Module : Machine Learning Appliqué

Dernière mise à jour : 30/03/2022

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

|  |  |
| --- | --- |
| **Responsable Module** | Jihen Hlel |
| **Enseignants – Intervenants** | Ghada Setti, Fares Hasni, Jihen Jebri, Olfa Laayouni, Oumayma Guesmi, Rahma Dhaouadi, Shema Essadi , Sarra zouari, Soumaya Nheri, wiem baazouzi, Wiem Trabelsi, Jihen Hlel |
| **Unité pédagogique** | IL |
| **Unité d’enseignement** |  |
| **Prérequis** | Sys. De Gestion de Bases de Données Calcul scientifique  Méthodes numériques pour l'ingénieur  Techniques d'estimation pour l'ingénieur |
| **Niveaux et Options** | 4ème BI  4ème TWIN  4ème SAE  4ème INFINI |

# 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 Appliqué) = CC\*50% + Projet \*50%.**

* Projet : Le projet consiste à une implémentation d’une solution intelligente.
* Le contrôle continu : La moyenne des Travaux pratiques tout au long de la formation (20%), le DS (30%).

**Acquis d’apprentissage :**

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

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

|  |  |  |
| --- | --- | --- |
|  | 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 de regression. | 3 & 5 & 6 |
| AA5 | Appliquer et évaluer des méthodes de classification | 3 & 5 |
| AA6 | Créer et évaluer des modèles basés sur les méthodes de segmentation | 5 & 6 |

**Contenu détaillé :**

**Chapitre 1 : Définition des concepts clés**

* Identifier les concepts de bases de l’apprentissage automatique
* 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 | * 3h |
| Rendu | * Cours |

**Chapitre 2 : 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 visualisation et de préparation des données :
  + Chargement de données sous forme de DataFrames dans le logiciel.
  + Visualisation de données (EDA)
  + Nettoyage de données
  + Corrélation
  + Outliers
  + Transformation de données

|  |  |
| --- | --- |
| Situation d’apprentissage | * Cours intégré + suivi |
| Durée | * 6h |
| Rendu | * Notebook 1 « Préparation des données » |

**Chapitre 3 : Modèles linéaires**

* Créer des modèles exploitant la relation de linéarité entre les données à travers la :
* Régression linéaire Simple et Multiple
* Régression polynomiale
* Evaluer les méthodes de régression en utilisant des méthodes de statistiques telles que R carré, Erreur quadratique moyenne (MSE)…

|  |  |
| --- | --- |
| Situation d’apprentissage | * Cours intégré |
| Durée | * 6h |

|  |  |
| --- | --- |
|  | * Notebook 2 « Régression Linéaire » |
|  | * Notebook 3 « Régression polynomiale » |

**Chapitre IV : Méthodes de classification**

* Appliquer des modèles supervisés exploitant les relations géométriques entre les données à travers :
* K Nearest Neighbors (KNN)
* Support Vector Machines (SVM)
* Arbre de décision
* XGboost
* Evaluer les performances de méthodes de classification en utilisant les métriques de performances (Accuracy, F1-score, Matrice de confusion, courbe ROC…)

|  |  |
| --- | --- |
| Situation d’apprentissage | * Cours intégré |
| Durée | * 12h |
|  | * Notebook 4,5,6 « Méthodes supervisées de classification » |

**Chapitre V : Méthodes de segmentationChers collègues,**

**Je tiens tout d’abord à m’excuser de ne pas vous avoir envoyé plus tôt le mail concernant la réunion de demain. Cela était dû au fait que je n’avais pas encore reçu la liste mise à jour des affectations.**

**Je profite de ce message pour vous transmettre à nouveau les informations relatives à la réunion prévue demain. Je vous invite à les consulter afin de bien vous préparer.**

**N’hésitez pas à me faire savoir si vous avez des questions ou des précisions à demander.**

**Cordialement,**

* Appliquer des modèles non supervisés exploitant les relations géométriques entre les données à travers :
* Méthodes de Clustering : k-means
* Analyse par Composante Principale (ACP)
* Evaluer les performances de méthodes de segmentation en utilisant les

Indicateurs de performances (Silhouette score, Coefficient de Jaccard…)

|  |  |
| --- | --- |
| Situation d’apprentissage | * Cours intégré |
| Durée | * 9h |
|  | * Notebook 7 et 8 |

**Les systèmes de recommandation**

* Distinguer entre les différents types des systèmes de recommandation
* Appliquer les algorithmes des systèmes de recommandation

|  |  |
| --- | --- |
| Situation d’apprentissage | * Cours intégré |
| Durée | * 3h |
|  | * Notebook 9 |

**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** |  | **X** |
| Préparer les données en utilisant les différentes librairies  Python. | **X** |  | **X** | **X** |
| Analyser, créer et évaluer des modèles de régression. | **X** | **X** | **X** | **X** |
| Appliquer et évaluer des modèles de classification | **X** |  | **X** | **X** |
| Créer et évaluer des modèles basés sur les modèles de segmentation | **X** |  | **X** | **X** |

**Références :**

* Introduction to Statistical Learning, Corrected 7th Printing. Gareth James, Daniela Witten, Trevor Hastie and

5



Robert Tibshirani, 2013. Springer. Available as PDF: [ISLR Seventh Printing.pdf](https://uwsto.instructure.com/courses/255375/files/12679545/download?wrap=1)

* Elements of Statistical Learning, 2nd ed. Trevor Hastie, Robert Tibshirani and Jerome Friedman, 2009. Springer. Available as PDF: [ESLII\_print12.pdf ](https://uwsto.instructure.com/courses/255375/files/12679544/download?wrap=1)
* Foundations of Data Science, Avrim Blum, John Hopcroft, and Ravindran Kannan, Thursday 4th January, 2018. Available as PDF: [BlumHopcraftKannan.pdf ](https://uwsto.instructure.com/courses/255375/files/12679447/download?wrap=1)
* Videos from ISL/ESL. [Available online and linked from ISL website. (Links to an external site.)](http://www-bcf.usc.edu/~gareth/ISL/)
* 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.),](https://www.anaconda.com/download) to program in Python and R.

6