

# Détection non supervisée de sons anormaux

Sylvain DEBIEU  
Geneviève FLEURY  
Quentin ROTT

Mentor :  
Thomas - DataScientest

Promotion : Bootcamp Data Scientist Mai-Juillet 2022



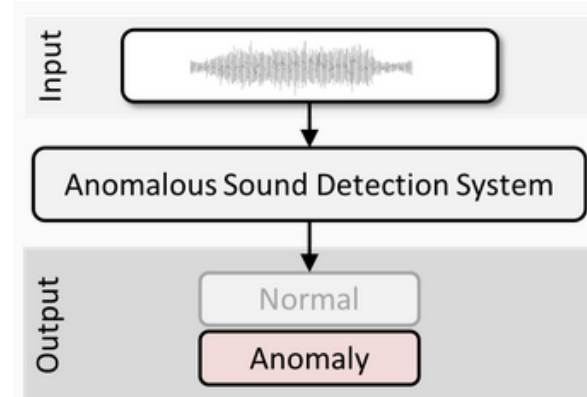
# Problématiques

- Amélioration de la maintenance préventive de machine industrielle
- Le son est un indicateur de l'état de la machine en temps réel
- L'humain n'est pas forcément le plus apte à juger de la conformité d'un son



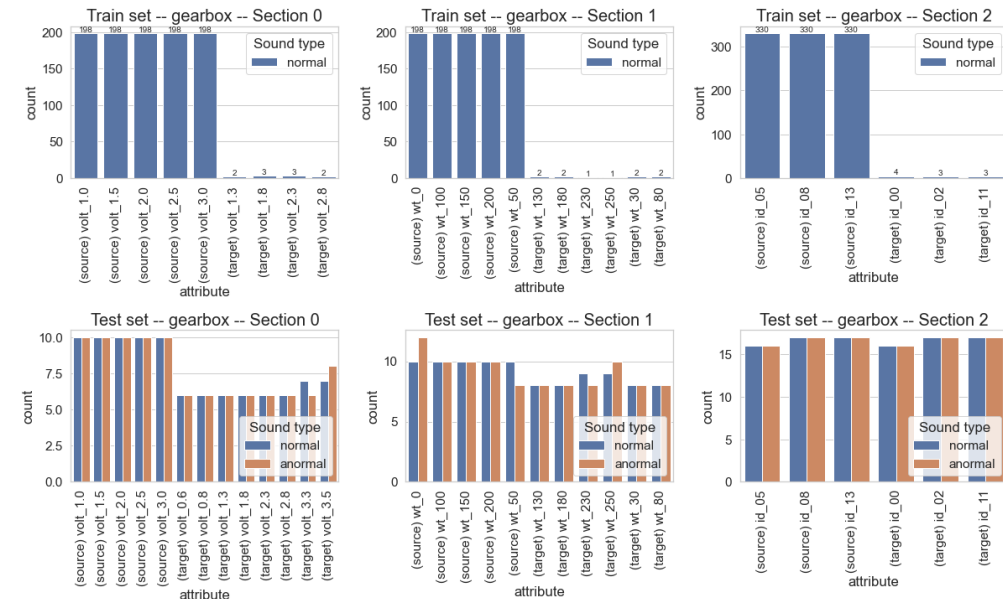
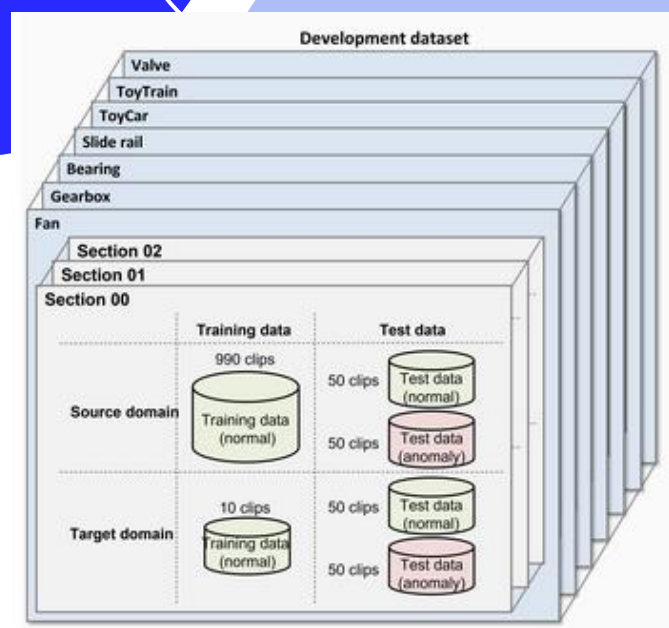
## Enjeux

- Concevoir un système de détection de sons anormaux (ASD)
- Peu de sons "anormaux" à disposition et beaucoup d'éléments à prendre en compte (bruits de fonds,...)
- Approche Deep-Learning non supervisé privilégiée



# Description du jeu de données

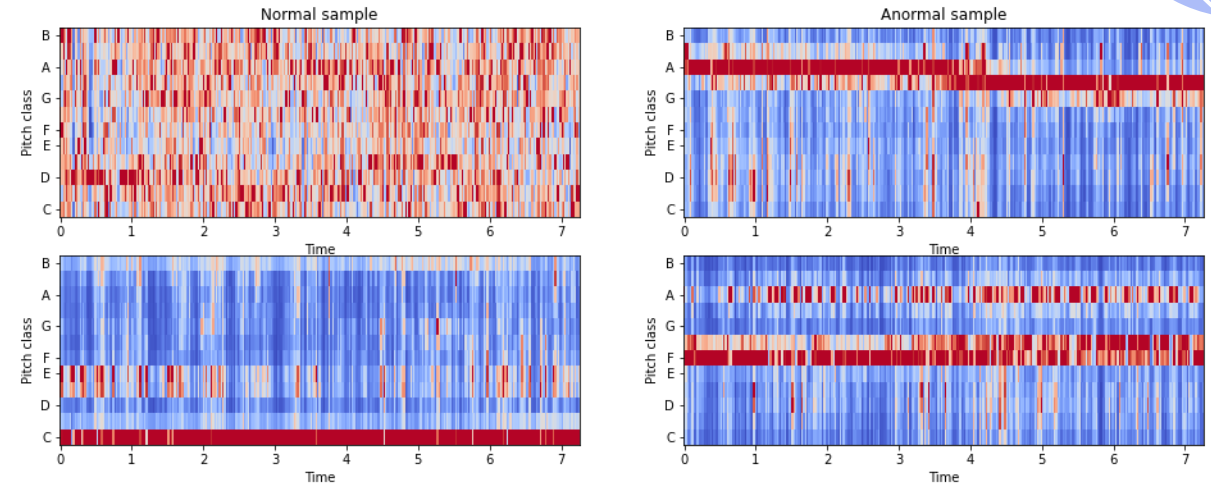
- 25200 clips sonores de 10 secondes en format .wav
- 7 types de machines ('fan', 'gearbox', 'bearing', 'slider', 'toy car', 'toy train', 'valve') mixé avec des bruits environnementaux provenant d'usines
- Par machines:
  - un jeu d'entraînement: 3000 clips normaux-domaine source /domaine cible(99/1)
  - un jeu de test : 300 clips normaux + 300 clips anormaux - domaine source /domaine cible (50/50)



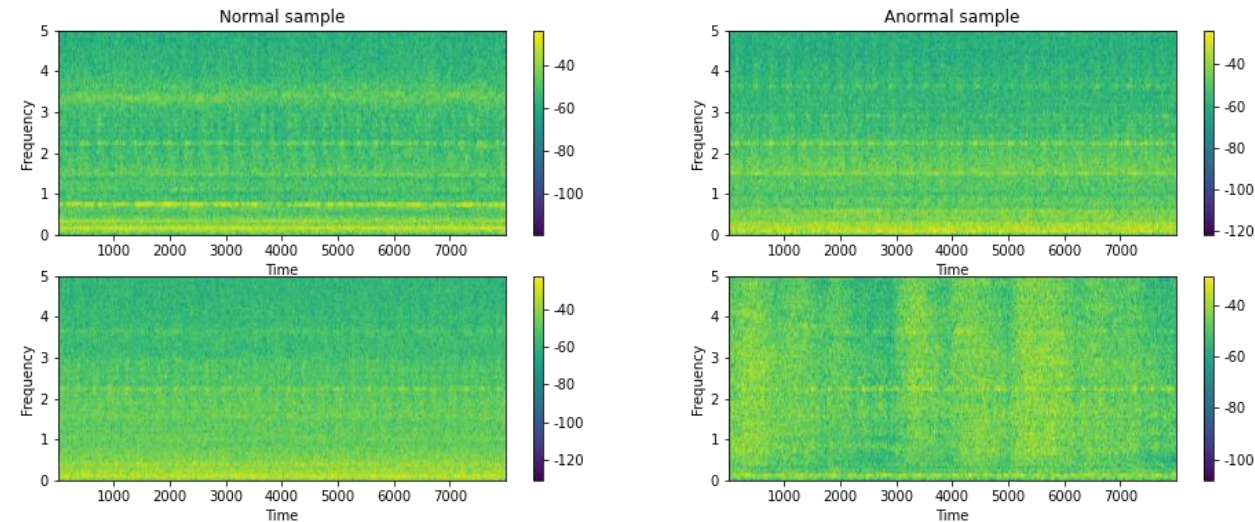
# Exploration des données

- Librairie spécifique utilisé: Librosa
- Ecoute des sons
- Extraction et Visualisation des attributs simples:
  - série temporelle
  - moyenne quadratique
  - spectre d'amplitude
- Extraction et Visualisation des attributs dans le domaine temporel:
  - Chromagramme (Tonalité en fonction du temps)
  - Spectrogramme (FFT, Mel, Wavelet)

Chromagrammes de gearbox volt\_1.0

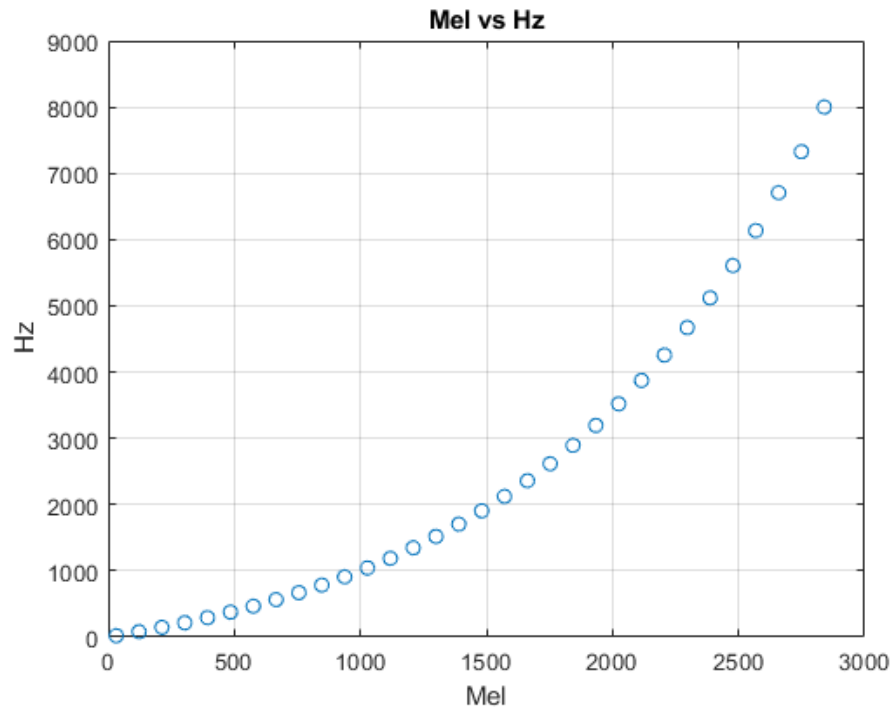


Approx. Coef. of the discrete wavelet transform (gearbox volt\_1.0)

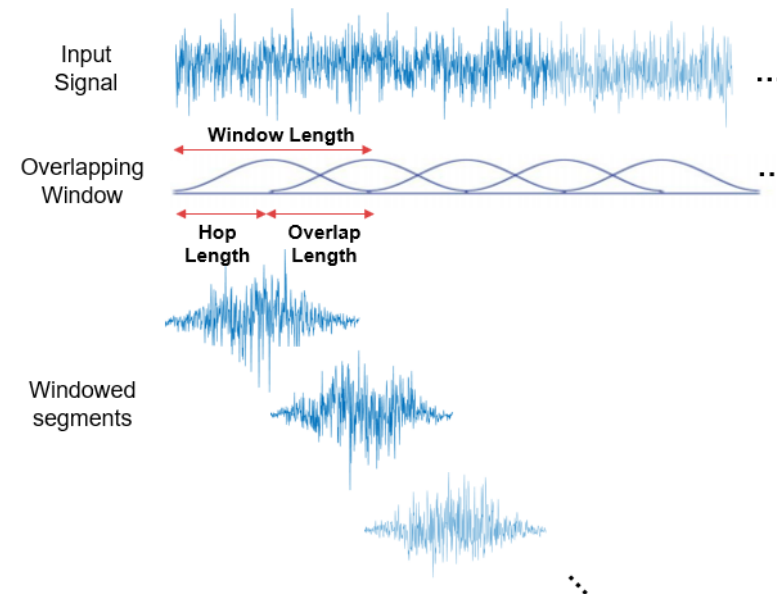


# Exploration des données

- Utilisation des Mel-spect: Échelle non linéaire de la perception des fréquences pour l'oreille humaine



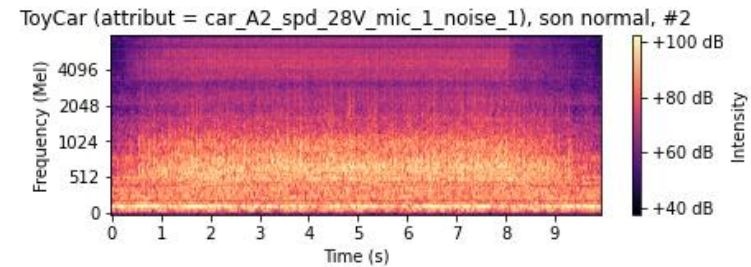
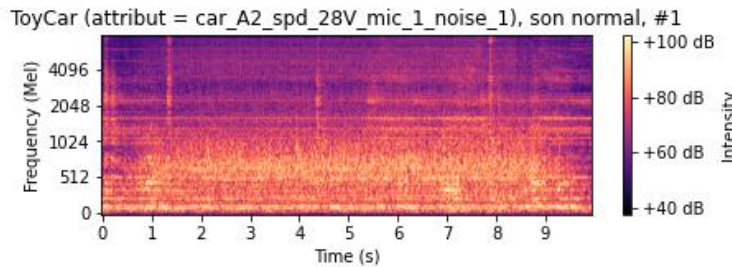
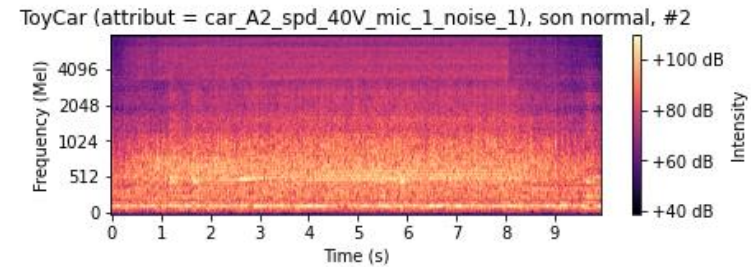
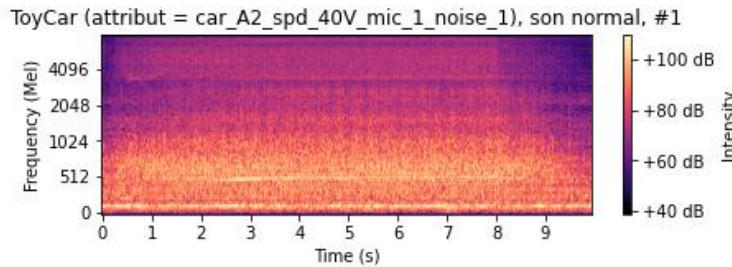
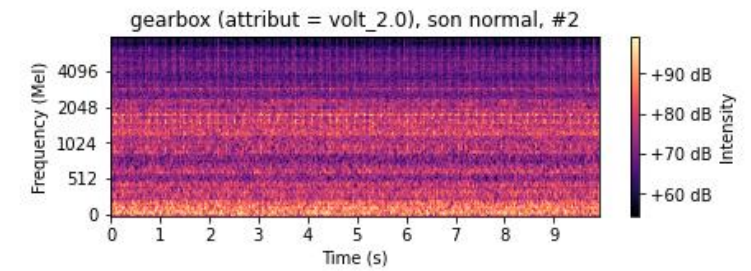
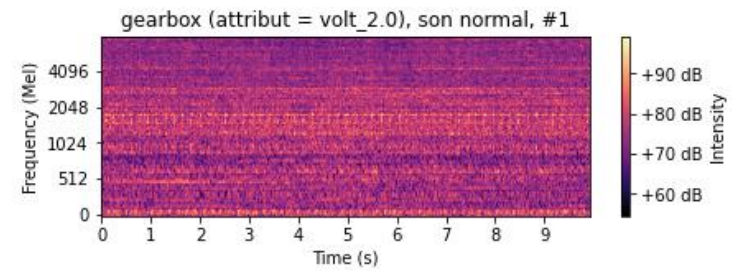
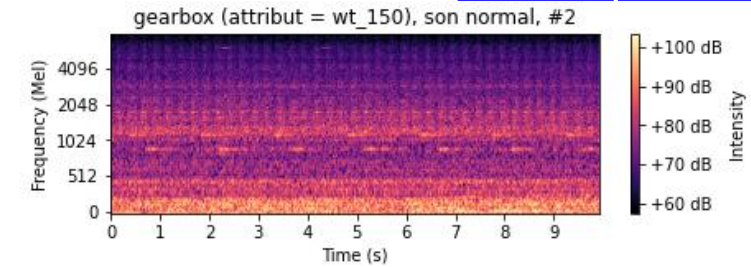
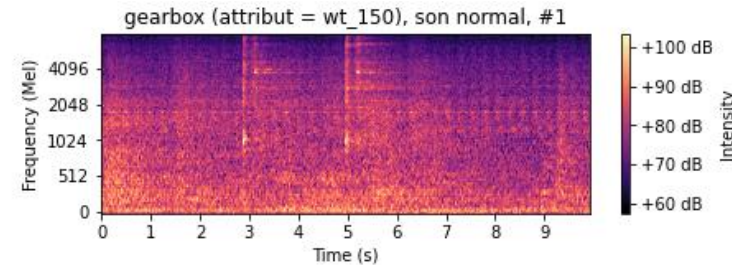
- Transformée de Fourier Discrète
  - Taille fenêtre : 1024 (64ms)
  - Saut de fenêtre : 512 (32ms)
  - Filtres mel: 128





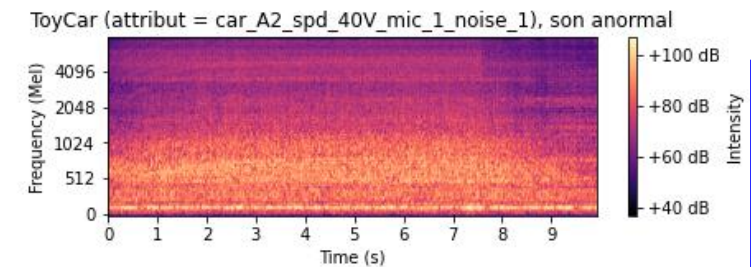
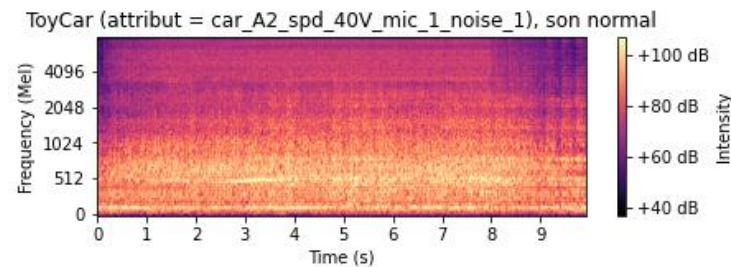
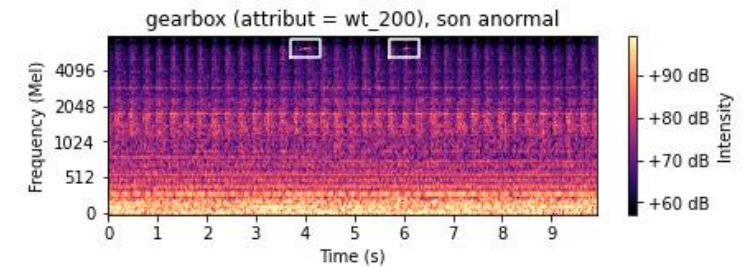
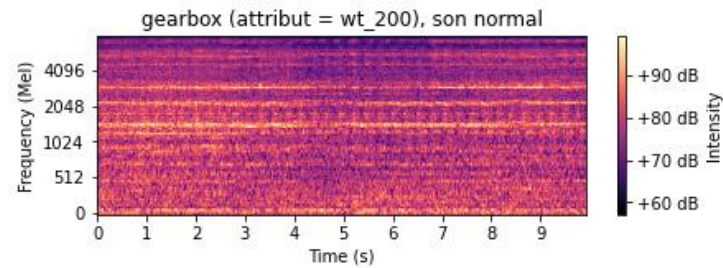
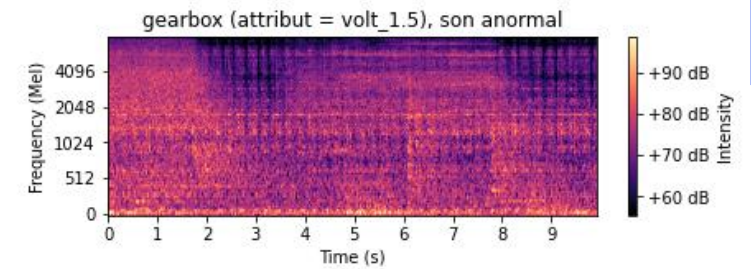
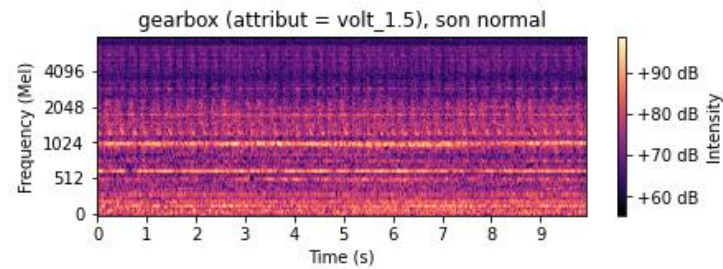
# Exploration des données

- Utilisation des Mel-spect: Échelle non linéaire de la perception des fréquences pour l'oreille humaine
- Comparaison entre sons normaux :
  - Des sons ayant le même attribut qui diffèrent
  - Des sons ayant un attribut différent qui se ressemblent



# Exploration des données

- Comparaison entre sons normaux/anormaux
- Conclusion EDA :
  - Des bruits environnementaux prédominants
  - Des anomalies visibles/non visibles
  - Des anomalies pouvant être ponctuelles



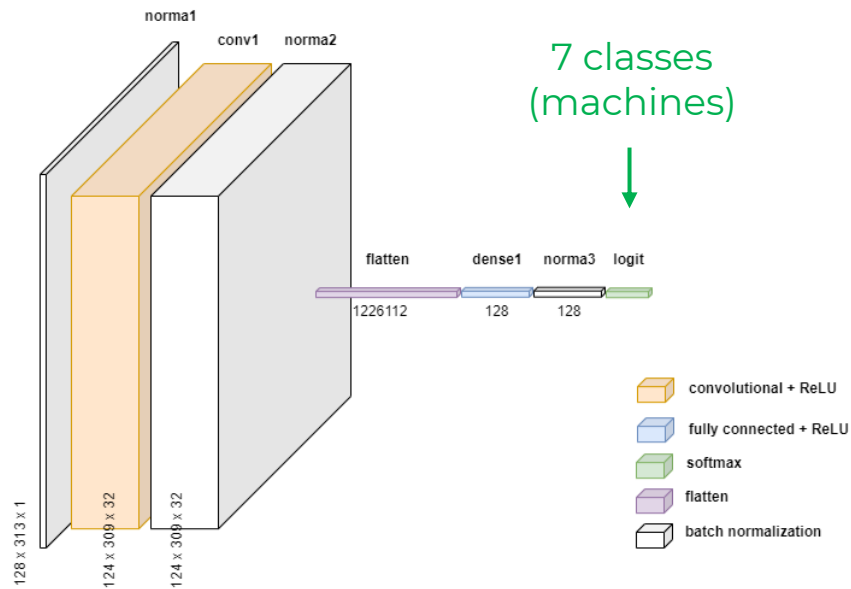
# Tableau des (principaux) modèles étudiés

	INPUT	DESCRIPTION	SCORE
ML supervisé	Spectrogrammes (échelle linéaire) aplatis en vecteurs	Réduction de dimension (SelectPercentile + PCA) + Gradient Boosting (xgboost)	$0.82 < \text{AUC} < 0.98$
DL supervisé	Spectrogrammes (échelle linéaire) 2D	CNN avec softmax / accuracy en métrique Modèle toutes machines	Arrive à séparer les différents type de machines $0.69 < \text{F1-score} < 0.98$
DL non supervisé	Mel-spectrogrammes 2D 128x313 (+ renormalization des pixels)	Embedding en vecteurs de taille 128 (FaceNet + triplet loss sur la section/la machine) + Random Forest	$0.45 < \text{AUC} < 0.78$
DL non supervisé	Mel-spectrogrammes 2D 128x313	Modèles DL basiques type CNN+DNN entraînés à reconnaître le type de machine	$0.45 < \text{AUC} < 0.78$
DL non supervisé	Mel-spectrogrammes 2D 128x313	Embedding via classifieurs de machines ou sections + autoencodeurs	$0.48 < \text{AUC} < 0.64$



# Description du modèle : DL non supervisé

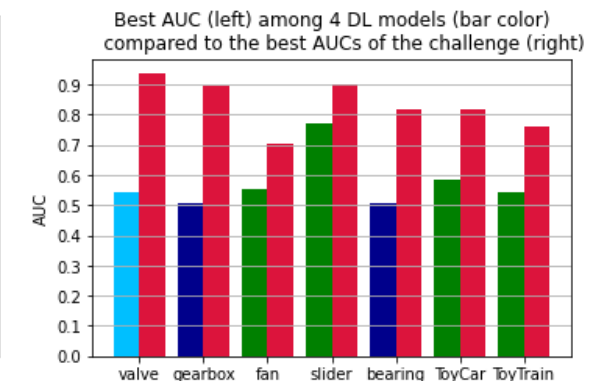
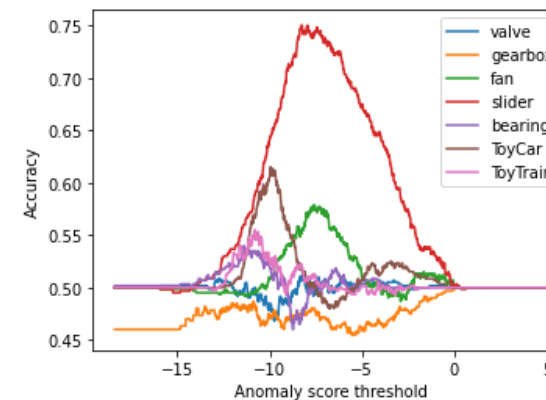
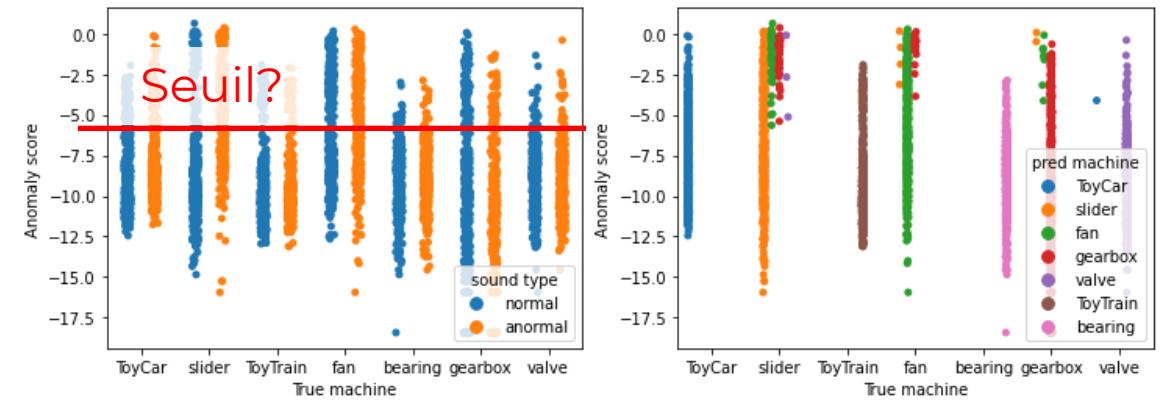
1) Construire un classifieur de machines



⇒ Output : probabilités  $p_{machine}$

2) Post-processing : extraire l'info normal/anormal

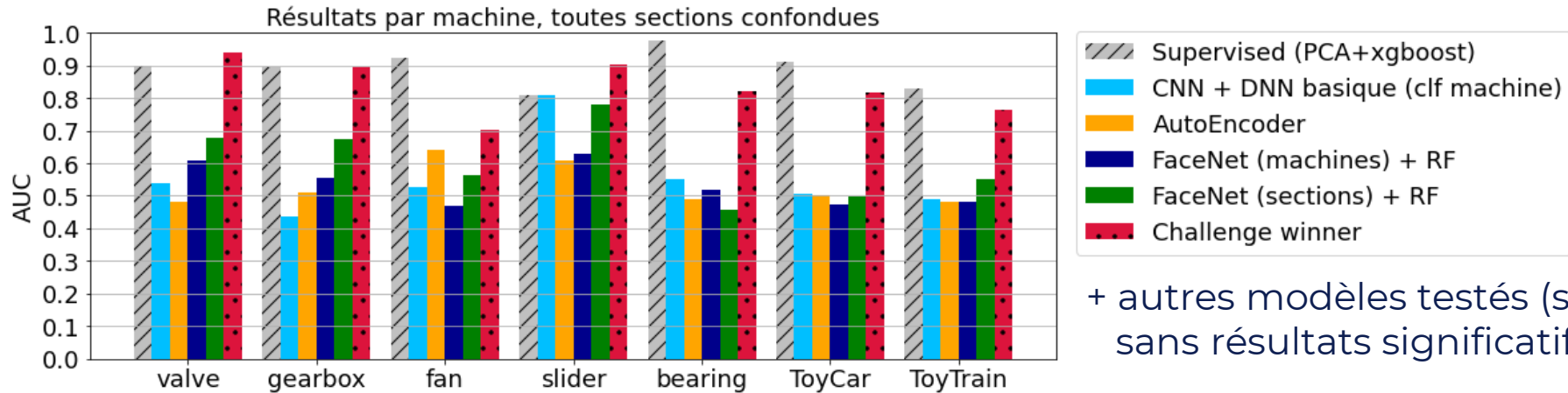
$$p_{anormal} \equiv 1 - p_{machine} \mapsto anomaly\ score \equiv g^{-1}(p_{anormal})$$



# Analyse rétrospective / Bilan

## Quelques résultats ...

- Preuve de faisabilité du problème supervisé en ML et DL :  
Bons résultats (au vu du temps limité et de l'étape de notre apprentissage)
- DL non supervisé: AUC dépendant du type de machines + sections

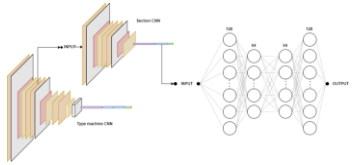


## ... mais très insatisfaisants. Pourquoi?

- Enregistrements bruyants : inadaptés pour auto-encodeurs
- Difficile de distinguer les anomalies à l'oreille/œil
- « Domain shift » : simplifier en considérant que le domaine « source »

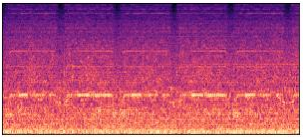
# Perspectives et suggestions d'amélioration

**Deep learning :** ➤ Approfondir l'approche FaceNet



- Classifieur de machines avec centerloss (pour l'approche AE)
- Transfer Learning
- Analyse GradCAM
- RNN

**Pre-processing :** ➤ Utilisation de filtres



- Varier ou combiner les formats des données d'entrées
- Spectrogrammes de phase (+ amplitude)

**Méthodologie :** ➤ Analyse des mauvaises prédictions

- Utiliser une approche d'ensemble
- Entraîner en utilisant les sons des autres machines comme des anomalies ou entraîner à reconnaître les attributs

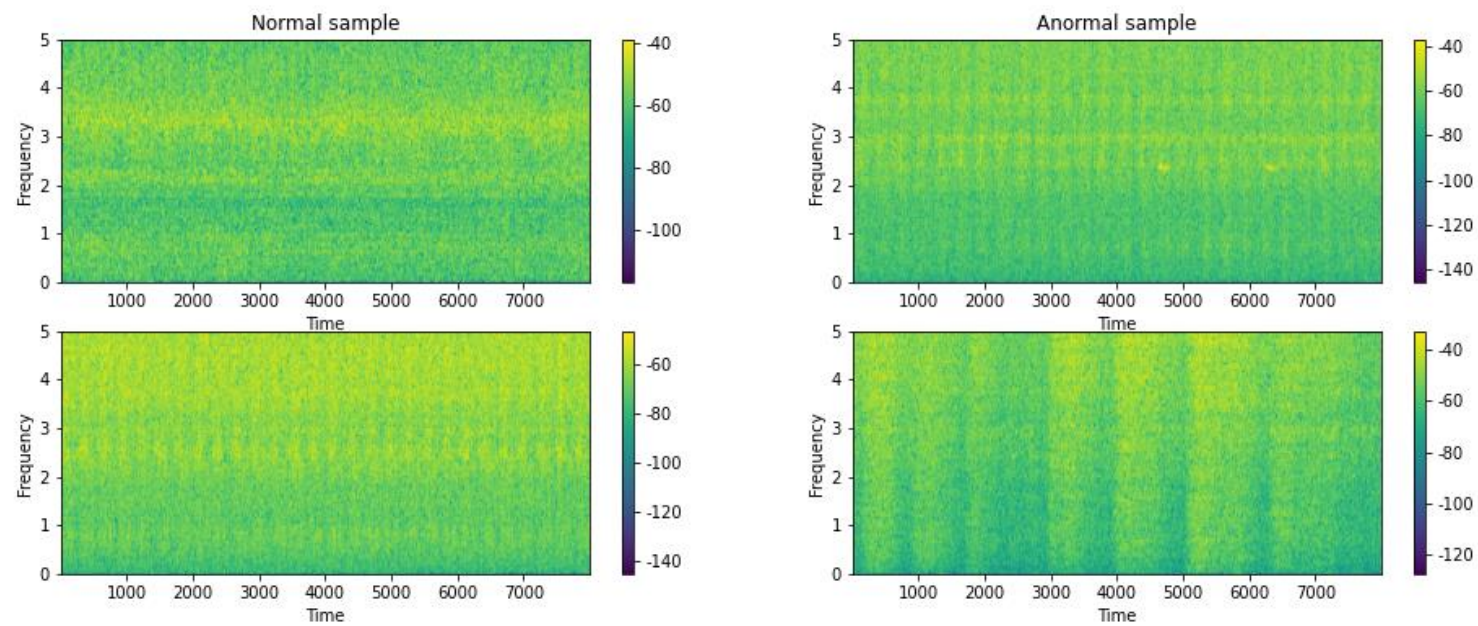
**AI or not AI ?** ➤ Empreinte carbone des codes

- Coût développement / infrastructure

# Présentation Streamlit



Details Coef. of the discrete wavelet transform (gearbox volt\_1.0)



# Approche FaceNet + classifieur RF

2D mel-spectros  
(128 x 313)

➔  
FaceNet  
(sur sections  
ou machines)

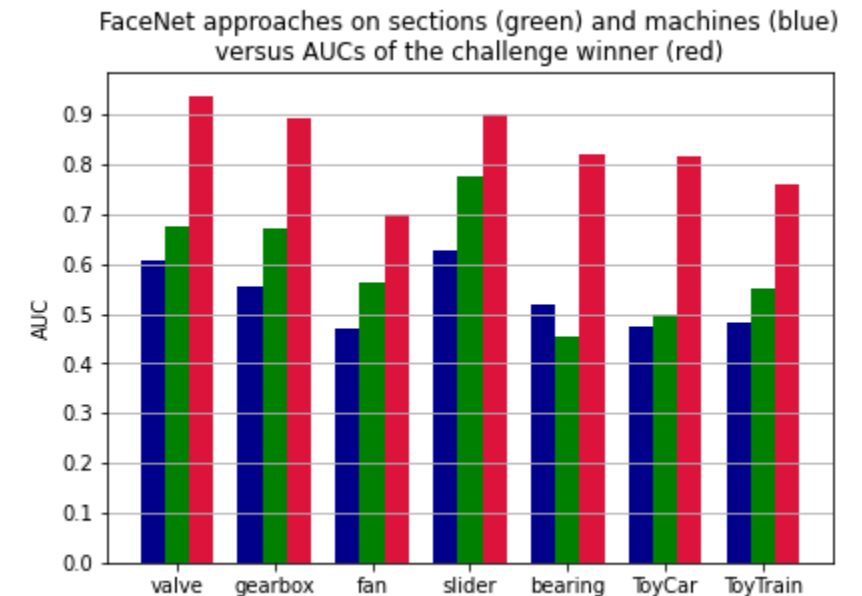
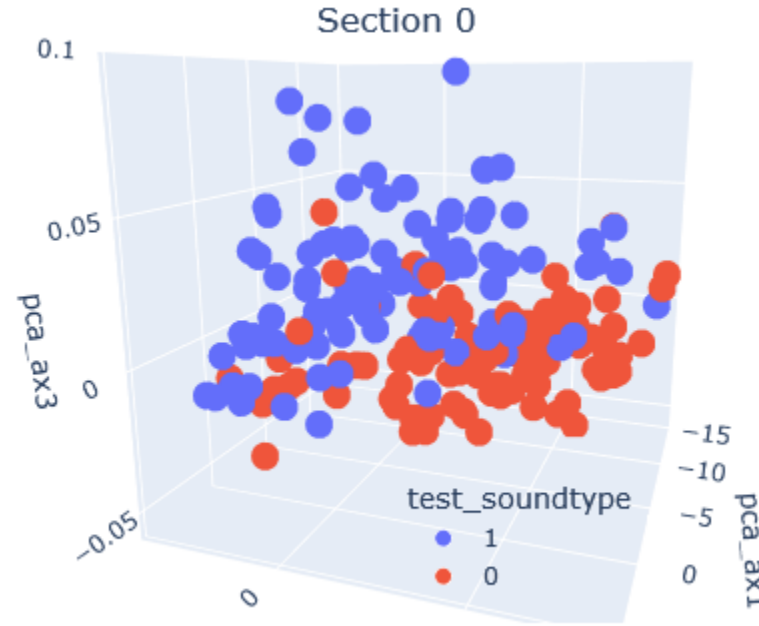
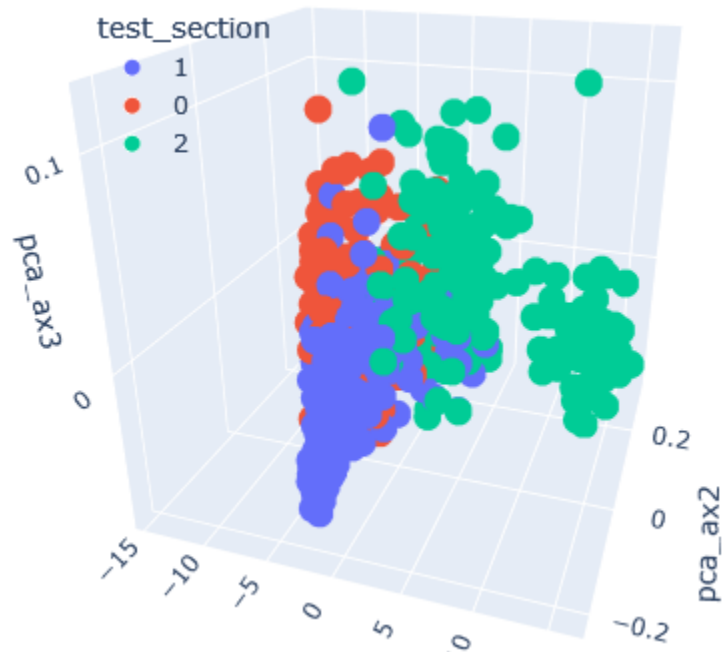
Vecteurs  
d'embedding  
de taille 128

➔  
Classifieur  
(RF)

Probas  $p_{sections}$   
ou  $p_{machines}$

➔  
Post-  
processing

Normal  
Anormal



# Dépendance de l'AUC aux sections

- Données du **domaine source** uniquement
- Classification non supervisée normal/anormal via modèle de classification des machines (CNN - Dense basique)

