
Guide d'exécution des fichiers de prétraitement et d'exécution du modèle

(Sur Windows)

Introduction

Cette documentation vise à fournir des instructions détaillées sur l'exécution des fichiers de prétraitement (pre_processing.py) et d'exécution du modèle (run_model.py) sur Windows. Dans un premier temps, le fichier de prétraitement est un code python permettant d'appliquer la fonction de prétraitement aux sets données 'test.csv' et 'train.csv' (voir concours Kaggle : [Titanic - Machine Learning from Disaster | Kaggle](#)) . Une fois les données prétraitées, le fichier d'exécution du modèle (run-model.py) permettra de faire tourner le modèle sur les données d'entraînement afin de tester son efficacité. De plus, un fichier de soumission sera créé. Ce fichier 'soumission.csv' donnera les prédictions du modèle sur les données test. Le format du fichier est celui demandé par Kaggle et permet donc d'obtenir un score.

Prérequis :

Avant de commencer, assurez-vous d'avoir les éléments suivants :

Prérequis hardware :

1. Un ordinateur avec un processeur compatible avec Python 3.10.
2. Un espace de stockage suffisant pour les fichiers de données prétraitées et les résultats de l'exécution du modèle.

Prérequis software :

1. Windows 10 version 17763.0 ou version supérieure.
 2. Python 3.10 ou supérieur, lien d'installation Windows store : [Python 3.10](#).
 3. Bibliothèques 'panda' et 'scikitlearn' (voir 'requirements.txt', Vous pouvez installer toutes les bibliothèques en utilisant la commande **pip install requirements.txt**)
-

Récupération des fichiers :

- L'ensemble des fichiers nécessaires sont dans le répertoire 'trunk'. Téléchargez le répertoire et placez-le dans le répertoire de votre choix sur votre système..


Exécution du fichier de prétraitement (pre_processing.py) :

1. Ouvrez l'invite de commande de Windows en appuyant sur la touche Windows + R, puis tapez "cmd" et appuyez sur Entrée. Naviguez jusqu'au répertoire où se trouve le fichier pre-processing.py en utilisant la commande "cd" (Change Directory). Par exemple, si votre fichier est dans le dossier C:\trunk, vous pouvez utiliser la commande suivante pour naviguer vers ce dossier : **cd C:\trunk**
2. Assurez-vous que le module pandas est installé en utilisant la commande suivante : **pip show pandas**
Si le module n'est pas installé, vous pouvez l'installer en utilisant la commande suivante : **pip install pandas**
3. Une fois que vous êtes dans le bon répertoire et que pandas est installé, vous pouvez exécuter le fichier pre_processing.py en utilisant la commande suivante : **python pre_processing.py**
4. Cela permettra de créer 3 nouveaux fichiers dans le répertoire 'Data' : 'test_preprocessed.csv', 'bonus_data_preprocessed.csv' et 'train_preprocessed.csv'.

Exécution du fichier d'exécution (run_model.py)

1. Ouvrez l'invite de commande de Windows en appuyant sur la touche Windows + R, puis tapez "cmd" et appuyez sur Entrée. Naviguez jusqu'au répertoire où se trouve le fichier run_model.py en utilisant la commande "cd" (Change Directory). Par exemple, si votre fichier est dans le dossier C:\trunk, vous pouvez utiliser la commande suivante pour naviguer vers ce dossier : **cd C:\trunk**
2. Assurez-vous que les modules scikit-learn et pandas sont installés en utilisant les commandes suivantes :
 - **pip show scikit-learn**
 - **pip show pandas**

Si les modules ne sont pas installés, vous pouvez les installer en utilisant la commande suivante : **pip install scikit-learn pandas**

- 
3. Une fois que vous êtes dans le bon répertoire et que les modules sont installés, vous pouvez exécuter le fichier `run_model.py` en utilisant la commande suivante :
`python run_model.py`
 4. Cela permettra de créer 3 nouveaux fichiers dans le répertoire 'Resultats' :
'precision_model.txt' , 'bonus_data_prediction.txt' et 'soumission.csv'.