ARBRES DE DÉCISION ET RANDOM FOREST : MODÈLE PRÉDICTIF

Durée & Format

Durée : 14 heures

Format disponible : Présentiel ou Distanciel

Tarifs

Tarif public intra : Consulter notre site Internet

Tarif public inter : Consulter notre site Internet

Public visé

Personne souhaitant maitriser l'utilisation d'un modèle prédictif basé sur la méthode des arbres de décisions et des forêts aléatoires (random forests).

Prérequis

- Une connaissance des outils statistiques de base est souhaitée.
- Dans le cas où la formation serait effectuée avec le logiciel R, une connaissance de base de ce logiciel est préconisée.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- Maitriser le vocabulaire spécifique à la construction des arbres de décisions et des forêts aléatoires
- Comprendre l'algorithme de construction des arbres de décisions
- Connaitre les objectifs et les différences entre les méthodes d'apprentissage supervisé
- Connaitre les paramètres à ajuster lors de la mise en œuvre des méthodes d'arbre de décision pour une prédiction quantitative
- Connaitre les paramètres à ajuster lors de la mise en œuvre des méthodes d'arbre de décision pour une prédiction qualitative
- Comprendre l'algorithme de construction des forêts aléatoires
- Maitriser les paramètres permettant d'estimer la qualité de prédiction des arbres de décisions

Programme et déroulement

Jour 1 – Matin

Explorer les différentes méthodes d'apprentissage supervisé

- Champs d'application des différentes méthodes
- Objectifs de l'apprentissage supervisé
- Objectifs de description
- Objectifs de prédiction
- Structure des jeux de données
- Présentation générale de l'éventail des méthodes
  - Arbre de décision
  - Forêts aléatoires

Jour 1 – Après-midi

Appliquer les méthodes d'arbre de décision

- Structure du jeu de données
- Principes, vocabulaire et objectifs
- Notion d'échantillon d'apprentissage, de validation et de test
- Comparaison de méthodes de type régression linéaire / logistique aux arbres de décision
- Principe de la segmentation selon le type de variable : Arbre de régression ou arbre de classification
- Définir les conditions d'arrêt de construction d'un arbre : Notion de pré-élagage
- Définition des groupes après construction de l'arbre
- Indicateurs de qualité et de pureté
- Comparaison d'arbre de décision selon un certain type d'algorithme : CHAID vs CART
- Avantages et inconvénients : limites des arbres de décision
- Mise en œuvre et interprétation des résultats obtenus après application d'une analyse par arbre de décision

Jour 2 – Matin

Utiliser la méthode des forêts aléatoires (random forest)

- Pourquoi avoir recourt aux forêts aléatoires ?
- Principes et objectifs
  - Instabilité de l'arbre
  - Notion de Bagging
  - Les erreurs liées à l'échantillonnage (Out-Of-Bag)
- Prédiction avec un algorithme de Forêt aléatoire : Les paramètres
- Evaluer l'importance des variables
  - Notion d'importance
  - Comportement de l'importance
  - Lien entre diversité des arbres et l'importance
  - Influence des paramètres
- Sélection de variables
  - Généralités et principes de la sélection
  - Procédure de sélection
- Les paramètres de sélection

Jour 2 – Après-midi

Optimiser le sous et surajustement

- Notions de biais d'un modèle
- Notions de variance d'un modèle
- Optimalité variance & biais
- Jugement de la qualité d'un modèle
- Méthode du Data Train / Data test
- Validation croisée
- Classique LOO (Leave One Out)
- Validation croisée par k fold

Réaliser des exercices pratiques

- Applications avec R
- Démonstration avec Excel

Cas d'études sur données des apprenants

Jeux de données

Afin de s'approcher au mieux des réalités quotidiennes des praticiens, nous suggérons de nous appuyer pour l'animation pratique de thématiques et surtout de jeux de données reflétant le quotidien des apprenants.

Cet élément est un facteur de réussite pour la formation. Elle permet aux apprenants de :

- Se "reconnaitre" dans les thèmes abordés,
- Mieux percevoir l'intérêt des notions étudiées
- S'approprier le contenu de la formation

Il sera donc pertinent que les apprenants puissent réfléchir en amont de la formation à des problématiques, jeux de données ou documents susceptibles d'être utilisés en support lors de la formation.

Outil logiciel

Cette formation n'est pas strictement dédiée à un logiciel. Les exercices et les illustrations se feront généralement sous R ou sous un autre logiciel partant de l'hypothèse que celui-ci intègre les outils techniques abordés.

Dans le cas où la formation serait effectuée avec le logiciel R, une connaissance de base de ce logiciel est préconisée.

Certification

Consulter le programme sur notre site internet pour identifier si cette formation est certifiante.

Modalités pédagogiques

- Explications théoriques suivies de pratiques guidées puis de mises en autonomie.
- Exercices autonomes et réguliers pour assurer l'assimilation

Moyens et supports pédagogiques

Votre formation se déroule à distance avec :

- 1 ordinateur
- 1 connexion Internet
- 1 adresse e-mail valide
- 1 équipement audio (micro et enceintes ou casque)
- 1 Webcam (facultatif – dans l'idéal)
- 1 deuxième écran (facultatif – dans l'idéal)

Evaluation et suivi

Les objectifs pédagogiques sont évalués et suivis grâce à différentes **méthodes adaptées aux acquisitions de compétences visées**. Nos formateurs réalisent ces évaluations **tout au long de la formation**, que ce soit pendant les séances synchrones ou asynchrones. Voici une liste non exhaustive des méthodes d'évaluation pouvant être utilisées en formation :

- **Questionnaire** de vérification de connaissances (Quiz)
- Réalisation de **Travaux Pratiques** : production ou amélioration d'un fichier
- **Mises en situation** et grilles d'analyse
- **Serious Game** : jeu de rôles et analyse
- **Activités interactives** à travers l'utilisation d'une plate-forme connectée
- …

Toutes nos formations intègrent **une auto-évaluation** via notre Extranet Stagiaire au début et à la fin de chaque action de formation. Cet outil offre à chacun la possibilité de mesurer sa progression par rapport aux objectifs pédagogiques visés et leurs atteintes.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire via son extranet stagiaire.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite. En savoir plus