

Cadrage et pilotage d'un projet de Data Science

DESCRIPTION

La Data Science, en tant que discipline scientifique et mathématique, est aujourd'hui bien connue de ceux qui la pratiquent. Mais savez-vous identifier les pièges et les embûches méthodologiques spécifiques aux projets de Data Science ?

Forts de nos nombreux retours d'expérience et des connaissances pratiques de nos consultants, nous vous convions à découvrir avec nous les enjeux et les approches qui permettront à vos projets de se dérouler en toute sérénité, du cadrage à la phase de pilotage de production.

OBJECTIFS PÉDAGOGIQUES

Savoir mener une phase de cadrage d'un projet de Data Science
Choisir les mesures de performance en accord avec un objectif métier
Piloter un projet de Data Science en exploration et production

PUBLIC CIBLE

Analyste
Statisticien
Développeur

PRÉ-REQUIS

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

MÉTHODE PÉDAGOGIQUE

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique du formateur, complétés de travaux pratiques et de mises en situation.

PROFILS 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.

POUR ALLER PLUS LOIN :

- Toutes nos formations Data Science
- Formation « Data Science niveau avancé »

Stage pratique en présentiel DATA SCIENCE

Code :
DSGDP

Durée :
2 jours (14 heures)

Exposés :
40%

Cas pratiques :
40%

Échanges d'expérience :
20%

Sessions à venir :

23 - 24 sept. 2019
Paris / 1 850 eur

Tarif & dates intra :

Sur demande

- Formation « Industrialisation d'un projet de Data Science »
- Ouvrage « Data Science : fondamentaux et études de cas » - Eric Biernat, Michel Lutz (Eyrolles)
- Ouvrage « The Elements of Statistical Learning: Data Mining, Inference, and Prediction » - Hastie, Tibshirani, Friedman
- Livre blanc "Hadoop, feuille de route"
- Toutes nos formations Hadoop
- Toutes nos formations Hadoop Cloudera

Programme pédagogique détaillé par journée

Jour 1

RAPPELS SUR LA DATA SCIENCE

- Qu'est-ce-que la Data Science ?
- Qu'est-ce qu'un modèle ?

LA PHASE DE CADRAGE

- Cadrage d'un sujet supervisé
 - Découverte des données
 - Compréhension des enjeux métier
 - Choix de l'approche
- Visualiser pour comprendre
- Echantillonnage
- La temporalité dans les données

PILOTER SON PROJET PAR LA MESURE

- L'importance du test
 - Lien avec les tests d'intégration
- Le choix des métriques à optimiser
 - Les métriques de régression
 - Les métriques de classification
- Le choix des métriques d'évaluation
- La validation croisée
 - Cas nominal
 - Cas des variables temporelles
- Monitoring des modèles

L'AGILE EN DATASCIENCE

- Livrer souvent
- Communiquer avec le client
- Garantir la qualité
- Amélioration continue

OPTIMISATION AVANCEE DE LA PERFORMANCE

- Sélection de variables
 - Eviter les fuites d'information
 - Cas des variables textuelles
- Choix des algorithmes
- Tuning des hyperparamètres
 - Grid search
 - Hyperopt
 - Parallélisme

AUDITER / COMPRENDRE UN MODELE

- Explicativité des algorithmes

- Feature importance
- Feature contribution

Jour 2

RAPPELS ET REVISION DU JOUR 1

CAS PRATIQUE

- Exposé du problème
- Découverte du dataset
- Choix des métriques
- Réalisation d'un premier modèle
- Optimisation des performances

CONCLUSION

- Synthèse des points abordés
- Partage sur la formation
- Questions/Réponses additionnelles