

PUBLIC

Ingénieurs et techniciens ayant besoin de méthodes d'apprentissage pour automatiser des tâches (prédiction, décision, etc.) de manière performante - Chefs de projets qui souhaitent mieux identifier les tâches que le machine learning permettrait d'automatiser - Toute personne ayant déjà été sensibilisée par le Data Mining et intéressée par les méthodes d'apprentissage

PRÉ-REQUIS

Aucun prérequis nécessaire

NIVEAU D'EXPERTISE

Fondamentaux

LES POINTS FORTS

Docaposte Institute propose plusieurs dispositifs pédagogiques adaptés aux apprenants :

Formation en présentiel

En groupe (inter-entreprises ou intra-entreprise)

En individuel (moniteur)

En journée ou en cours du soir (sur demande spécifique)

Formation en distanciel

Distanciel synchrone

Distanciel asynchrone

MOYENS PÉDAGOGIQUES

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel

Code
701982

Durée
Sur mesure / 7 heures

Tarif
Nous contacter

*Repas inclus (en présentiel)

Objectifs pédagogiques

- Comprendre les concepts d'apprentissage automatique et l'évolution du big data vers l'apprentissage automatique
- Comprendre les défis de l'utilisation de l'apprentissage automatique, y compris les avantages attendus et les cas d'utilisation
- Identifier la place du machine learning dans la chaîne de traitement des données
- Connaître les principaux outils et acteurs du marché
- Passer en revue les algorithmes clés et les approches de projet à appliquer en fonction de votre cas d'utilisation métier
- Identifier les clés du succès des projets qui intègrent le machine learning

Programme de la formation

Introduction et historique Processus complet d'élaboration d'un modèle prédictif

- Prétraitement des données (données manquantes, aberrantes, sélection de variables pertinentes...)
- Introduction à l'apprentissage supervisé et non supervisé
- Introduction à la classification et la régression
- Le principe du partitionnement des données en données d'apprentissage et données de test
- Évaluation de l'apprentissage, matrice de confusion et métrique

Arbres de décision

- L'apprentissage par partitionnement
- Construction d'un arbre de décision sur un jeu de données synthétique
- Principaux points à considérer lors de l'induction d'un arbre de décision à partir de données
- CHAID, C4.5 et CART : Les trois principales méthodes d'induction d'arbres proposés dans les logiciels
- Les différences et les points communs Arbres de régression

dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

SATISFACTION ET EVALUATION

- En amont de la formation
 - ▶ Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
 - ▶ Auto-positionnement des apprenants afin de mesurer le niveau de départ.
- Tout au long de la formation
 - ▶ Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...
- A la fin de la formation
 - ▶ Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
 - ▶ Evaluation par le formateur des compétences acquises par les apprenants.
 - ▶ Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
 - ▶ Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

ACCOMPAGNEMENT FORMATION À DISTANCE

En cas de nécessité, une assistance technique et pédagogique est joignable entre 8h30 et 18h (jours ouvrés):

- par téléphone : 01 83 10 10 10
- par mail : care-formation@lefebvre-dalloz.fr

Une réponse immédiate est apportée ; si besoin, le demandeur est mis en relation avec un expert dans un délai maximum de 48h.

La régression par arbres

- Une méthode non-linéaire de régression
- Rapprochement avec les arbres de décision
- Analyse discriminante prédictive
- Un modèle paramétrique de discrimination

Analyse discriminante de Fisher

- Évaluation globale du modèle
- Évaluation individuelle des variables

Régression logistique

- Le modèle LOGIT
- Estimation, évaluation globale et évaluation individuelle des variables
- Régression logistique polytomique
- Régression logistique lorsque la variable dépendante Y prend plus de 2 valeurs
- Régression logistique multinomiale lorsque Y est nominale
- Régression logistique polytomique ordinale lorsque Y est ordinale (odds adjacents, odds cumulatifs proportionnels)

Classifieur bayésien naïf (modèle d'indépendance conditionnelle)

- Principe de l'indépendance conditionnelle
- Cas des prédicteurs continus
- Cas des prédicteurs discrets
- Construction d'un modèle explicite linéaire
- Sélection de variables

Bagging, random forest, boosting

- Méthodes ensemblistes basées sur des apprentissages répétés
- Boosting : principe et variantes
- Principe du Bagging
- Random Forst : principe et variantes
- Mesures d'importance des variables – Impact sur le biais et la variance des classifieurs

Gradient boosting

- Généralisation du boosting avec l'introduction explicite de fonctions de coûts
- Importance du paramétrage

Support Vector Machine (SVM)

- Principe de Machines à vecteurs de support ou Séparateur à Vaste Marge
- Principe de la maximisation de la marge
- Marge souple (soft margin)
- Classifieurs linéaires et classifieurs non-linéaires avec l'astuce du noyau (kernel trick)
- Fonction Noyau

Réseaux de neurones – perceptron simple et multicouches

- Introduction aux réseaux de neurones artificiels pour l'apprentissage supervisé
- La Perceptron
- Passage du modèle linéaire au modèle non-linéaire : le perceptron multicouches
- Apprentissage non supervisé
- Principe du clustering (K moyennes ou k means)
- Principe de la Classification Ascendante Hiérarchique (CAH)
- Règles d'association

Discrétisation des variables quantitatives

- Découpage en classe d'une variable quantitative
- Méthodes non-supervisées et supervisées (chi-merge, mdlpc)

Filtrage des variables

- Approche FILTRE préalable à l'apprentissage supervisé
- Techniques de classement (ranking)
- Techniques de sélection basées sur la corrélation
- Information mutuelle, entropie de Shannon, rapport de corrélation, lambda de Wilks

Induction de règles prédictives

- Construction de bases de règles en analyse prédictive
- Conversion des arbres en règles et algorithmes génétiques pour l'induction de règles
- Scoring – le ciblage marketing
- Le ciblage clientèle
- Construction et lecture de la courbe LIFT (GAIN CHART)

Analyse RFM (récence – fréquence – montant)

- Segmentation RFM (récence-fréquence-montant)
- Finalité, démarche, usage, variantes, limites

Grille de score

- Élaboration de la grille de score à partir des résultats de la régression logistique
- Méthode Disqual et scoring

Intégration des coûts de mauvais classement en apprentissage supervisé

- Prise en compte des coûts pour l'évaluation et la construction des modèles prédictifs
- Correction des règles d'affectation, techniques intégrées, bagging, la méthode MetaCost
- Courbe ROC
- Évaluation d'un classifieur à l'aide de la courbe ROC
- Le critère AUC

Quelques méthodes non-paramétriques de discrimination

- Deux méthodes non-paramétriques de classement dérivés du schéma Bayésien la méthode des K-plus proches voisins et le modèle d'indépendance conditionnelle

Tirage rétrospectif et redressement des résultats échantillonnage non-représentatif

- Modification du seuil d'affectation basé sur le score
- Utilisation de la courbe ROC à cet effet

Normalisation des scores

- Rendre comparable des scores fournis par des méthodes différentes
- Ramener dans l'intervalle $[0 ; 1]$ et harmonisation des distributions
- Diagramme de fiabilité

Méthodes de ré-échantillonnage pour l'évaluation des performances Inadéquation de l'évaluation en re-substitution

- Le schéma apprentissage-test
- Les techniques de ré-échantillonnage : validation croisée, leave-one-out, bootstrap
- Comparaison des performances des techniques supervisées
- Diagramme de fiabilité

A noter

En amont et en aval de la formation, le positionnement pédagogique sera effectué à l'aide d'un questionnaire d'auto-positionnement.