



An End-to-End Data Science Project

Deena Gergis
Lead Data Scientist @ Bayer



Workshop overview:

Session 1 Preparation 10.04.2022

Start with the business problem, find data source, preprocess data, set up team process and tech

Session 2 Analytics 17.04.2022

Analyze and understand your data. Gain insights and prepare for the predictive modeling

Session 3 Machine learning 08.05.2022

Build and evaluate prediction model(s), use Mlflow to keep track of the various experiments

Session 4 Production 15.05.2022

Create prediction functions and production class, develop an API, create a dashboard that the user will access and call the API

What you will do:

- **Form a team of 3 members**
- **During the sessions:** You will get tasks to be done
- **After the sessions:**
 - You will complete the whole covered phases
 - Dig deeper into the various technologies discussed



Part 1: Modelling Training

- 1. Clean your data*
- 2. Decide on your modelling strategy*
- 3. Decide about the evaluation metric*
- 4. Train a baseline model*
- 5. Train more sophisticated models*
- 6. Decide which model will be used*



MLflow

Summary

Track with Mlflow

```
In [9]: # Initialize client and experiment
client = MlflowClient()
mlflow.set_experiment(EXPERIMENT_NAME)
exp = client.get_experiment_by_name(EXPERIMENT_NAME)

In [10]: # Start a new run and track
with mlflow.start_run(experiment_id=exp.experiment_id):
    mlflow.log_param("pca_var", PCA_VAR)      # Track model parameter
    mlflow.log_metric("MSE", mse_test)        # Track error value
    mlflow.log_artifact(EXPORT_MODEL_PATH)    # Track exported model
```

Retrieve experiment

```
In [3]: # Initialize client
client = MlflowClient()

# Get experiment
exp = client.get_experiment_by_name(EXPERIMENT_NAME)
```

```
In [4]: # Get all runs
runs = mlflow.search_runs([exp.experiment_id])
runs
```

	run_id	experiment_id	status	artifact_uri	metrics.MSE	params.pca_var
0	41e1628508fc4a2f83651cecede6d8a	1	FINISHED	file:///home/deena_gergis/mlflow_illustration/...	57.884313	1
1	d39dec7e183450d87f679fb044c6e66	1	FINISHED	file:///home/deena_gergis/mlflow_illustration/...	31.675892	0.95
2	77c97ff7b2fe46d9becba4c230dd3193	1	FINISHED	file:///home/deena_gergis/mlflow_illustration/...	31.831237	0.8
3	f11b91bf1c7441c5bacc7e73adfbbe59	1	FINISHED	file:///home/deena_gergis/mlflow_illustration/...	57.884313	0.3

Tutorial: <https://www.linkedin.com/pulse/mlflow-better-way-track-your-models-deena-gergis/>

Repo: https://github.com/Deena-Gergis/mlflow_tracking



Assignment:

*Train, evaluate and track
your models*



Questions?