

TENDANCES & USAGES

Data Science et machine Learning - Les fondamentaux

REF : SIDS101

DUREE : 7h

Présentiel Classe virtuelle

PUBLIC

Cette formation s'adresse à tout public : membres de DSI, Direction études : Directeurs, managers, chefs de projets, consultants, chargés d'études, responsables commerciaux, responsables marketing ...

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Evoluer dans un service Informatique, marketing, études ou dans une société de service

MODALITES PEDAGOGIQUES

1 poste et 1 support par stagiaire

8 à 10 stagiaires par salle

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

Cette formation Data Science et Machine Learning, les fondamentaux vous permettra de :

- Savoir reconnaître une problématique business qui profiterait de l'apport de la data science
- Mettre en place une gouvernance projet dédié
- Choisir les bons outils en fonction de la problématique

PROGRAMME

CHAPITRE 1 : data science et data mining

- Introduction au data mining
- Différence entre data science, data mining, machine learning
- Les principales méthodes en data science :
 - Le data mining : régression, arbre de décision, k-means
 - Le SVM,
 - Le machine learning : boosting et bagging, gradient boosting, random forest, xgboost.
 - Introduction au deep learning : principes des réseaux de neurone, algorithme de Kohonen, le perceptron multicouche.

CHAPITRE 2 : principe du big data

- L'architecture distribuée
- Principe d'un cluster Hadoop : map reduce, Yarn, HDFS,
- Présentation de Spark

CHAPITRE 3 : Overview des principaux outils

- R, Python
- Les solutions payantes : SAS, Dataiku, Spark avec pyspark, les solutions Cloud (zoom sur AWS).

CHAPITRE 4 : Présentation d'un use case en Python (le code sera fourni en format Jupyter).

- Construction du data frame d'études
- Présentation des résultats d'un arbre de décision, d'une forêt aléatoire, d'un xgboost et d'un perceptron multicouche.

Version du : 25/01/2024



