

Statistique

METHODES D'APPRENTISSAGE SUPERVISE : ARBRES DE DECISION ET FORETS ALEATOIRES

Référence : 2020-STAT-N2-ArbreRandomForest

2 jours (14h) – Présentiel ou À distance

Population visée : Statisticiens – Data scientists – Ingénieurs et chargés d'études/d'analyses – Chercheurs – Doctorants

Public concerné : Toute personne souhaitant comprendre le contexte d'utilisation, les concepts, et la mise en œuvre des méthodes d'apprentissage supervisé de type arbres de décision et forêts aléatoires.

La formation en quelques mots

Cette formation s'adresse à des personnes souhaitant découvrir le contexte d'utilisation des méthodes algorithmiques d'apprentissage supervisé de type arbres de décision et forêts aléatoires. Ces méthodes ont pour principal objectif d'extraire des informations de données massives. Les méthodes d'arbres de décision et forêts aléatoires sont mises en pratique dans un objectif de prise de décision métier.

Ces types de méthodes s'appliquent en présence de données :

- 🌀 Volumineuses en nombre de variables,
- 🌀 Volumineuses en nombre d'individus,
- 🌀 Non structurées,
- 🌀 Comportant des variables redondantes (confusions entre variables),
- 🌀 ...

Thèmes principaux

- 🌀 Arbre de décision
- 🌀 Forêt aléatoire

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.

Ce souhait est un facteur de réussite pour la formation. Elle permet aux apprenants de :

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

Prérequis

IL EST INDISPENSABLE que les participants aient de bonnes connaissances sur les outils statistiques de base : corrélation, écart-type, variance, intervalles de confiance, tests d'hypothèses.









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

Outil logiciel

Les mises en applications pourront se faire autour du logiciel R si les apprenants sont autonomes, sous XlStat ou sous JMP.



Objectifs pédagogiques

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






-  Maîtriser le vocabulaire spécifique aux méthodes d'apprentissage supervisé
-  Identifier le contexte et les conditions d'application des méthodes d'apprentissage supervisé
-  Connaître les objectifs et les différences entre les méthodes d'apprentissage supervisé
-  Décrire la méthodologie inhérente à ces méthodes
-  Mettre en œuvre et interpréter les résultats des méthodes d'apprentissage supervisé
-  Connaître les principaux indicateurs de cohérence liés aux méthodes d'apprentissage supervisé
-  Connaître les paramètres à ajuster lors de la mise en œuvre des méthodes d'arbre de décision et de forêt aléatoire
-  Maîtriser les paramètres permettant d'estimer la qualité de ces analyses

Programme

Généralités sur 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
 - Forêt aléatoire

Contexte d'application des 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

Page 2/4 - METHODES D'APPRENTISSAGE SUPERVISE : ARBRES DE DECISION ET FORETS ALEATOIRES - Mis à jour le 06/01/2022

Concepteur : Groupe ARKESYS – Diffuseur : Groupe ARKESYS



La certification qualité a été délivrée au titre de la ou des catégories d'actions suivantes : actions de formation

ARKESYS.NET

70 rue Bergson - 42000 Saint-Etienne
T. 04 37 24 36 78 - info@arkesys.fr

SAS AU CAPITAL DE 60 000 € - SIRET : 501 033 609 00030
ETS secondaire : 24 espace H.Vallée - 69007 Lyon - SIRET : 501 033 609 00014

AVANTEAM FORMATION

650 chemin de l'Aumône Vieille - 13400 Aubagne
T. 04 37 24 36 78 - info@arkesys.fr

SAS AU CAPITAL de 10 000 € - SIRET : 749 889 713 00022

- 🔄 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
- 🔄 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

Contexte d'application des méthodes de forêt aléatoire

- 🔄 Pourquoi avoir recours 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

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

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 par e-mail ou via la plate-forme FOAD.

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.

Page 4/4 - METHODES D'APPRENTISSAGE SUPERVISE : ARBRES DE DECISION ET FORETS ALEATOIRES - Mis à jour le 06/01/2022

Concepteur : Groupe ARKESYS – Diffuseur : Groupe ARKESYS



La certification qualité a été délivrée au titre de la ou des catégories d'actions suivantes : actions de formation

ARKESYS.NET

70 rue Bergson - 42000 Saint-Etienne
T. 04 37 24 36 78 - info@arkesys.fr

SAS AU CAPITAL DE 60 000 € - SIRET : 501 033 609 00030
ETS secondaire : 24 espace H.Vallée - 69007 Lyon - SIRET : 501 033 609 00014

AVANTEAM FORMATION

650 chemin de l'Aumône Vieille - 13400 Aubagne
T. 04 37 24 36 78 - info@arkesys.fr

SAS AU CAPITAL de 10 000 € - SIRET : 749 889 713 00022