

# Machine learning, l'état de l'art les solutions du marché et leurs mises en oeuvre

Séminaire de 2 jours - 14h

Réf : MLE - Prix 2024 : 2 090€ HT

Ce séminaire détaille les enjeux liés au traitement de la donnée par l'Intelligence Artificielle, et en particulier par les algorithmes du Machine Learning. Il montre aux décideurs, les principaux algorithmes du domaine, les solutions concrètes et la démarche de projet à appliquer selon les cas d'usages en entreprise.

## OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Positionner le Machine Learning dans la chaîne de traitement de la donnée

Distinguer les compétences nécessaires ou les profils à recruter

Identifier les clés de réussite d'un projet autour du Machine Learning

Comprendre les concepts d'apprentissage automatique et l'évolution du Big Data vers le Machine Learning

Appréhender les enjeux de l'utilisation du Machine Learning, incluant les bénéfices attendus et des exemples d'usage

## MÉTHODES PÉDAGOGIQUES

Illustration par des cas concrets.  
Présentation des principaux cas d'usage selon les secteurs d'activités (Automobile, industrie, biens de consommation, finance, santé énergie, agriculture, transports, télécommunication...).

## LE PROGRAMME

dernière mise à jour : 11/2022

### 1) Histoire du Machine Learning et contexte du Big Data

- Replacer à leur échelle les concepts d'Intelligence Artificielle, apprentissage automatique (machine learning)...
- Le lien avec les mathématiques, les statistiques (inférentielles), le data mining et la data science.
- Passer de l'analyse descriptive à l'analyse prédictive puis prescriptive.
- Les applications du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).
- La typologie des algorithmes de Dominique CARDON.
- La communauté Data Science et les challenges Kaggle (ex. de Netflix).

*Etude de cas : Etudes d'applications concrètes du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).*

### 2) Les données à disposition : collecte et préparation

- Données structurées, semi-structurées et non structurées.
- Nature statistique des données (qualitatives ou quantitatives).
- Objets connectés (IoT) et streaming.
- Opportunités et limites de l'Open Data.
- Identification des corrélations, problème de la multicollinéarité.
- Réduction des dimensions par Analyse des Composantes Principales.
- Détection et correction des valeurs aberrantes.
- Les ETL (Extract Transform Load).

## FINANCEMENT

Ce cours fait partie des actions collectives Atlas.

## PARTICIPANTS

Dirigeants d'entreprise (CEO, COO, CFO, SG, DRH...), DSI, les CDO, responsables informatiques, consultants, responsables de projets Big Data.

## PRÉREQUIS

Posséder une culture informatique générale, des notions de probabilités et statistiques sont recommandés.

## COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

## MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- Le Web scraping.

*Démonstration* : Démonstration d'un ETL (Extract Transform Load). Recueil de données Web.

### 3) Les outils du marché pour le traitement de la donnée et le Machine Learning

- Les logiciels traditionnels (SAS, SPSS, Stata...) et leur ouverture à l'Open Source.
- Choisir entre les deux leaders Open Source : Python et R.
- Plateformes Cloud (Azure, AWS, Google Cloud Platform) et solutions SaaS (IBM Watson, Dataiku).
- Nouveaux postes en entreprises : data engineer, data scientist, data analyst, etc.
- Associer les bonnes compétences à ces différents outils.
- Les API en ligne (IBM Watson, Microsoft Cortana Intelligence...).
- Les chatbots (agents conversationnels).

*Démonstration* : Démonstration d'un chatbot (agent conversationnel) et d'Azure Machine Learning.

### 4) Les différents types d'apprentissage en Machine Learning

- Apprentissage supervisé : répéter un exemple.
- Apprentissage non supervisé : découvrir les données.
- Online (Machine) Learning par opposition aux techniques batch.
- Reinforcement learning : optimisation d'une récompense.
- Autres types d'apprentissage (par transfert, séquentiel, actif...).
- Illustrations (moteurs de recommandation...).

*Démonstration* : Démonstrations sur les différents types d'apprentissage Machine Learning possibles.

### 5) Les algorithmes du Machine Learning

- Régression linéaire simple et multiple. Limites des approches linéaires.
- Régression polynomiale (LASSO). Séries temporelles.
- Régression logistique et applications en scoring.
- Classification hiérarchique et non hiérarchique (KMeans).
- Classification par arbres de décision ou approche Naïve Bayes.
- Random Forest (développement des arbres de décision).
- Gradient Boosting. Réseaux de neurones. Machine à support de vecteurs.
- Deep Learning : exemples et raisons du succès actuel.
- Text Mining : analyse des corpus de données textuelles.

*Démonstration* : Démonstration des différents algorithmes de base sous R ou Python.

### 6) Procédure d'entraînement et d'évaluation des algorithmes

- Séparation du jeu de données : entraînement, test et validation.
- Techniques de bootstrap (bagging).
- Exemple de la validation croisée.
- Définition d'une métrique de performance.
- Descente de gradient stochastique (minimisation de la métrique).
- Courbes ROC et de lift pour évaluer et comparer les algorithmes.
- Matrice de confusion : faux positifs et faux négatifs.

*Démonstration* : Démonstration du choix du meilleur algorithme.

### 7) Mise en production d'un algorithme de Machine Learning

- Description d'une plateforme Big Data.
- Principe de fonctionnement des API.
- Du développement à la mise en production.
- Stratégie de maintenance corrective et évolutive.
- Evaluation du coût de fonctionnement en production.

*Démonstration* : Démonstration d'API de géolocalisation et d'analyse de sentiments.

### 8) Aspects éthiques et juridiques liés à l'Intelligence Artificielle

- Missions de la CNIL et évolutions à venir.

- Question du droit d'accès aux données personnelles.
- Question de la propriété intellectuelle des algorithmes.
- Nouveaux rôles dans l'entreprise : Chief Data Officer et Data Protection Officer.
- Question de l'impartialité des algorithmes.
- Attention au biais de confirmation.
- Les secteurs et les métiers touchés par l'automatisation.

*Réflexion collective* : Réflexion en commun pour identifier les clés de réussite.

## LES DATES

---

CLASSE À DISTANCE  
2024 : 04 juil., 15 oct., 10 déc.

PARIS  
2024 : 27 juin, 08 oct., 03 déc.