

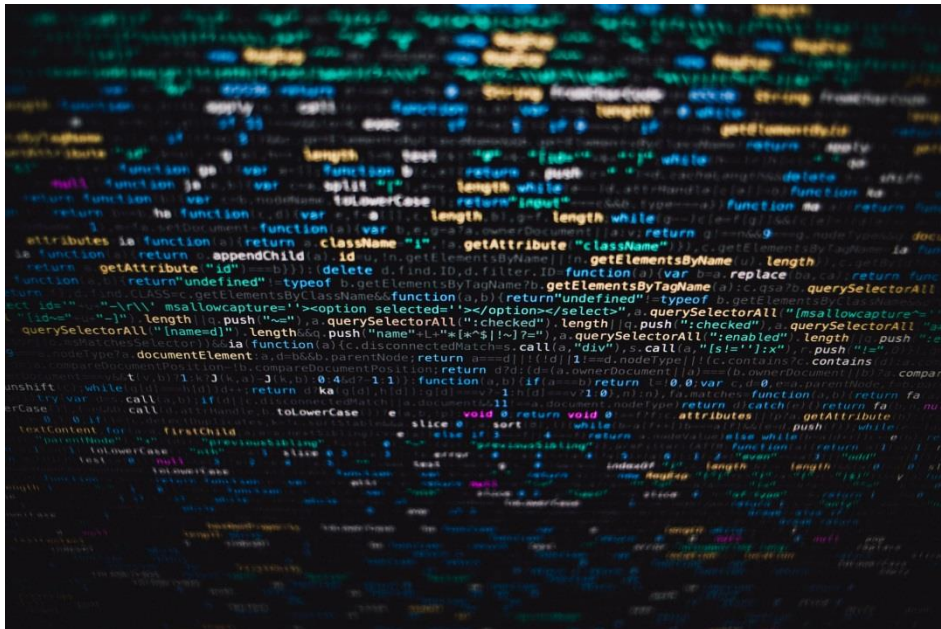
## Département Informatique

MASTER EN SCIENCES ET TECHNIQUES



SYSTÈMES D'INFORMATION DÉCISIONNELS ET IMAGERIE

FACULTÉ DES SCIENCES ET TECHNIQUES ERRACHIDIA



# TP2 - Machine Learning Prétraitement et visualisation de données

ZEKKOURI Hassan

Vendredi 06 Decembre 2019

Responsable du module :

Prof. OUANAN Mohamed

# REMERCIEMENT

---

Nous tenons à vous remercier monsieur **Mohamed OUANAN** pour votre formation et vos services.

Nous sommes également reconnaissant de nous avoir donné l'occasion de s'ouvrir sur un nouvel aspect de technologie qui est le Machine Learning et y mettre en œuvre nos compétences dont nous avons obtenu au cours du module.

Hassan ZEKKOURI

---

# PLAN

## Contents

<i>TP2 - Machine Learning</i> .....	1
<i>Prétraitement et visualisation</i> .....	1
<i>de données</i> .....	1
<i>REMERCIEMENT</i> .....	2
<b>INTRODUCTION</b> .....	4
<b>I. Scikit-learn</b> .....	5
<b>II. Jeux de données en Scikit-Learn</b> .....	5
<b>III. TP1/TP2</b> .....	5
<b>IV. Jupyter notebook   TP2</b> .....	6
Partie 1 : Imputation de données manquantes .....	6
Partie 2 : Manipuler les variables catégoriques dans un jeu de données .....	6
Partie 3 : Normalisation de données .....	6
<b>FIN</b> .....	6

# INTRODUCTION

## ➤ Objectif :

Dans ce TP nous allons découvrir le module scikit-learn qui spécialisé en Machine Learning et on va essayer de découvrir et tester ses fonctionnalités !

# I. Scikit-learn

Scikit-learn est un logiciel écrit en Python, qui nécessite l'installation préalable du langage Python et des librairies NumPy1 et SciPy2 (pour le calcul scientifique), dans des versions qui doivent vérifier certaines contraintes de compatibilité. Le plus simple est d'installer une distribution de Python complète, comme Anaconda3, qui comprend la plupart des librairies courantes développées en Python, dont les trois citées plus haut. Le site officiel du logiciel Scikit-learn est :

<http://scikit-learn.org/stable/index.html>

La documentation en ligne est complète, et devra être consultée chaque fois que nécessaire :

<http://scikit-learn.org/stable/documentation.html>

Des tutoriaux sont disponibles à l'adresse suivante :

<http://scikit-learn.org/stable/tutorial/index.html>

## II. Jeux de données en Scikit-Learn

Un certain nombre de jeux de données sont disponibles dans scikit-learn. Il est également possible de générer des données artificielles ou de récupérer des données externes.

Documentation relative au chargement de jeux de données :

<http://scikit-learn.org/stable/datasets/>

Les jeux de données disponibles dans scikit-learn sont : iris, boston, diabetes, digits, linnerud, sample images, 20newsgroups. Chacun de ces jeux de données se récupère à l'aide de la commande `load_nom-jeu` qu'il faut dans un premier temps charger. Par exemple, pour récupérer le jeu iris :

1. Importation d'un jeu de données

```
Entrée [6]: from sklearn.datasets import load_iris  
iris = load_iris()
```

Alors la variable iris contient maintenant notre jeu de données.

## III. TP1/TP2

Pour bien documenter le travail au fur et à mesure nous avons choisie de travailler avec l'utile jupyter-notebook (IDE).

Tout le travail réalisé peut être consulté en ligne sur :

<https://notebooks.azure.com/DgrinderHZ/projects/machine-learning>

il suffit de l'ouvrir sur Azur Microsoft ou la télécharger pour Anaconda Jupyter.

# IV. Jupyter notebook | TP2

## Partie 1 : Imputation de données manquantes

Lien de notebook :

<https://machinelearning-dgrinderhz.notebooks.azure.com/j/notebooks/Master%20MST%20SIDI%20-%20FST%20Errachidia/TP2/Machine%20Learning%20-%20TP2%20-%20Partie%201%20.ipynb>

## Partie 2 : Manipuler les variables catégoriques dans un jeu de données

Lien de notebook :

<https://machinelearning-dgrinderhz.notebooks.azure.com/j/notebooks/Master%20MST%20SIDI%20-%20FST%20Errachidia/TP2/Machine%20Learning%20-%20TP2%20-%20Partie%202.ipynb>

## Partie 3 : Normalisation de données

Lien de notebook :

<https://machinelearning-dgrinderhz.notebooks.azure.com/j/notebooks/Master%20MST%20SIDI%20-%20FST%20Errachidia/TP2/Machine%20Learning%20-%20TP2%20-%20Partie%203.ipynb>

# FIN

Merci