Tracking Experiments with MLflow

Matt Dancho

# Overview

|  |
| --- |
| MLFlow by Databricks – for machine learning lifecycle management (experimentation to production) |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
| UMAP is like PCA (dimensionality reduction) <https://umap-learn.readthedocs.io/en/latest/> |
|  |
| There are sevaral features (atributes) for each dot (company). These are computed from the income and balance sheet.  The model is not embedded in the app. Every time you click the submit button, it generates a post request that runs a model on a separate server. |
| MLOps is similar to DevOps |
| <https://link.medium.com/SipcEZpOk8> & <https://aws.amazon.com/batch/> |
| Concerned with reproducibility, docker, servers, APIs, etc. |
|  |
| We will focus on tracking here. |
|  |
|  |
| Less software on users machine. |
| 2 sessions that talk to each other. |

# Code

|  |
| --- |
| Installing and Checking MLFlow |
| Bankrupcy Classifier  Carrier – for connecting models to mlflow  fs - filesystem |
|  |
| Attr1-xx are features |
| Dictionary of what the attributes are. |
|  |
|  |
| Attr 37 as many missing values |
|  |
|  |
|  |
| Correlation Funnel: It shows which feature has the highest correlation to the output class |
|  |
|  |
| You can compate to Correlation Funnel as well, but trust the ML model more (for model accuracy). Correlation Funnel is better for talking points to execs. |
|  |
|  |
| MLFlow URI 🡪 top level folder 🡪 Then Setup an experiment 🡪 Check that it is there |
| View UI |
| As we add models, it will be added here. |
|  |
| Create a model, save the model, then make a prediction. |
| Calculate metrics (CV) |
| Log the parameters, model (as a predictor), metrics, artifacts (path to model). |
| Now we have those models and entries here. |
| Artifacts has the model which can be used to deploy. |
| Serve this model on an API. |
|  |
|  |
|  |
| Paste your code    Add atributes and click “Execute” 🡪 It will return prediction. You can also connect a Shiny App and call this API (better than adding attributes manually. |
| Full Stack Data Science |