

Documentation technique

Le projet est séparé en deux parties, d'abord nous avons dû nettoyer les données et entraîner les modèles de classification et de régression. Puis la partie déploiement et utilisation des modèles.

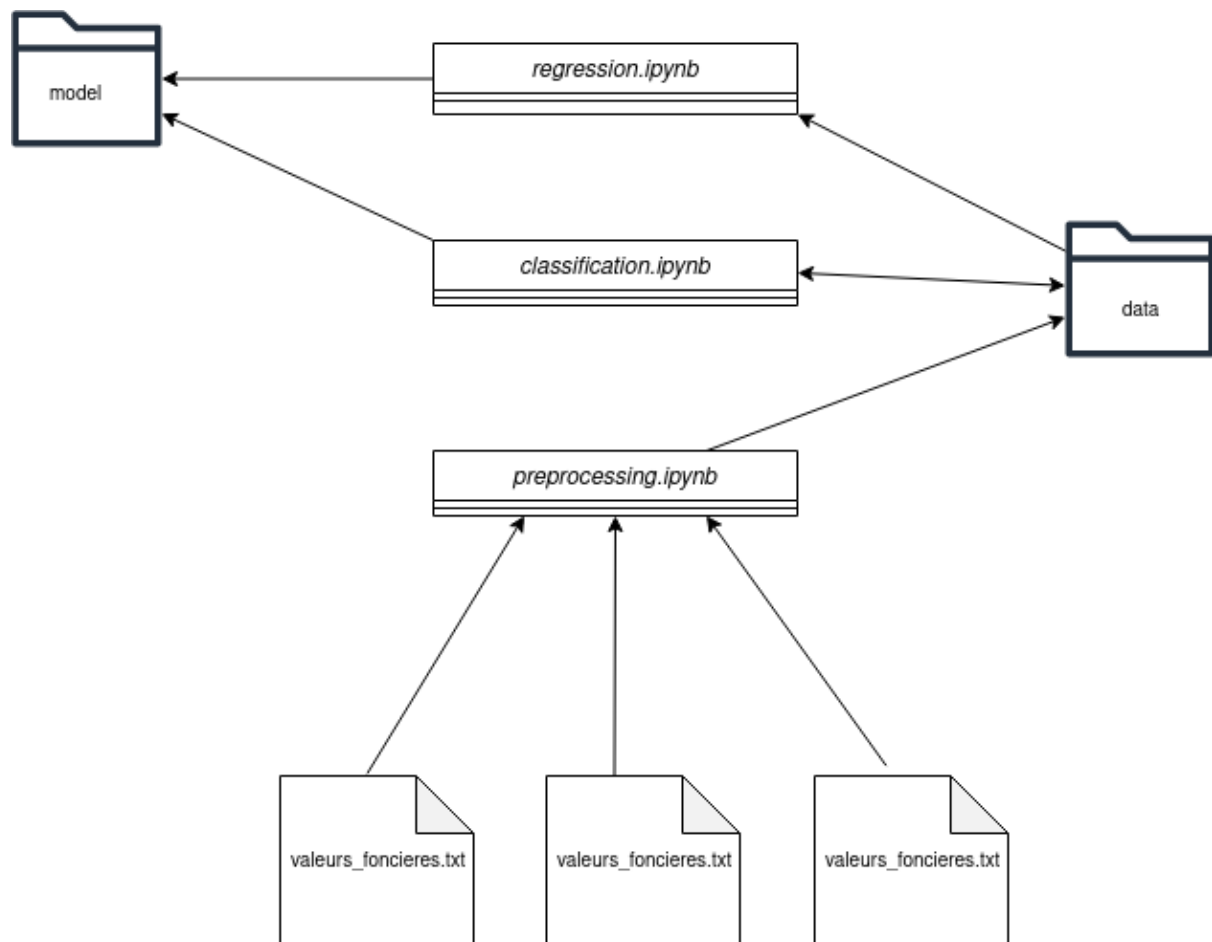
Entraînement du modèle :

1

Généralisation / Déploiement :

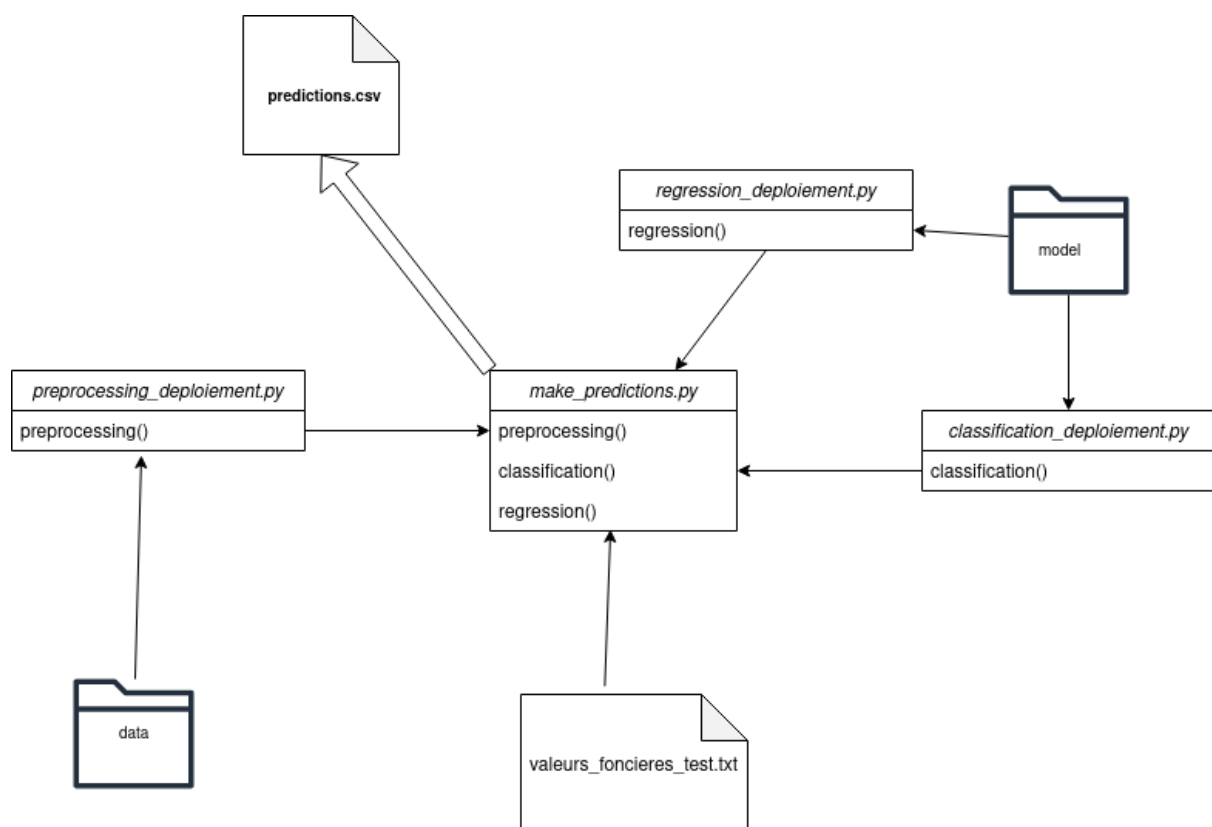
2

Entraînement du modèle :



Le notebook preprocessing explore et nettoie les données, tout en ajoutant des variables supplémentaires au jeu de données. Puis le fichier de classification prend ces données nettoyées, crée un modèle de prédiction du code type local des individus lorsqu'il est inexistant et sauvegarde ces données dans le dossier 'model'. Le modèle est ensuite sauvegardé dans un dossier pour le déploiement par la suite. Enfin le fichier régression prend les données finales, crée et entraîne un modèle de régression pour chacun des code type local. Les modèles sont alors aussi sauvegardés dans le dossier 'model', pour être utilisés par la suite.

Généralisation / Déploiement :



Pour être utilisé notre projet comporte un ensemble de scripts et de fonctions python, l'utilisateur peut donner en entrée un ensemble de biens à prédire. Le programme s'occupe alors de nettoyer les données et d'appliquer les modèles de machine learning générés lors de la phase d'entraînement.

Le fichier 'make_predictions.py' retourne enfin un fichier de prédiction, avec les colonnes 'ID' et 'TARGET', indiquant pour chaque biens sa valeur foncière prédite.