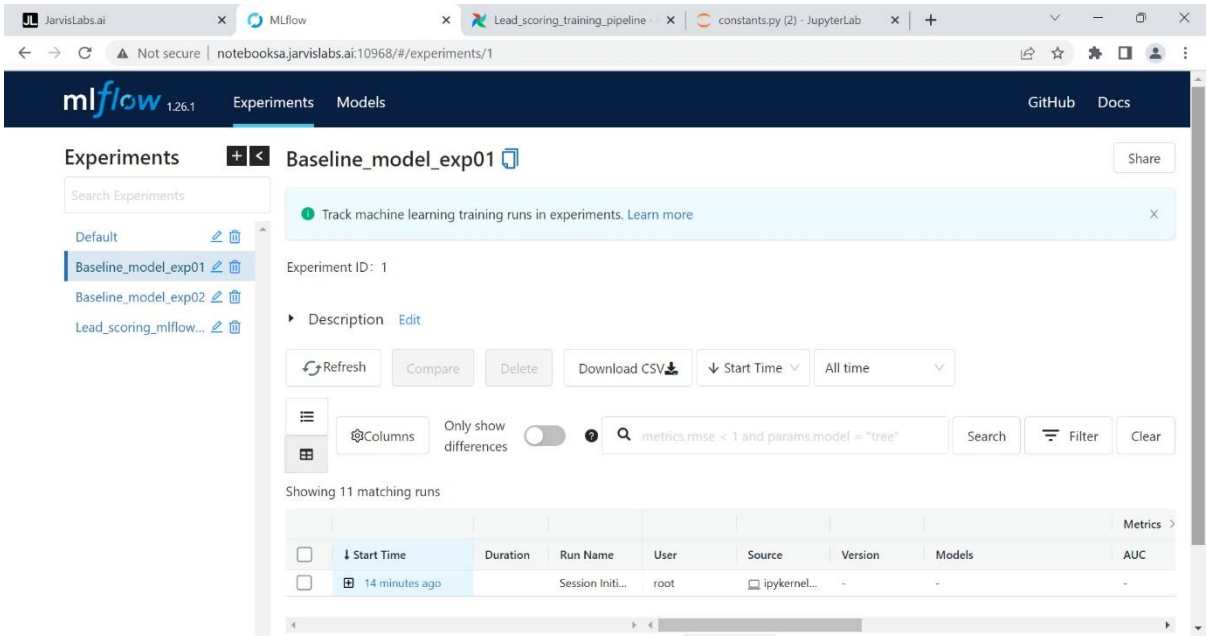

Figure 12: Main MLFlow UI of Baseline_model_exp01

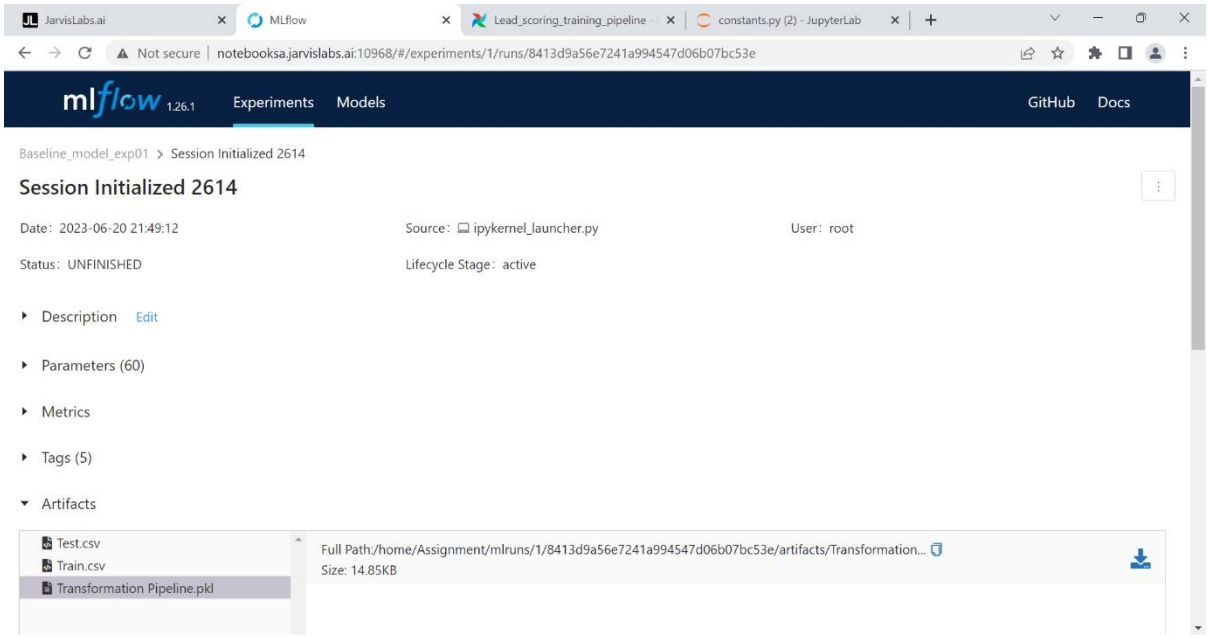

Figure 13: MLFlow Artifacts of Baseline_model_exp01

MLFlow of Baseline_model_exp02

Baseline_model_exp02 is the pycaret setup experiment done for training model **after removing certain features**.

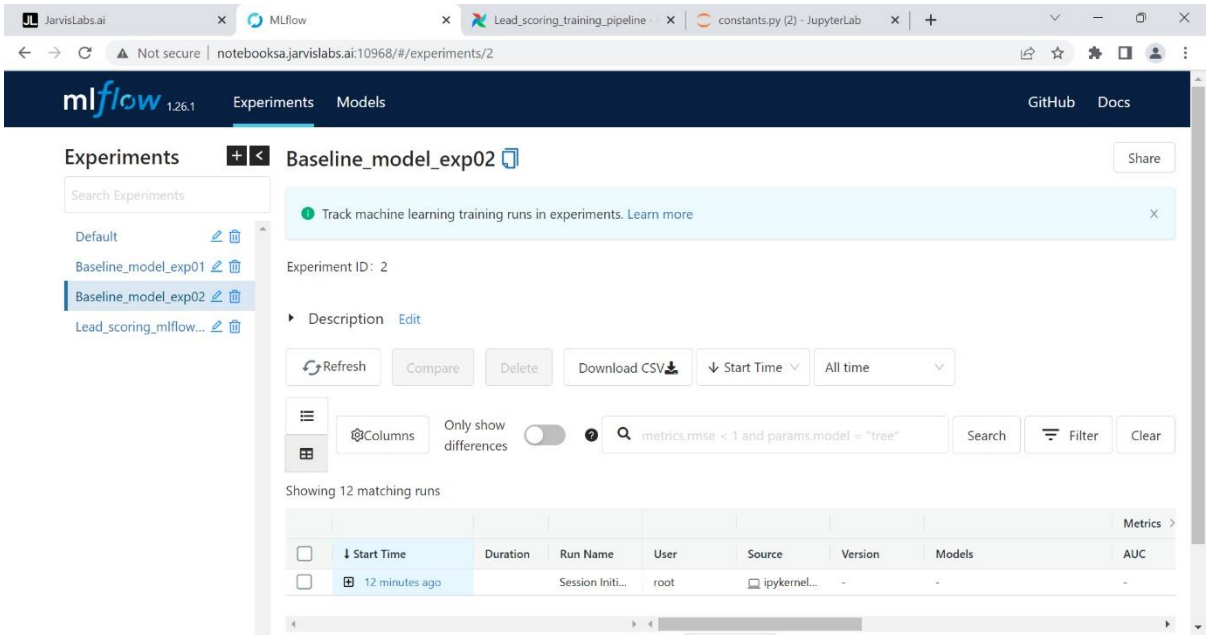

Figure 14: Main MLFlow UI of Baseline_model_exp02

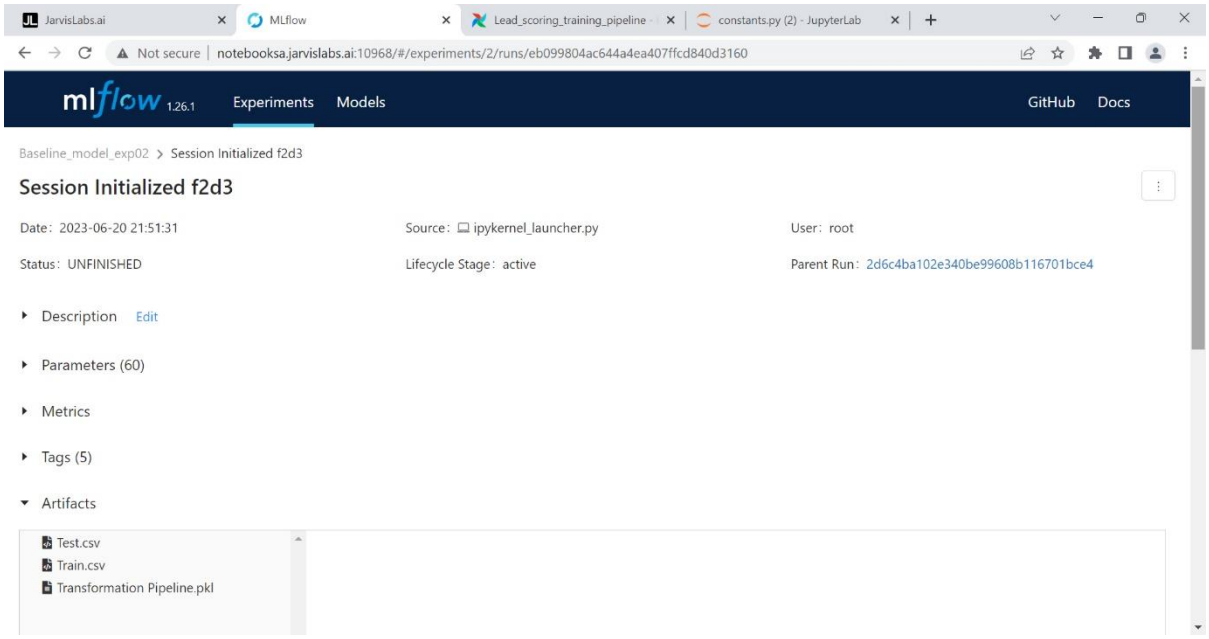

Figure 15: MLFlow Artifacts of Baseline_model_exp02

MLFlow of Lead_scoring_mlflow_production

Lead_scoring_mlflow_production is the final model we have selected. We can see 3 models are generated:

- **Version 1** - One was created when dummy notebook was run to check the working of utils.py functions.
- **Version 2** - One was created during schedule run of airflow (automatic).
- **Version 3** - One was created manually from airflow UI. This is our final model and we have moved this version to **PRODUCTION** stage.

Please check the figure captions to understand the following screenshots.

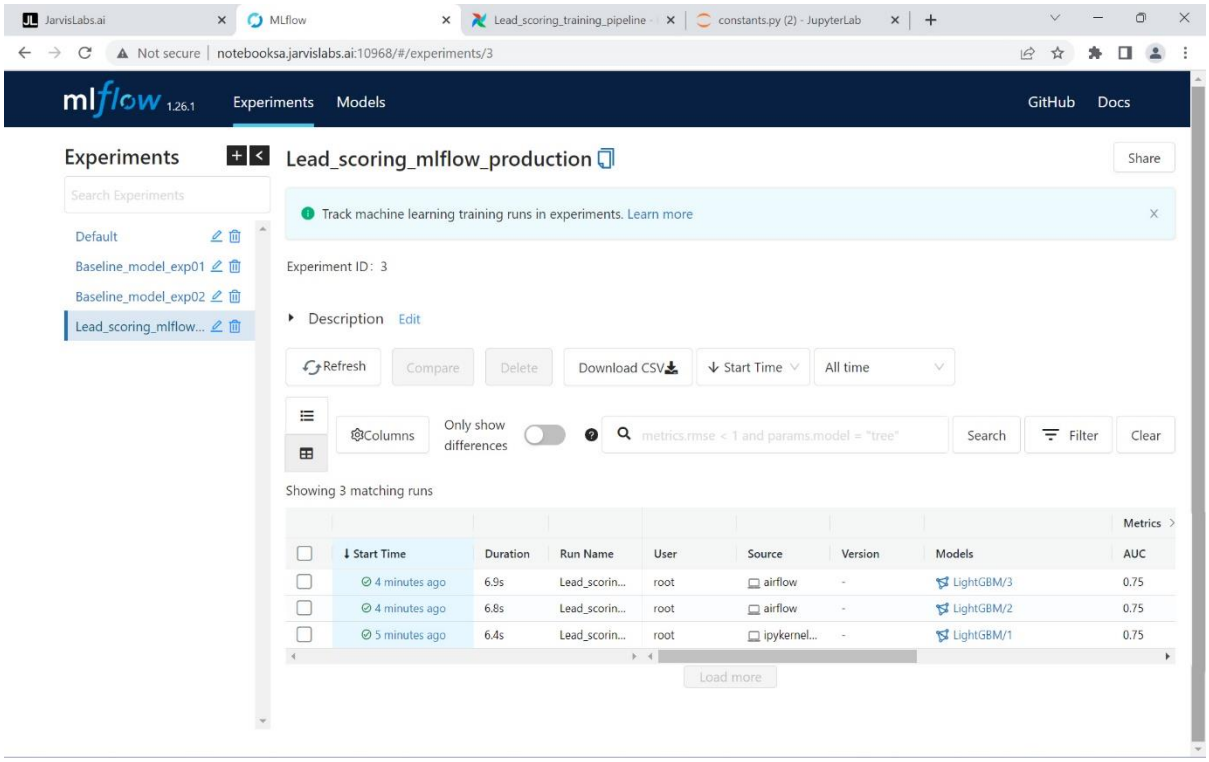

Figure 16: Main MLFlow UI of Lead_scoring_mlflow_production

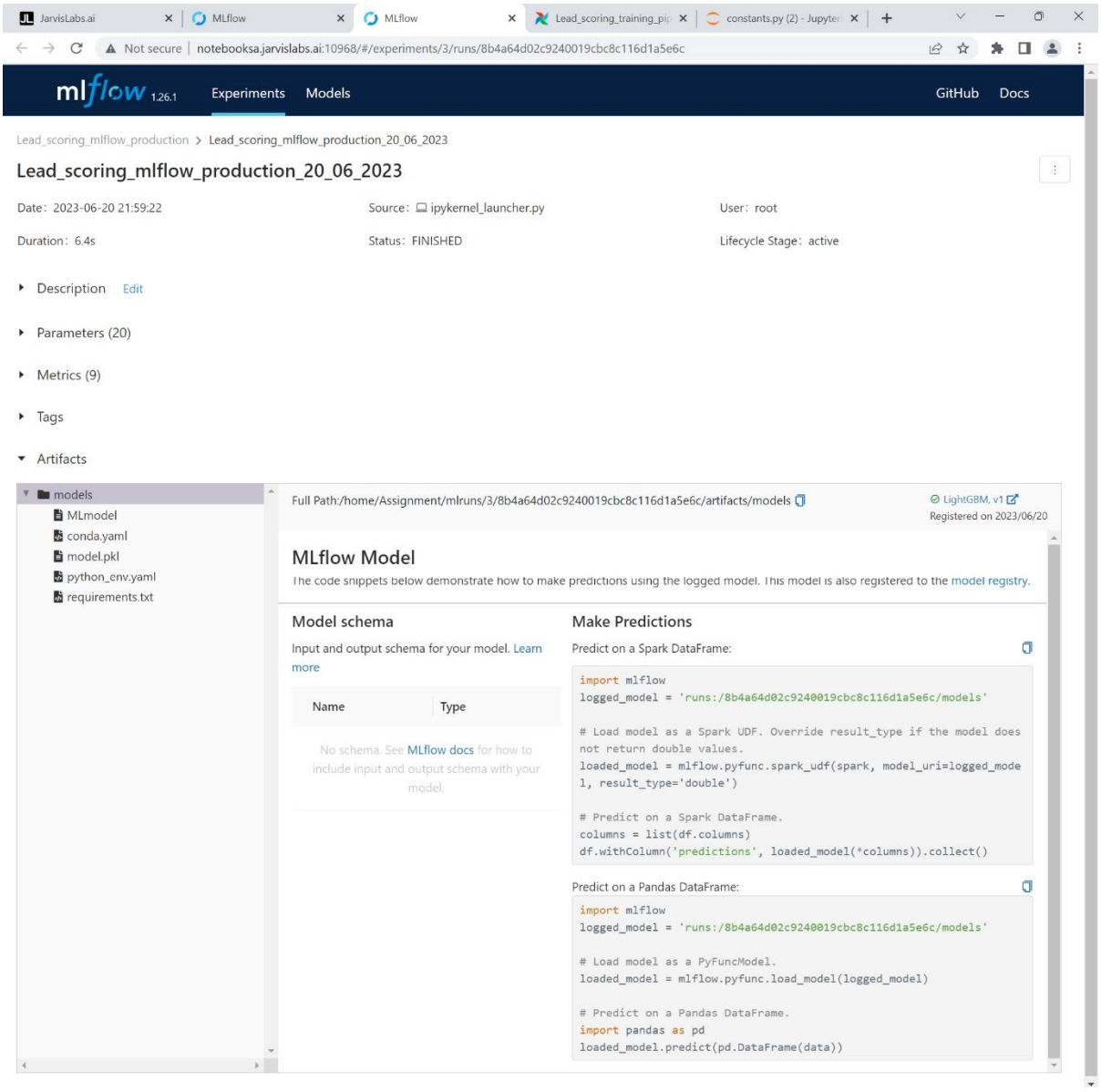

Figure 17: MLFlow Artifacts of Version 1 model

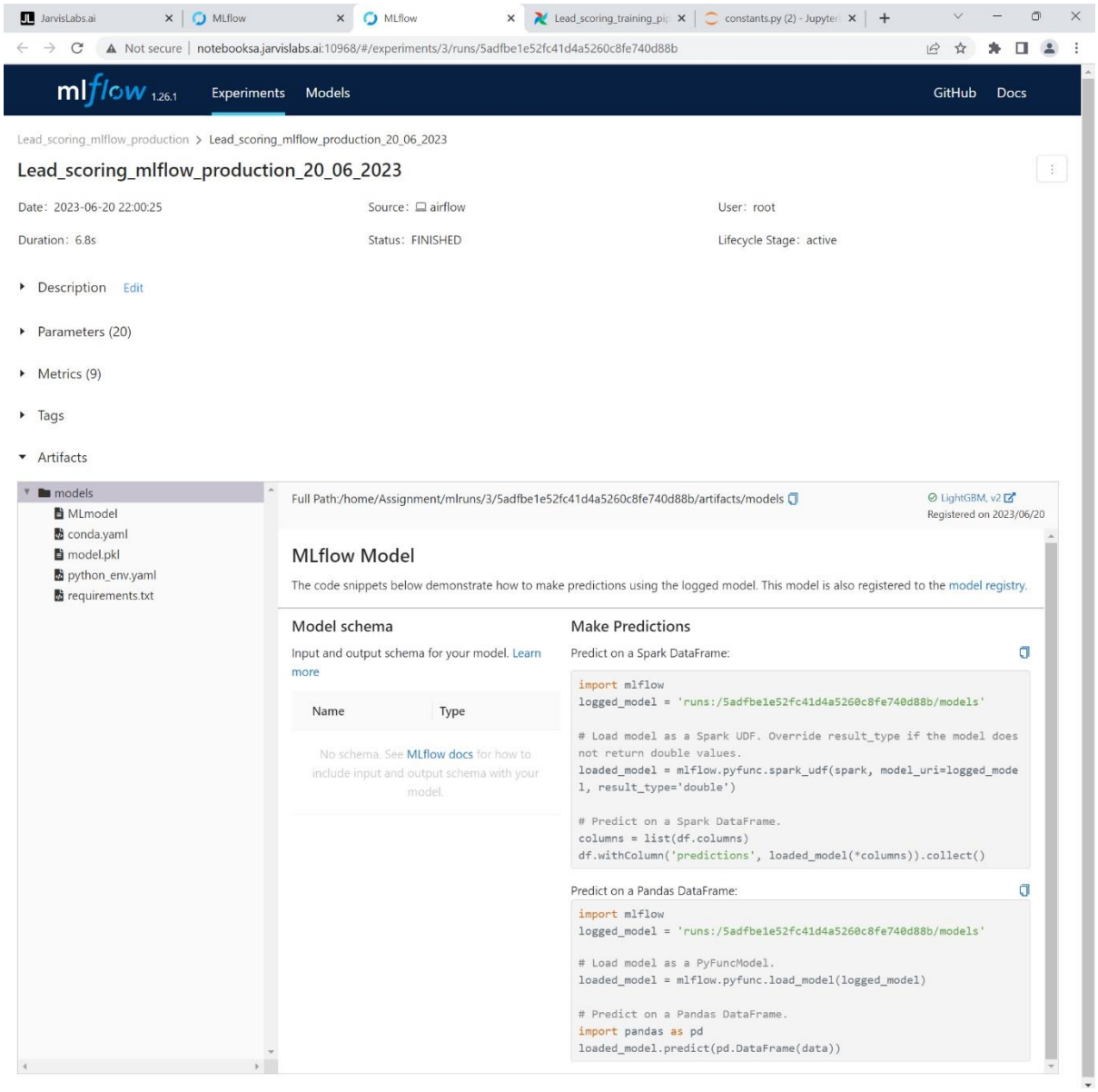

Figure 18: MLFlow Artifacts of Version 2 model

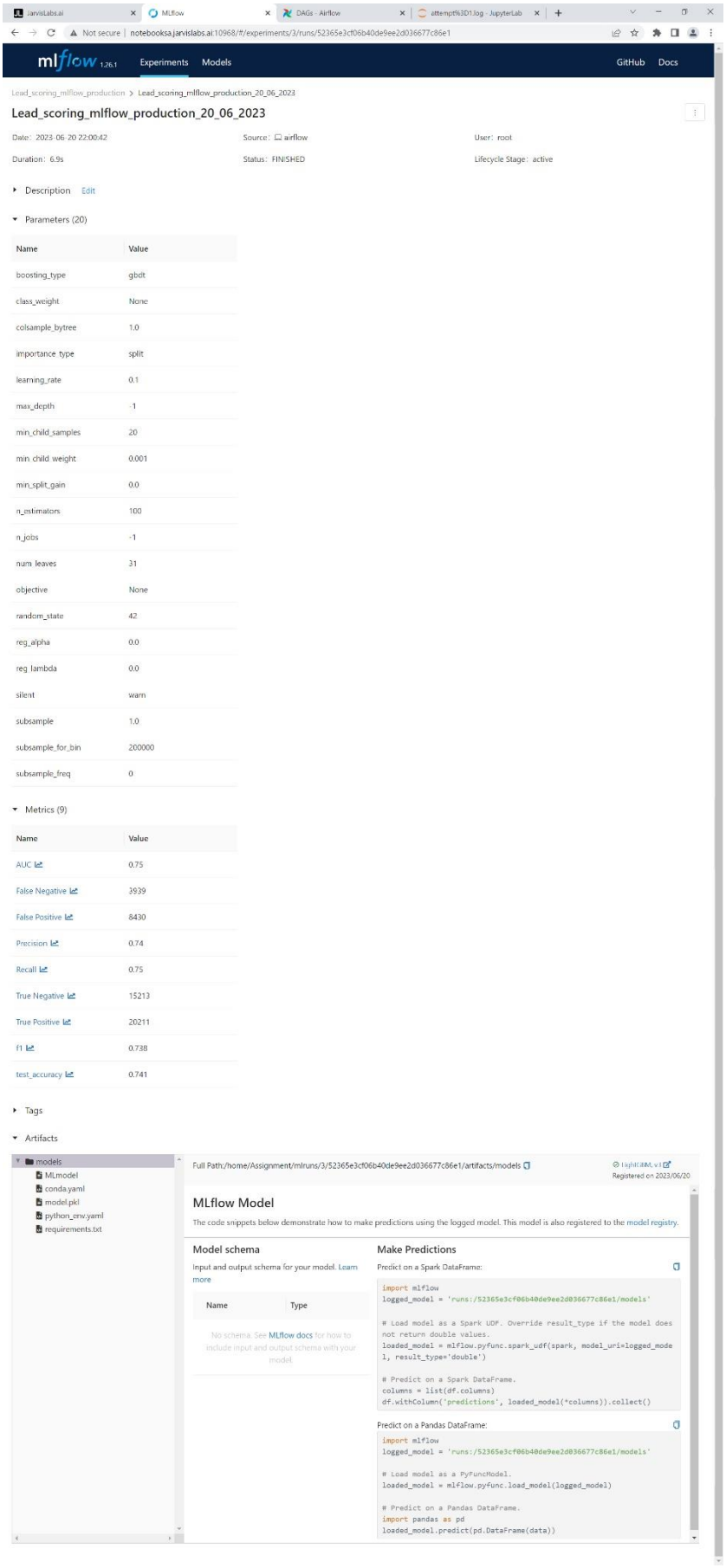

Figure 19: MLFlow Artifacts of Version 3 model

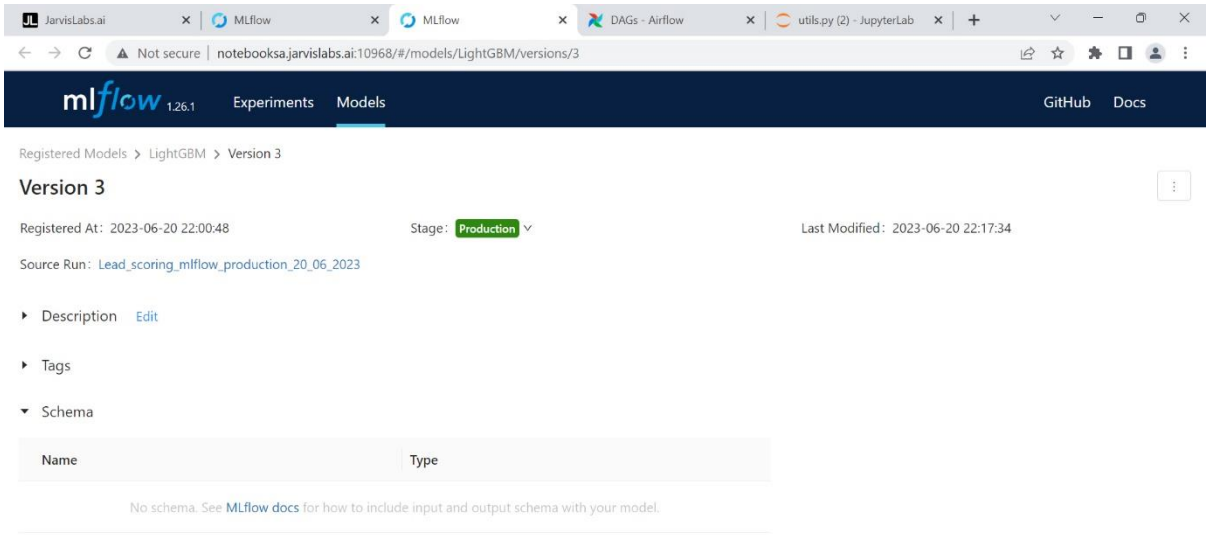

Figure 20: Version 3 model registered as Production Stage

Pytest of Data Pipeline

Screenshot of pytest done for data pipeline.

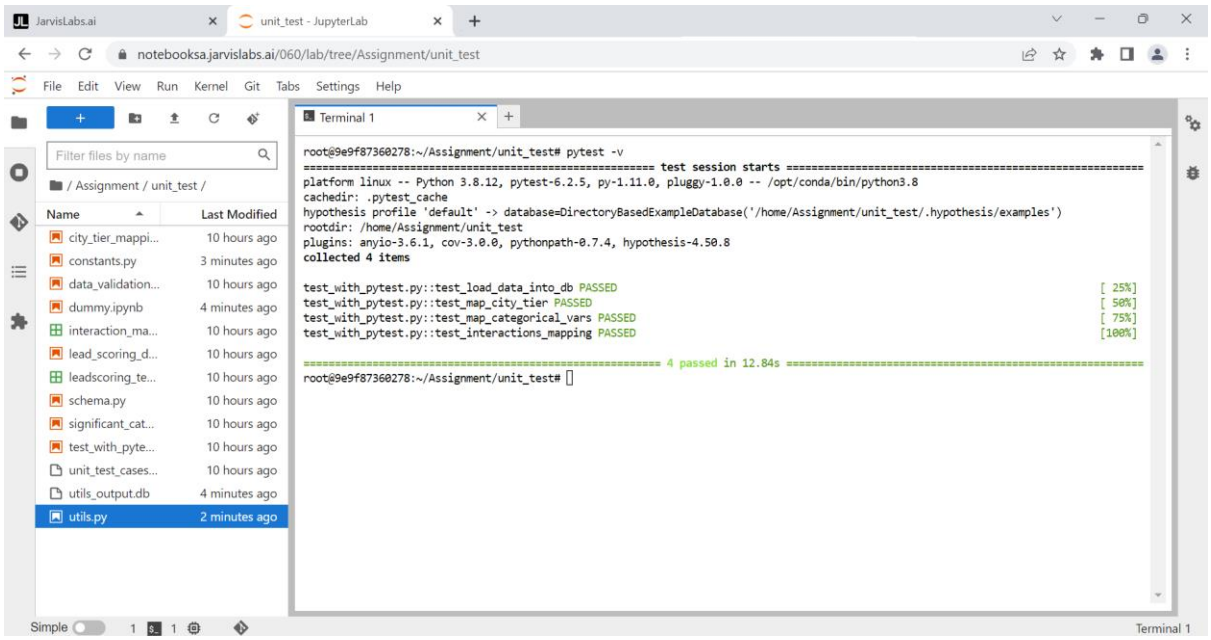

Figure 21: Pytest Test Results