

## **Data Science**

### **Mini-projet N°2 : Machine Learning (ML)**

*Pr. Abdelhamid FADIL*

#### **Objectifs et données du projet ML :**

L'objectif du projet ML est de mettre en pratique les notions traitées dans le cours en les appliquant sur une problématique réelle à travers la mise en œuvre d'un pipeline ML complet allant de la présentation et la compréhension de la problématique au déploiement de la solution ML.

La dataset du mini-projet (RT-IoT2022.csv) est un jeu de données issu d'une infrastructure IoT réelle, qui combine du trafic normal (usage légitime d'appareils IoT) et du trafic malveillant «adverse» (scénarios d'attaque : SSH brute-force, DDoS, scans Nmap, etc.). Il contient 123 117 instances (observations) et 83 caractéristiques (features). Les données sont de type mixte (réel / catégorique), adaptées à des tâches variées comme la classification, le clustering, la détection d'anomalies...

L'objectif principal du dataset est de développer et évaluer des modèles de détection d'intrusion (IDS) dans des environnements IoT réels.

Dans ce projet, on s'intéresse à la classification des types d'attaques.

Pour plus d'informations : <https://archive.ics.uci.edu/dataset/942/rt-iot2022>

#### **Questions et barème de notation :**

N°	Axe	Intitulé	Points
1	Compréhension de la problématique	- Présentation de la problématique et le rôle du ML dans l'étude du thème (détection des attaques)	3
2	Analyse exploratoire des données (EDA)	Compréhension des données, compréhension des variables, du target, exploration des données (types des variables, spécification de la variable Target, corrélations, répartition statistique des variables, tendances...) avec commentaires et conclusions tirées de l'EDA	3
3	Pré-processing des données	Pré-processing des données (traitement des données manquantes, encoding, normalisation, outliers...) & Feature Selection avec commentaires et justification des choix effectués	3
4	Construction des modèles ML	Application des algorithmes ML (au moins 10) avec la présentation des résultats de chaque algorithme (indicateurs d'évaluation, graphes de performance...)	3
5	Tuning des modèles candidats et choix du modèle final	- Choix de 2 modèles ML les plus adéquats (argumentation, indicateur d'évaluation...) puis tuning de ces modèles candidats - Choix du modèle final avec la mention des résultats finaux du test de ce modèle	3
6	Déploiement du modèle	- Préparation du déploiement du projet avec Streamlit (ou autre outil) - Déploiement de l'application sur le cloud (Streamlit Cloud Community)	3

		- Test du modèle déployé avec des nouvelles données	
7	Présentation du projet et ergonomie de l'application web	2 pts seront réservés pour l'évaluation de la forme de la présentation du travail, du rapport et de la qualité des commentaires ainsi que de l'ergonomie de l'application web de déploiement du modèle	2

**Outils à utiliser :** Les participants sont libres à choisir les outils et les librairies qu'ils souhaitent utiliser (scikit-learn, pycaret, knime...)

**Deadline de remise du projet :** Dimanche 4 janvier 2026 à 23H

**Livrables du projet :**

Les livrables sont à déposer sur la **plateforme Teams** dans l'espace « **Devoirs** » du mini-projet 2 sous forme d'**un seul fichier zippé** (rar ou zip) contenant les livrables suivants ;

1. Rapport synthétique (pdf) décrivant le processus de mise en œuvre du projet ML et les réponses aux questions citées ci-dessus avec la mention du **lien URL de l'application déployée** et le **lien vers votre espace github** où sont stockés les fichiers de l'application déployée (script python, fichier.pkl...)
2. Fichier notebook complet avec des commentaires traçant les différentes activités effectuées pour la réalisation du projet en donnant l'ensemble des résultats (traitement de données, construction des modèles, graphiques et métriques d'évaluation, choix effectués, graphiques...)