

DP100 - Synthèse modules 1-4 : base

Notions essentielles

workspace

espace de travail qui contient les expériences, les modèles, les datasets, etc..

experiments

permet de réaliser des run qui sont consultables dans le studio

script

permet de rassembler la logique à exécuter dans un fichier (plutôt que dans le notebook).
On peut ensuite préciser des paramètres à utiliser comme entrées de ce script.

estimator

permet d'exécuter des experiments dans un container avec un environnement spécifié (soit on indique les dépendances dont on a besoin (à l'aide d'un environnement), soit on utilise un estimator préconfiguré).

model

les modèles habituels de ML (sous format.pkl), on peut les enregistrer dans un workspace.
Les modèles enregistrés sont versionnés et possèdent des propriétés qui permettent de choisir lequel on veut utiliser.

mlflow

librairies opensource qui donnent des fonctions pour logger des informations sur des expériences de machine learning. S'interface avec azure.

datastore

magasin de données qui permet de conserver les informations de connexion. Par défaut les workspaces contiennent deux datstores : filestore et [blobstore](#) (images, vidéos, textes, ...)

[dataset](#)

un sous-ensemble du datastore qui peut être de deux types : tabular dataset ou filedataset.

[dataref](#)

un chemin dans le datastore. Cette notion n'est plus recommandée, il faut utiliser les datasets.

compute target

permet d'indiquer la cible sur laquelle sera exécuté un calcul. On pourra voir [ComputeInstance](#), [AmlCluster](#) et AKS.

Questions