

PROJET 7

Méthodes de prévision des séries chronologiques

LO Ousmane , Machine Learning Engineer

Openclassrooms

November 27, 2022



SOMMAIRE

- 1 Contexte et objectif
- 2 Prévion à l'aide de modèles classiques
- 3 Prévion à l'aide de modèles récents
- 4 Comparaison de Modèles
- 5 Conclusion

Contexte et Objectif

Améliorer les prévisions sur les méthodes classiques avec cette démarche :

- Récupérer un jeu de données récent et plus réel
- Utiliser quelques méthodes classiques et récentes sur ces données
- Analyser les performances des modèles
- Comparer ces deux méthodes en fonction de leur performances

Jeu de données Energydataset : disponible dans Darts Library

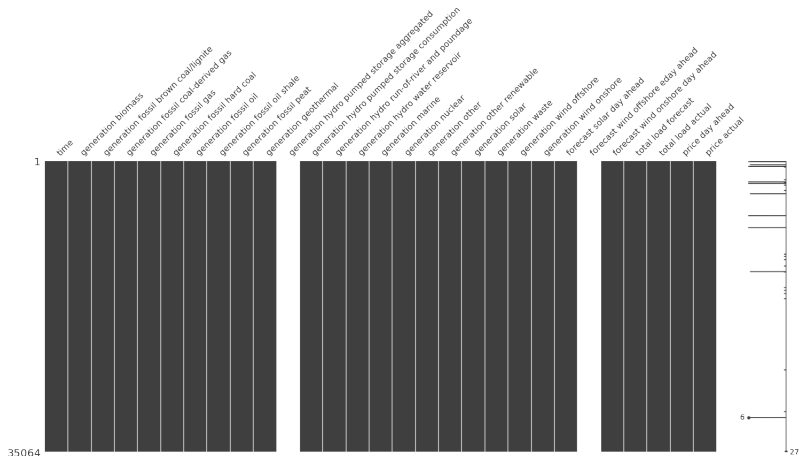


Figure: Plot dataset's train an test



Nombre ligne	Nombre de colonne	Nombre de types	
35064	39	objet	float
		1	28

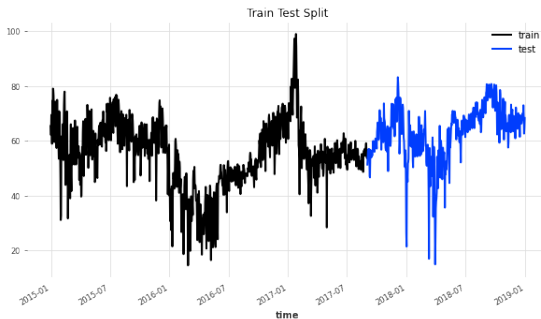


Figure: Plot dataset's train and test

Prévision à l'aide de modèles classiques

On effectue des prévisions sur les modèles classiques avec ses quatre modèles:

- Auto ARIMA
- Prophet
- Tbeats
- Xgboost

Auto ARIMA

Dans le modèle ARIMA de base, nous devons fournir les valeurs p , d et q qui sont essentielles. Des techniques statistiques sont utilisés pour générer ces valeurs en traçant l'ACF et le PACF. Dans Auto ARIMA, le modèle lui-même générera les valeurs p , d et q .

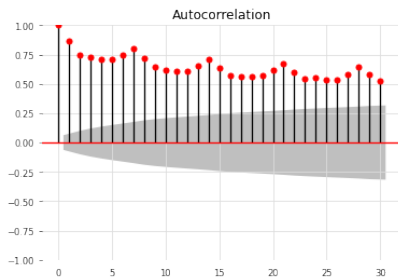


Figure: ACF

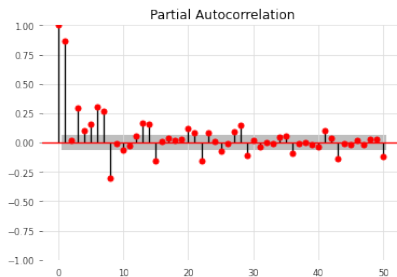


Figure: PACF

Resultats Auto ARIMA

Ce modèle auto arima donne SARIMAX(1,0,2) comme modèle optimale avec une erreur RMSE de 12.10%

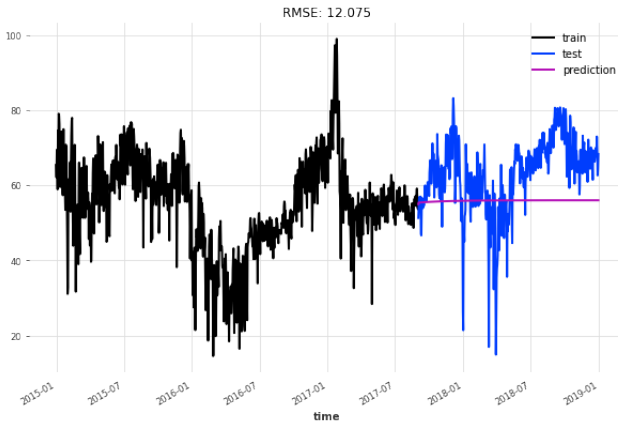


Figure: Predictions by Auto arima model



Prophet

Facebook prophet est basée sur un modèle additif. Le Modèle additif se présente ainsi : $y(t) = g(t) + s(t) + \epsilon(t)$ où

- $y(t)$ correspond à la modélisation de la série temporelle,
- $g(t)$ la tendance,
- $s(t)$ la composante saisonnière,
- $\epsilon(t)$ la composante aléatoire ou erreur.

Facebook prophet a pour but de résoudre les problématiques suivantes :

- La difficulté de créer des modèles de prévision fiables
- La rigidité et le manque de robustesse des techniques de prévisions automatiques



Resultats Prophet

Les predictions sont un peu éloignés de la validation avec une erreur RMSE très élevée de 30.51% comparé à Auto arima qui est de 12.10%.

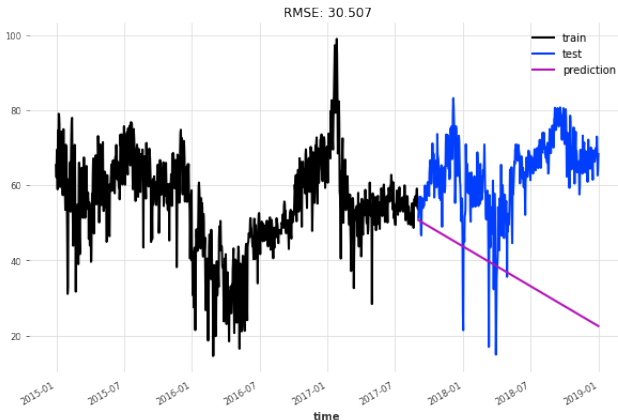


Figure: Predictions by Prophet model

TBATS est une méthode de prévision pour modéliser des données de séries chronologiques. L'objectif principal est de prévoir des séries chronologiques avec des modèles saisonniers complexes en utilisant le lissage exponentiel

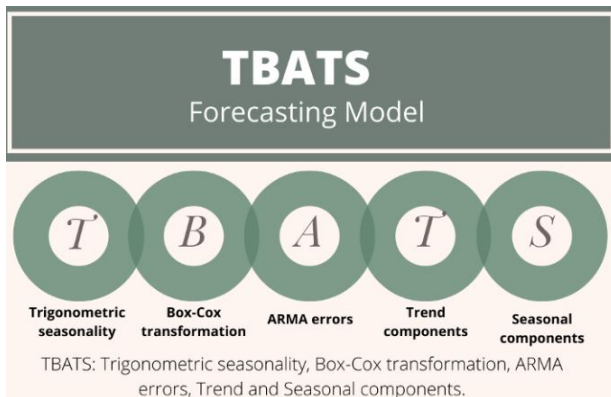


Figure: TBATS

How Tbats choose final model

TBATS envisagera diverses alternatives et s'adaptera à un certain nombre de modèles. Il considérera des modèles :

- Le modèle final sera choisi en utilisant le critère AIC

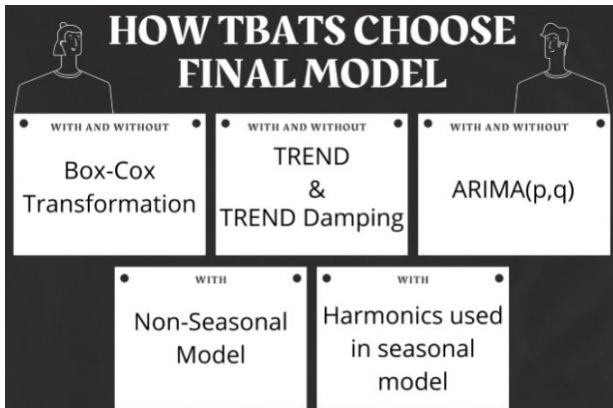


Figure: How choose final model

Resultats Tbeats

Les prédictions obtenus avec le modèle Tbeats semble légèrement correctes comparé aux modèles Auto arima et prophet. Ce modèle fournit une erreur RMSE de prévision de 13.03.

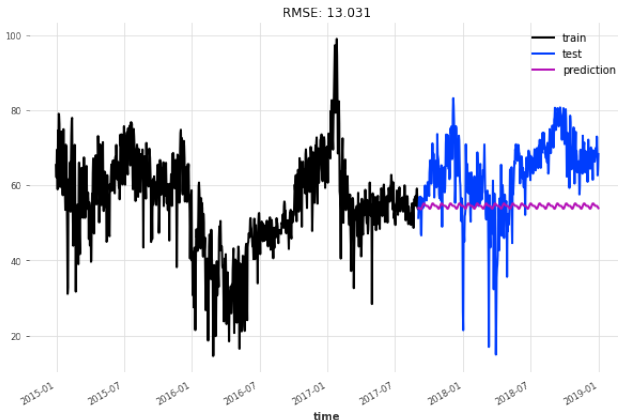


Figure: Predictions by Tbeats model

XGBoost (Extreme Gradient Boosting) est une implémentation de gradient boosting pour les problèmes de classification et de régression. Cela peut être utilisé pour la prévision de séries chronologiques en restructurant l'ensemble de données d'entrée pour ressembler à un problème d'apprentissage supervisé.

time	price actual	Target
2017-09-04	56.404167	56.906667
2017-09-05	56.906667	54.146250
2017-09-06	54.146250	55.586250
2017-09-07	55.586250	56.970000
2017-09-08	56.970000	47.089583
...
2018-12-26	69.054167	72.959583
2018-12-27	72.959583	69.230417

Figure: dataframe using in XGB



Resultats Xgboost

La prévision avec le modèle Xgboost fournit de très bon résultat avec une erreur de prévision de 6.42% comparé aux modèles Auto arima et prophet

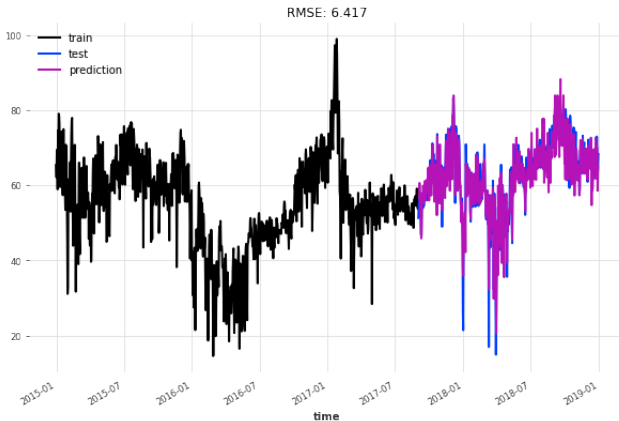


Figure: Predictions by Xgboost model

Prévision à l'aide de modèles récents

On essayera ces 3 modèles récentes basées sur la formation d'un modèle de réseau récurrent auto-régressif sur un grand nombre de séries temporelles connexes:

- Nbeats
- Tft
- DeepAR

NBEATS est une étape intéressante dans l'application de l'apprentissage en profondeur aux séries chronologiques, car il crée une architecture dédiée aux séries chronologiques en utilisant une approche complètement différente :

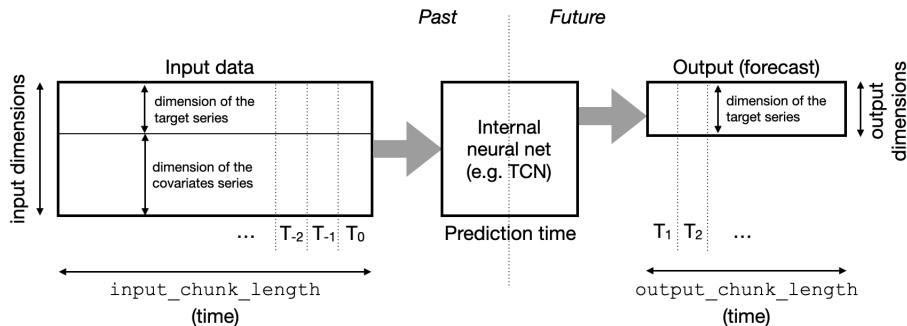


Figure: Architecture Nbeats model



Resultats Nbeats

Les prévisions obtenues avec ce modèle sont meilleurs par rapports aux modèles classiques hormis Xgboost avec une erreur RMSE de 8.50%.

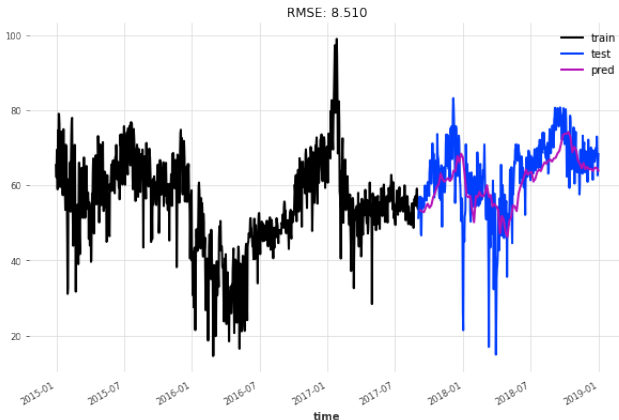


Figure: Predictions by Nbeats model

Tft (Temporal Fusion Transformer), une nouvelle architecture basée sur l'attention qui combine une prévision multi-horizon de haute performance avec des informations interprétables sur la dynamique temporelle en utilisant:

- IL utilise des couches récurrentes pour le traitement local et des couches d'auto-attention interprétables pour apprendre les dépendances à long terme.
- Le Tft utilise également des composants spécialisés pour la sélection judicieuse des fonctionnalités pertinentes et une série de couches de déclenchement.

Resultats Tft

Les résultats avec Tft sont encore mieux avec une erreur de prévision RMSE de 7.12% hormis Xgboost. On voit qu'aussi Tft est un modèle qui controle bien les prévisions dans le temps avec des intervalles de confiance.

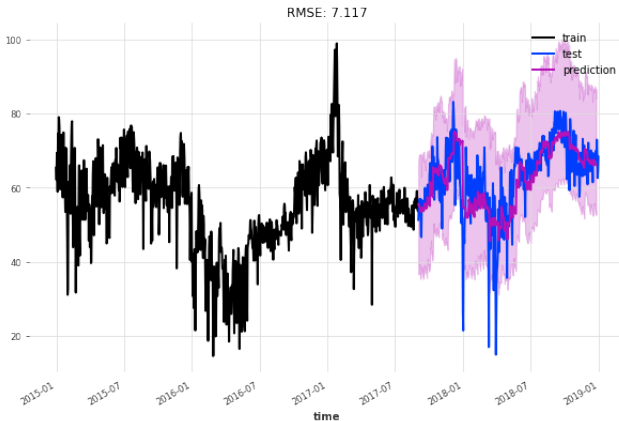


Figure: Predictions by Tft model

DeepAR est un réseau neuronal récurrent (RNN) de type réseau neuronal bien adapté aux séries chronologiques. DeepAR est entièrement récurrent dans le sens où, au moment de la prédiction, une sortie est calculée à l'aide de ces entrées :

- valeur cible précédente, qui sera définie sur la dernière valeur cible connue pour la première prédiction, et pour toutes les autres prédictions de manière auto-régressive
- l'état caché précédent,
- les covariables au temps t pour prévoir la cible au temps t

Resultats DeepAR

Ce modèle fournit de très performances. Les prévisions sont contrôlées dans le temps avec des intervalles de confiances et que l'erreur RMSE obtenue avec ce modèle est 6.15%. C'est donc un intéressant modèle.

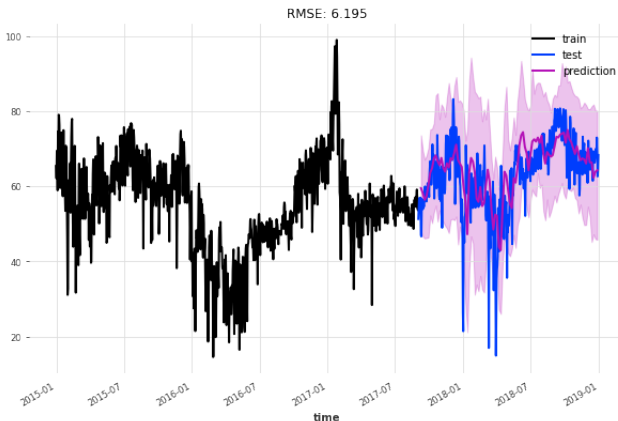


Figure: Predictions by DeepAR model

Comparaison de Modèles

Les resultats de prévisions sur les modèles récents fournissent de meilleures performances par rapport aux modèles classiques. Et que DeepAR le modèle conçu sur les réseaux récurrents offre le meilleur résultat parmi tous les modèles étudiés.

	Modèle	RMSE	Training_time
0	FPROPHET	30.507053	0.262511
1	XGBoost	6.416534	0.541170
2	TBATS	13.030874	42.681676
4	NBeats	8.510379	21.736298
5	TFT	7.116940	23.739106
7	DeepAR	6.195211	15.736126

Figure: Resultats metrics and times training



Comparaison de Modèles

DeepAR est meilleur que Xgboost mais très couteux en temps.

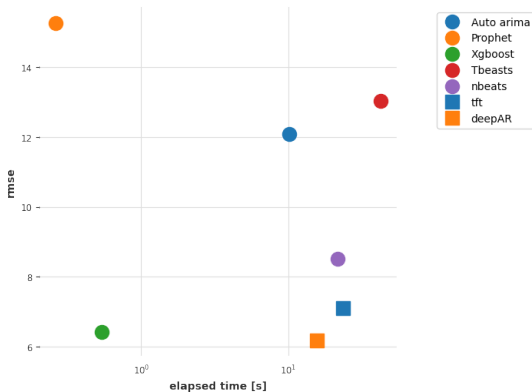


Figure: Models with rmse/training time

Conclusion

Les prévisions obtenues avec les méthodes récentes sont meilleures que celles des méthodes classiques. Cependant ces derniers sont moins coûteuses et plus faciles à implémenter contrairement aux modèles récents qui sont très coûteux et difficile à implémenter.

Maintenant, quelle méthode convient le mieux à votre cas ? Comme toujours, cela dépend:

- les méthodes classiques, si vous avez affaire principalement à des séries isolées qui ont un historique suffisant.
- si vous avez affaire à un plus grand nombre de séries ou à des séries de dimensionnalités plus élevées, les méthodes ML et les modèles globaux seront souvent la voie à suivre