

Statistique

MACHINE LEARNING – DECOUVERTE : CLASSIFICATION, K-MEANS, ANALYSE FACTORIELLE DISCRIMINANTE, ARBRES DE DECISION ET FORETS ALEATOIRES

Référence : 2020-STAT-N2-MachineLearningDcv

1 jours (7h) – Présentiel ou À distance

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

Public concerné : Toute personne souhaitant découvrir le contexte et les concepts d'utilisation des méthodes de machine learning de type classification, k-means, analyse factorielle discriminante, arbres de décision et forêts aléatoires.

La formation en quelques mots

Cette formation s'adresse à des personnes souhaitant découvrir le contexte d'application et les concepts des méthodes de machine learning (classification, k-means, analyse factorielle discriminante, arbre de décision et forêt aléatoire). Ces méthodes ont pour principal objectif d'extraire des informations de données massives. Les méthodes de types classification, k-means sont utilisées dans un contexte d'analyse exploratoire. On entend par analyse exploratoire la « découverte » des données dont l'objectif est de se faire une première idée sur celles-ci : statistiques descriptives, recherche de tendances, de corrélations. 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),
- 🌀 ...

La formation conviendra tout à fait à un public venant chercher du savoir statistique sur :

- 🌀 Les concepts des différentes méthodes descriptives d'analyse exploratoire, de classification, d'arbre de décision et de forêt aléatoire
- 🌀 Les contextes d'application de chaque méthode

Thèmes principaux

- 🌀 K-means
- 🌀 Classification ascendante hiérarchique (CAH)
- 🌀 Analyse factorielle discriminante (AFD)
- 🌀 Arbre de décision
- 🌀 Forêt aléatoire

Page 1/3 - MACHINE LEARNING – DECOUVERTE : CLASSIFICATION, K-MEANS, ANALYSE FACTORIELLE DISCRIMINANTE, 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

Outil logiciel

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

Prérequis

Il est souhaitable d'avoir des connaissances de bases en statistiques.

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

Objectifs pédagogiques

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

- Connaître les objectifs et les différences entre les méthodes d'apprentissage supervisé et non supervisé
- Maîtriser le vocabulaire spécifique aux méthodes d'apprentissage supervisé et non supervisé
- Identifier le contexte et les conditions d'application des méthodes d'apprentissage supervisé et non supervisé
- Connaître les démarches des méthodes d'apprentissage supervisé et non supervisé

Programme

Généralités sur les différentes méthodes de machine learning

- Limites des statistiques classiques
- Champs d'application des différentes méthodes
- Introduction sur le machine learning - Les objectifs et problématiques

Découverte de la méthode d'Analyse discriminante

- Structure du jeu de données
- Contexte d'application, objectifs et méthodologie de l'analyse discriminante
- Notions de classement et de discrimination
- Méthodologie de l'AFD
- Interprétation des sorties logiciel : cercle factoriels, corrélations variables x axes
- Qualité de l'AFD (de la discrimination obtenue)
 - Tests univariés et multivariés (lambda de Wilks)
 - Graphique des individus
 - Matrice de confusion (et éventuellement courbe ROC)
- Les confusions et erreurs à ne pas commettre

Découverte des méthodes de classification : classification ascendante hiérarchique (CAH) et k-means

- Structure du jeu de données
- Contexte d'application, objectifs et méthodologie de la CAH et des k-means
- Lecture d'un dendrogramme

- Différents types de classification : sur les individus, sur les variables, sur les modalités
- Avantages et inconvénients des méthodes de classification
- Interprétation des sorties logiciel

Découverte des 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

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.