

Audit « Python pour la Data Science »

Client : L'ORÉAL

Formateurs : Romain Rincé (Formation en anglais) et François Giraud

Contact audit :

Lieu de la formation : A distance ou présentiel sur demande

Nombre de stagiaires : 6x2

Durée de la formation : 5 jour(s) -

Les connaissances préalables et les objectifs des stagiaires sont en adéquation avec le contenu et la durée de la formation : ☒ Oui ☐ Non

Pré-requis techniques, puissance machine : Pas de besoin technique spécifiques. Les participants doivent pouvoir accéder sans blocage à <https://colab.research.google.com/> et disposer d'un compte google chacun (temporaire ou non)

Pré-requis d'installation logicielle : Pas de prérequis

Version du produit :

Support formateur : ☒ Oui ☐ Non

Si non, support recommandé :

Profil des stagiaires : ☐ Débutant ☐ Intermédiaire ☒ Avancé

Attentes des stagiaires : Uniformisation des pratiques de codes en python pour le traitement des données et du développement d'algorithmes machine learning/Deep learning ainsi que la capacité de déployer leurs modèles pour la mise en production. Pratique du code sur les bibliothèques suivantes : Numpy, Pandas, Matplotlib, Scikit-learn, XGBoost, Tensorflow/Keras

Complément d'information : Détails des modèles et cas d'utilisations modernes des modèles sur le traitement d'image au-delà de la classification tels que : offuscation, détection d'erreur et segmentation. Une des formations sera dispensée en anglais (2 jours de préparation sont à prévoir pour la partie XGBoost et l'adaptation des supports en anglais)

Objectif de la formation :

Cette formation vous permettra d'utiliser les librairies les plus populaires dédiées à l'analyse de données, la data science et l'apprentissage automatique.

Découvrez les librairies classiques de la data science telles que numpy et pandas, de la visualisation avec matplotlib, du Machine learning avec Scikit-learn et XGBoost et finalement du machine learning avec Tensorflow/ Keras.

Plan de formation : *(merci d'apporter toute modification utile au présent plan de cours en ajoutant ou barrant tout chapitre utile ou exclus du projet de formation. Vous pouvez également copier-coller votre propre plan de cours si vous jugez qu'il est plus en adéquation avec les attentes et les objectifs des stagiaires).*

Outils Python pour l'analyse et le traitement de données

- Calcul scientifique avec NumPy
- Traitement de données tabulaires avec Pandas
- Visualisation des données avec Matplotlib et seaborn

Machine learning avec scikit-learn et XGBoost

Utilisation des modèles populaires implémentés sur scikit-learn et XGBoost :

- Régression (linéaire, arbres et forêt)
- Classification (régression logistique, SVM, forêt)
- Clustering
- Réduction de dimension (Rappel sur l'utilité de la réduction de dimension,

utilisation de PCA et démonstration de méthodes non linéaire moderne)

- Techniques d'arbres boostées avec XGBoost

Deep Learning avec Tensorflow et Keras

- Utilisation de la librairie Keras pour la mise en place de réseaux de neurones pour classification et régression : Réseaux classiques (MLP) et traitement de l'image (CNN)
- Utilisation de Keras et Tensorflow pour le développement de modèles personnalisés via les outils de spécialisation de modèle :
 - Création de couches personnalisées
 - Utiliser le GradientTape pour l'implémentation de modèles récents

Modèles avancés pour le traitement d'image

- Techniques d'occlusion pour la détection d'erreur et l'interprétation des modèles
- Segmentation d'image et détection de contours
- Modèle transformeur pour le traitement d'image

Mise en production

Développement et déploiement d'une API de ML

LE PLAN DE COURS SUIVANT A-T-IL ETE MODIFIE ?

☒ Oui ☐ Non