# MACHINE LEARNING : MÉTHODES DE CLASSIFICATION ET PRÉDICTION

## Durée & Format

**Durée :** 28 heures

**Format disponible :** Présentiel ou Distanciel

## Tarifs

**Tarif public intra :** Consulter notre site Internet

**Tarif public inter :** Consulter notre site Internet

## Public visé

- Toute personne souhaitant comprendre le contexte d'utilisation, les concepts, et la mise en œuvre des méthodes de classements et de prédiction d'une variable qualitative.

## Prérequis

- Il est nécessaire que les participants aient des 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.

## Objectifs pédagogiques

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

- Maîtriser le vocabulaire spécifique aux méthodes d'apprentissage à finalité de classement
- Identifier le contexte et les conditions d'application des méthodes d'apprentissage supervisé et non supervisé
- Connaître les objectifs et les différences entre les méthodes de classement
- 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
- Maitriser les paramètres permettant d'estimer la qualité de ces analyses

## Programme et déroulement

### Jour 1 – Matin

VOLET 1 : LES ALGORITHMES

Généralités sur les différentes méthodes d'apprentissage supervisé

- Différences entre méthodes supervisées et non supervisées
- Objectifs de l'apprentissage supervisé
  - Objectifs de description
  - Objectifs de prédiction

### Jour 1 – Après-midi

- Structure des jeux de données
- Présentation générale de l'éventail des méthodes

### Jour 2 – Matin

La méthode knn

- Principe de la méthode des plus proches voisins
- Algorithme de calcul
- Distances entre individus
- Choix des proximités
- Définition du paramètre k

**Jour 2 – Après-midi**

La régression logistique

- Variable explicative et variable expliquée (continue / binaire)
- Différences entre la régression linéaire classique et la régression logistique
- Variables explicatives qualitatives, variables explicatives quantitatives
- Objectifs de la régression logistique
- Définition du modèle Logit (courbe sigmoïde)
- Conditions d'utilisation à respecter
- Estimation et interprétation des coefficients du modèle
- Test de significativité du modèle (validation du modèle)
- Tests d'apport d'une variable (test de Wald, tests sur les rapports de vraisemblance)
- Interprétation du Khi² de Wald
- Odds-ratios
- Parallèle odds ratios et risques relatifs
- Mise en œuvre et analyse des résultats d'un modèle de régression logistique multiple
- Estimation et interprétation des coefficients du modèle multiple

**Jour 3 – Matin**

L'analyse factorielle discriminante

- Structure du jeu de données et contexte d'application
- Objectifs détaillés de l'AFD
- Notions de classement et de discrimination
- Méthodologie de l'AFD
- Comparaison avec l'ACP
- 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

**Jour 3 – Après-midi**

Les supports vecteurs machines (SVM)

- Démarche des svm
- Notions de marge
- Séparation linéaire
- Séparation non linéaire
- Fonction noyau

**Jour 4 – Matin**

Les arbres 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
- 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 4 – Après-midi**

De l'arbre à la foret - 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
  - Validation

VOLET 2 : VALIDATIONS DES METHODES, MESURE DES PERFORMANCES

- Partitionnement des données disponibles
  - Jeu d'entrainement
  - Jeu de validation
  - Tests sur le jeu d'entrainement
  - Tests sur le jeu de validation
- La validation croisée
  - Leave one out
  - K fold
  - Leave v out
- Compromise biais / variance
- Mesures des performances des classifications
  - Matrices de confusions
  - Courbe Roc
  - Aire sous la courbe (AUC)
  - Sensibilité & spécificité

## 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 a lieu en présentiel :**

- 1 vidéoprojecteur par salle
- 1 ordinateur

**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)

**Votre formation se déroule sur notre plate-forme de formation avec :**

- 1 ordinateur
- 1 connexion Internet
- 1 adresse e-mail valide
- 1 équipement audio (micro et enceintes ou casque - facultatif)

**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.

## 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