

Formation Machine learning, État de l'art

Présentation

Cette formation vous permettra de maîtriser les bonnes pratiques du Machine Learning.

Durée : 14,00 heures (2 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

- 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

Prérequis

- Avoir une culture informatique générale,
- Posséder des notions de probabilités et statistiques est recommandé.

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

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

- 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 multi-couches

- 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

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Moyens pédagogiques et techniques

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.

- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

NB : dans le cadre d'une Action collective, chaque stagiaire bénéficiaire sera contacté par un prestataire choisi par l'Opco Atlas afin d'évaluer « à chaud » la qualité de la formation suivie.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.