

Cahier thématique 8

Configuration de projet ML CI/CD avec l’outil DVC

**Manipulation 1**: Création d’un projet CI avec Git et DVC

Change control

Version 1 : Aout 2020

**Préparé par :** Hafed Benteftifa  
  
© Hafed Benteftifa 2014-2020

Ce document ne peut être utilisé dans le cadre d’une formation, publication papier, site internet ou tout support sans mon accord express.

Aucune reproduction, même partielle, ne peut être faite de ce document et de l'ensemble de son contenu : textes, images, etc. sans mon autorisation express. Pour toutes informations, communiquer avec moi sur [info@degenio.com](mailto:info@degenio.com).

**Manipulation 1:** Création d’un projet CI avec Git et DVC

**Objectif**

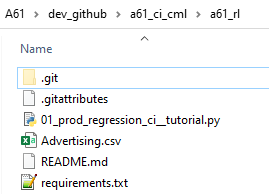
Procéder à la mise en da place du projet Dev ML avec l’outil Git et intégration avec DVC

**Préliminaire**

* Accès à GitHub

**Démarche**

1. On commence par créer un repository local . Procéder à sa création avec GitHub Desktop, soit



1. Au niveau de la racine du projet, effectuer un init avec dvc, soit :



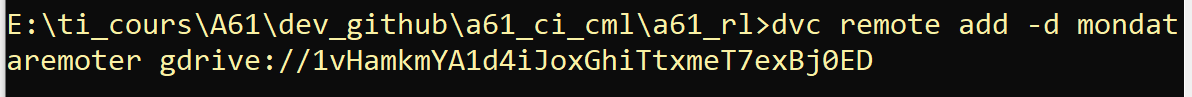
1. Sur google drive créer un repertoire /data/versionage



Prendre le lien google url et faire une copie

1. Dans cmd

Dvc remote add -d mondataremote gdrive://url\_code

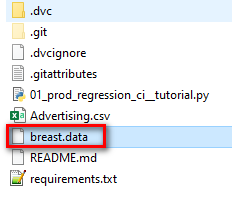


Git add

Git commit -m “premier commit »

Git push

1. Obtenir le dataset breast.data et le copier dans le repertoire courant, soit :



1. Ajouter le dataset en utilisant DVC, soit :

dvc add breast.data

1. Lors de l’ajout de breast.data au niveau de dvc, le fichier breast.data.dvc est créé. Avant de commencer à tracker le changement sur le fichier breast.data, on utilise la commande suivante afin de ne pas suivre le changement au niveau du fichier breast.data.dvc sur GitHub, soit:

Git add breast.data.dvc .gitignore

1. Effectuer un commit sur Git:

Git commit -m “debut du tracking du dataset”

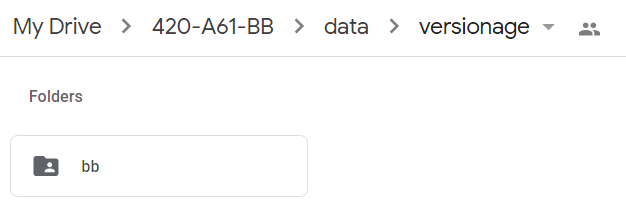
Git push

1. On fait maintenant la partie DVC

Dvc push

Note : Si un message apparait concernant une authentification, allez sur chrome et utiliser le lien qui vous est fourni

1. À la fin, un code vous est fourni que vous allez utilisez dans CMD
2. Une fois terminé, vérifier dans votre google drive que le fichier a bien été chargé



1. Ouvrir une autre session et cloner votre projet github.

Git clone https :url du projet github

1. Procéder à la récupération du dataset avec la commande :

Dvc pull breast.data

Références :

Dvc.org

<https://dvc.org/doc/user-guide/setup-google-drive-remote>

<https://dvc.org/doc/command-reference/remote/add>

<https://www.youtube.com/watch?v=kZKAuShWF0s&t=33s&ab_channel=DVCorg>