

Quantmetry

L'estimation d'incertitudes avec

MAPIE

Meetup Labelia
07/03/2022



Grégoire Martinon
Reliable AI Lead
gmartinon@quantmetry.com



Vianney Taquet
Senior Data Scientist
vtaquet@quantmetry.com



Vincent Blot
Data Scientist
vblot@quantmetry.com

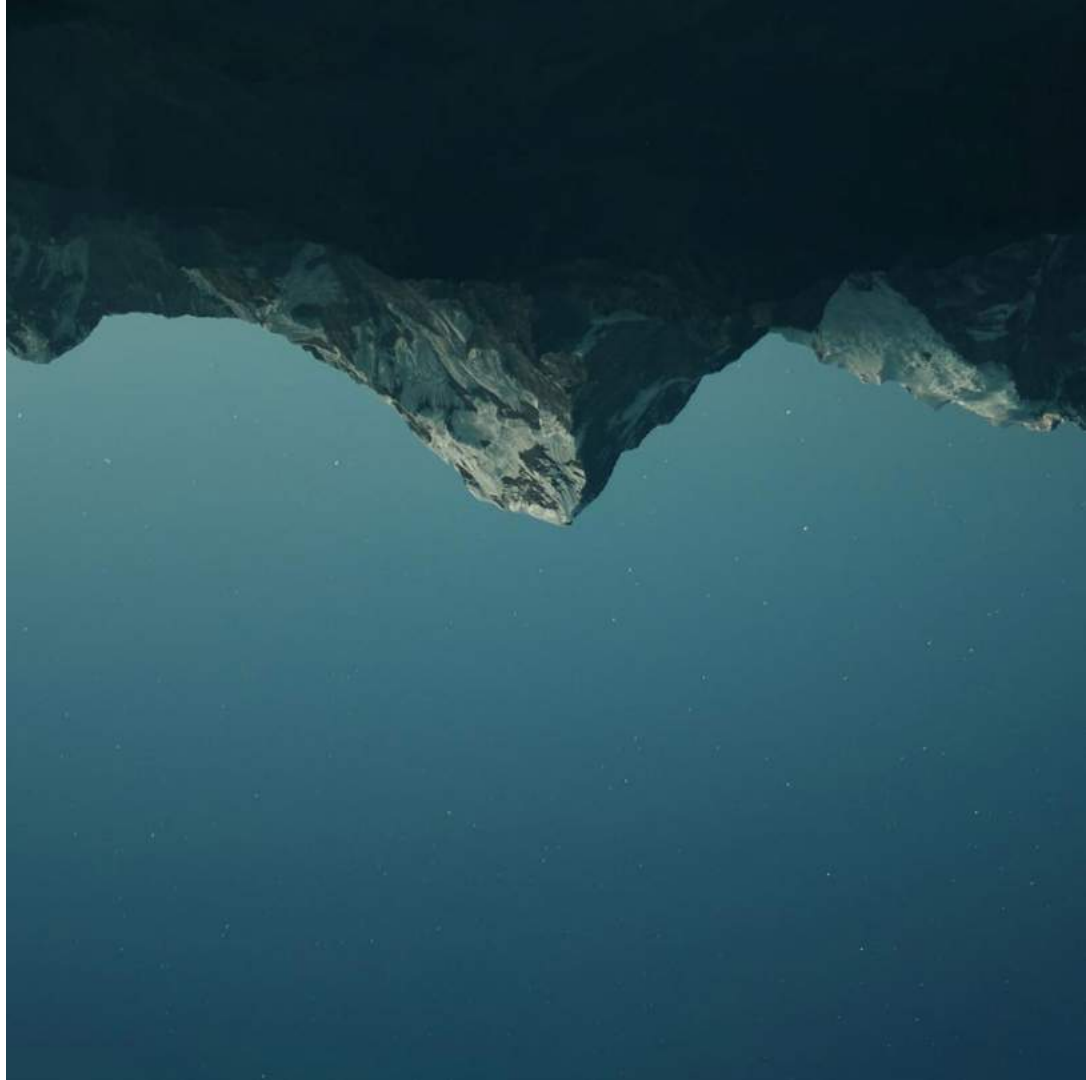


Thomas Morzadec
Data Scientist
tmorzadec@quantmetry.com



1

Contexte et enjeux



Quantmetry : Qui sommes nous ?



UN CABINET
Français

100%

indépendant
depuis 2011

20%

de la masse
salariale vers
la recherche

+50

publications et
articles de blog
par an

15

prix innovation
et recherche

Nos expertises complémentaires nous permettent de développer et déployer des systèmes d'IA critiques sur tout type de cas d'usage

Conseil data



Ai Strategy

Modélisation



Time Series



NLP



Computer Vision



Reliable Ai

Architecture &
implémentation

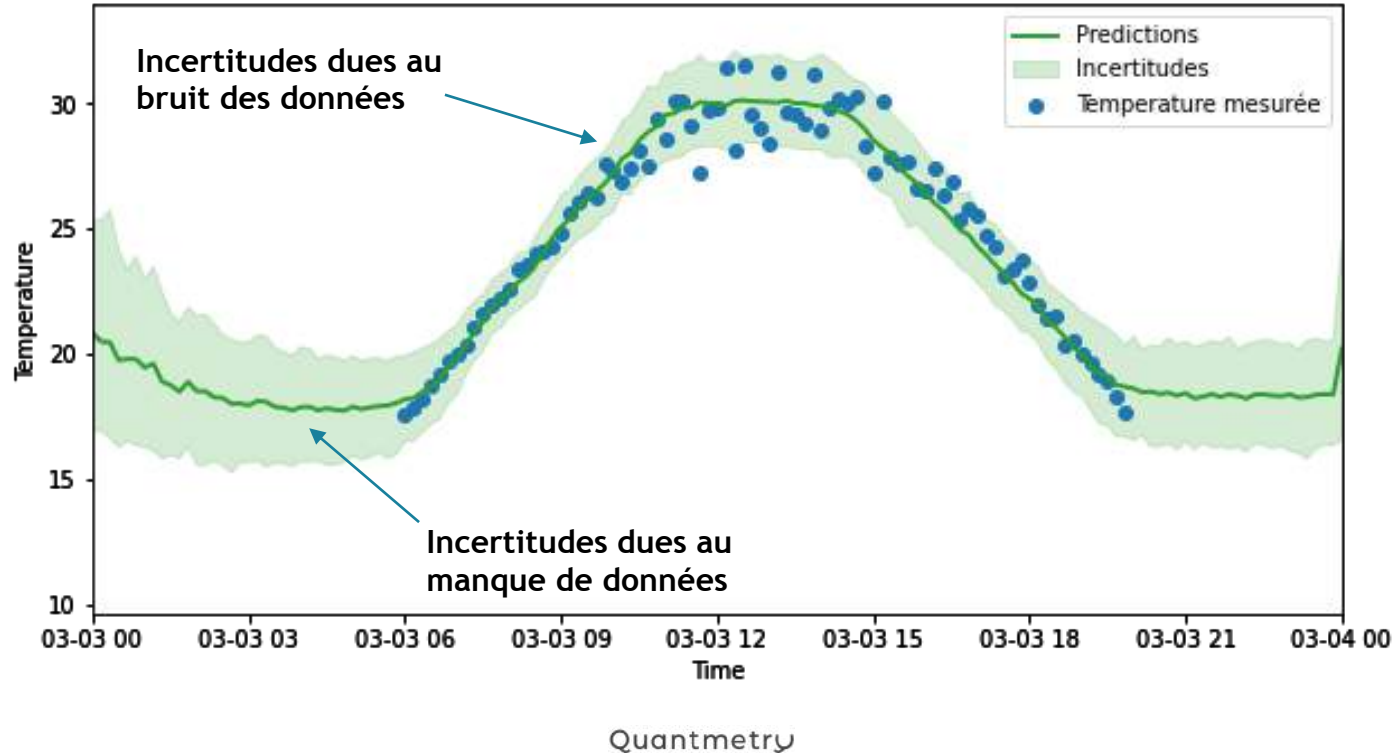


Ai Product

Quantmetry

Qu'est-ce qu'une incertitude en regression ?

Exemple : mesure de la température au cours du temps



Pourquoi estimer les incertitudes ?

L'estimation des incertitudes est cruciale pour le développement et le déploiement de modèles d'IA fiables et implique toutes les personnes liées à l'IA.

Meilleure compréhension des prédictions de l'IA

Quand est-ce que mon modèle *sait* ou *ne sait pas* ?



Concepteurs d'IA

Meilleure gestion des risques

Puis-je utiliser ces prédictions pour prendre ma décision business ?



Opérateurs d'IA

Obligatoire dans certains domaines

Les modèles d'IA sont-ils conformes à la réglementation ?



Régulateurs

Rendre l'IA plus digne de confiance

La décision prise par l'IA à mon sujet est-elle fiable ?



Personnes impactées



Pourquoi avoir développé MAPIE ?

Car aucune librairie python ne garantissait trois critères clés :

Garanties
mathématiques

Agnostique des
modèles d'IA

Implémentation
open-source

Méthode	Garanties mathématiques	Agnostique des modèles	Implémentation open-source
Régression quantile	✗	✓	✓
Perturbation des données (Bootstrap, Jackknife)	✗	✓	✓
Prédiction conforme	✓	✓	✗
Perturbation des modèles (Random seed, MC Dropout)	✗	✓	✓
Bayesian inference	✓	✗	✓

Quantmetry

L'histoire de MAPIE

2021									
Mars	Avril	Mai	Juin	Juillet	Aout	Septembre	Octobre	Novembre	Décembre



Genèse du
package



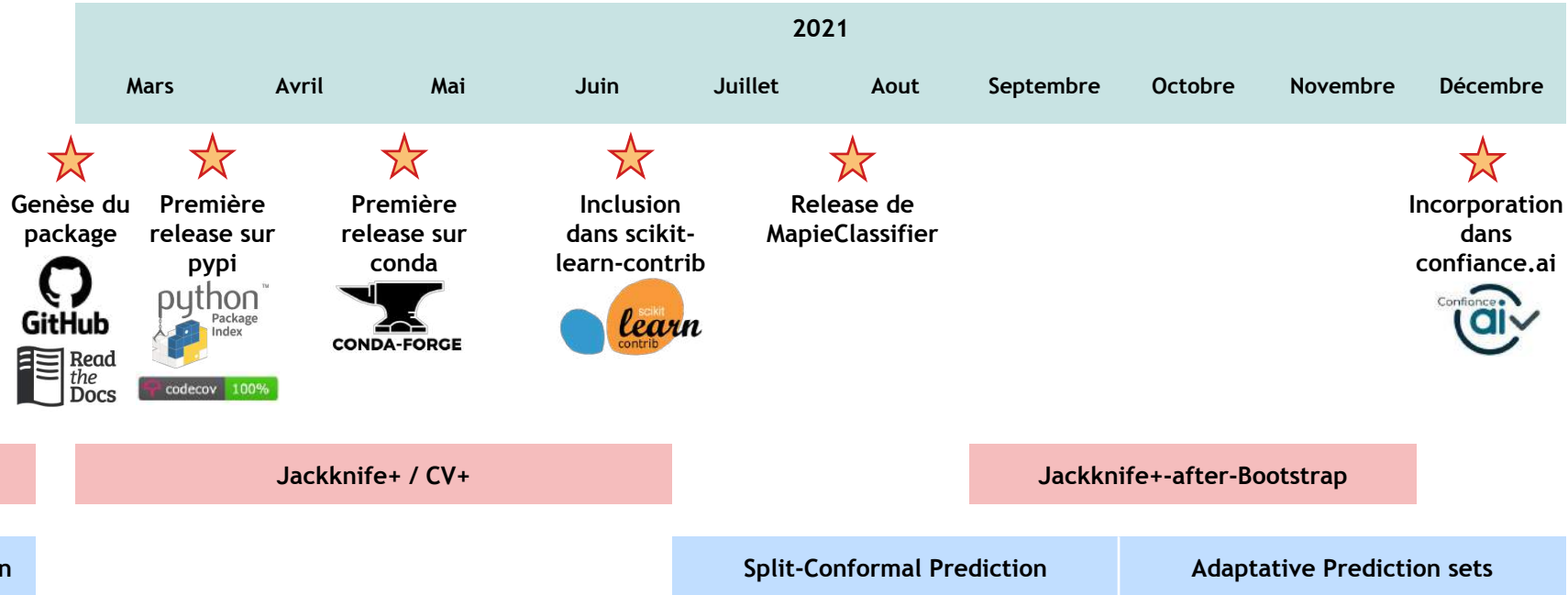
GitHub



Double ambition scientifique

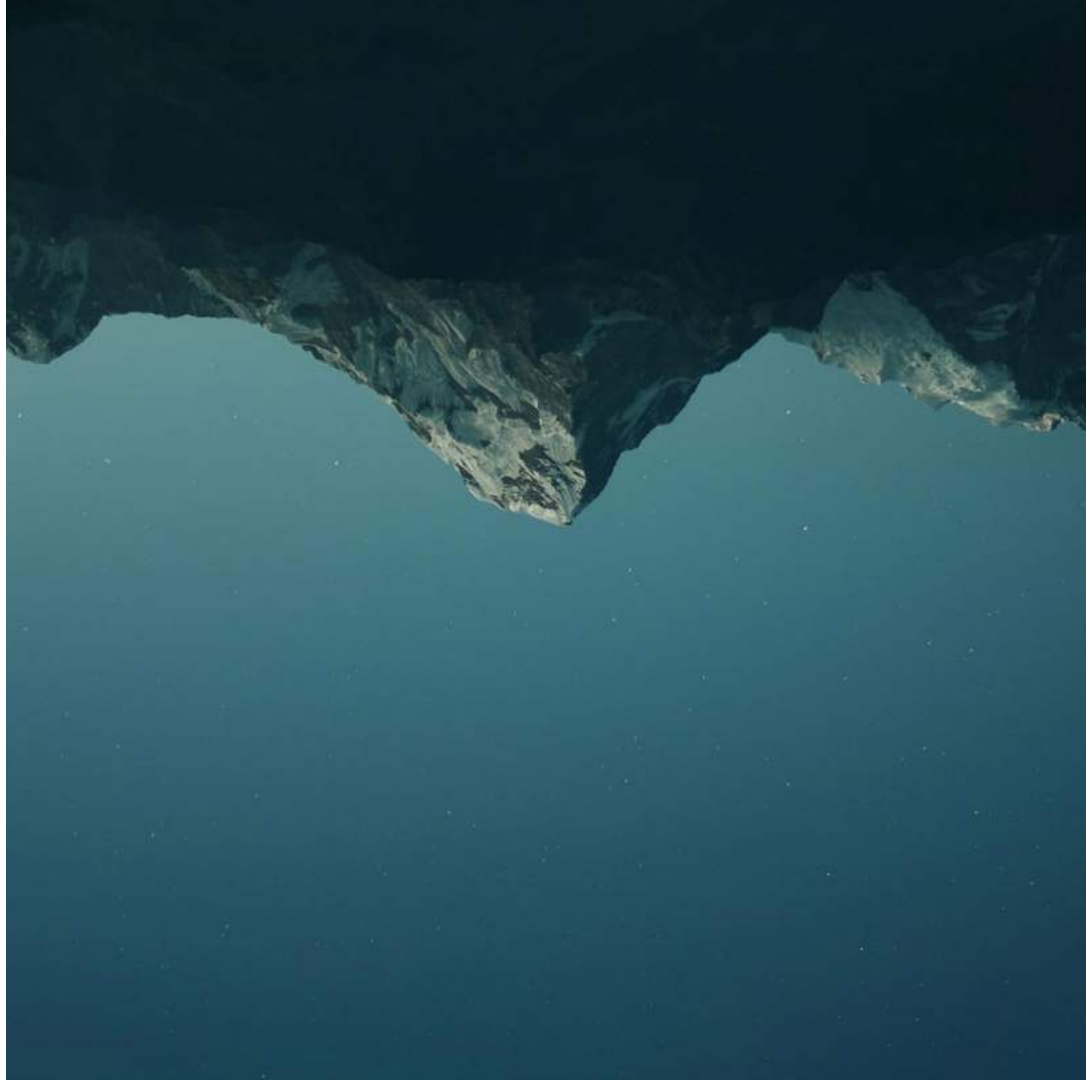
- **Quantification des incertitudes pour tout type de modèle de ML et cas d'usage**
 - Compatible avec n'importe quel régresseur ou classificateur "scikit-learn compatible"
- **Garanties mathématiques sur la couverture marginale des incertitudes**
 - Implémentation de méthodes SOTA de prediction conforme

L'histoire de MAPIE



2

MAPIE pour la Régression



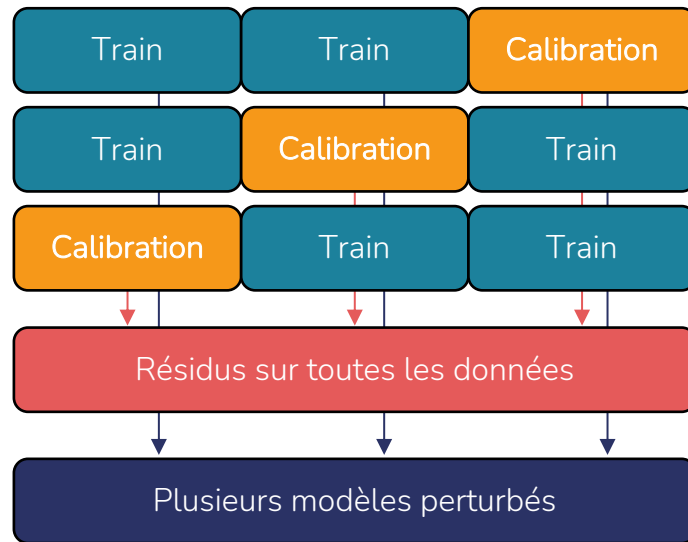


Model Agnostic Prediction Interval Estimator

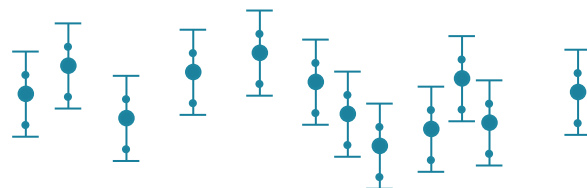
- 👍 Calcul d'incertitudes pour tout problème de régression
- 👏 Implémente l'état de l'art du calcul d'incertitudes*
- 😍 Bâti sur un socle théorique solide et élégant*
- 🍷 Compatible avec n'importe quel estimateur type scikit-learn
- 📁 Publié et documenté en open-source
- 🚀 Packagé sur PyPI et conda-forge : `pip install mapie`
- 🔥 Inclus dans scikit-learn-contrib
- 📊 100% de couverture des tests unitaires

« Avec MAPIE, je suis certain de mes incertitudes. »

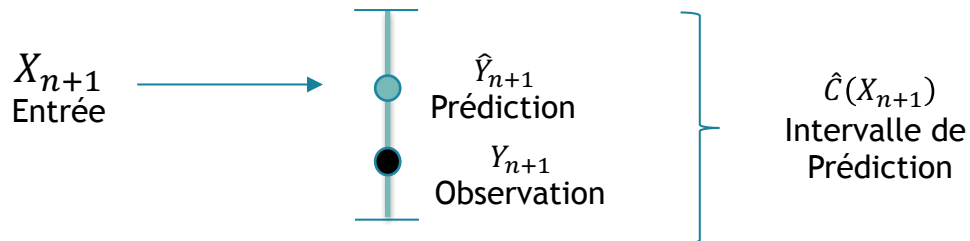
* Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani.
Predictive inference with the jackknife+. Ann. Statist., 49(1):486-507, 022021



Prédictions multiples avec erreurs mesurées
par validation croisée



MAPIE pour la Régression



→ **Garantie sur la couverture marginale :**

$$\mathbb{P}\{Y_{n+1} \in \hat{C}(X_{n+1})\} \geq \textit{couverture cible}$$

Exemple: couverture cible de 90% = MAPIE garantit qu'au moins 90% des observations sont dans leur intervalle de prédiction.

Remarque: C'est une notion **globale** et non **locale**:

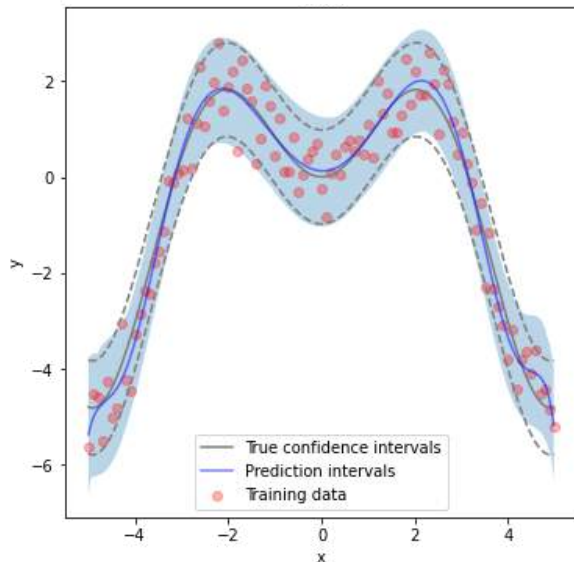
Je peux prédire un cours avec 90% de certitude, mieux si l'indicateur est haut, moins bien s'il est bas

Exemple : Calculer une valeur en fonction d'un indicateur

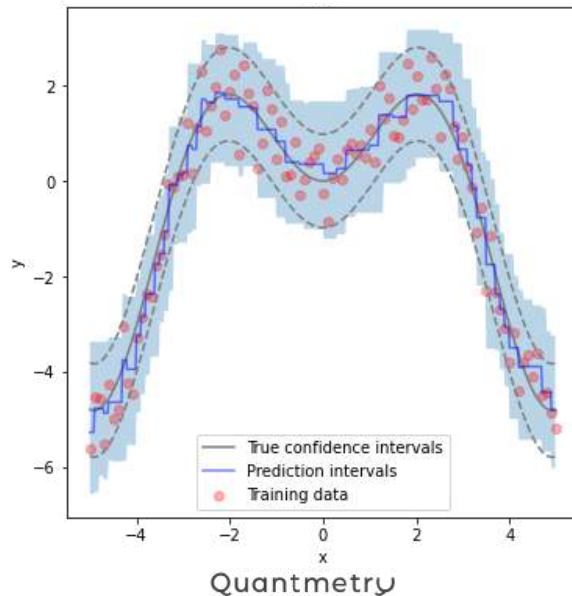
Je ne connais pas la fonction mais des échantillons sur lesquels entraîner un modèle

- **MAPIE permet de reconstruire la fonction avec intervalles d'incertitude**
- **MAPIE fonctionne avec tout type de modèle scikit-learn compatible**

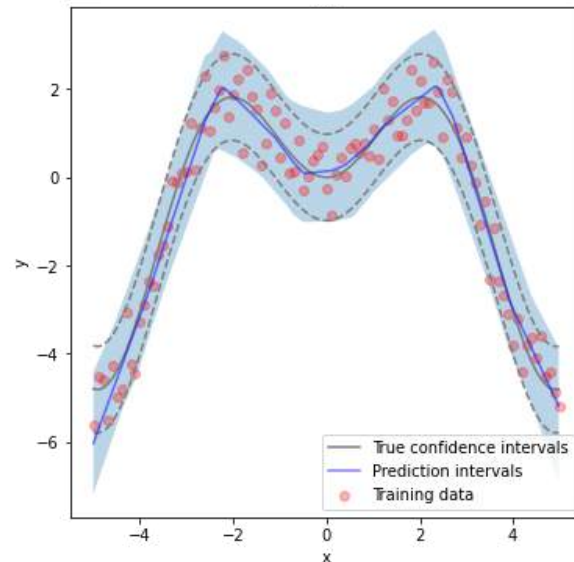
Polynomial function



XGBoost

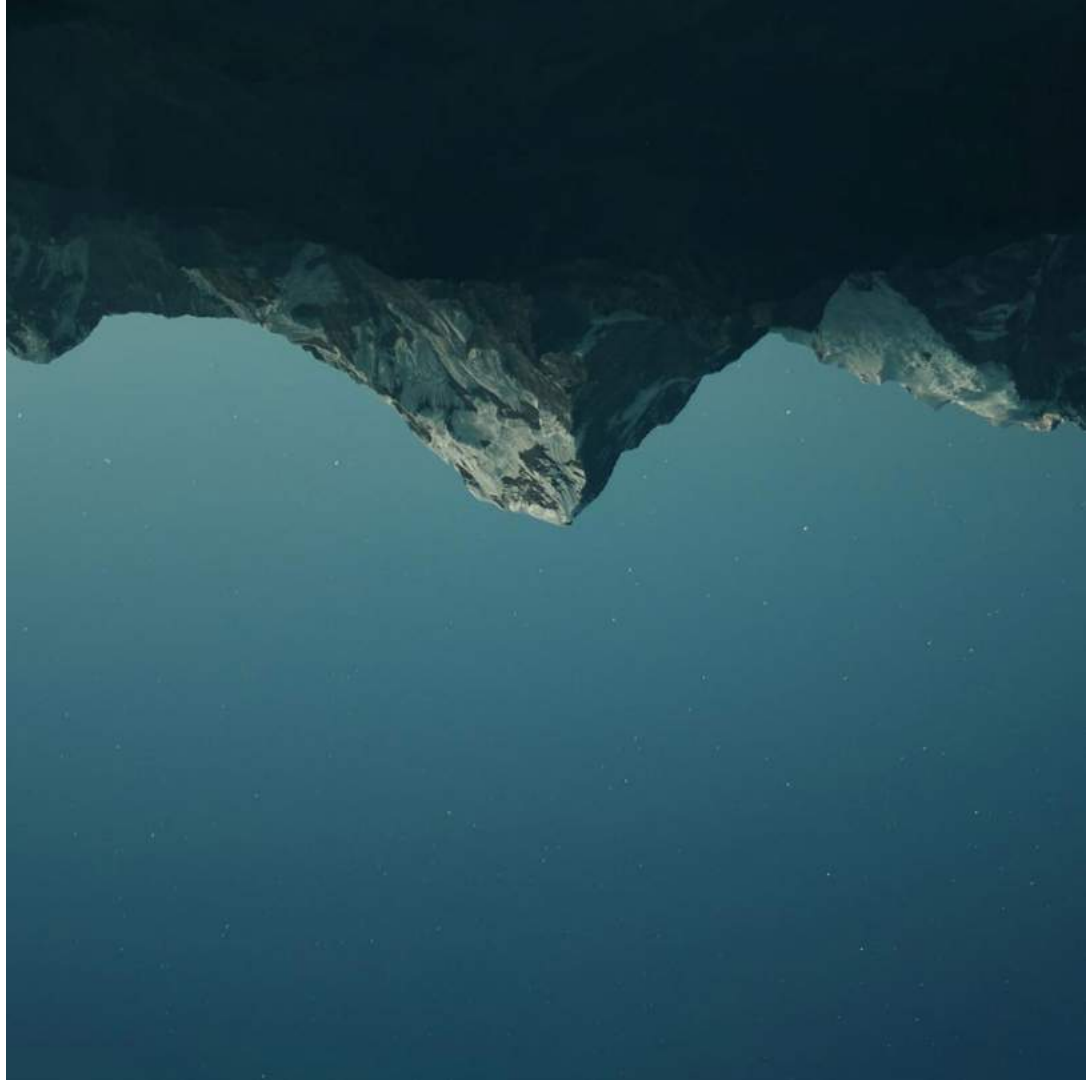


Keras Multilayer Perceptron



3

MAPIE pour les Séries
Temporelles



MAPIE pour les Séries Temporelles

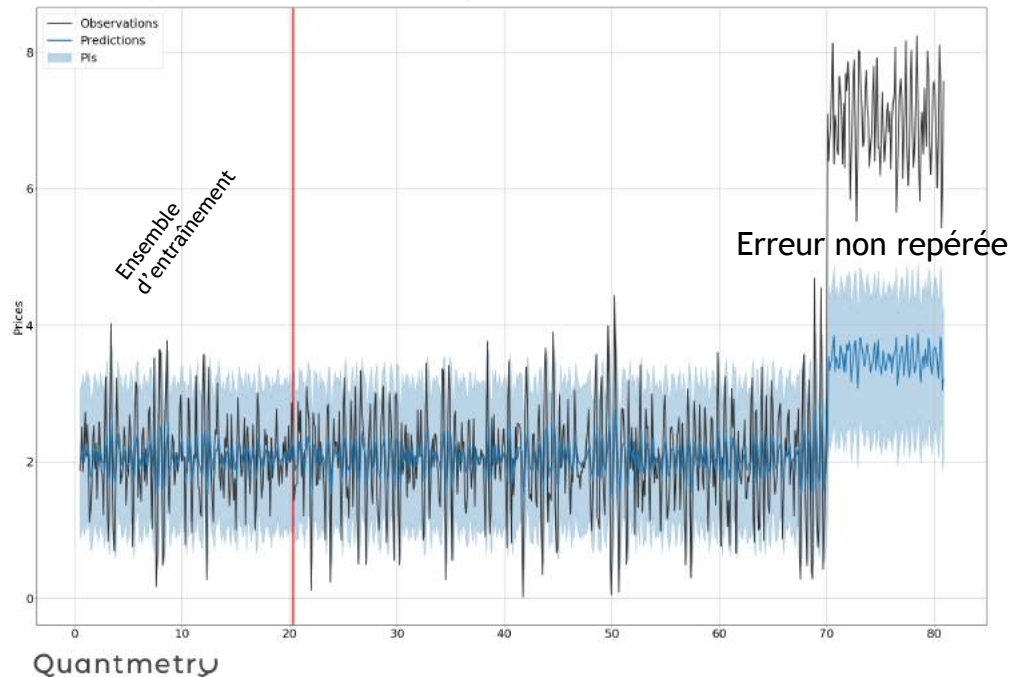
(Implémentation en cours)

- **Points clés:**
 - Les garanties de MAPIE supposent que les **données sont échangeables** (~indépendantes).
 - Ce n'est pas toujours valable, notamment pour les séries temporelles (auto-corrélations)

Exemple:

- Je cherche à mesurer un cours avec des phénomènes dynamiques
 - Relation de corrélations entre les données
 - Changements, parfois brutaux

La classe MAPIE usuelle ne s'applique pas



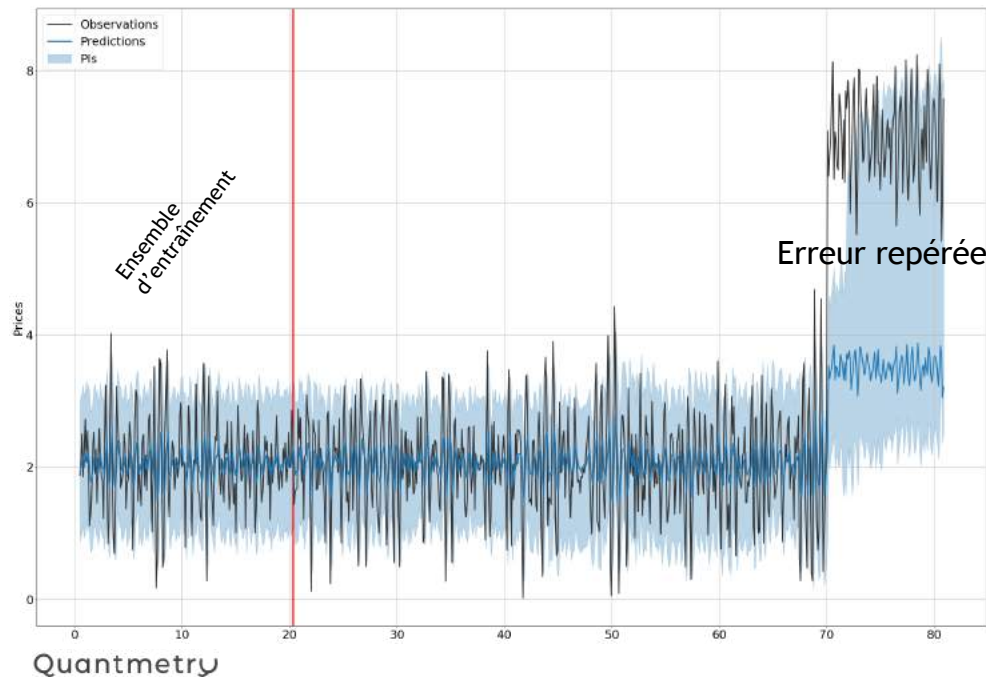
MAPIE pour les Séries Temporelles

(Implémentation en cours)

- **MAPIE Séries Temporelles:**

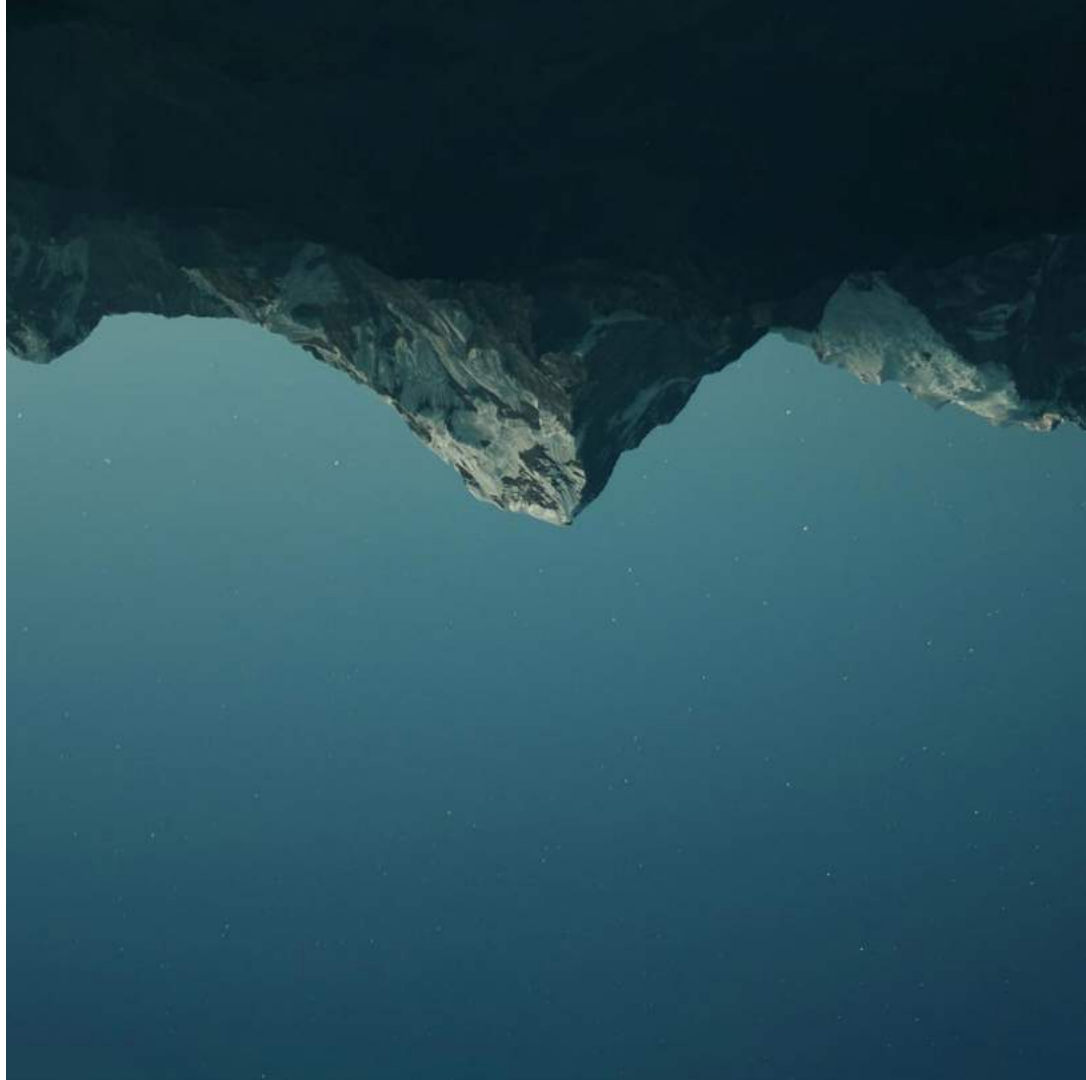
- **Hypothèses plus fortes:**
 - Qualité des estimations
 - Distribution des erreurs
- **Mise à jour dynamique**
des résidus du modèle au cours
des prédictions

Après un temps d'inertie,
**MAPIE intègre les nouvelles
erreurs**



4

MAPIE pour la classification



Que sont la calibration et les prédictions conformes?

Commençons par un exemple :



Vous voulez créer un algorithme de conduite autonome.



Vous vous fixez les contraintes suivantes :



Vous devez être sûr à 90% que votre algorithme vous donne une bonne prédiction.



Si votre algorithme n'est pas sûr de lui, vous voulez savoir s'il hésite plutôt entre un arbre et une voiture ou bien entre un arbre et une personne âgée.

Comment la calibration et la prédictions conformes peuvent nous apporter des réponses ?



Si les scores de prédictions sont calibrés, nous pouvons nous assurer d'avoir ces 90% de certitude



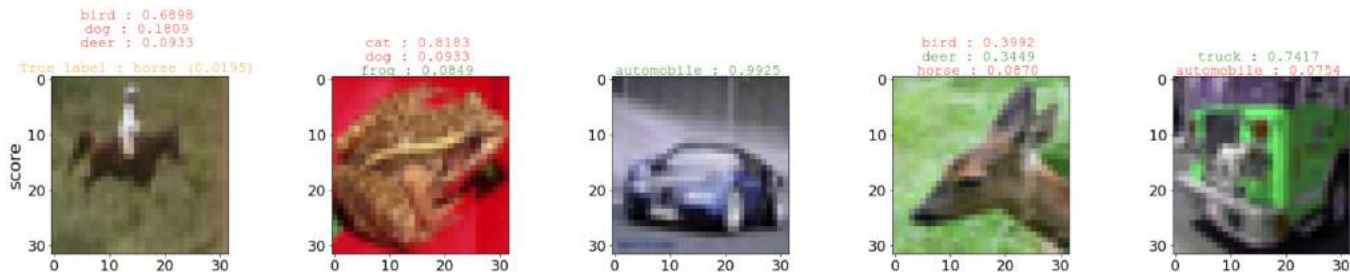
Grâce aux prédictions conformes, nous pouvons savoir si notre algorithme hésite entre un arbre et une personne âgée

Comment MAPIE peut nous aider à atteindre cet objectif ?

Point clé : les sorties d'un modèle de classification ne sont pas des probabilités !

Comment MAPIE peut nous aider à résoudre ce problème ?

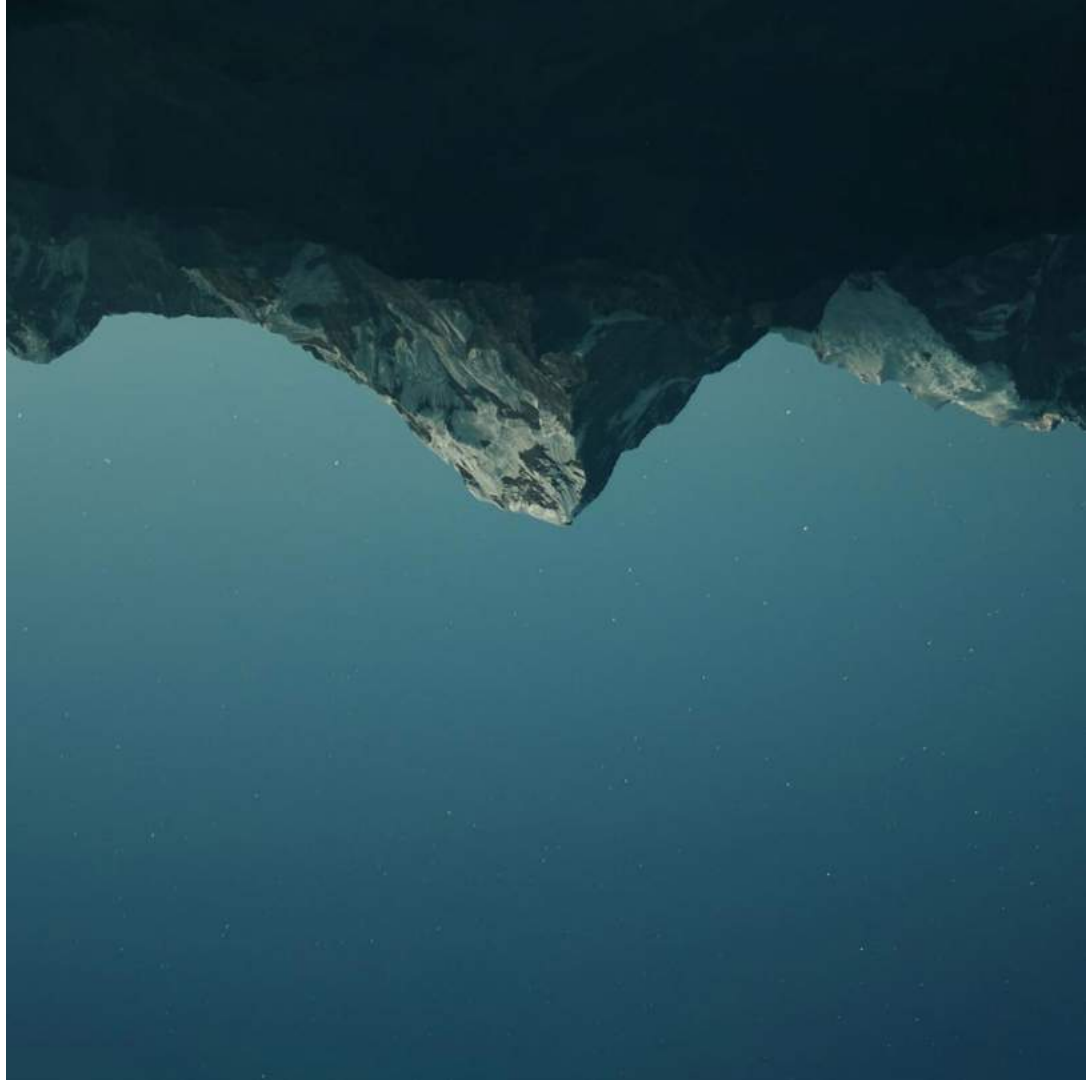
- Ensemble de labels possibles pour une observation avec une garantie de couverture (c'est-à-dire que MAPIE garantie qu'une proportion X de vos jeux de prédictions contiendront le vrai label)
- 🔒 Si votre jeu de prédiction ne comprend qu'un label, alors vous êtes sûrs à X% que ce label est le vrai
- 🚶 Si votre algorithme hésite entre une personne âgée et un arbre, MAPIE inclura les deux labels dans votre jeu de prédiction.



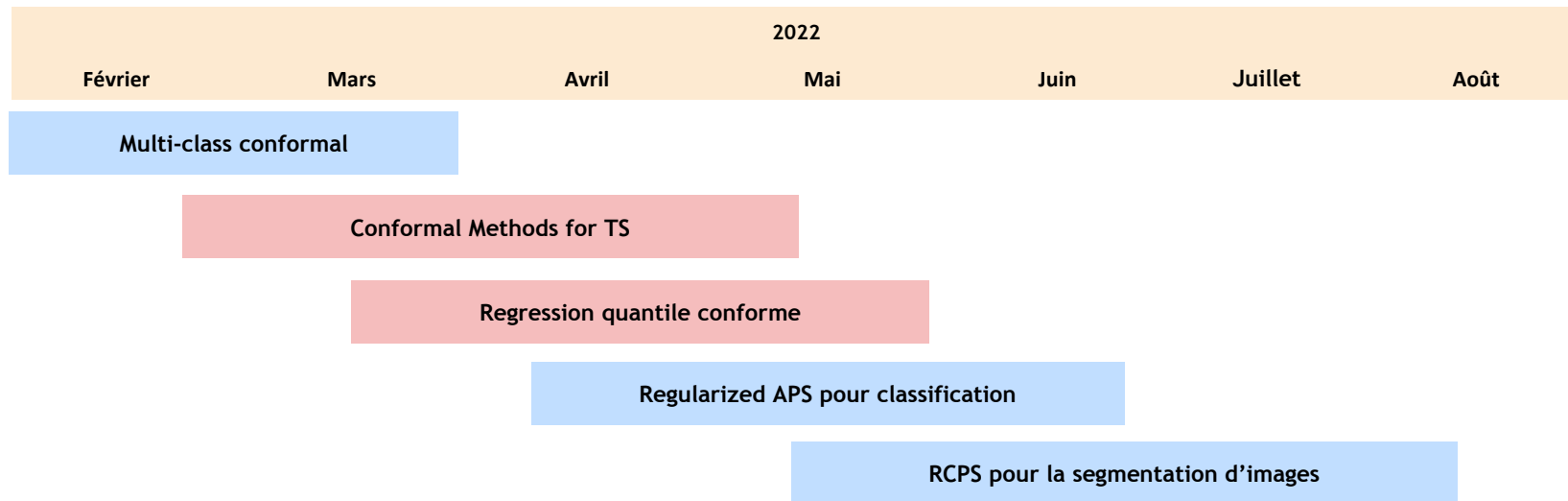
Jeux de prédictions avec un niveau de couverture cible de 90 %

5

Quelles suites pour MAPIE
en 2022 ?



Les nouvelles étapes pour MAPIE en 2022



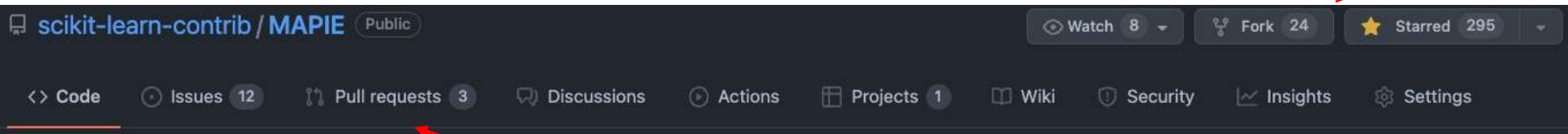


Comment puis-je contribuer à MAPIE ?

<https://github.com/scikit-learn-contrib/MAPIE>

```
pip install mapie conda install mapie
```

Ajouter une étoile



Proposer une nouvelle implémentation



Quantmetry



MERCI !



Merci à nos partenaires



école —————
normale —————
supérieure —————
paris-saclay —————

université
PARIS-SACLAY

