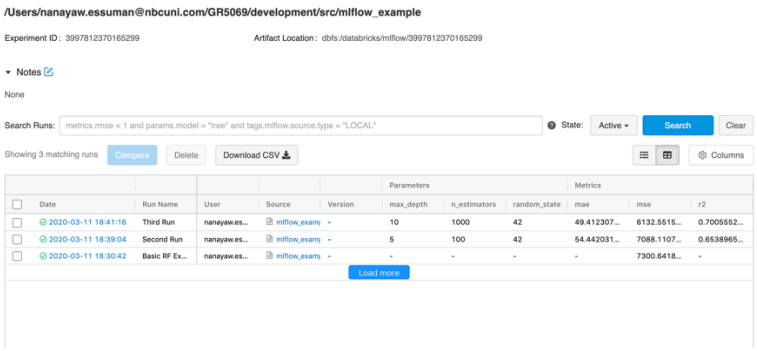
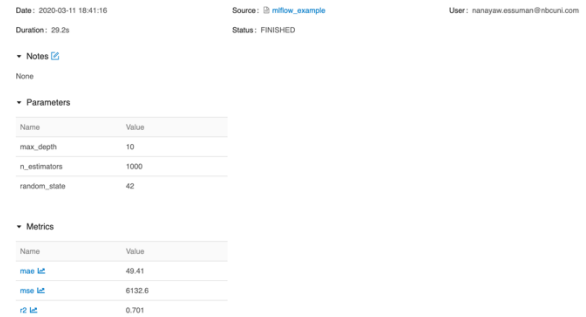
Assignment 3 – Building a Model and Model Tracking

For this assignment, we'd like you to use the F1 Datasets we have been using for the class and build any ML model of your choice and track the model for each run.

1. Select any of the F1 datasets in AWS S3 to build your model. You are allowed to join multiple datasets.
2. Build any model of your choice
3. Log the parameters used in the model in each run
4. Log the model
5. Log every possible metric from the model
6. Log at least two artifacts (plots, or csv files)
7. Track your MLFlow experiment and run at least 10 with different parameters
8. Select your best model run and explain why
9. Take a screenshot of your MLFlow Homepage as part of your assignment submission



10. Take a screenshot of your detailed run page



Screenshot 1

Screenshot 2

11. Push your model code to GitHub and add the screenshots to the resources folder in your GitHub repo

Resources to help

https://github.com/marco-morales/QMSS-GR5069_Spring2020/blob/master/week_07/resources/mlflow_example.py - My Example in Class

https://pages.databricks.com/rs/094-YMS-629/images/financial-fraud-detection-decision-tree.html?_ga=2.108783938.1703692841.1584145486-301162589.1569349546 – Real Life Example

https://docs.databricks.com/_static/notebooks/mlflow/mlflow-quick-start-python.html - Python Getting Started Notebook

https://docs.databricks.com/_static/notebooks/mlflow/mlflow-quick-start-r.html - R Getting Started Notebook

https://docs.databricks.com/applications/mlflow/index.html - MLFlow Documentation