



# Analyse de données :

## Les tremblements de terre sur la période 1990-2023

Groupe "Room One" Andreea / Loic / Xavier / Zacharia

**Comment analyser et utiliser les données  
sismiques historiques pour identifier des  
tendances et prédire les séismes à fort impact ?**



# Aperçu du dataset



## Contenu

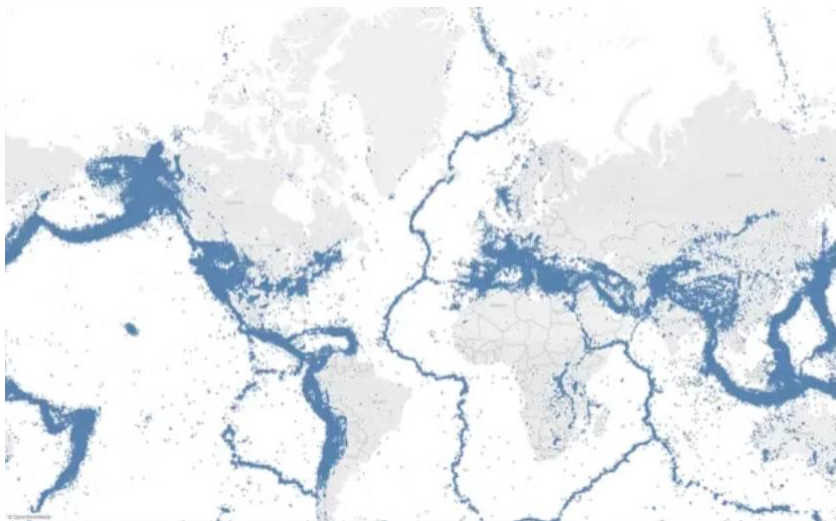
- 12 colonnes
- 3,4 millions de lignes (toutes les mesures mondiales) entre 1990 et 2023



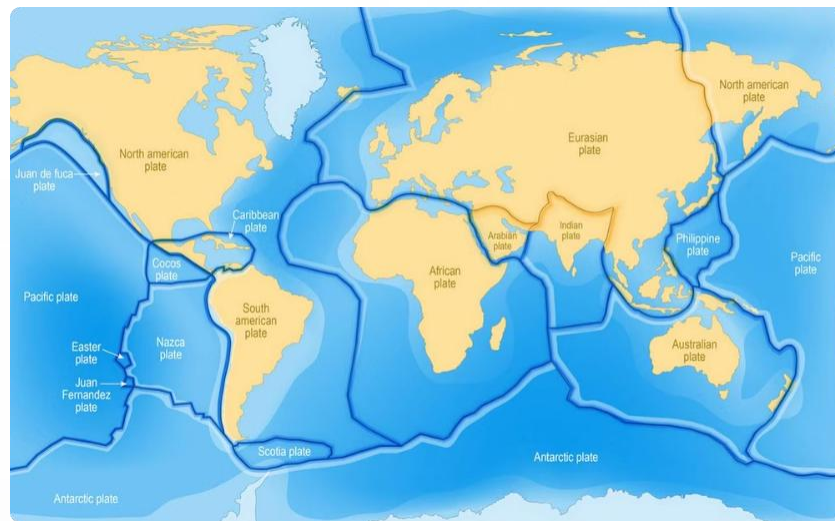
## Informations clés

- Date
- Géographie de l'épicentre
- Caractéristiques: Magnitude, Longitude
- Conséquences : Significativité, Tsunami

# Visualisation globale des épicentres



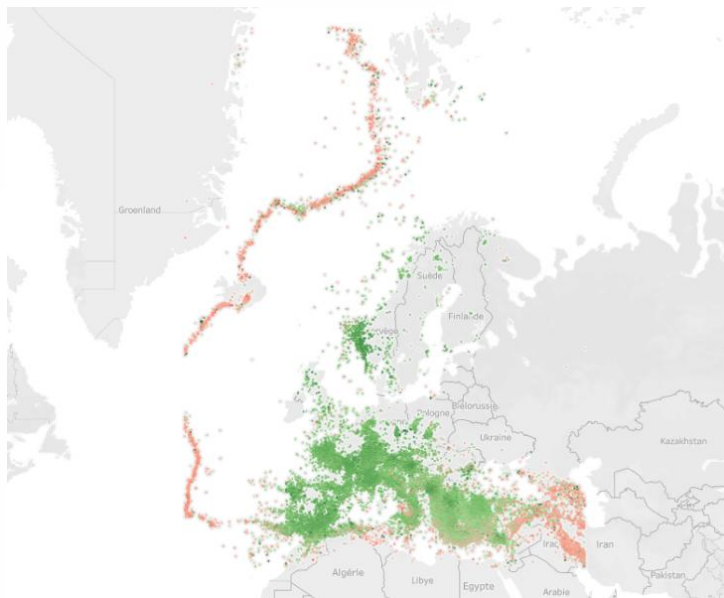
Visualisation du dataset sous Tableau



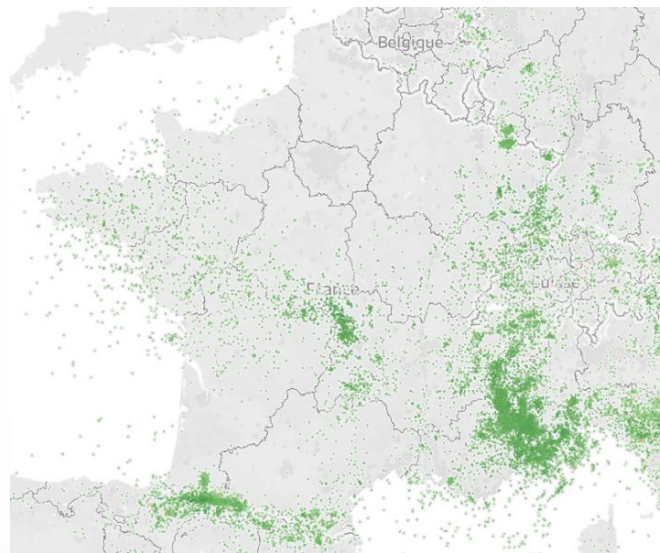
Carte des plaques tectoniques

*[source futura-sciences]*

# Visualisation locale des épicentres

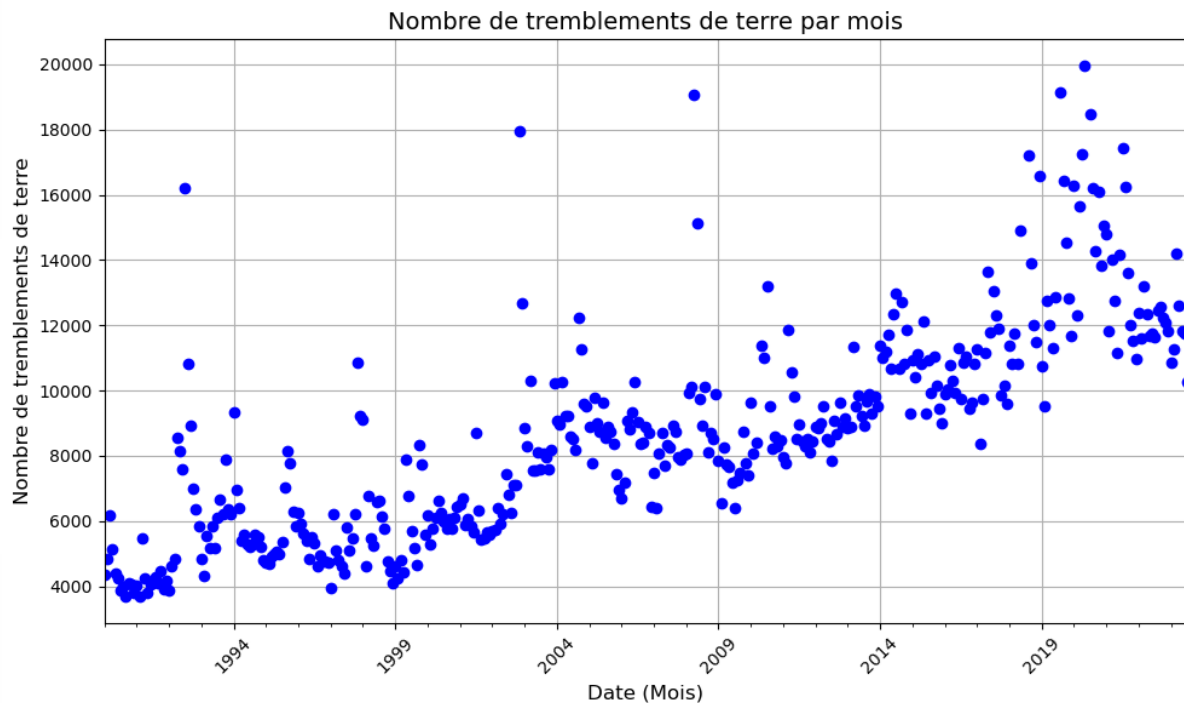


**Filtre Europe sous Tableau**



**Zoom sur la France**

# Analyse du nombre de séismes



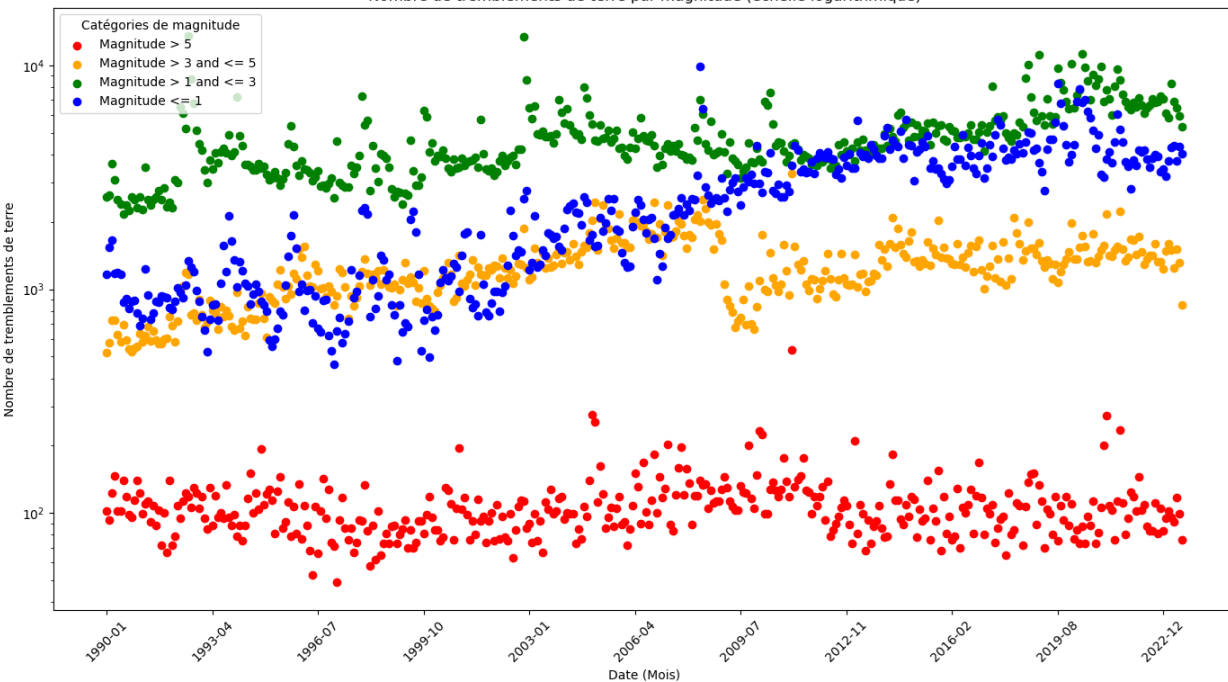
☐ Volume important

☐ Croissance régulière

Devons-nous nous inquiéter de cette croissance ?

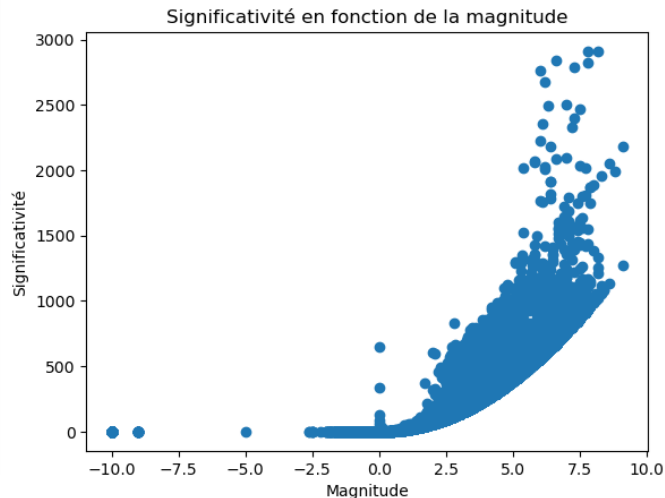
# Analyse du nombre de séismes

Nombre de tremblements de terre par magnitude (échelle logarithmique)

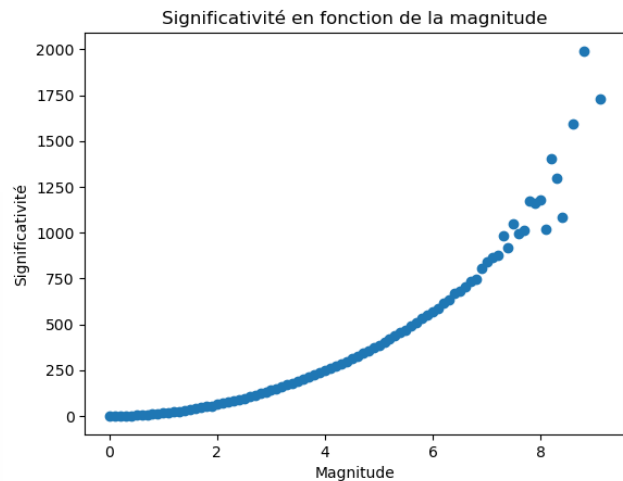


Une croissance soutenue  
par les micro-séismes

Probablement de meilleures techniques  
de détection et/ou d'enregistrement ?



Filtrage / Nettoyage



# Significativité

Mesure les conséquences (impacts humains et environnementaux) d'un séisme

*Exemple : Valeur > 600 -> Événement significatif*

Très forte relation avec la magnitude



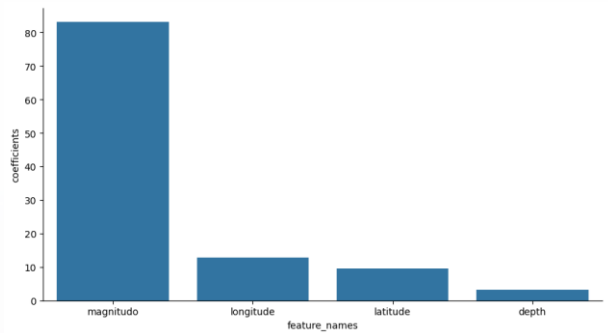
# Significativité et Machine Learning

Objectif : Modèle de prédiction de la significativité (conséquences)  
en fonction des caractéristiques initiales (lieu, magnitude, profondeur)

0.9  
R<sup>2</sup> score



1  
Facteur principal  
Magnitude



Prédiction d'une valeur numérique :  
Modèle de régression linéaire

# Significativité et Machine Learning

Application du même modèle en réduisant  
le dataset aux significativités > 600



**0.15**

**R<sup>2</sup> score Test**

Performance trop faible



Test en ajoutant des colonnes de magnitude  
(puissance 2, puissance 3...)

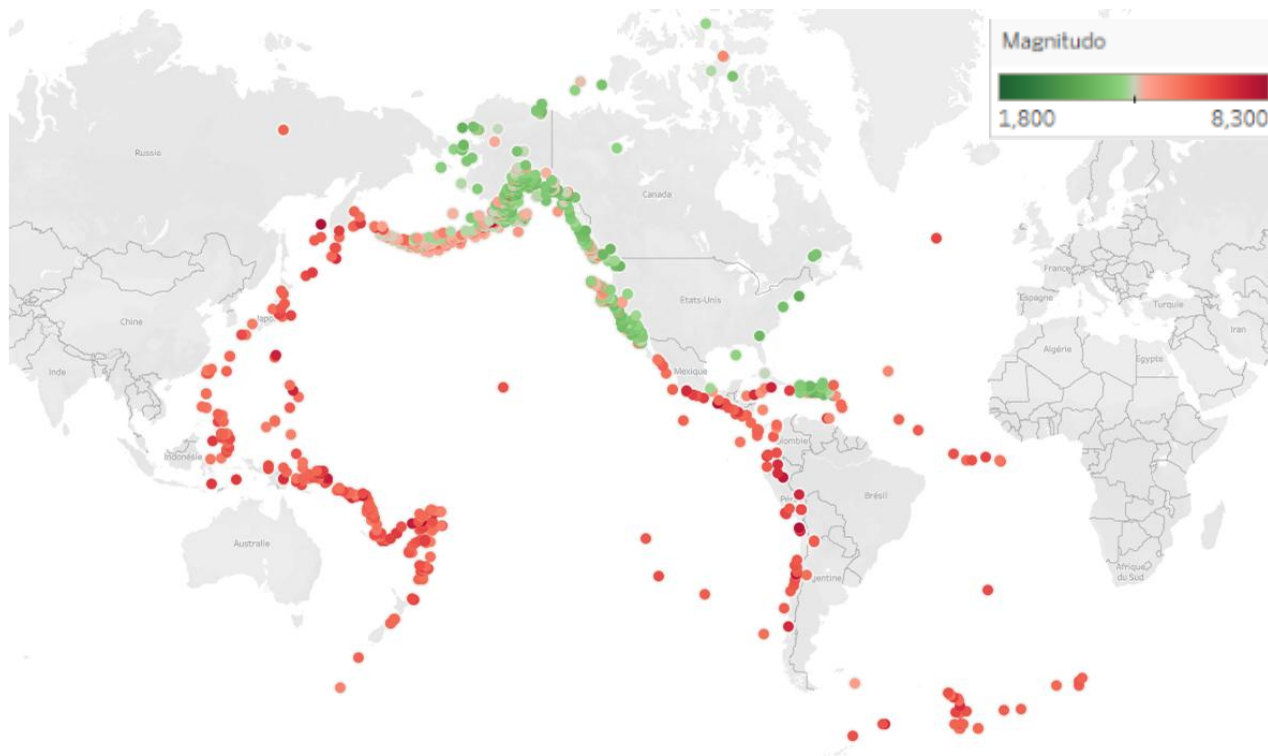


Amélioration légère du modèle, mais pas  
suffisante (maximum atteint : 0.28)

**Conclusion : Le modèle ne performe pas suffisamment pour la prédiction des grosses catastrophes**

**Néanmoins il peut être utilisé sur des données spécifiques (par exemple : l'Europe)**

# Localisation des Tsunamis



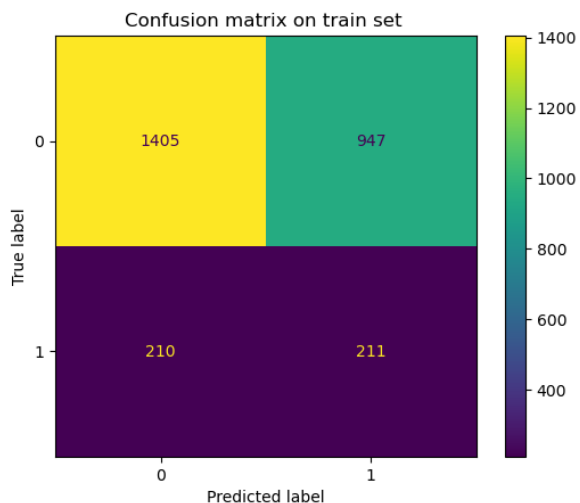
**Modèle de prédiction des tsunamis selon :**

- > Magnitude
- > Profondeur
- > Localisation



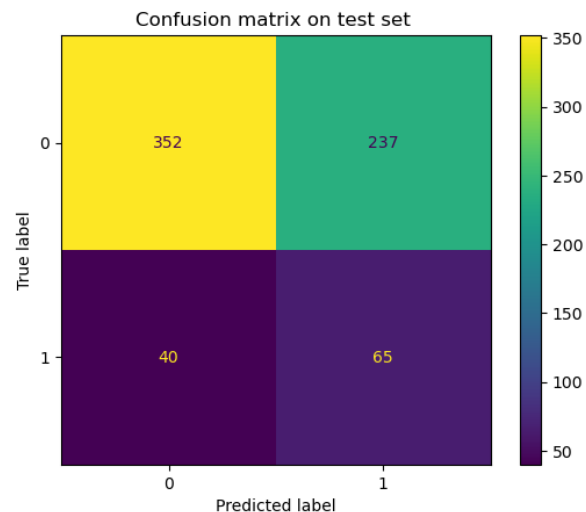
# Performance du modèle de régression logistique

## Données d'entraînement



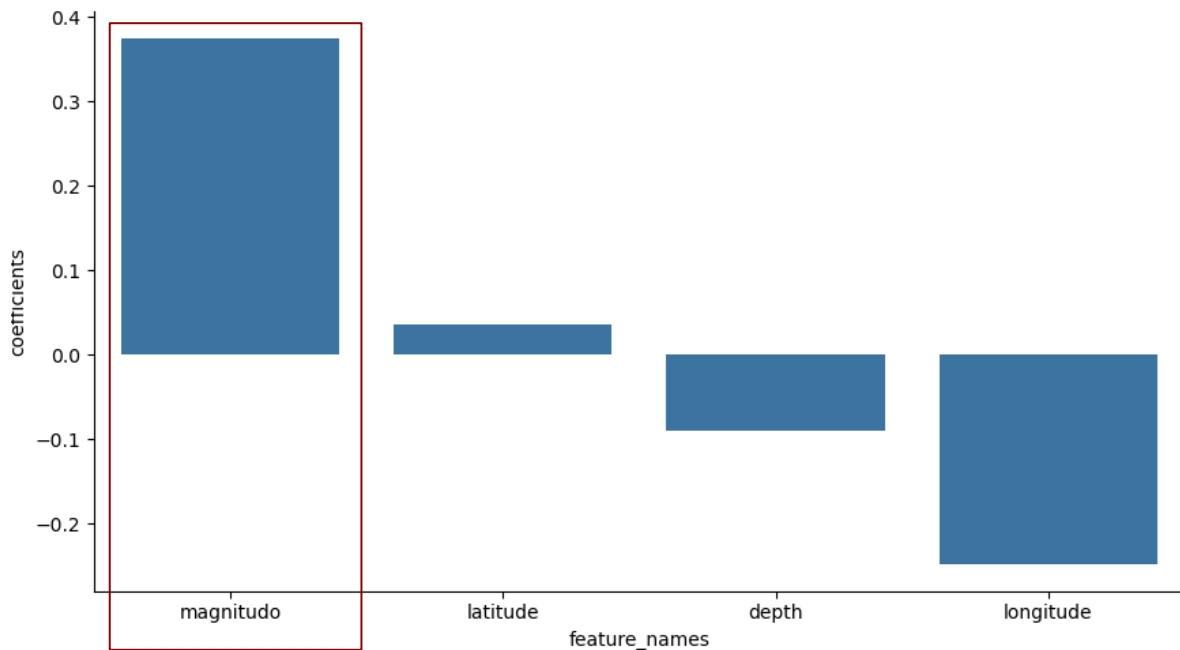
58% VS 60%  
prédiction

## Données de test

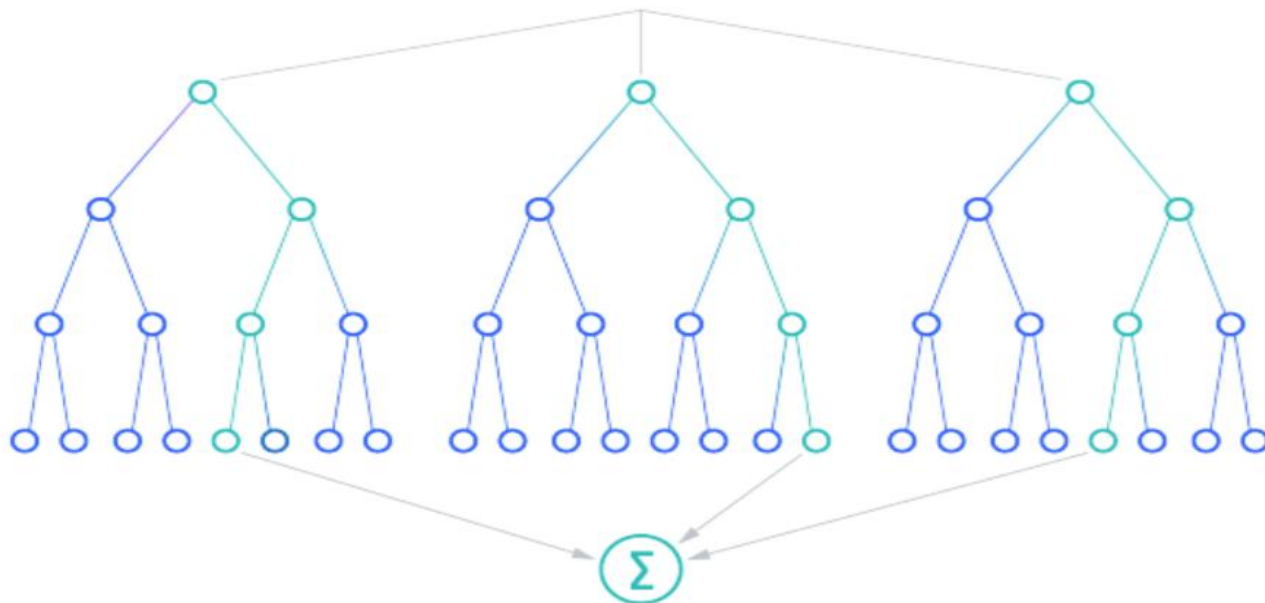


Tsunami non prédit qui ne se produit pas	Tsunami prédit qui ne se produit pas
Tsunami non prédit qui se produit	Tsunami prédit qui se produit

# Impact des coefficients sur ce modèle

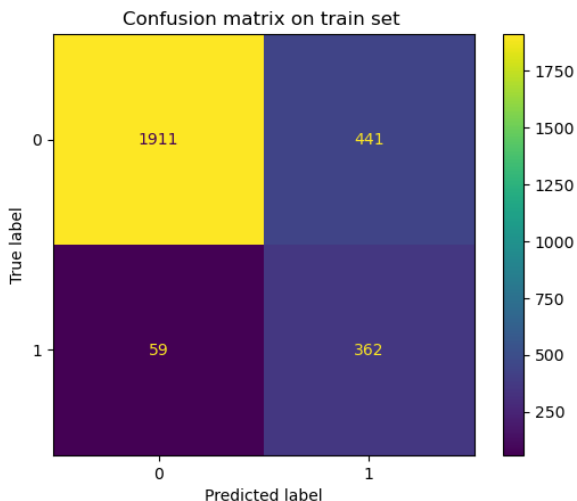


# Modèle de forêt aléatoire pour la prédiction de tsunami



# Performance du modèle de la forêt

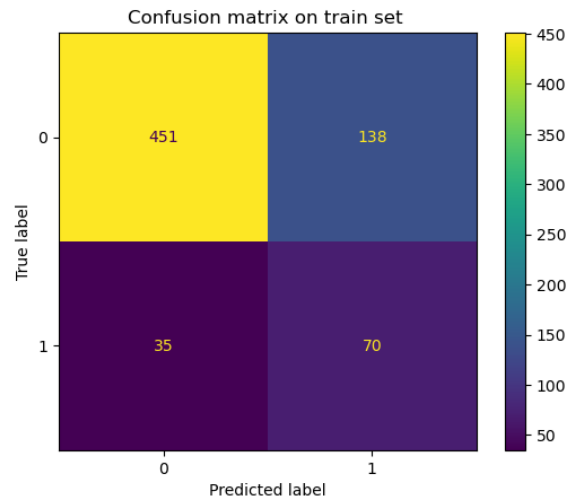
## Données d'entraînement



82% VS 75%  
prédiction

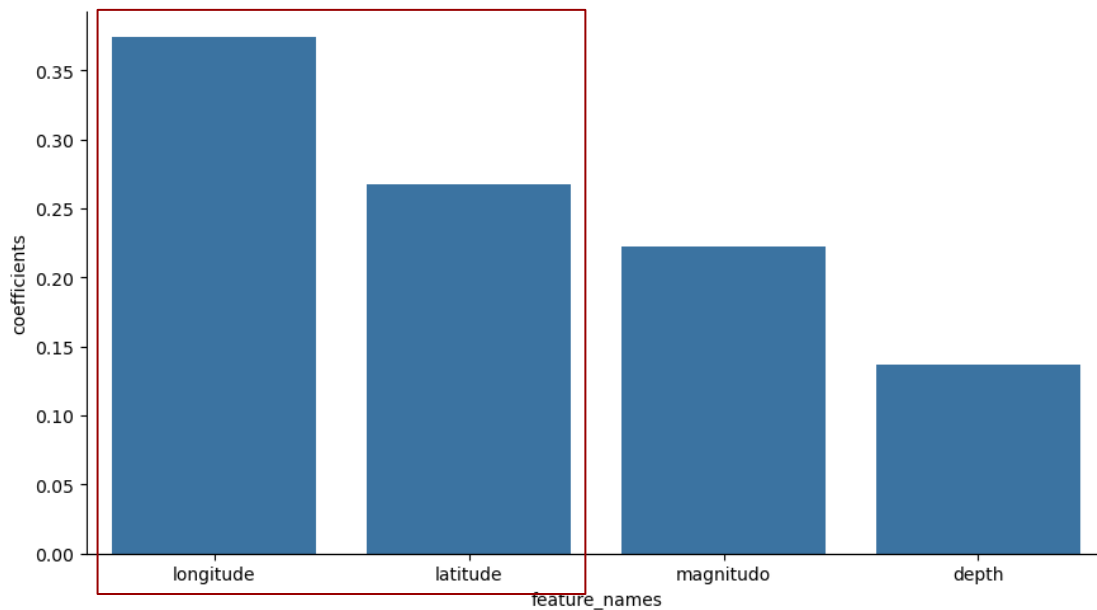
F1 score : 78%

## Données de test



Tsunami non prédit qui ne se produit pas	Tsunami prédit qui ne se produit pas
Tsunami non prédit qui se produit	Tsunami prédit qui se produit

# Impact des coefficients sur ce modèle





# Conclusion

## Prédire l'impact d'un séisme

- Est complexe
- Demande plus !

