

PYTHON POUR DATA-SCIENCES

CODE STAGE : BD060

OBJECTIFS

Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python, Savoir extraire des données d'un fichier et les manipuler, Mettre en place un modèle d'apprentissage simple

DURÉE

3 jours

PUBLIC

Les développeurs en Python.

PRÉ-REQUIS

Maîtrise de la programmation Python.

PROGRAMME

Positionnement Python

Besoins des data-scientist: calculs, analyse d'images, machine learning, interface avec les bases de données

Apports de python: grande variété d'outils, expertise dans le domaine du calcul scientifique

Tour d'horizon des outils: pandas, agate, bokeh, scikit-learn, pybrain, tensorflow, keras, mxnet, caffe

Calculs et graphiques

NumPy : Base du calcul sur des tableaux

SciPy : Scientific Tools for Python, couche scientifique

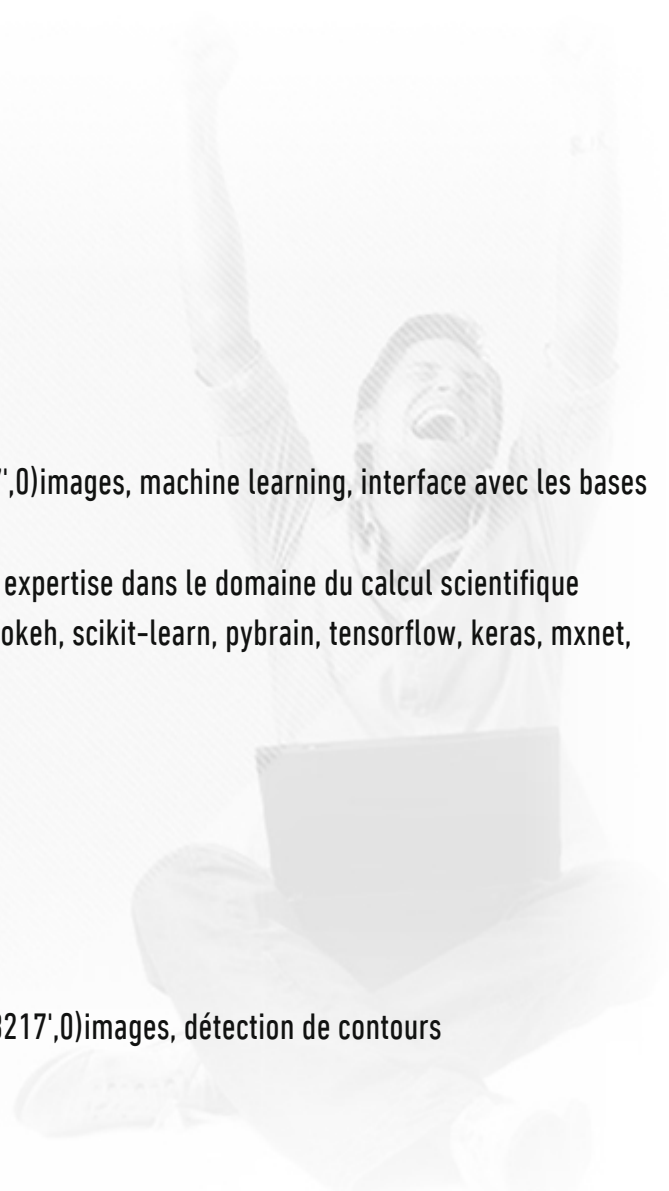
Manipulation de tableaux, fonctions mathématiques.

Représentation graphique avec basemap et matplotlib.

Mise en oeuvre de SciPy/NumPy : manipulation d'images, détection de contours

Manipulation de données relationnelles

Pandas : manipulation de tables de données



Tableaux avec Pandas: indexation, opérations, algèbre relationnelle

Stockage dans des fichiers: CSV, h5py, netCDF

Comparaison et performances Pandas / NumPy

Machine learning et deep learning

Présentation de TensorFlow,scikit-learn,keras,mxnet,caffe

TensorFlow:principe de fonctionnement, plate-formes supportées, distribution,

APIs fournies en standard, modèles dcode2utf('8217',0)apprentissage

Projet scikit-learn:classification, régression, validation de modèles prédictifs.

Démonstrations avec les modèles fournis par scikit-learn

Positionnement et comparaison avec Keras, mxnet, caffe

