

Industrialisation d'un projet de Data Science

Adopter les bonnes pratiques de développement pour la Data science

DESCRIPTION

La maturité du marché sur les sujets de la Data Science ne cesse de grandir chaque jour. C'est pourquoi, il nous semble à présent important d'être capable de déployer les POCs (Proof of Concept) les plus pertinents en production, et enfin voir les applications de data science apporter de la valeur aux entreprises.

Durant cette formation, vous profiterez de la riche expérience de nos consultants pour découvrir la Data Science industrialisée, et vous pratiquerez les meilleures méthodes pour faire passer votre projet du stade de la recherche à celui de l'exploitation en évitant les pièges de l'effet « big bang ».

OBJECTIFS PEDAGOGIQUES

- Savoir emmener son modèle en production
- Apprendre à gérer les nouvelles contraintes
- Minimiser le coût de transfert de l'exploration à la production
- Découvrir les concepts de Software Craftmanship appliqués à la Data Science

PUBLIC CIBLE

- Analyste
- Statisticien
- Développeur

PRE-REQUIS

- Connaissance des bases de la Data Science (modèles, biais, variance, etc.)
- Connaissance des bibliothèques de manipulation de donnée en python (pandas, numpy, etc.)

Stage pratique

Data Science

Code :

DSIND

Durée :

2 jour(s) (14,00 heures)

Exposés : **10 %**

Cas pratiques : **80 %**

Echanges d'expérience : **10 %**

Inter-entreprises :

Prochaines sessions disponibles [sur notre site web](#).

Tarif : 1 850,00 € HT / participant

Intra-entreprise :

Tarifs et dates sur demande.

- Connaissance de la ligne de commande linux (bash par exemple)

METHODE PEDAGOGIQUE

Une formation pratique durant laquelle les participants partent d'un code d'exploration préparé par le formateur et qu'ils vont amener en production, étape par étape, en démontrant que la transition doit se faire de manière la plus fluide possible.

PROFIL DES INTERVENANTS

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

PROGRAMME PEDAGOGIQUE DETAILLE

Jour 1

INTRODUCTION

Rôle du data scientist

Convictions

INSTALLATION DES ENVIRONNEMENTS

Jupyter Notebook

PyCharm

Environnement anaconda

CLEAN CODE ET NOTEBOOK

Nettoyer son notebook

Noms des variables

Utilisation de fichiers de configuration

Programmation fonctionnelle

Immutabilité

Créer des fonctions

Mise en pratique : "Appliquer les techniques de Clean code à un notebook"

TESTER SON CODE

Découverte du TDD

Initiation à unittest

- Les classes de test
- Écrire son premier test
- Les méthode setup et teardown

Intégration à setuptools

Mise en pratique : "Écrire ses premiers tests"

Jour 2

RAPPELS ET RÉVISION DU JOUR 1

DOCUMENTER SON PROJET

Découverte de Sphinx

Découverte de sphinx-quickstart

Mise en pratique : "Documenter son projet"

GÉRER LES VERSIONS

Introduction à git

Gérer les versions de code

Utilisation des tags

Stratégies de versionning

Gérer des datasets et des modèles

RENDRE SON CODE DÉPLOYABLE

Définition du packaging

Introduction à setuptools

Installer son code en local

Gestion des dépendances

Règles simples de démarrage

Mise en pratique : "Créer et installer son package"

CONCLUSION

Synthèse des points abordés

Partage sur la formation

Questions/Réponses additionnelles

