

Machine learning, l'état de l'art

DS010

Durée: 2 jours

1 460 €

6 au 7 février

20 au 21 mars

15 au 16 mai

10 au 12 juillet

31 août au 1er septembre

20 au 21 novembre

Public :

toute personne souhaitant comprendre les mécanismes et les bénéfices du Machine Learning : dirigeants d'entreprise, DSI, responsables informatiques, consultants, responsables de projets Big Data...

Objectifs :

A l'issue de la formation, le stagiaire sera capable d'intégrer les principes fondamentaux du Machine Learning pour une mise en oeuvre adaptée d'un projet d'Intelligence Artificielle.

Connaissances préalables nécessaires :

avoir une culture informatique générale. Posséder des notions de probabilités et statistiques est recommandé.

Objectifs pédagogiques :

les concepts d'apprentissage automatique et l'évolution du Big Data vers le Machine Learning
Appréhender les enjeux de l'utilisation du Machine Learning, incluant les bénéfices attendus et des exemples d'usage
Identifier le positionnement du Machine Learning dans la chaîne de traitement de la donnée
Connaitre les outils et les acteurs leaders du marché
Découvrir les principaux algorithmes et la démarche projet à appliquer selon les cas d'usages en entreprise
Identifier les clés de réussite d'un projet intégrant du Machine Learning

Programme :

les concepts d'apprentissage automatique et l'évolution du Big Data vers le Machine Learning

Définition. De la statistique à l'apprentissage automatique.
Apprentissage automatique : comprendre ou prédire?
Besoin en puissance de calcul et de stockage.
Intégration de l'apprentissage automatique dans les fermes de Big Data.
Les valeurs d'observation, et les variables cibles.
Ingénierie des variables.

Appréhender les enjeux de l'utilisation du Machine Learning, incluant les bénéfices attendus et des exemples d'usage

Comment automatiser les processus métier. Attentes. Création de valeur à partir de la donnée. Problème du ratio pertinence/volume.
Les risques et écueils. Importance de la préparation des données. L'écueil du "surapprentissage".
Les erreurs d'architecture à éviter.

Atelier : mise en évidence d'erreurs d'apprentissage sur des données non qualifiées.

Modélisation automatique. Le rôle du data scientist.

Atelier : démonstration de reconnaissance d'images.

Identifier le positionnement du Machine Learning dans la chaîne de traitement de la donnée

Le pattern MapReduce. Exemple d'utilisation.
Gouvernance des données. Qualité des données.
Transformation de l'information en donnée. Qualification et enrichissement.
Sécurisation et étanchéité des lacs de données.
Flux de données et organisation dans l'entreprise. De la donnée maître à la donnée de travail. MDM.
Mise en oeuvre pratique des différentes phases :
nettoyage, enrichissement, organisation des données.
Zoom sur les données : format, volumes, structures.
Zoom sur les requêtes, attentes des utilisateurs.
Étapes de la préparation des données.
Définitions, présentation du data munging

Connaitre les outils et les acteurs leaders du marché

Comparatifs des outils d'apprentissage automatique. Les outils en mode local, en mode distribué.
Les acteurs. Leurs outils.

Atelier : utilisation de scikit learn et de SparkML. Comparatif.

Apprentissage profond : introduction aux réseaux de neurones.
Réseaux de neurones à convolution. Modèles de CNN.
L'approche du Deep Learning. Deeplearning4j sur Spark. TensorFlow sur rig, sur Spark.

Atelier : mise en oeuvre d'une reconnaissance automatique avec TensorFlow

Découvrir les principaux algorithmes et la démarche projet à appliquer selon les cas d'usages en entreprise

Apprentissage supervisé/non supervisé, classification ou régression.
Algorithme paramétrique ou non-paramétrique, linéaire ou non-linéaire.
Les méthodes : apprentissage supervisé et non supervisé
Classification des données,
Algorithmes : régression linéaire, k-moyennes, k-voisins, classification naïve bayésienne, arbres de décision, forêts aléatoires, ...

Atelier : classification automatique d'un jeu de données à partir d'une régression logistique

Création de jeux d'essai, entraînement et construction de modèles.
Prévisions à partir de données réelles. Mesure de l'efficacité des algorithmes. Courbes ROC.
Parallélisation des algorithmes. Choix automatique.

Atelier : Mise en évidence des erreurs d'apprentissage en fonction des hyper-paramètres

Identifier les clés de réussite d'un projet intégrant du Machine Learning

Choix des architecture. Comment définir le besoin métier?
Extraction et organisation des classes de données.
Applications aux fermes de calculs distribués. Problématiques induites. Approximations. Précision des estimations.
Analyse factorielle.
Visualisation des données. L'intérêt de la visualisation. Outils disponibles.