An End-to-End Data Science Project

Deena Gergis
Lead Data Scientist @ Bayer

#ACCELERATE

Workshop overview:

Session 1
Prep & Analytics
12.09.2021

Start with the business problem, set the foundation up, find data source, preprocess Start the descriptive analytics pipeline)

Session 2
Machine learning
19.09.2021

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

Session 3
Deployment in Prod
03.09.2021

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:

- **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

i.e.: No Spoon-feeding :-)

#ACCELERATE

& let’s get started
and pick-up where we’ve left off

Part 1
Machine learning hands-on

Part 2
Productionizing the model

#ACCELERATE

Part 1
Machine learning hands-on

#ACCELERATE

MLflow recap

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
```

Out[4]:

	run_id	experiment_id	status	artifact_uri	metrics.MSE	params.pca_var
0	41e1628508fc4a2f83651ceceede6d8a	1	FINISHED	file:///home/deena_gergis/mlflow_illustartion/...	57.884313	1
1	d39decb7e183450d87f679fb044c6e66	1	FINISHED	file:///home/deena_gergis/mlflow_illustartion/...	31.675892	0.95
2	77c97ff7b2fe46d9becba4c230dd3193	1	FINISHED	file:///home/deena_gergis/mlflow_illustartion/...	31.831237	0.8
3	f11b91bf1c7441c5bacc7e73adfbbe59	1	FINISHED	file:///home/deena_gergis/mlflow_illustartion/...	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*

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

#ACCELERATE

Part 2: Production

1. Develop prediction functions
2. Refactor your code to clean script
3. Create an API on top
4. Integrate the API with the product

Assignment:

Create a product using Dash or Streamlit and call the API from within

Questions?

#ACCELERATE