



UNIVERSITÉ DE SHERBROOKE

PROJET IFT 712  
INFORMATIQUE - DEUXIÈME CYCLE

---

# MÉTHODES DE CLASSIFICATION PAR SKLEARN

---

*Réalisé par :*  
THOUIN Kevin  
BENNANI Kaoutar

*Supervisé par :*  
Pr. Pierre-Marc JODOIN

Année universitaire 2019 - 2020

# Table des matières

<b>Table des figures</b>	<b>2</b>
<b>1 Présentation du projet</b>	<b>3</b>
1 Présentation . . . . .	3
2 Choix de la base de données . . . . .	3
3 Choix du design . . . . .	3
4 Gestion du projet . . . . .	4
<b>2 Les algorithmes et démarche utilisés</b>	<b>5</b>
1 Algorithmes . . . . .	5
1.1 Régression logistique . . . . .	5
1.2 SVM . . . . .	5
1.3 Réseaux de neurones . . . . .	5
1.4 Bagging . . . . .	5
1.5 AdaBoost . . . . .	5
2 Démarche scientifique . . . . .	5
<b>3 Analyse des résultats</b>	<b>6</b>
1 . . . . .	6

## **Table des figures**

# Chapitre 1

## Présentation du projet

### 1 Présentation

Ce projet de session fait partie des travaux du cours IFT712, il a pour objectif de tester quelques méthodes de classification sur une base de données Kaggle ([www.kaggle.com](http://www.kaggle.com)) avec la bibliothèque Sklearn .

### 2 Choix de la base de données

Nous avons choisi comme base de données : "Wine Dataset".

A propos de la base de données, les données sont le résultat d'une analyse chimique de vins cultivés dans une des régions d'Italie mais issus de trois cultivars différents.

Parmi ses caracteristiques :

- 178 instances
- 13 variables
- les valeurs des attributs sont des Integer et des Float
- Pas de valeurs manquantes
- Généralement utilisé pour les tâches de classification

### 3 Choix du design

Le projet contient 8 classes en total.

- La classe 'Classification\_main.py'
- La classe 'Classification\_neural\_net.py' : Décrit l'algorithme de classification par réseau de neurones.
- La classe 'Classification\_logistique.py' : S'agit de la classification par régression logistique.
- La classe 'Classification\_bagging.py' : Classification par l'algorithme Bagging.
- La classe 'Classification\_adaboost.py' : Classification par l'algorithme AdaBoost.
- La classe 'Classification\_svm.py' : Il s'agit du code correspondant au SVM.

- La classe 'Classification\_hyperparameter.py' : Il s'agit des techniques de la recherche des hyperparamètres pour chacun des algorithmes de classification.
- La classe 'Classification\_io.py' : Cette classe crée des données d'entraînement et de test à partir du "Wine Dataset" et a une fonction permettant l'affichage graphique.

## 4 Gestion du projet

Pour la réalisation du projet, nous avons utilisé le gestionnaire de version de code "git" via la plateforme "gitHub"

# **Chapitre 2**

## **Les algorithmes et démarche utilisés**

Cette partie consiste à présenter les algorithmes utilisés, ainsi que la démarche scientifique suivie.

### **1 Algorithmes**

#### **1.1 Régression logistique**

#### **1.2 SVM**

#### **1.3 Réseaux de neurones**

#### **1.4 Bagging**

#### **1.5 AdaBoost**

### **2 Démarche scientifique**

# Chapitre 3

## Analyse des résultats

Dans cette partie nous allons analyser les résultats de chacun des algorithmes.

### 1