

Prévision de collectes

Projet scientifique 2022



epalia

#desPalettes
etdesHommes

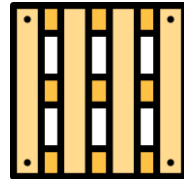


École des Ponts

ParisTech

But du Projet

Création d'un algorithme de prévision de collectes de palettes.



1. Présentation d'Epalia

2. Clustering

3. EMD et Méthode LSTM

4. Méthode ARIMA

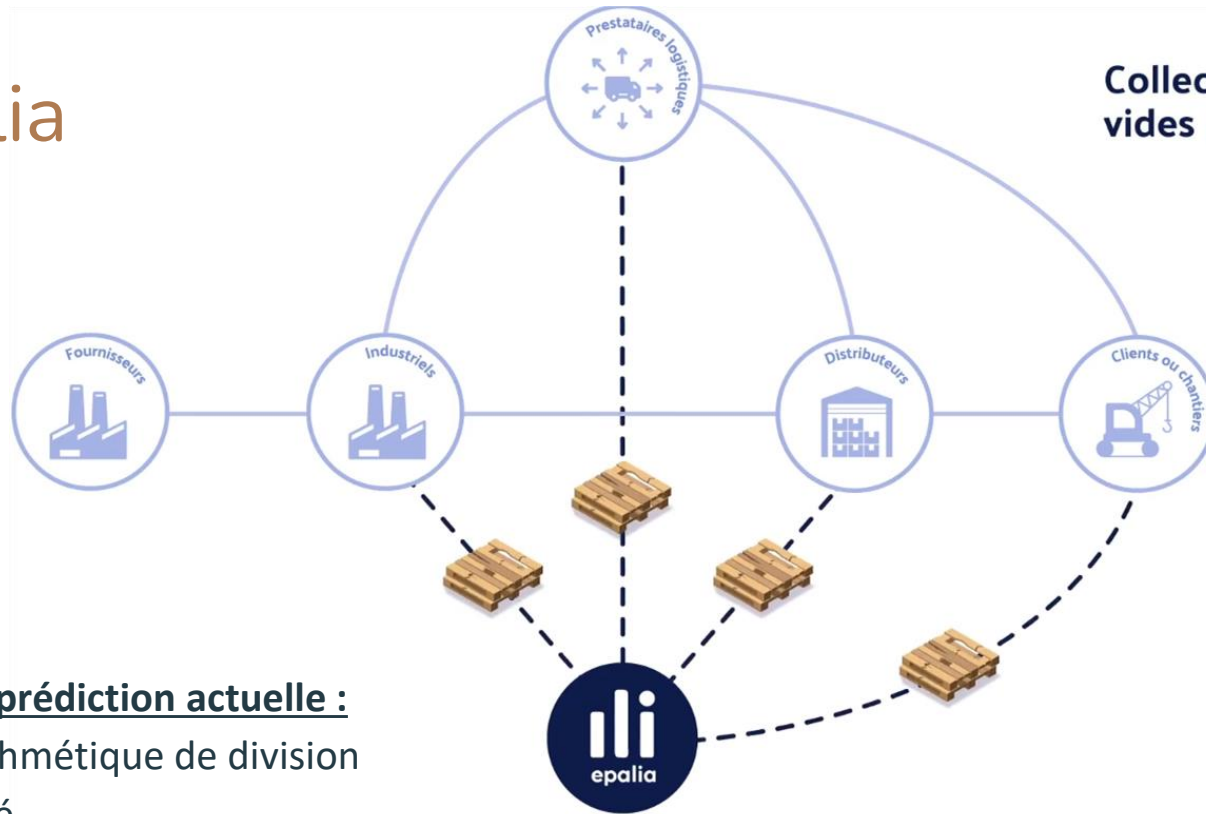
5. Utilisation de l'historique des négoce

6. Conclusion



Epalia

Collecte des palettes vides par Epalia



Méthode de prédiction actuelle :
Méthode arithmétique de division
par jour ouvré

Présentation des données utilisées



Historique de collecte des palettes 2020, 2021, 2022 :

Données de 2020 : Covid + Croissance

→ Seules les données de 2021 ont été utilisées



Historique des livraisons fournisseur-négoce depuis 2015 :

Transmis par les fournisseurs

→ Utilisation des données de 2021

1. Présentation d'Epalia
2. **Clustering**
3. EMD et Méthode LSTM
4. Méthode ARIMA
5. Utilisation de l'historique des négoce
6. Conclusion



Clustering : Objectif



Problème : Par dépôt, les **données sont trop peu nombreuses** pour permettre une prédiction fiable.

Solution : On **agglomère** les dépôts par zone géographique pour faire la prédiction sur les sous-ensembles. Une fois la prédiction faite, on pourra **désagréger** en attribuant à chaque dépôt sa proportion en volume à l'année.

→ Clustering

Utilisation de plusieurs méthodes différentes :

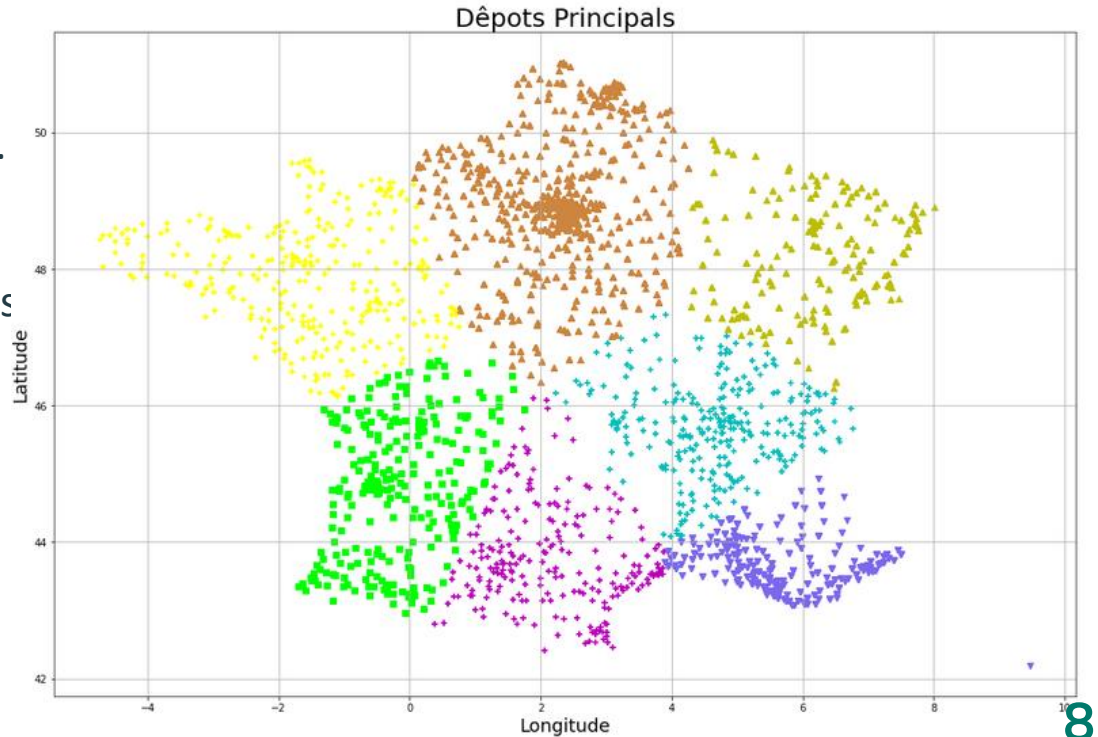
K-means, DBSCAN et HDBSCAN, regroupement par site de collecte

Clustering : K -means

Rapprochement **géographique**.

On associe chaque dépôt au cluster dont le centre est le plus **proche**.

Utilisation de la méthode du **coude** : → 7 zones distinctes

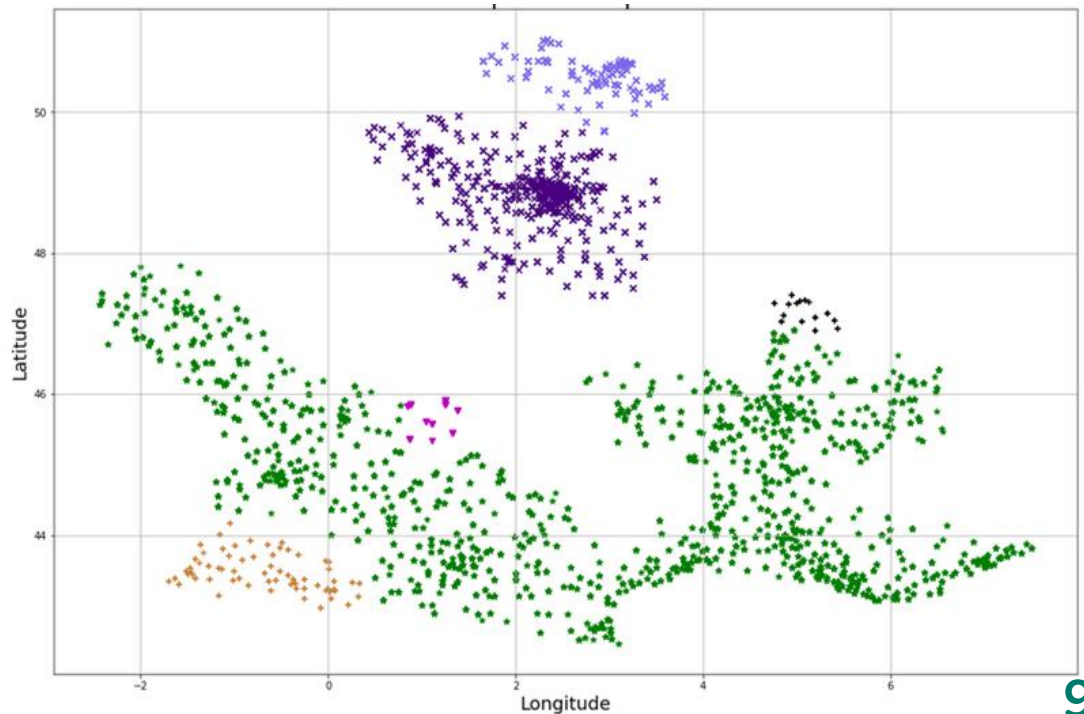


Clustering : DBSCAN

Rapprochement par **densité**
des dépôts

Problèmes :

- Des dépôts pas pris en compte
- Clusters trop étalés géographiquement

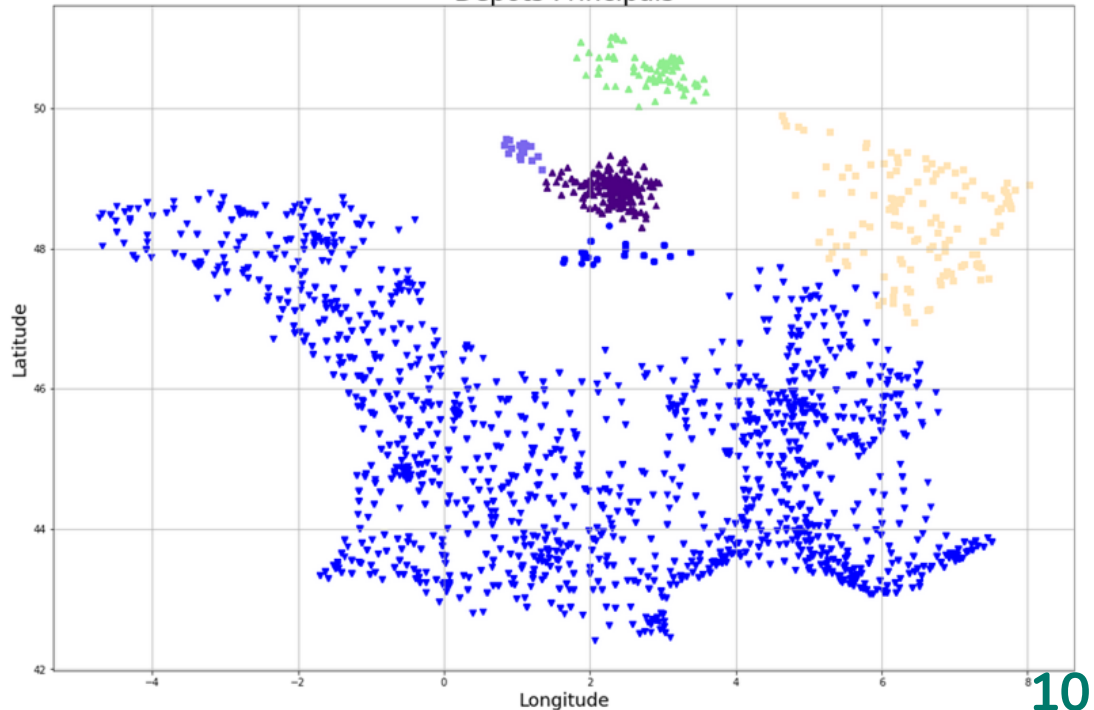


Clustering : HDBSCAN

Rapprochement par **densité**

→ Mêmes problèmes, zones encore plus larges.

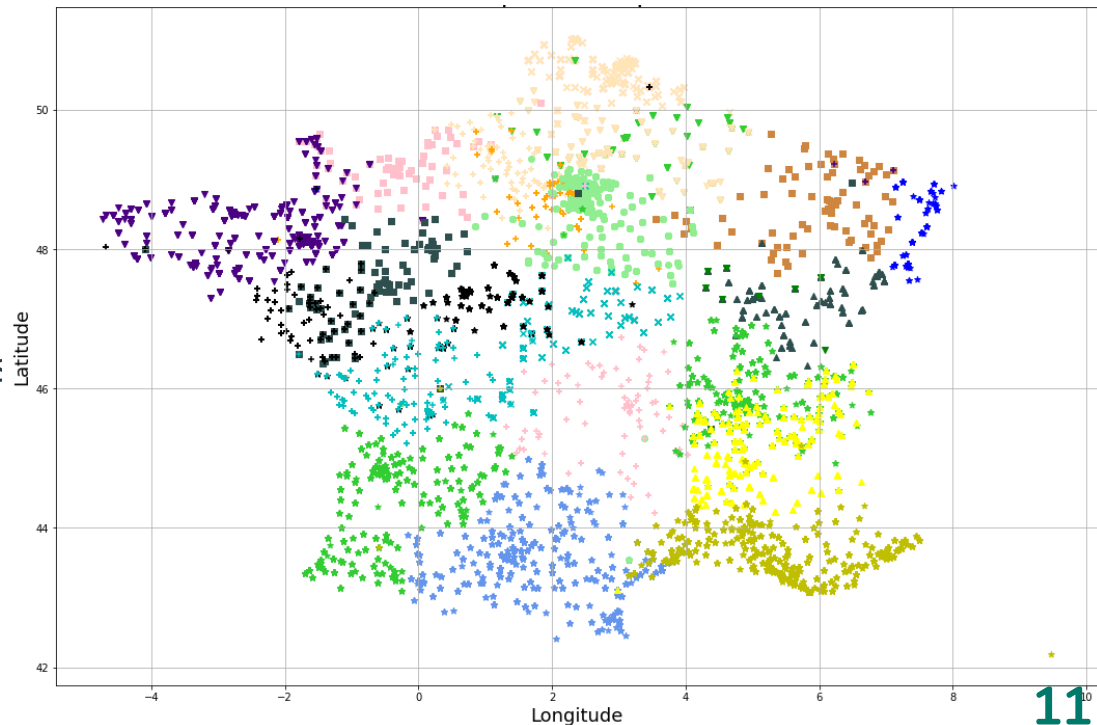
→ Isolation de la zone parisienne qui peut être intéressante pour compléter notre K-means



Clustering par site

On regroupe ensemble les dépôts desservis par un même **site de collecte**.

Malgré quelques recouvrements et dépôts isolés, le clustering semble **pertinent** géographiquement.



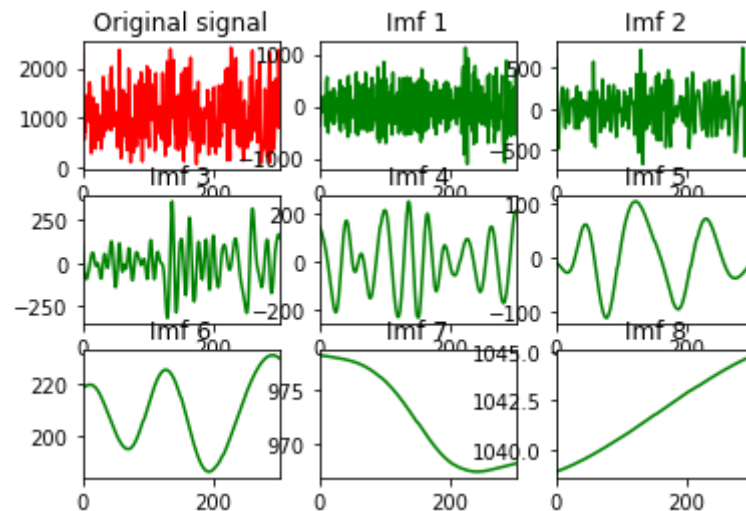
1. Présentation d'Epalia
2. Clustering
- 3. EMD et Méthode LSTM**
4. Méthode ARIMA
5. Utilisation de l'historique des négoce
6. Conclusion



EMD

Empirical Mode Decomposition

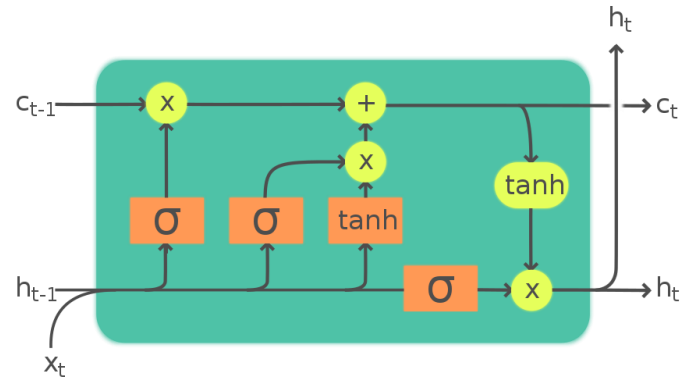
Une méthode pour décomposer le signal en différentes courbes.



LSTM

Long short-term memory

C'est un neural network basé sur le RNN (recurrent neural network). On l'utilise souvent dans la prédiction du temps.



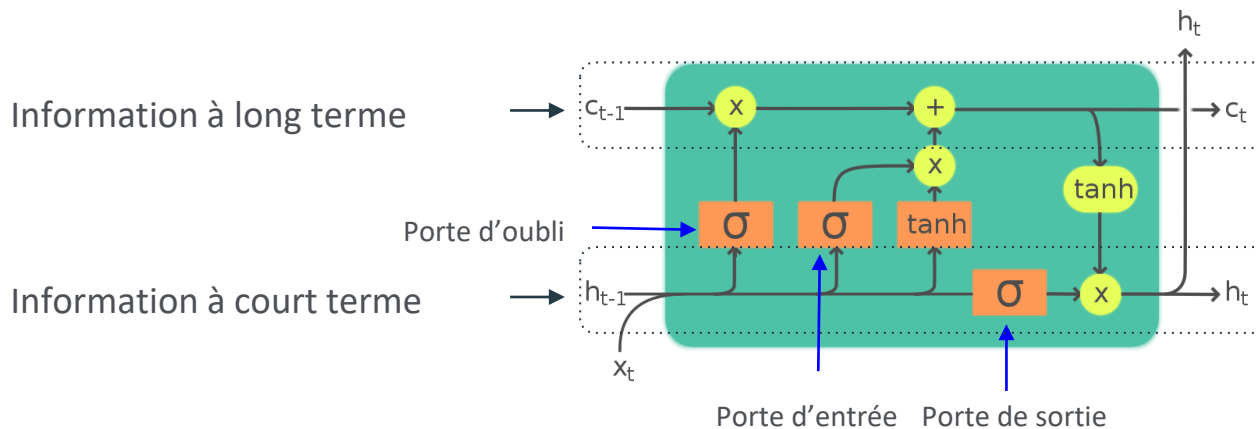
Legend:

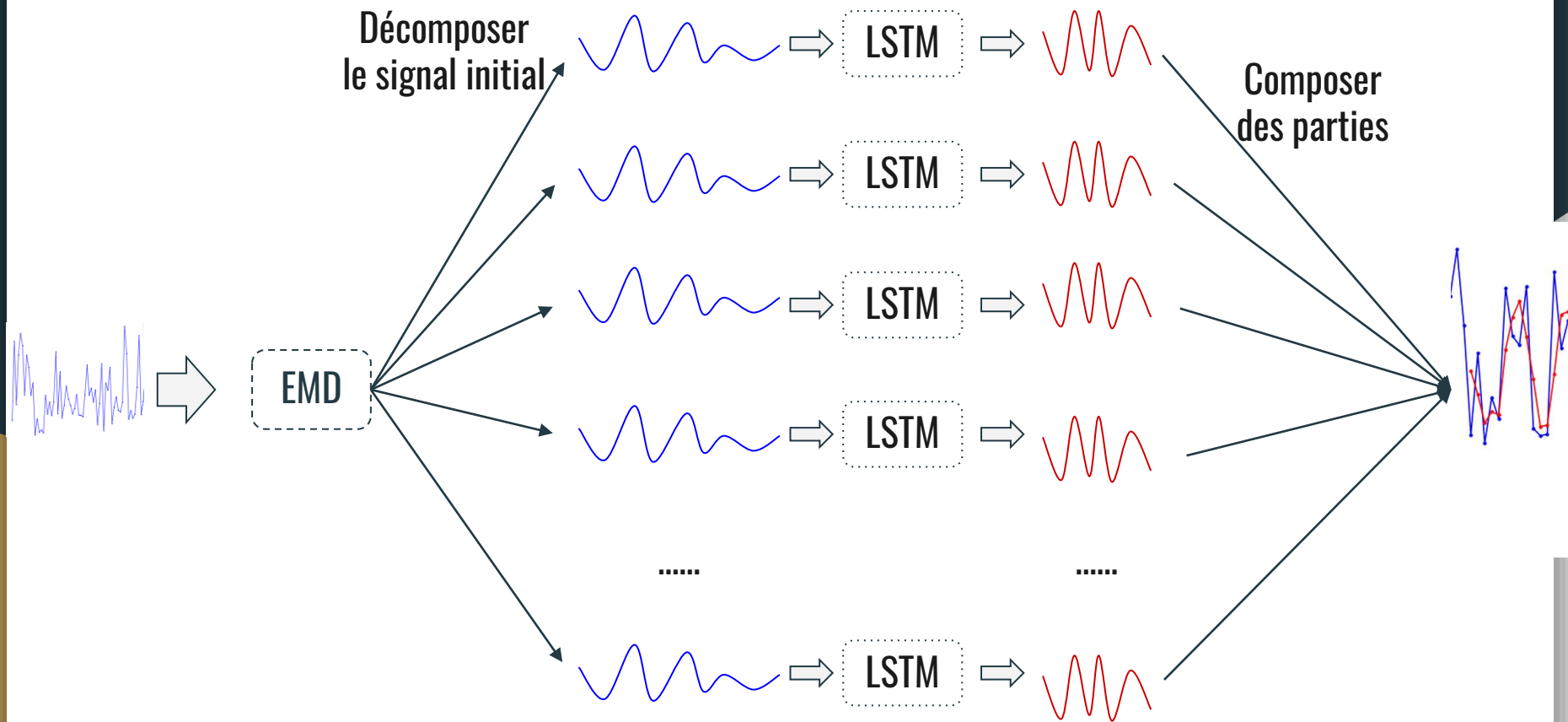
Layer	ComponentwiseCopy	Concatenate

LSTM

Toutes les cellules peuvent conserver les informations à long terme et à court terme. Et la cellule peut les renouveler par trois porte:

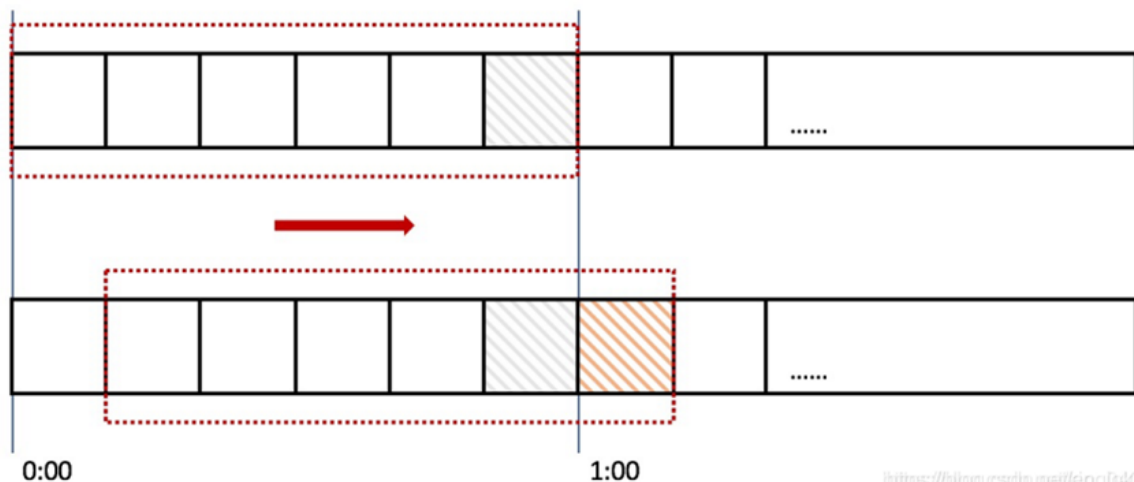
Porte d'oubli, Porte d'entrée et Porte de sortie





Cross Validation

Cross-validation: on la réalise avec la fenêtre glissante.



<https://blog.csdn.net/engjoke6>

Le choix de paramètre: Bayesian Optimization

Prévision

On fait la prévision sur trois échelles différentes :

Prévision quotidienne

Prévision hebdomadaire

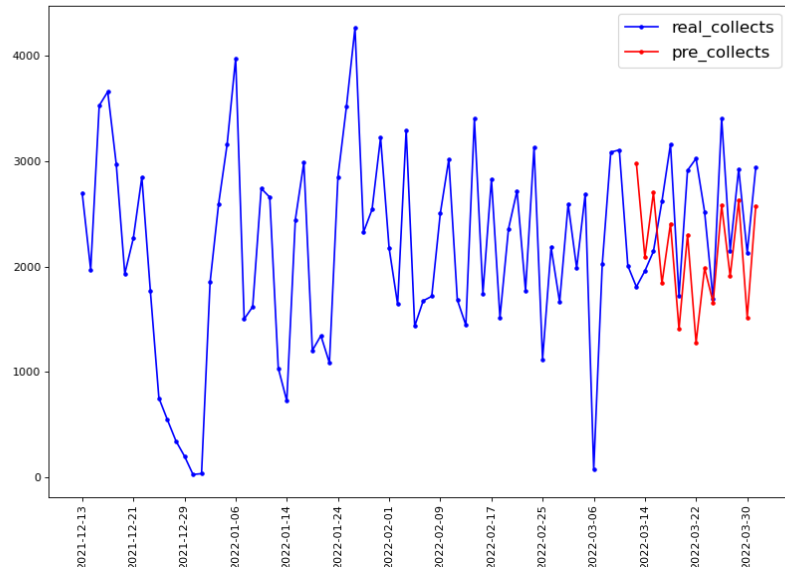
Auto-prévision sur un mois

Prévision quotidienne

On prévoir “demain” basé sur des données “aujourd’hui”

Date	Réel	Prévision
2022-03-05	172	-
2022-03-06	3	70
2022-03-07	390	290
2022-03-08	388	510
...

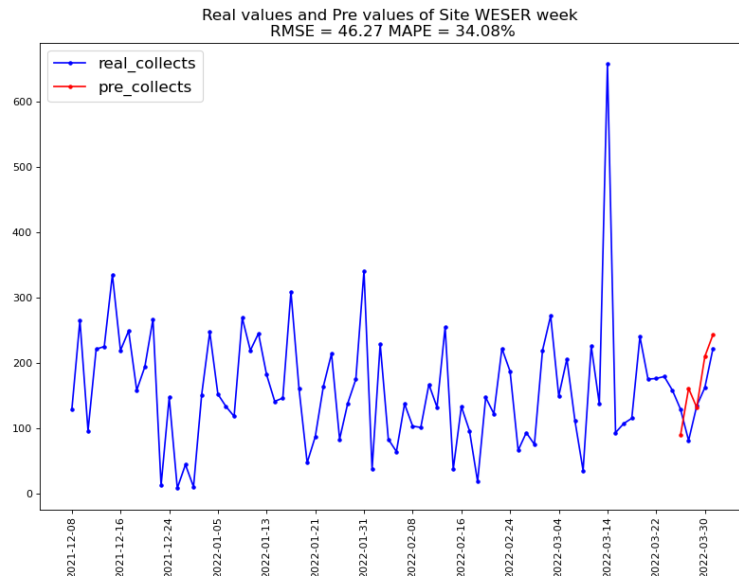
Real values and Pre values of Site BOUYER LEROUX
RMSE = 732.33 MAPE = 23.87%



Prévision hebdomadaire

On prévoir “la semaine prochaine” basé sur des données “cette semaine”

Date	Réel	Prévision
2022-03-22 ~ 2022-03-26	[221, 182, 183, 185, 151]	-
2022-03-29 ~ 2022-04-02	[120, 87, 127, 155, 201]	[108, 157, 131, 197, 223]
...

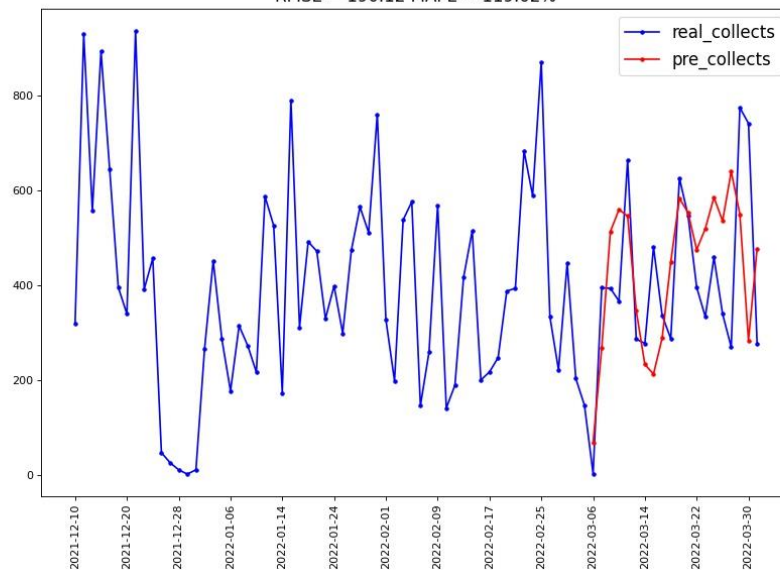


Auto-prévision sur un mois

Dans cette version de prévision, on prévoit les collectes par les données “prévues” (à part la première prévision quotidienne).

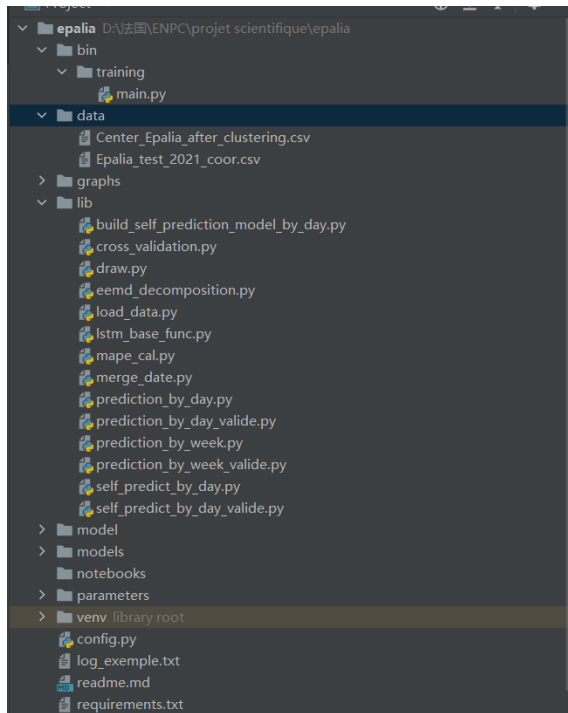
Date	Réel	Prévision
2022-03-05	172	-
2022-03-06	3	70
2022-03-07	390	290
2022-03-08	388	510
...

Real values and Pre values of Site XELLA THERMOPIERRE FRANCE for 1 months mean
RMSE = 190.12 MAPE = 119.02%

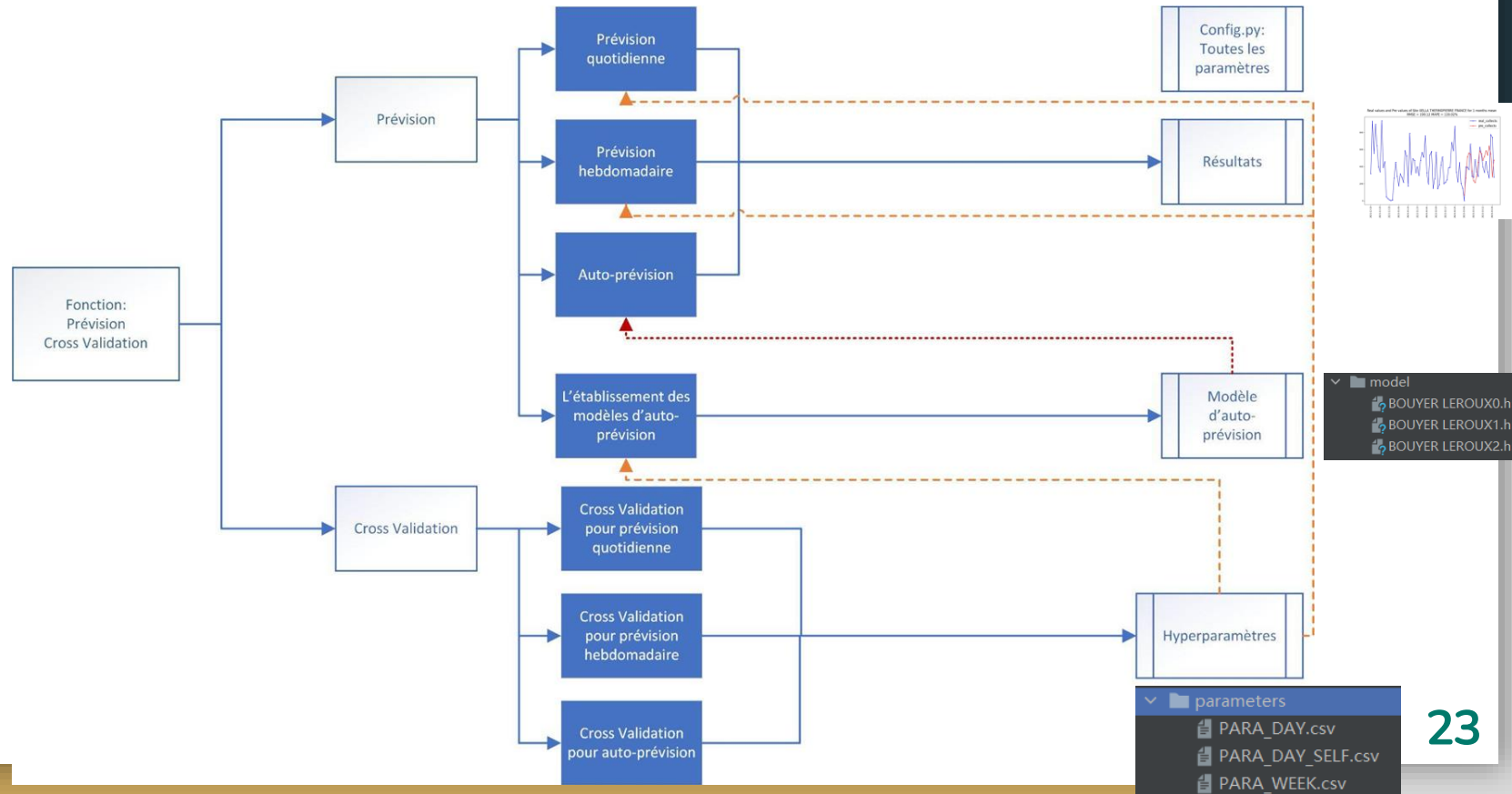


L'emballage des codes

Pour que l'entreprise puisse réutiliser ce que l'on a fait et le modifier, on a emballé notre code.



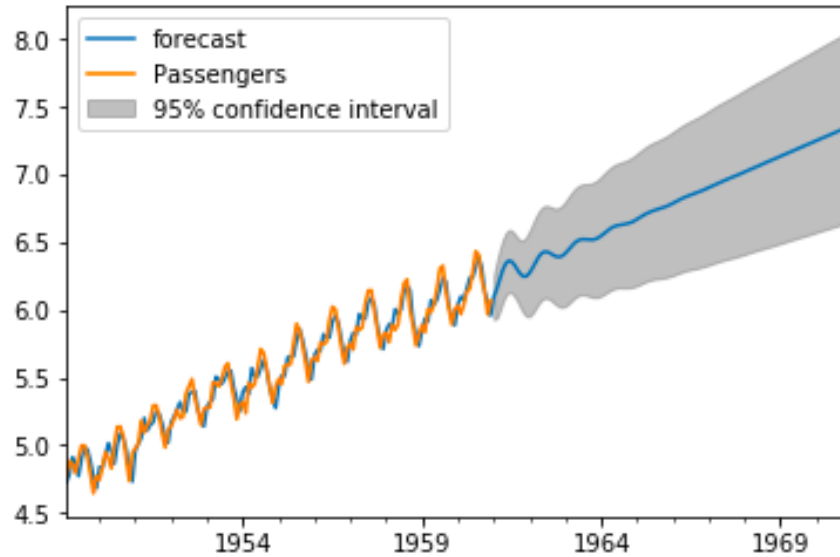
L'emballage des codes



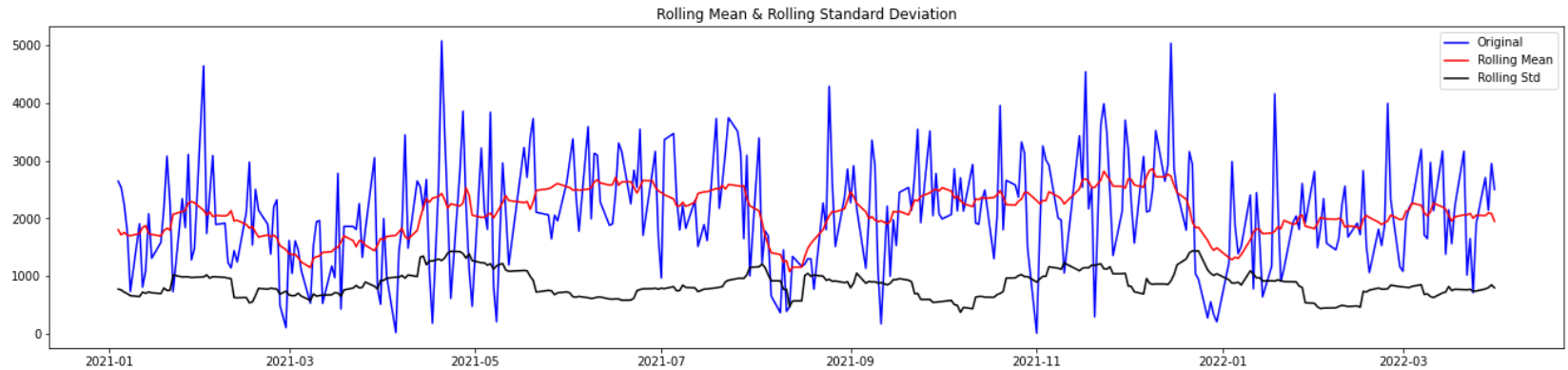
1. Présentation d'Epalia
2. Clustering
3. EMD et Méthode LSTM
- 4. Méthode ARIMA**
5. Utilisation de l'historique des négoce
6. Conclusion



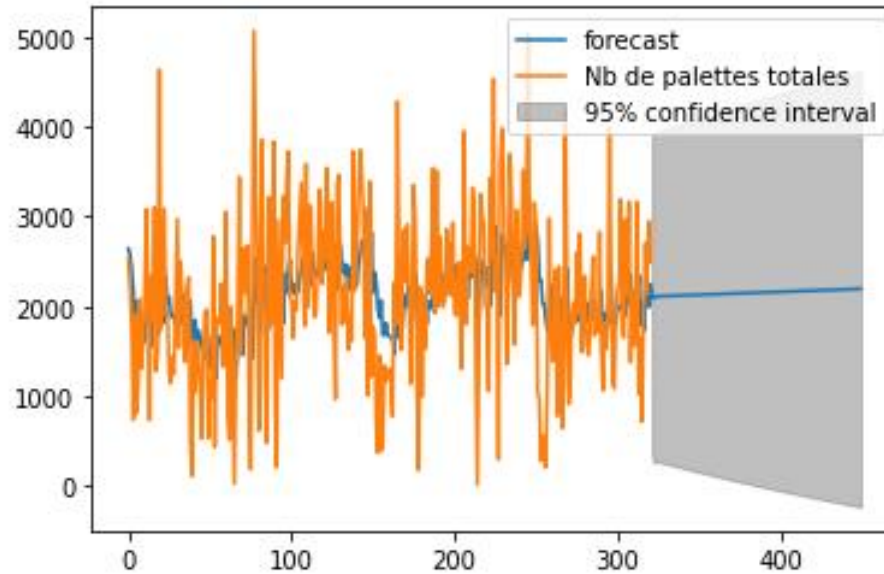
Modèle ARIMA



Moyenne roulante et variance



Prédictions



1. Présentation d'Epalia
2. Clustering
3. EMD et Méthode LSTM
4. Méthode ARIMA
- 5. Utilisation de l'historique des négoces**
6. Conclusion

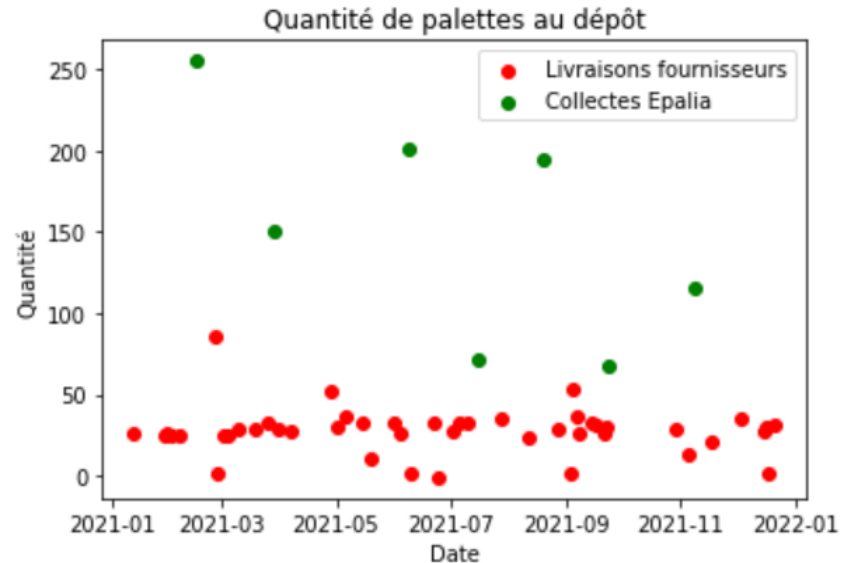


Utilisation de l'historique fournisseur-négoce

Base de données sur les palettes livrées par les fournisseurs aux négociants.



Objectif : avoir une vision avant leur utilisation des palettes qui vont être récupérées par Epalia.

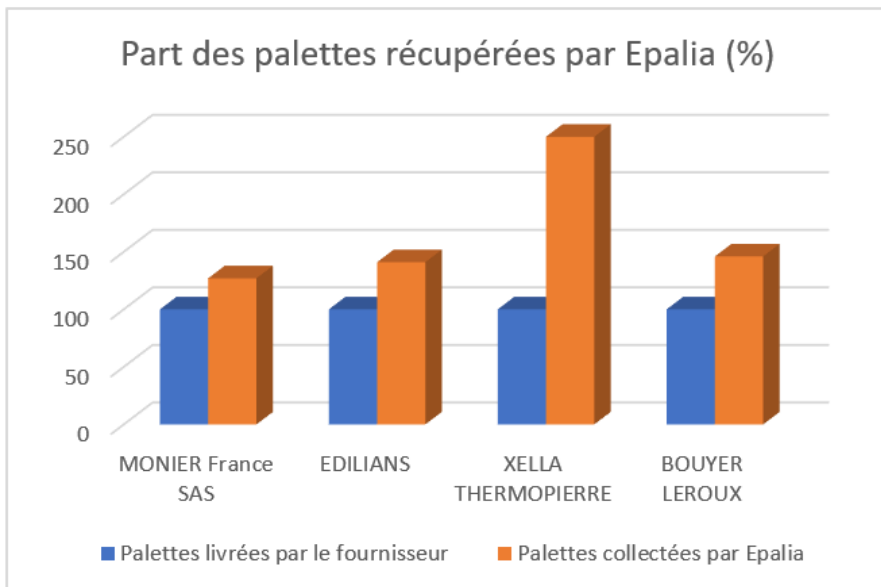


Observation réalisée

Le nombre de palettes reçues par les fournisseurs dont on a les données est toujours **plus important** que le nombre de palettes livrées.

Explications possibles :

Les livraisons ne sont pas toutes **communiquées**, ou les dépôts reçoivent des palettes provenant d'un chargeur par **l'intermédiaire** d'autres fournisseurs.



1. Présentation d'Epalia
2. Clustering
3. EMD et Méthode LSTM
4. Méthode ARIMA
5. Utilisation de l'historique des palettes
6. **Conclusion**



Conclusion

- Signaux temporels très **variables** à tous les niveaux, ce qui rend la prédiction très difficile.
- La **désagrégation** par dépôts après prédiction est compliquée : certains dépôts ne sont collectés qu'une à deux fois par an.
- Nécessité de comparer nos résultats avec la prédiction réalisée par Epalia

Conclusion

KPI : MAPE, RMSE, Somme des écarts

Méthodes de base : moyenne, moyenne actualisée, méthode Epalia

Méthode	Moyenne	Moyenne actualisée	Epalia
Somme écarts pondérée	174%	174%	30700%
RMSE	52	52	83

Conclusion

Nouvel indicateur : RMSE restreint aux livraisons > 0

Nouvelles méthodes : Moyenne par livraison, mix Epalia/moyenne par livraison

Méthode	Moyenne	Moyenne actualisée	Epalia	Moyenne livraison	Mix Epalia - moyenne
PRMSE (à minimiser)	173	178	128	127	110