



FORMATION EN PRESENTIEL ET/OU EN DISTANCIEL (SELON LE THEME)

# MACHINE LEARNING AVANCÉE

Dernière mise à jour le 19/02/2024.

## OBJECTIF

Comprendre la nature des problèmes traités par les techniques d'apprentissage automatique utilisées dans le Big Data

## COMPÉTENCES DÉVELOPPÉES

- Comprendre la nature des problèmes traités par les techniques d'apprentissage automatique utilisées dans le Big Data
- Comprendre les problématiques sous-jacentes et les solutions apportées pour résoudre certains types de problèmes
- Cette formation vous permettra de maîtriser les bonnes pratiques du Machine Learning.

## PUBLIC CONCERNÉ

Ingénieurs et techniciens, Chefs de projet et toute personne ayant déjà été sensibilisée par la machine learning.

## PRÉ-REQUIS

- Des connaissances solides en statistiques et en algorithmes sont exigées.

## MODALITÉS ET DÉLAIS D'ACCÈS

Vous pouvez vous inscrire pour suivre une de nos formations jusqu'à la veille de la date de démarrage si la formation est financée directement par votre entreprise ET si le nombre maximum de participants n'est pas atteint. Si la formation est financée via un OPCO, vous devez au préalable avoir obtenu un accord de ce dernier

## ACCESSIBILITÉ AUX PERSONNES EN SITUATION DE HANDICAP

Nos locaux sont accessibles aux Personnes à Mobilité Réduite PMR. De plus, nos conseillers sont disponibles pour vous accompagner dans vos démarches à travers nos partenaires. Nous sommes en mesure de mobiliser les expertises, les outils nécessaires pour vous accueillir, vous accompagner et vous former.

## MODALITÉS DE DÉROULEMENT DE L'ACTION DE FORMATION

Formation présentielle ou distancielle dispensée par un formateur expérimenté. La formation alterne des exposés théoriques, des démonstrations et la mise en pratique au travers d'exercices et de cas concrets.



Niveau

Autres formations  
professionnelles continues



Modalité

Présentiel



Effectif par session

2 mini > 8 maxi



Durée

14 heures / 2 jour(s)

Tarif

1390 € HT  
par participant

Réf.

MACHLEARNav

Contact

0465260114  
nathalie.husson@univlearn.fr  
<https://univlearn.fr/>



## PROGRAMME

### Module 1 : INTRODUCTION ET HISTORIQUE

- Processus complet d'élaboration d'un modèle prédictif

### Module 2 : 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

### Module 3 : ARBRES DE RÉGRESSION

- La régression par arbres
- Une méthode non-linéaire de régression
- Rapprochement avec les arbres de décision

### Module 4 : ANALYSE DISCRIMINANTE PRÉDICTIVE

- Un modèle paramétrique de discrimination
- Analyse discriminante de Fisher
- Évaluation globale du modèle
- Évaluation individuelle des variables

### Module 5 : 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)

### Module 6 : CLASSIFIEUR BAYESIEN 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

### Module 7 : 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

### Module 8 : GRADIENT BOOSTING

- Généralisation du boosting avec l'introduction explicite de fonctions de coûts



- Importance du paramétrage

#### **Module 9 : 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

#### **Module 10 : 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 multicouche

#### **Module 11 : APPRENTISSAGE NON SUPERVISÉ**

- Principe du clustering (K moyennes ou k means)
- Principe de la Classification Ascendante Hiérarchique (CAH)
- Règles d'association

#### **Module 12 : DISCRÉTISATION DES VARIABLES QUANTITATIVES**

- Découpage en classe d'une variable quantitative
- Méthodes non-supervisées et supervisées (chi-merge, mdlpc)

#### **Module 13 : 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

#### **Module 14 : 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

#### **Module 15 : SCORING – LE CIBLAGE MARKETING**

- Le ciblage clientèle
- Construction et lecture de la courbe LIFT (GAIN CHART)

#### **Module 16 : ANALYSE RFM (RÉCENCE – FRÉQUENCE – MONTANT)**

- Segmentation RFM (récence-fréquence-montant)
- Finalité, démarche, usage, variantes, limites

#### **Module 17 : GRILLE DE SCORE**

- Élaboration de la grille de score à partir des résultats de la régression logistique
- Méthode Disqual et scoring

#### **Module 18 : 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

**Module 19 : 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

**INTERVENANT(S)**

Consultant formateur ayant la double compétence pédagogique et technique

**ÉVALUATION**

Chaque module de cours est concrétisé par un TP afin de permettre l'acquisition d'un vrai savoir-faire sur tous les points abordés, Questions posées par le formateur tout au long de la formation à l'oral ou à travers un QCM

**ATTRIBUTION FINALE**

Attestation individuelle de formation