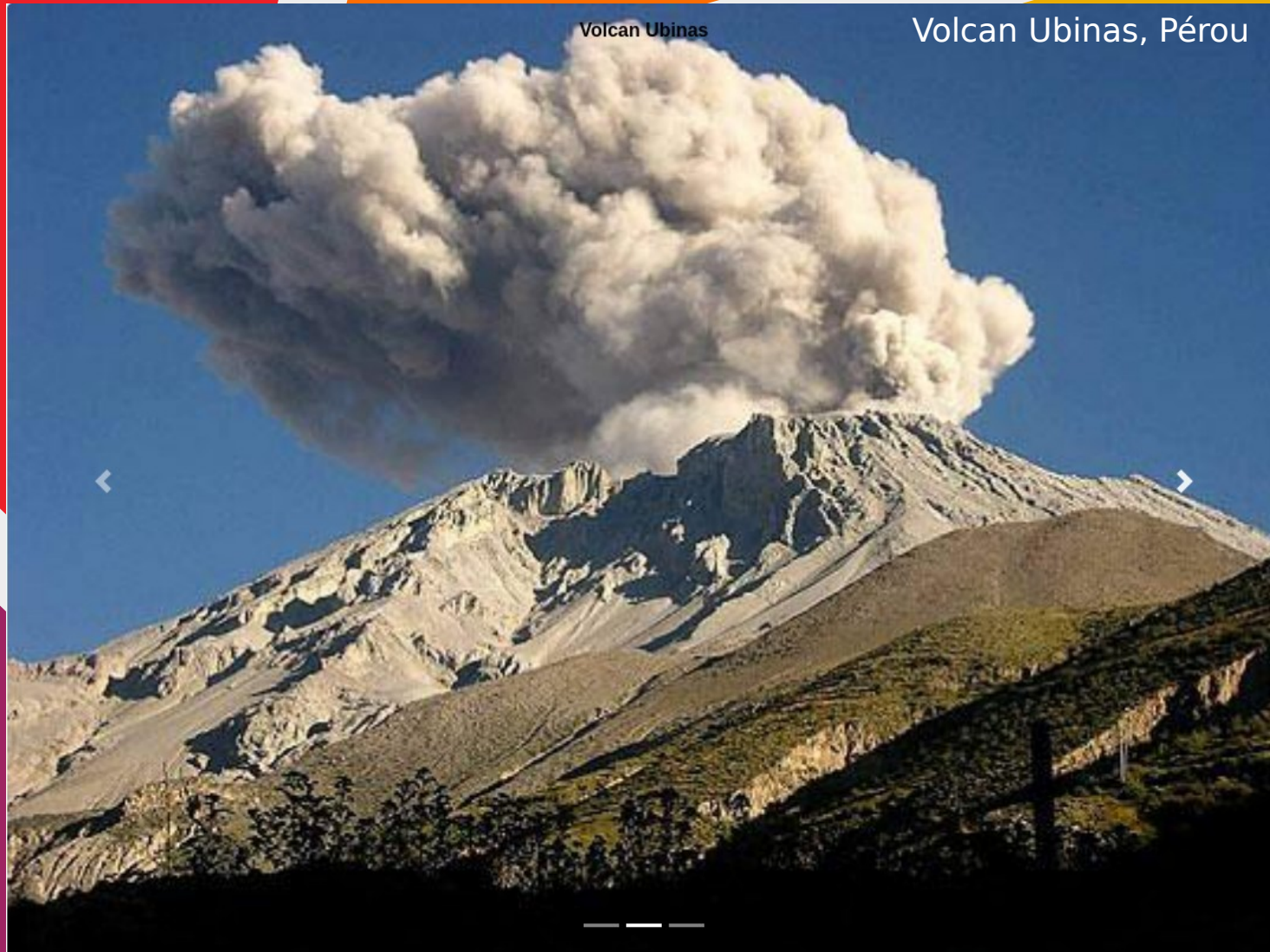


Projet volcan

- Séverine
- Jean-Michel
- Sylvan
- Rubal

Suivi - 19/11/2021

Le contexte



- Actif
- Plusieurs types d'événements
- Grosse quantité de données



Traitement des données difficile

Le projet



Utilisateurs	Vulcanologues & chercheurs en séismologie
Besoin identifié	<ul style="list-style-type: none">→ Classification de types d'événements volcaniques→ Identification d'un éventuel motif dans l'enchaînement des classes d'événements
Produit proposé	Algorithmes entraînés & rapports
Valeur du produit	Facilitation de l'extraction d'informations pertinentes de l'activité volcanique

Les user stories



En tant qu'utilisateur,

1	Je veux prédire la catégorie d'événement associée à chaque signal
2	Je veux savoir si mes labels sont biaisés ou pas
3	Je veux savoir s'il y a une autre façon de classifier mes données
4	Je veux obtenir les informations pertinentes extraites de mes données
5	Je veux identifier des possibles motifs dans l'enchaînement des événements sismiques

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Préparation des données :

- équilibrage du jeu de données



100 événements par classe
= 700 observations

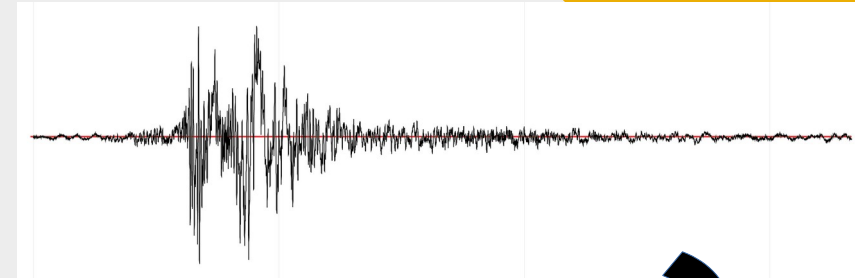
Classe	Compte
EXP	152
HIB	457
LP	97386
PIS	1641
TOR	102
TR	16139
VT	1539

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Préparation des données :

- équilibrage du jeu de données,
- sélection d'informations pertinentes



	Duration	Sampling rate	Overlap	npts	energy	variance	mean	median	maximum	amplitude
91057	37	100.0	False	360001.0	0.33601	1.249415e+03	214.37514	215.00000	562.00000	613.00000
94878	27	50.0	False	180001.0	0.37573	1.372030e+04	70.64000	72.00000	481.00000	902.00000
100776	43	50.0	False	180001.0	0.66832	1.532403e+04	53.30977	55.00000	825.00000	1703.00000
67302	34	40.0	False	143999.0	41.71327	2.963555e+07	-196.81839	-226.50000	20491.00000	38204.00000
3561	29	50.0	False	180000.0	0.02370	7.584260e+03	298.54346	299.96945	612.47302	644.30743
...
109371	32	50.0	False	180001.0	42.12105	8.778938e+01	53.15813	53.00000	89.00000	70.00000
121642	25	100.0	False	360001.0	0.00654	5.156741e+02	0.20072	0.26109	115.38883	233.11829
104584	46	50.0	False	180001.0	0.42680	9.138694e+03	53.83870	54.00000	566.00000	925.00000
76191	31	100.0	False	360001.0	0.28696	8.865084e+04	-39.37097	-36.00000	1544.00000	2890.00000
94312	18	50.0	False	180001.0	0.17511	9.608813e+03	66.37889	66.00000	511.00000	903.00000

700 rows × 10 columns

Exemple

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Préparation des données :





- équilibrage du jeu de données,
- sélection d'informations pertinentes,
- regroupement de classes d'événements.

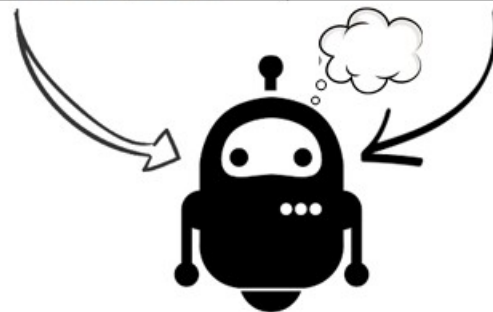
Classes	Sous-classes
EXP	/
HIB	/
LP	LP1 LP2 LPP
PIS	/
TOR	/
TR	TR1 TR2 TRA
VT	VT1 VTP

L'avancement

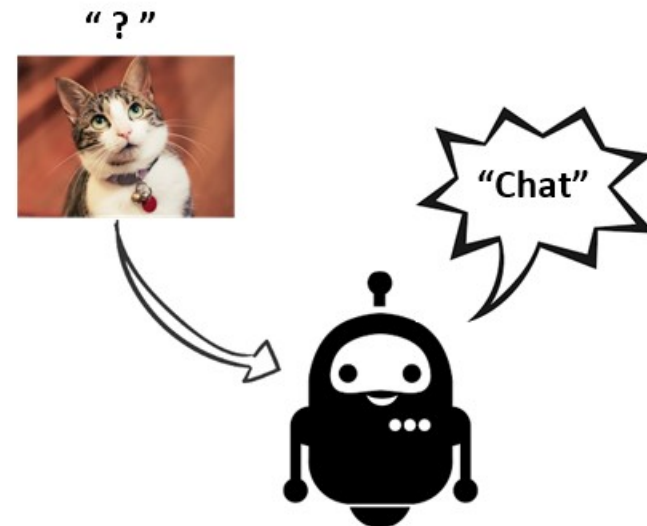
Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation d'apprentissage automatique supervisé (IA)

x	y
	"Chien"
	"Chien"
	"Chat"
	"Chien"



Apprentissage Supervisé



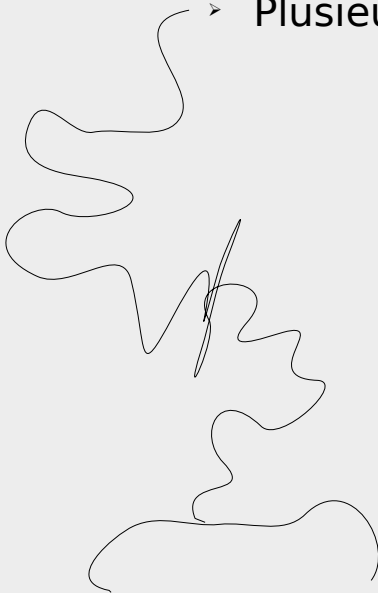
Utilisation finale

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation d'apprentissage automatique supervisé

- Plusieurs modèles (algorithmes) différents



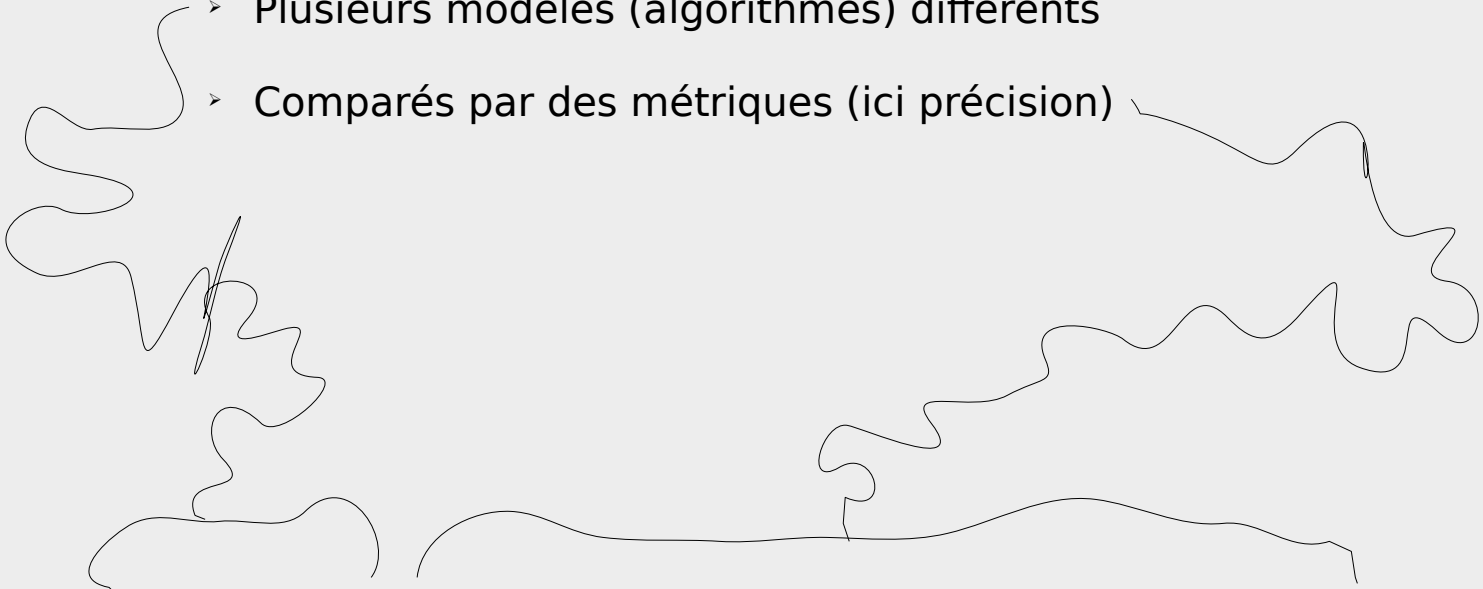
Modèle	Accuracy w/ selection	Accuracy selection1
SVC		
RF	0.73	0.69
KNN		
XGBoost	/	0.66

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation d'apprentissage automatique supervisé

- Plusieurs modèles (algorithmes) différents
- Comparés par des métriques (ici précision)



Modèle	Accuracy w/ selection	Accuracy selection1
SVC	0.19	0.19
RF	0.73	0.69
KNN	0.46	0.46
XGBoost	/	0.66

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation d'apprentissage automatique supervisé

- Plusieurs modèles (algorithmes) différents
- Comparés par des métriques (ici accuracy)

Modèle	Accuracy w/ selection	Accuracy selection1
SVC	0.19	0.19
RF	0.73	0.69
KNN	0.46	0.46
XGBoost	/	0.66

Pas mal !

Et si on visualisait tout ça ?

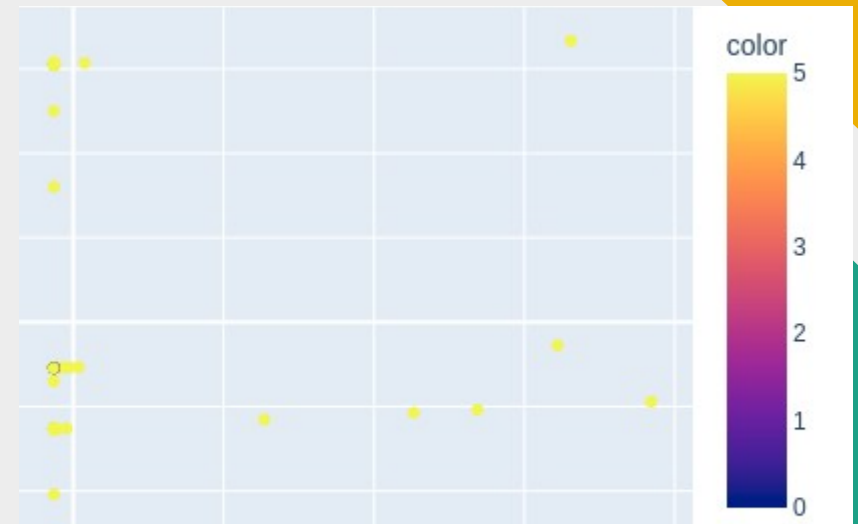
L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Vraies classes



Prédictions



Modèle	Accuracy w/ selection	Accuracy selection1
SVC	0.19	0.19
RF	0.73	0.69
KNN	0.46	0.46
XGBoost	/	0.66

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Vraies classes



Prédictions

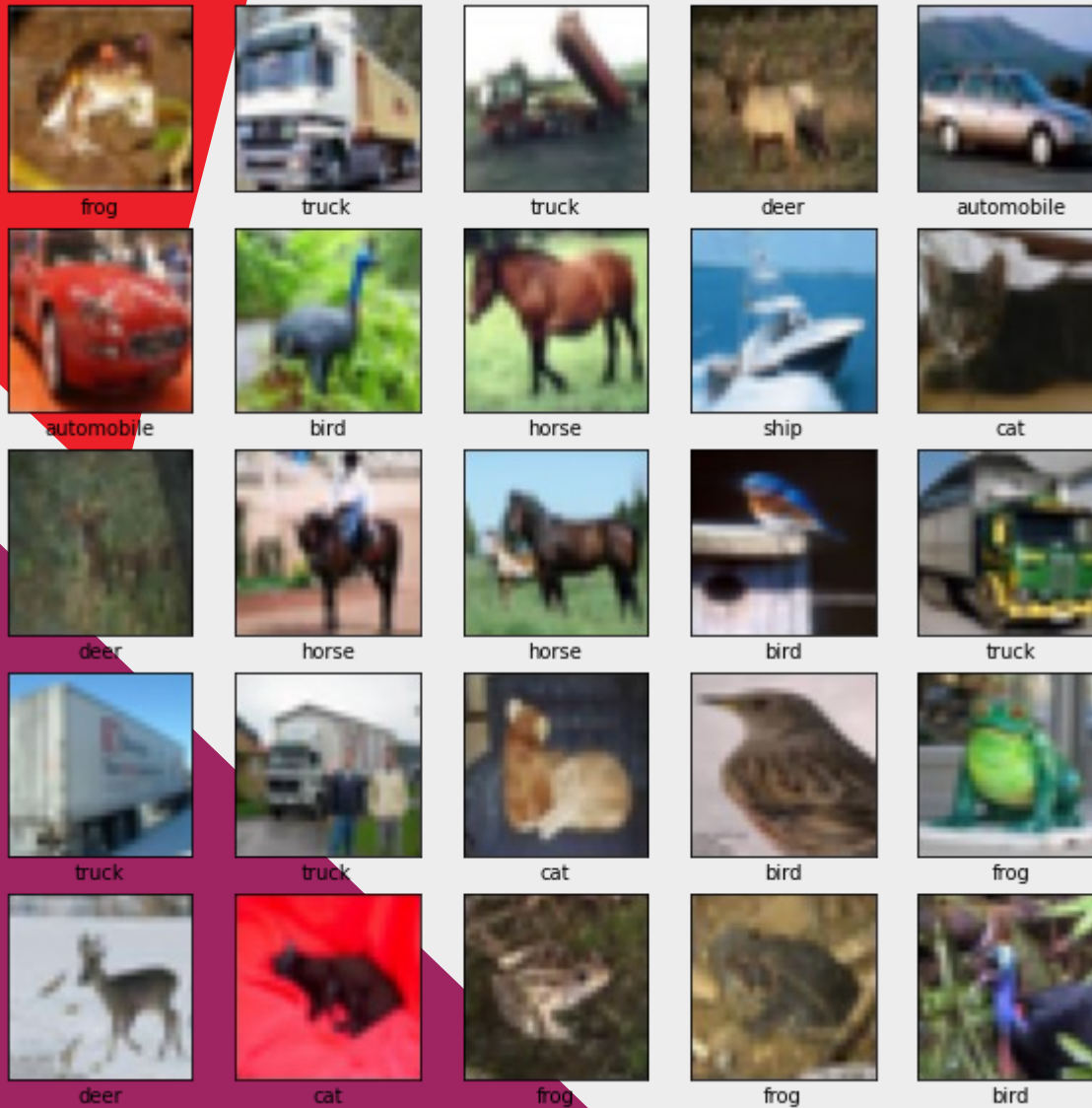


Modèle	Accuracy w/ selection	Accuracy selection1
SVC	0.19	0.19
RF	0.73	0.69
KNN	0.46	0.46
XGBoost	/	0.66

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation de réseaux de neurones

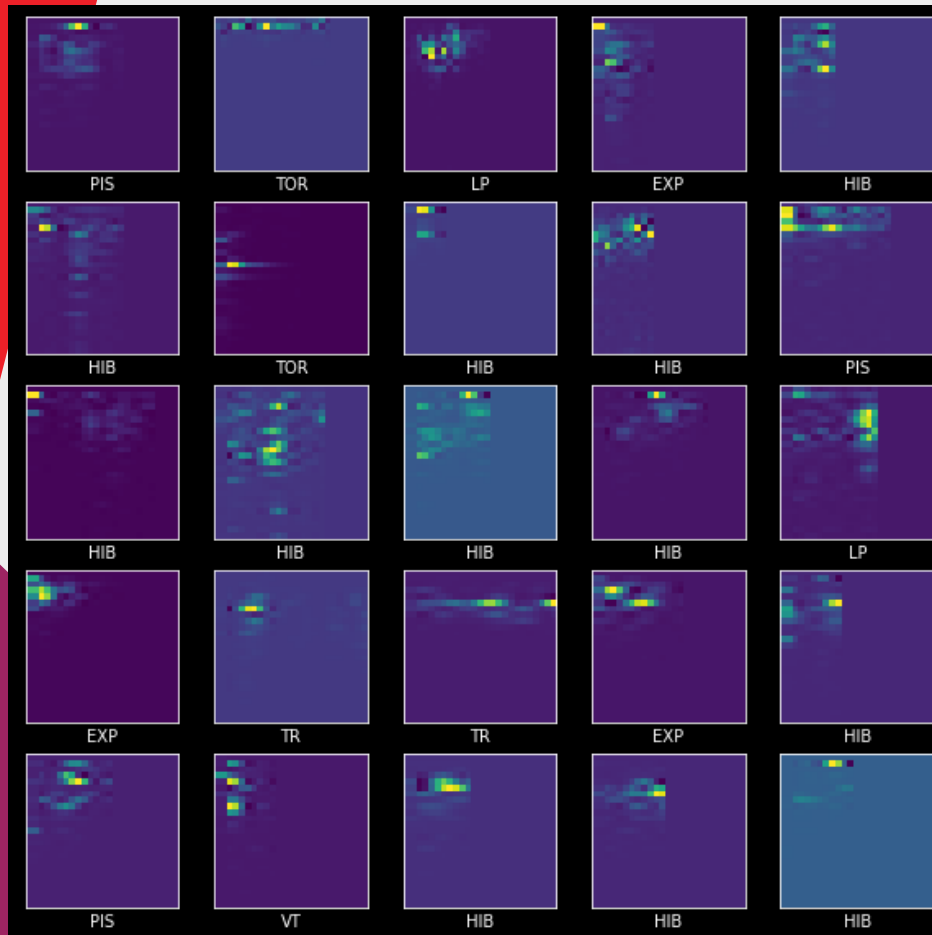


Modèle entraîné sur des **images**

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation de réseaux de neurones

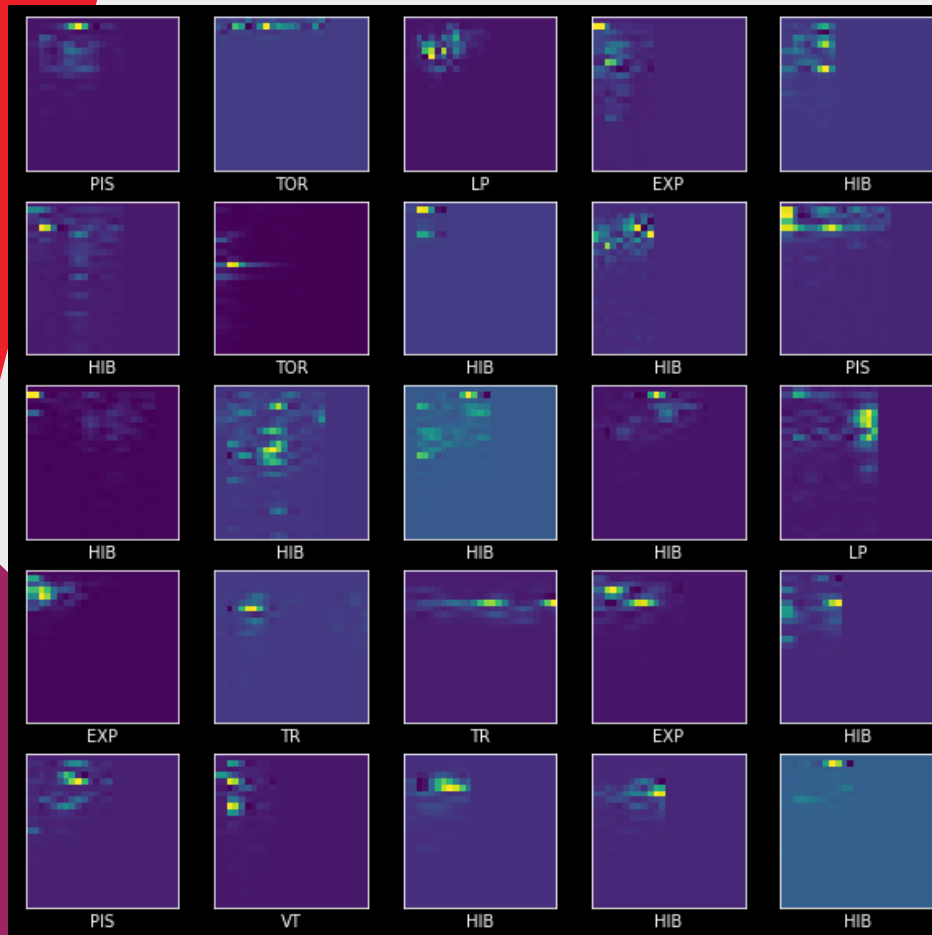


Modèle entraîné sur des **images**
(spectrogrammes)

L'avancement

Je veux prédire la catégorie d'événement associée à chaque signal

Utilisation de réseaux de neurones



Modèle entraîné sur des **images**
(spectrogrammes)

Modèle	Accuracy
CNN	0.05

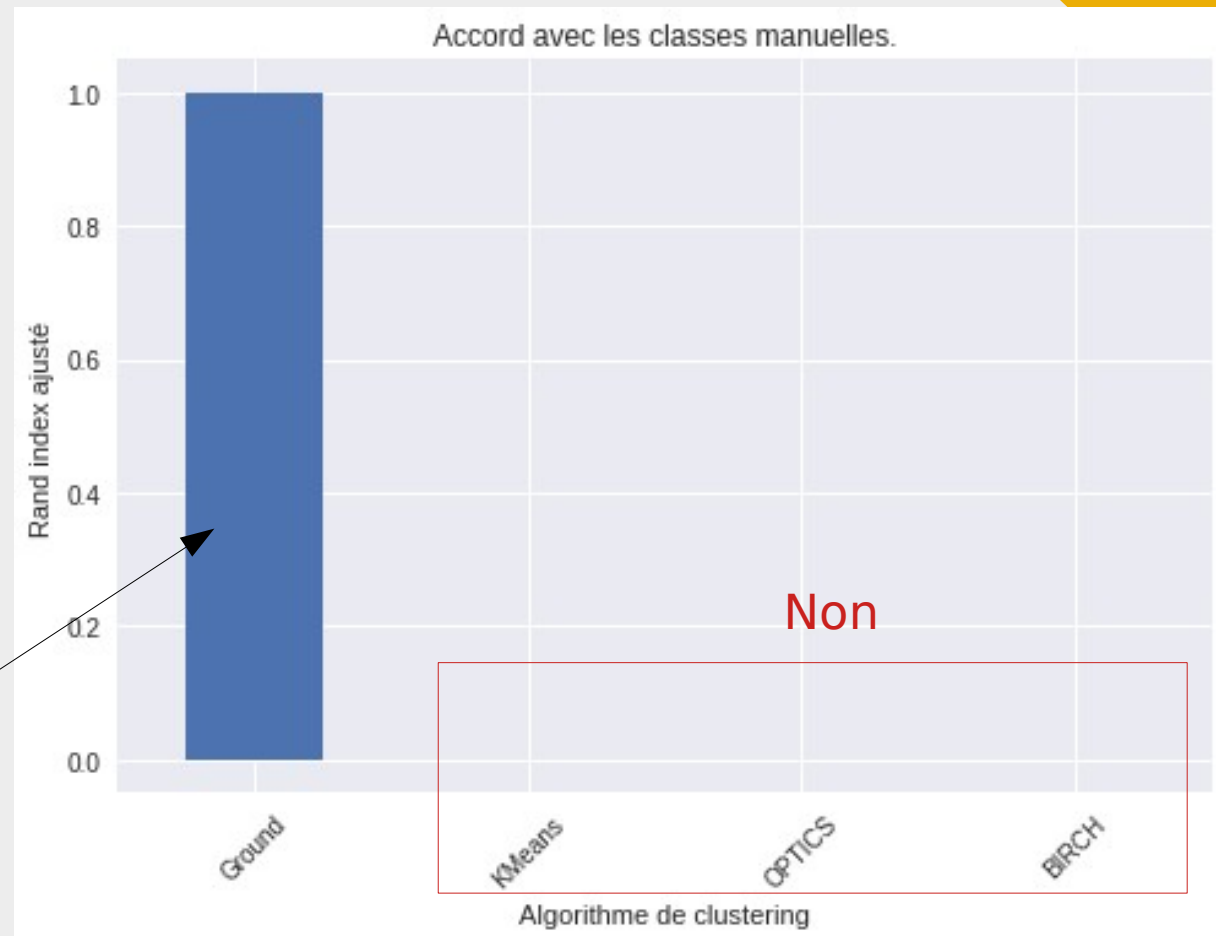
Ca marche pas du tout ...

L'avancement

Je veux savoir si mes catégories sont biaisés ou pas
Je veux savoir s'il y a une autre façon de classer mes données

Utilisation d'apprentissage automatique non-supervisé

L'algorithme prédit-il les
mêmes classes que l'humain ?

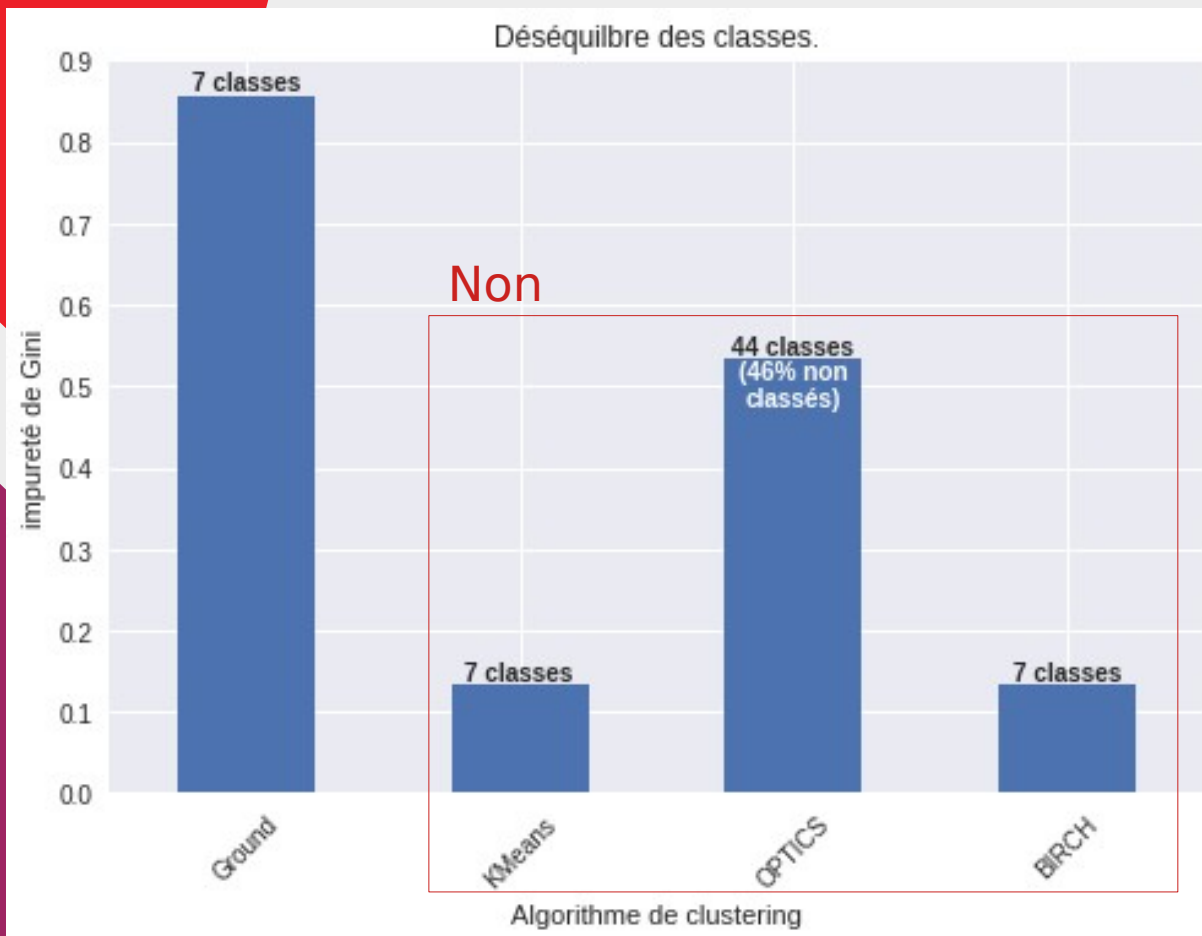


Classes "manuelles"

L'avancement

Je veux savoir s'il y a une autre façon de classer mes données

Utilisation d'apprentissage automatique non-supervisé



Les classes proposées par l'algorithme sont-elles équilibrées ?

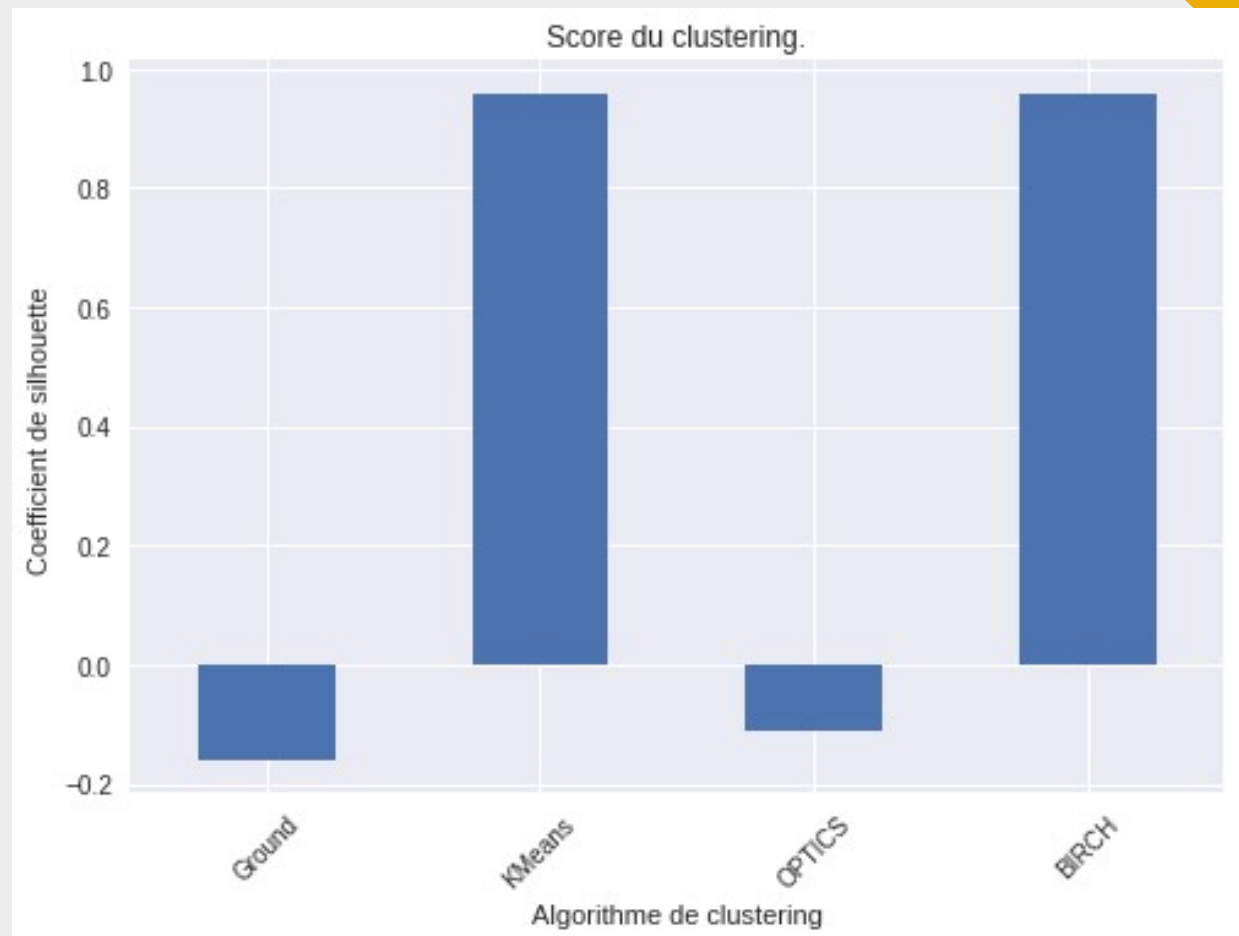
L'avancement

Je veux savoir si mes catégories sont biaisés ou pas

Utilisation d'apprentissage automatique non-supervisé

Quelle est la qualité
du clustering proposé
par l'algorithme ?

Oui mais ...

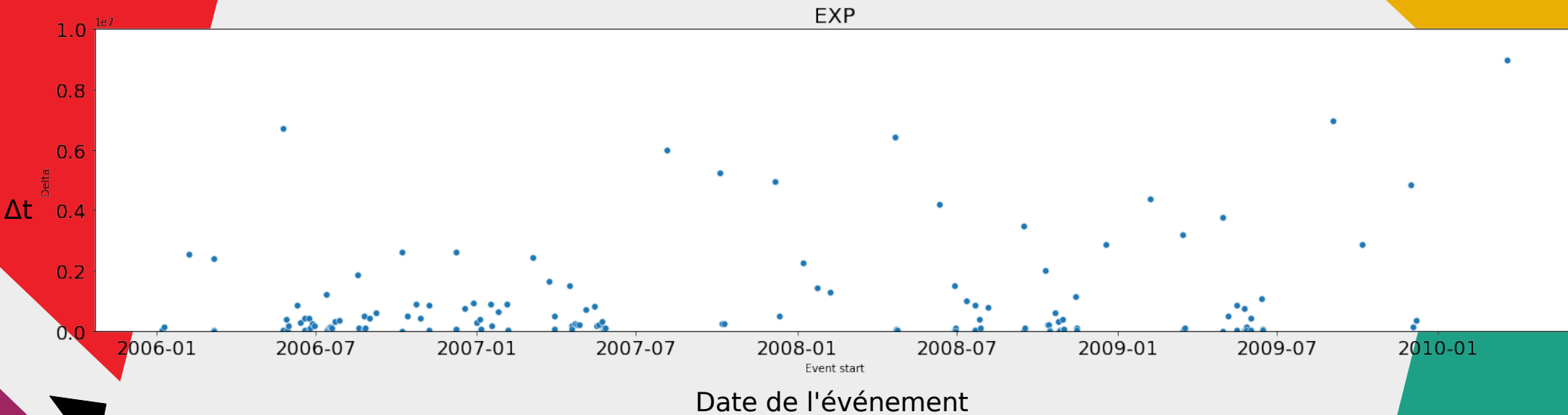


L'avancement

Sismik

Je veux identifier des possibles motifs dans l'enchaînement des événements sismiques

Objectif : prédire le prochain événement



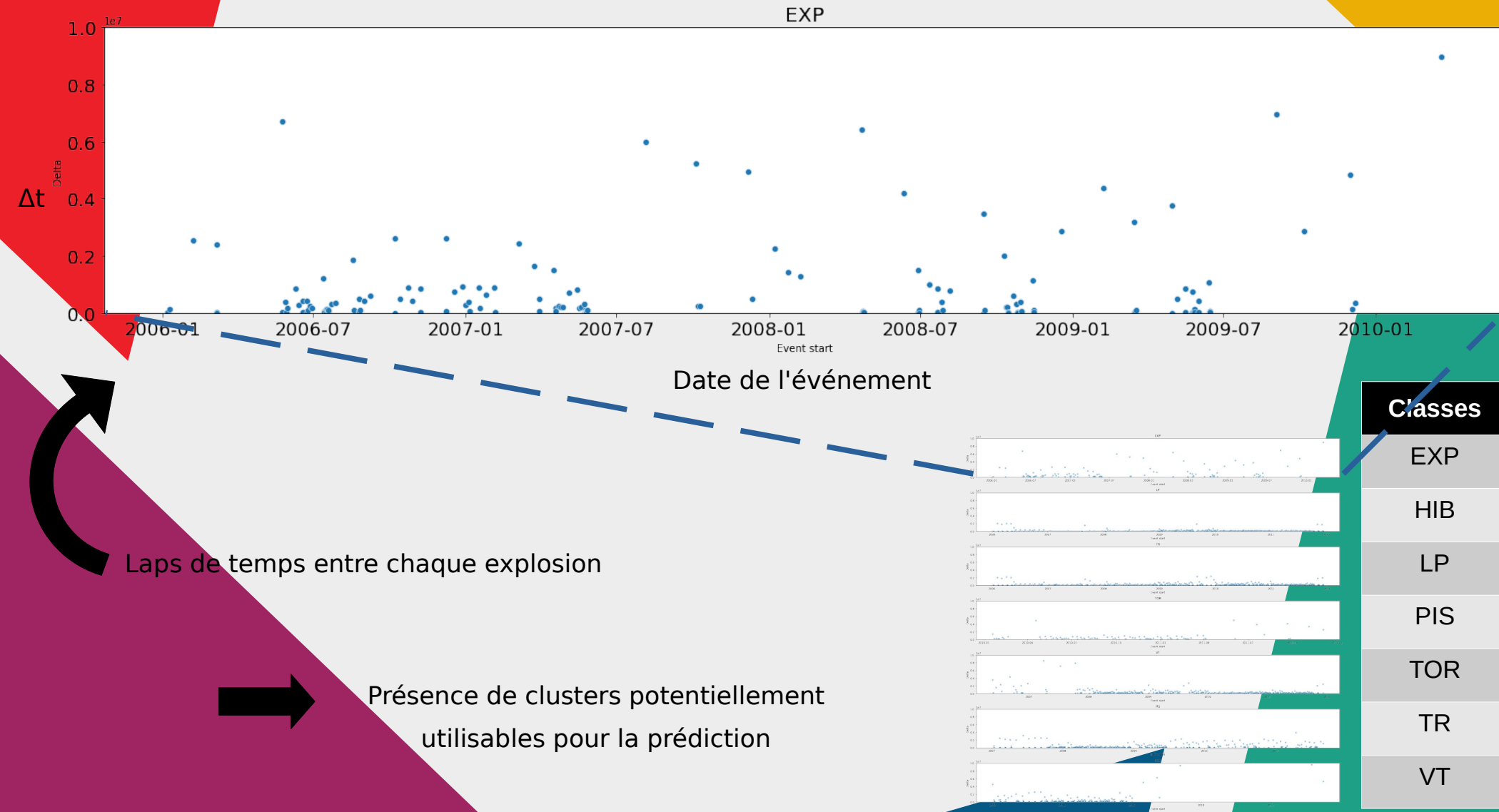
Laps de temps entre chaque explosion

Présence de clusters potentiellement utilisables pour la prédiction

L'avancement

Je veux identifier des possibles motifs dans l'enchaînement des événements sismiques

Objectif : prédire le prochain événement



Merci

- Séverine
- Jean-Michel
- Sylvan
- Rubal