



TÉLÉCOM PHYSIQUE STRASBOURG

RAPPORT DE PROTECTION DES DONNÉES

Analyse de données cyberphysiques sur la distribution d'eau

Nathan CERISARA

Clément DESBERG

Ysée JACQUET

Lucas LEVY

21 janvier 2026

Table des matières

1	Données	1
1.1	Visualisation des données	1
1.1.1	Présentation générale des données	1
1.1.2	Analyse des données	1
1.2	Prétraitement	1
2	Application des algorithmes	1
2.1	KNN	1
2.2	Random Forest	1
2.3	XGBoost	1
2.4	MLP	1
2.5	Modèles de transformers	1
2.5.1	Tab Transformer	1
2.5.2	FT Transformer	1
2.5.3	MLP avec attention	1
3	Évaluation	1

Table des figures

1	Matrice de corrélation des erreurs entre les modèles	2
---	--	---

Abréviations employées

- CNN : Convolutional Neural Network

Introduction

TO COMPLETE

1 Données

1.1 Visualisation des données

1.1.1 Présentation générale des données

1.1.2 Analyse des données

1.2 Prétraitement

2 Application des algorithmes

2.1 KNN

2.2 Random Forest

2.3 XGBoost

2.4 MLP

2.5 Modèles de transformers

2.5.1 Tab Transformer

2.5.2 FT Transformer

2.5.3 MLP avec attention

3 Évaluation

Rang	Modèle	Précision	F1 (macro)	Précision équilibrée	MCC	Temps (s)
1	small_knn	0.9774	0.8007	0.8235	0.9402	0.0
2	small_tab_transformer	0.9683	0.8001	0.8250	0.9188	40.7
3	small_attention_mlp	0.9048	0.7259	0.8106	0.7946	35.2
4	small_mlp	0.8786	0.6889	0.7988	0.7515	22.7
5	small_ft_transformer	0.8401	0.6593	0.7970	0.7004	56.1
6	small_xgboost	0.8676	0.6412	0.7906	0.7287	0.6
7	small_random_forest	0.6101	0.4687	0.7242	0.4779	0.2

TABLE 1 – Comparaison des expériences sur le jeu de données physiques physical_small



FIGURE 1 – Matrice de corrélation des erreurs entre les modèles

Entrée	Label réel	Nombre d'erreurs	Prédiction majoritaire
5	normal	7	MITM
35	normal	7	MITM
36	scan	7	normal
210	normal	7	MITM
466	normal	7	DoS
598	normal	7	scan
768	normal	7	physical fault
848	normal	7	physical fault
864	normal	7	scan
892	normal	7	DoS
927	normal	7	MITM
1104	normal	7	MITM
1203	normal	7	scan
1257	normal	7	physical fault
1549	scan	7	normal
1559	normal	7	physical fault
1628	normal	7	MITM

TABLE 2 – Échantillons systématiquement mal classés lors des différentes exécutions

Conclusion