

Industrialisation avancée d'un projet de Data Science

Orchestration, déploiement et monitoring d'un système de Data Science

DESCRIPTION

Une fois que le code du projet de Data Science a été industrialisé, il reste un certain nombre d'étapes pour industrialiser le système au complet. Il s'agira notamment de mettre en place un mécanisme d'orchestration, mais aussi d'être capable de déployer régulièrement, et de suivre le comportement du système.

Tout au long de cette formation avancée, vous profiterez de la riche expérience de nos consultants-formateurs pour découvrir les systèmes de Data Science industrialisés, et vous pratiquerez les meilleures méthodes pour faire passer votre projet en production.

OBJECTIFS PEDAGOGIQUES

- Découvrir le découpage d'un projet de Machine Learning en différentes briques et les organiser
- Savoir construire l'architecture applicative d'un projet Data Science
- Déployer son application de Machine Learning en production et interagir avec
- Faire consommer des données chaque minute à une application de Machine Learning
- Mesurer les performances de son application de Machine Learning

PUBLIC CIBLE

- Data Scientist
- Data Engineer
- Machine Learning Engineer
- Développeur d'applications de Data Science
- Analyste
- Statisticien

Stage pratique
Data Science

Code :
DSIN2

Durée :
3 jour(s) (21,00 heures)

Exposés : **40 %**
Cas pratiques : **50 %**
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.

PRE-REQUIS

- Connaissances de base en programmation et en scripting
- Avoir suivi la formation "Industrialisation de la Data Science" (DSIND) ou savoir tester une application python
- Avoir suivi la formation "Fondamentaux de la Data Science" (DSFDX) ou savoir modéliser et développer un problème de Data Science simple

METHODE PEDAGOGIQUE

Une formation alternant théorie et pratique durant laquelle les participants partent d'un modèle de machine learning construit par le formateur qu'ils vont amener en production, étape par étape, en découvrant l'ensemble des briques techniques à mettre en place.

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 À LA CI / CD

- Les différents environnements nécessaires

- L'intérêt de l'intégration continue et du déploiement continu
- Les outils de CI / CD existants

PYRAMIDE DE TESTS

- Description des différents types de tests
- Focus sur les tests d'intégrations et fonctionnels
- Présentation de framework de tests

MISE EN PRATIQUE : "METTRE EN PLACE UN OUTIL D'INTÉGRATION CONTINUE"

CLOUD & INFRASTRUCTURE AS CODE

- Introduction au cloud et aux différents providers
- Choix d'un cloud provider pour les travaux pratiques
- Infrastructure as code, principe et découverte de Docker, Ansible et Terraform

ORCHESTRATION

- Le besoin d'orchestration dans un système de Data Science
- Présentation de différents orchestrateurs
- Mise en pratique : "Passer d'un script Python à une série de tâches orchestrées avec Airflow"

Jour 2

ARTEFACTS

- Patterns d'architecture technique pour favoriser la mise en production et la maintenabilité du projet
- Identification des artefacts (modèles, données, code...)
- Cycles de vie des artefacts
- Mise en pratique : mise en place de MLFlow

DÉPLOIEMENT

- La démarche
- Le couplage / découplage
- Les stratégies de déploiement

EXPOSITION

- Méthodes pour exposer un modèle de Data Science
- Mise en pratique : “Interagir avec un modèle existant”

Jour 3

MISE EN PRATIQUE : “PASSER D’UN MODÈLE ONE-SHOT À UN MODÈLE QUI PRÉDIT TOUTES LES MINUTES”

- A partir d’un flux de données disponibles, mettre en place les outils nécessaires pour faire une prédiction par minute

FEEDBACK LOOPS

- L’intérêt des boucles de feedback
- Identifier les différentes boucles de feedbacks
- Mettre en place les différentes boucles de feedbacks

MONITORING

- L’enjeu du monitoring
- Les indicateurs à suivre
- Les pratiques pour mettre en place un monitoring utile tout en respectant les contraintes de SLA

AMÉLIORATION CONTINUE

- Choisir le bon moment pour réentraîner un modèle
- Les stratégies à mettre en place pour s’assurer que le modèle en production est toujours pertinent
- MLOps
- Qualité de données pour un projet de *Machine Learning*

BILAN ET CLÔTURE

- Partage et retour d’expérience
- Questions / réponses
- Évaluation de la session

