

# Présentation\_notebook

May 29, 2024

## 1 Waterflow Project

Développement d'un outil visant à estimer la potabilité de l'eau.

### 1.1 Data Exploration

pourcentage d'entrées avec un ph compris entre 6.5 et 8.5 : 40%

### 1.2 Préparation de la donnée

- Supression des Outliers
- Imputation des valeurs manquantes pas la médiane
- Oversampling sur la classe 1 Potable
- Robust Scaling

### 1.3 Feature Selection

SelectKBest

Elimine :

- ph
- turbidité

On décide de conserver ces variables.

Multicollinearity

VIF

### 1.4 Deep Learning

### 1.5 Data Engineering

### 1.6 Conclusion

- Dataset Inexploitable
- Peak d'accuracy du model à 0.68

L'outil n'est pas suffisamment fiable pour être utilisé pour estimer avec suffisamment de fiabilité la potabilité de l'eau.