

Machine Learning : État de l'art et bonnes pratiques

Code : NTE05

Durée : 2 jours

Classe : Présentiel / à distance

Public

- Ce cours Machine Learning s'adresse essentiellement à des dirigeants, des directeurs ou responsables SI, directeurs techniques, responsables informatique qui doivent comprendre les mécanismes du Machine Learning pour envisager un projet d'optimisation et transformation massive de données.

Prérequis

- Pour suivre cette formation Machine Learning, il est recommandé de connaître les principaux algorithmes du Machine learning et d'avoir des notions de probabilité ou statistiques (scolaire), ainsi qu'une bonne culture informatique générale.

Objectifs

- Comprendre les concepts de l'IA et la place du Machine Learning par rapport au Big Data
- Appréhender les apports concrets du Machine Learning pour les entreprises
- Positionner le Machine Learning dans les applications qui manipulent les données
- Identifier les principaux outils et acteurs du marché
- Classifier les différents algorithmes selon les cas d'usage
- Adopter une démarche projet en fonction des cas d'usages
- Identifier les risques et les éléments de réussite d'un projet basé sur le machine Learning

Programme détaillé

1- Introduction au Machine Learning

- Situer le Machine Learning dans l'évolution et l'histoire du Big Data
- Définir les concepts d'Intelligence Artificielle et d'apprentissage automatique (Machine Learning)
- Comprendre les différences entre l'analyse descriptive, l'analyse prédictive puis prescriptive.
- Exemples d'application du Machine learning pour divers directions ou métier : marketing, vente, logistique, RH, santé, transport, la sécurité, l'énergie, la distribution, le luxe, le tourisme...
- Déterminer les résultats et bénéfices attendus du Machine Learning

2- Machine Learning, outils et acteurs du marché

- Identifier les bibliothèques les plus utilisées (TensorFlow, Keras, Pytorch, Scikit learn).
- Choisir entre les deux leaders Open Source : Python et R.
- Plateformes Cloud (Azure, AWS, Google Cloud Platform) et solutions SaaS (IBM Watson, Dataiku).
- Connaître les API en ligne (IBM Watson, Microsoft Cortana Intelligence, IA AWS ..) et les chatbots

3- La donnée et les apprentissages en Machine Learning

- Identifier les données structurées, semi-structurées et non structurées
- Choisir la nature statistique des données (qualitatives ou quantitatives)
- Définir la typologie des algorithmes :
 - Apprentissage supervisé : répéter un exemple-
 - Apprentissage non supervisé : découvrir les données
 - Reinforcement Learning : optimisation d'une récompense
 - Les autres types d'apprentissage (par transfert, séquentiel, actif...)
- Faire le lien entre les mathématiques, le Big Data, l'IA et le machine Learning



Machine Learning : État de l'art et bonnes pratiques

Code : NTE05

Durée : 2 jours

Classe : Présentiel / à distance

4- Les principaux algorithmes du machine Learning

- Utiliser la régression linéaire simple et multiple
- Tester la régression polynomiale
- Définir les séries temporelles
- Comprendre la régression logistique et applications en scoring
- Identifier la classification hiérarchique et non hiérarchique (KMeans)
- Définir une classification par arbres de décision ou approche Naïve Bayes
- Utiliser le Random Forest (développement des arbres de décision)
- Gradient Boosting
- Réseaux de neurones
- Machine à support de vecteurs

5- Démarche Machine Learning dans le traitement de la donnée

- Prévoir la collecte et la préparation des données
- Identification des corrélations
- Réduire la complexité d'un problème pour le résoudre par analyse des composantes principales
- Comment réduire la dimension et sélectionner les variables pertinentes ?
- Détection et correction des valeurs aberrantes
- Data augmentation : création de nouvelles variables pour aider à résoudre le problème

6- Mise en pratique : Procédure d'entraînement et d'évaluation des algorithmes

- Séparation du jeu de données en plusieurs : entraînement, test et validation
- Techniques de bootstrap (bagging)
- Exemple de la validation croisée
- Définition d'une métrique de performance
- Descente de gradient stochastique (minimisation de la métrique)
- Courbes ROC et de lift pour évaluer et comparer les algorithmes
- Matrice de confusion : faux positifs et faux négatifs

7- Envisager la mise en production d'un algorithme de machine Learning

- Description d'une plateforme Big Data
- Principe de fonctionnement des API
- Du développement à la mise en production
- Stratégie de maintenance corrective et évolutive
- Évaluation du coût de fonctionnement en production

8- Mise en œuvre et gestion de projet impliquant du Machine Learning

- Les spécificités d'un projet Machine Learning, son cycle de vie
- Identifier les différents acteurs d'un projet et post-projet
- Définir les nouveaux rôles dans l'entreprise : chief data officer, data protection officer, data engineer, data scientist, data analyst, data miner...
- Déterminer les prestataires externes et l'écosystème interne à l'entreprise
- Préparer sa roadmap de la mise en œuvre du machine Learning (avant, pendant et après le projet) Mise en place d'un Proof Of Concept
- Prévoir rétro planning, Checklist et bonnes pratiques d'un projet Machine Learning
- Protection et droit d'accès aux données personnelles
- L'accompagnement aux changements nécessaire (formation, communication, management)

