**Fiche de Challenge Machine Learning avec Scikit-learn**

**Délais**

* **Deadline :** Lundi prochain

**Jour 6 : Introduction au machine learning supervisé**

* **Objectif :** Comprendre les concepts de base du machine learning supervisé.
* **Activités :**
  + Différencier apprentissage supervisé et non supervisé.
  + Découvrir les étapes de construction d’un modèle (préparation des données, entraînement, test).
* **Challenge :** Préparer un dataset pour le machine learning en créant des features (caractéristiques) et des labels (étiquettes).
  + **Ressource :**
    - Supervised vs Unsupervised Learning (documentation scikit-learn)

**Jour 7 : Régression linéaire**

* **Objectif :** Créer un modèle de régression linéaire simple.
* **Activités :**
  + Créer un modèle de régression linéaire avec scikit-learn.
  + Séparer les données en ensemble d’entraînement et de test.
* **Challenge :** Entraîner un modèle de régression linéaire sur un dataset de prédiction de prix de maisons.
  + **Ressources :**
    - Linear Regression Tutorial
    - California Housing Prices Dataset pour un exemple de dataset.

**Jour 8 : Classification binaire**

* **Objectif :** Apprendre à classifier des données avec des algorithmes simples.
* **Activités :**
  + Utiliser la régression logistique ou KNN pour des problèmes de classification.
  + Comprendre les métriques d’évaluation (accuracy, matrice de confusion, précision, rappel).
* **Challenge :** Construire un modèle de régression logistique pour prédire si un passager du Titanic a survécu ou non.
  + **Ressources :**
    - Titanic Dataset (Kaggle)
    - Logistic Regression Tutorial

**Jour 9 : Prétraitement des données**

* **Objectif :** Comprendre l’importance de la normalisation et de la standardisation.
* **Activités :**
  + Normaliser et standardiser les données.
  + Utiliser des pipelines pour automatiser le prétraitement.
* **Challenge :** Appliquer la normalisation et la standardisation à un dataset, puis observer les différences de performance.
  + **Ressources :**
    - Data Preprocessing with Pipelines
    - StandardScaler Documentation

**Jour 10 : Algorithmes avancés de classification**

* **Objectif :** Découvrir des modèles plus performants pour la classification.
* **Activités :**
  + Introduction aux arbres de décision, Random Forest et SVM.
  + Entraîner un modèle Random Forest pour un dataset.
* **Challenge :** Comparer les performances de plusieurs modèles (régression logistique, KNN, Random Forest) sur un problème donné.
  + **Ressources :**
    - Random Forest Documentation
    - SVM Tutorial