

Module : Machine Learning

Dernière mise à jour : 01/10/2022

Code	Cours intégré	HNE	ECTS
SI-17	42h	35h	3

Responsable Module	Sonia MESBAH
Enseignants – Intervenants	Ahlem Marzouk, Dorra Trabelsi, Ines Channoufi, Mohamed Hedi Riahi, Rahma Ferjani, Shema Essadi, Sonia Mesbah.
Unité pédagogique	GL-BD
Unité d'enseignement	INFBD0001 Analyse de données
Prérequis	<p>Sys. De Gestion de Bases de Données</p> <p>Calcul scientifique</p> <p>Méthodes numériques pour l'ingénieur</p> <p>Techniques d'estimation pour l'ingénieur</p>
Niveaux et Options	4ème année DS

Objectif du module :

L'objectif de ce cours est de familiariser les étudiants avec la modélisation des données en utilisant les méthodes de Machine Learning, tout en respectant les phases d'un projet Machine Learning.

Mode d'évaluation :

- La moyenne de ce module est calculée comme suit :

Moyenne (ML) = Contrôle Continu*40% + Projet *60%.

- **Projet** : Le projet consiste à une reproductibilité d'un article scientifique.
- **Le contrôle continu** : La moyenne des Travaux pratiques présentiels, les Travaux individuels non présentiels tout au long de la formation.

Acquis d'apprentissage :

A la validation de ce module, l'étudiant sera capable de :

	Acquis d'apprentissage	Niveau d'approfondissement (*)
AA1	Identifier expliquer les concepts clés du machine Learning	1 & 2 & 3
AA2	Différencier les principales phases d'un projet machine Learning	2
AA3	Préparer les données en utilisant les différentes librairies Python.	2 & 3
AA4	Analyser, créer et évaluer des modèles linéaires.	3 & 5 & 6
AA5	Appliquer et évaluer des méthodes basées sur des relations géométriques entre les données	3 & 5
AA6	Créer et évaluer des modèles basés sur les méthodes de théorie de graphe	5 & 6

* : (1 : Mémoriser, 2 : Comprendre, 3 : Appliquer, 4 : Analyser, 5 : évaluer, 6 : Créer).

Contenu détaillé :

Chapitre 0 : Définition des concepts clés

- Situer la science des données (DS) et l'Intelligence Artificielle (IA).
- Distinguer la différence entre DS et IA.
- Définir les attentes des entreprises de l'IA et la science des données.
- Différencier les étapes d'un projet Data Science.

Situation d'apprentissage

- Cours intégré

Durée	<ul style="list-style-type: none"> • 3h
Rendu	<ul style="list-style-type: none"> •

Chapitre 1 : Processus de préparation des données

- Manipuler les différentes librairies Python destinées à la science des données à savoir Numpy, scipy, pandas, sklearn.
- Définir le processus de préparation des données :
 - Chargement de données sous forme de DataFrames dans le logiciel.
 - Visualisation de données
 - Nettoyage de données
 - Exploration des données
 - Transformation de données
 - Stockage de données...
- Générer des données de dimension réduite en utilisant la méthode « Analyse en composantes principales ».

Situation d'apprentissage	<ul style="list-style-type: none"> • Cours intégré
Durée	<ul style="list-style-type: none"> • 9h
Rendu	<ul style="list-style-type: none"> • Notebook 1 « Préparation des données »
	<ul style="list-style-type: none"> • Lab 1 : préparation des données

Chapitre 2 : Modèles génératifs

- Appliquer des modèles non supervisés exploitant les relations géométriques entre les données à travers l’algorithme **K-means**
- Appliquer des modèles probabilistes exploitant la distribution des probabilités des données à travers l’algorithme GMM (Gaussian Mixture Models)
- Evaluer les performances des modèles génératifs

Situation d'apprentissage	<ul style="list-style-type: none"> • Cours intégré
Durée	<ul style="list-style-type: none"> • 6h
	<ul style="list-style-type: none"> • Notebook 2 « Modèles génératifs »
	<ul style="list-style-type: none"> •

Chapitre 3 : Modèles discriminatif

- Appliquer des modèles supervisés exploitant les relations géométriques entre les données à travers :
 - K-plus proche voisin KNN
 - Support Vector Machines SVM
 - Régression Logistique
- Evaluer les performances de la modélisation en utilisant les indicateurs de performance (Matrice de confusion, courbe ROC...).

Situation d'apprentissage	<ul style="list-style-type: none"> • Cours intégré
Durée	<ul style="list-style-type: none"> • 12 h
	<ul style="list-style-type: none"> • Notebook 3 « Méthodes supervisées de classification »
	<ul style="list-style-type: none"> • Lab 2

Chapitre 4 : Méthodes Ensemblistes

- Créer des modèles basés sur les graphes de décision à travers les arbres de décision
- Comprendre la différence entre le Bagging et le Boosting.
- Créer des modèles de Bagging à travers « Random Forest »
- Evaluer les performances de méthodes ensemblistes en utilisant les indicateurs de performances (Matrice de confusion, courbe ROC...)

Situation d'apprentissage	<ul style="list-style-type: none"> • Cours intégré
Durée	<ul style="list-style-type: none"> • 9h
	<ul style="list-style-type: none"> • Notebook 4 « Méthodes de théorie de graphe »
	<ul style="list-style-type: none"> • Lab 2




Evaluation :

	Report/ Homework	Présentation	TP	Project
Identifier expliquer les concepts clés du machine Learning		X		
Différencier les principales phases d'un projet machine Learning				X
Préparer les données en utilisant les différentes librairies Python.	X		X	X
Analyser, créer et évaluer des modèles linéaires.	X		X	X

Appliquer et évaluer des méthodes basées sur des relations géométriques entre les données	X		X	X
Créer et évaluer des modèles basés sur les méthodes de théorie de graphe	X		X	X

S

Références :

- Introduction to Statistical Learning, Corrected 7th Printing. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, 2013. Springer. Available as PDF: [ISLR Seventh Printing.pdf](#) 
- Elements of Statistical Learning, 2nd ed. Trevor Hastie, Robert Tibshirani and Jerome Friedman, 2009. Springer. Available as PDF: [ESLII_print12.pdf](#) 
- Foundations of Data Science, Avrim Blum, John Hopcroft, and Ravindran Kannan, Thursday 4th January, 2018. Available as PDF: [BlumHopcraftKannan.pdf](#) 
- Videos from ISL/ESL. [Available online and linked from ISL website. \(Links to an external site.\)](#)
- A computer (Mac, Linux, or Windows) with an up-to-date operating system and a modern web browser and your favorite code editor (ViM, emacs, Notepad++,...) and ability to install Anaconda <https://www.anaconda.com/download> (Links to an external site.), to program in Python and R.