

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

/Users/nanayaw.essuman@nbcuni.com/GR5069/development/src/mlflow_example

Experiment ID: 3997812370165299 Artifact Location: dbfs/databricks/mlflow/3997812370165299

▼ Notes [🔗](#)

None

Search Runs: State: Active ▼ Search Clear

Showing 3 matching runs Compare Delete Download CSV [📄](#) Columns

	Date	Run Name	User	Source	Version	Parameters			Metrics		
						max_depth	n_estimators	random_state	mse	mse	r2
<input type="checkbox"/>	2020-03-11 18:41:16	Third Run	nanayaw.ess...	mlflow_exam	-	10	1000	42	49.412307...	6132.5515...	0.7005562...
<input type="checkbox"/>	2020-03-11 18:39:04	Second Run	nanayaw.ess...	mlflow_exam	-	5	100	42	54.442051...	7088.1107...	0.6538965...
<input type="checkbox"/>	2020-03-11 18:30:42	Basic RF Ex...	nanayaw.ess...	mlflow_exam	-	-	-	-	7300.6418...	-	-

[Load more](#)

10. Take a screenshot of your detailed run page

Date: 2020-03-11 18:41:16 Source: [mlflow_example](#) User: nanayaw.essuman@nbcuni.com

Duration: 29.2s Status: FINISHED

▼ Notes [🔗](#)

None

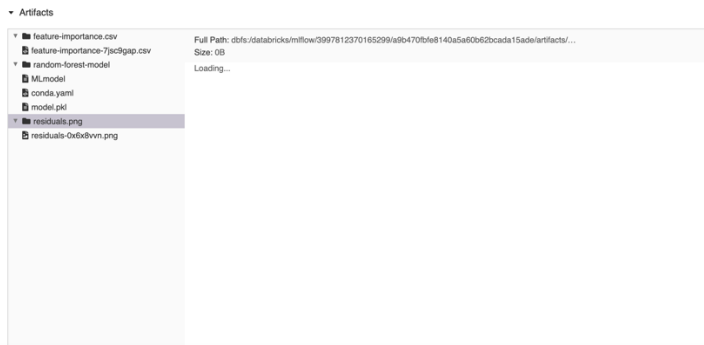
▼ Parameters

Name	Value
max_depth	10
n_estimators	1000
random_state	42

▼ Metrics

Name	Value
mse 📄	49.41
mse 📄	6132.6
r2 📄	0.701

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