
Magasins Rossman

— Projet OpenClassrooms —

Par Xavier Montamat

Problématique

Prévoir avec précision le chiffre d'affaire de chaque magasin Rossmann.

Ce à partir de l'historique des ventes de chaque magasin, et divers informations additionnelles.

La prévision doit être faite de manière journalière sur une période de 2 mois.

Axes d'approche envisagés

Approche en réseau de neurones

- Peu de transformation de données
- Réseau de neurones existant
- Configurer les couches

Approche machine learning

- Algorithmes plus classiques de régression
- Utiliser des techniques de pointe
- Optimiser les paramètres

Plan de réalisation

Analyse et traitements

- Chiffre d'affaire
- Analyse temporelle
- Intemporelle
- Transformations

Modélisation

- Prophet
- Modèle XGboost

Résultats

- Résultats
- Autres approches
- Conclusion
- Améliorations

Analyse et traitements

Réalisons une analyse des données afin d'avoir une bonne vision du problème, et des informations disponibles.

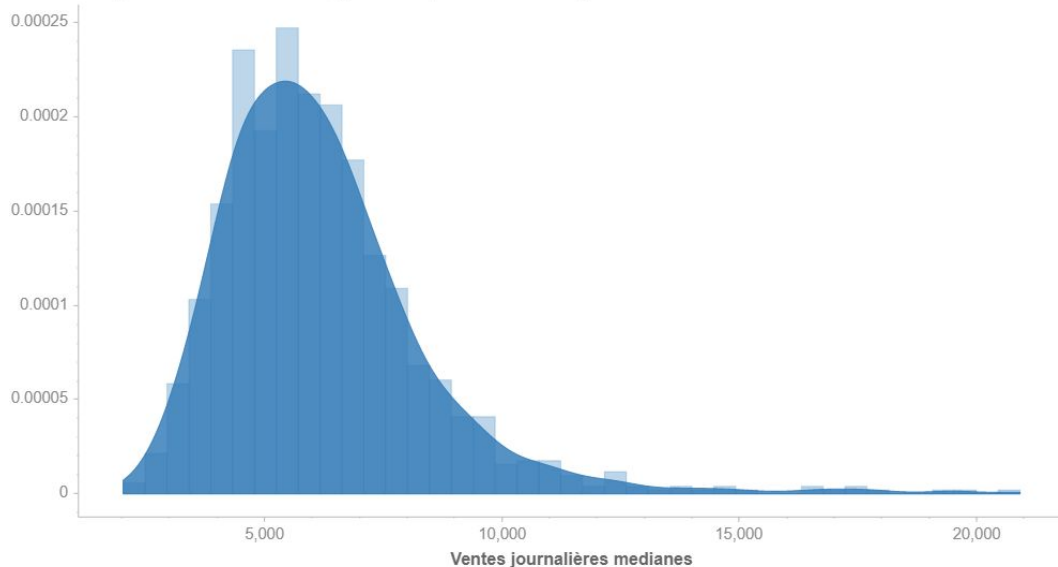
Puis exposons les transformations effectuées sur nos features

Chiffre d'affaire

La valeur cible du sujet

- 1115 Magasins Rossmann

Répartition des magasins par ventes journalières



Observations

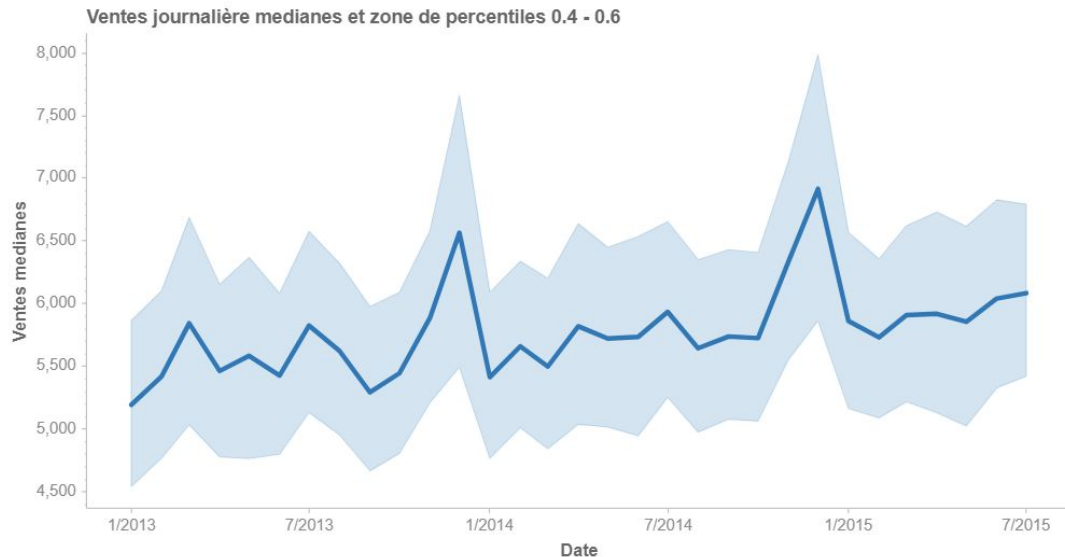
- Distribution Gaussienne des magasins
- Mediane 6000€
- Minimale 2000€
- Maximale 21000€

Variations importantes

Analyse temporelle

Tendance globale

Tendance des ventes sur 2.5 ans



- Historique 2.5 ans
- Prédictions sur 48 jours

Observations

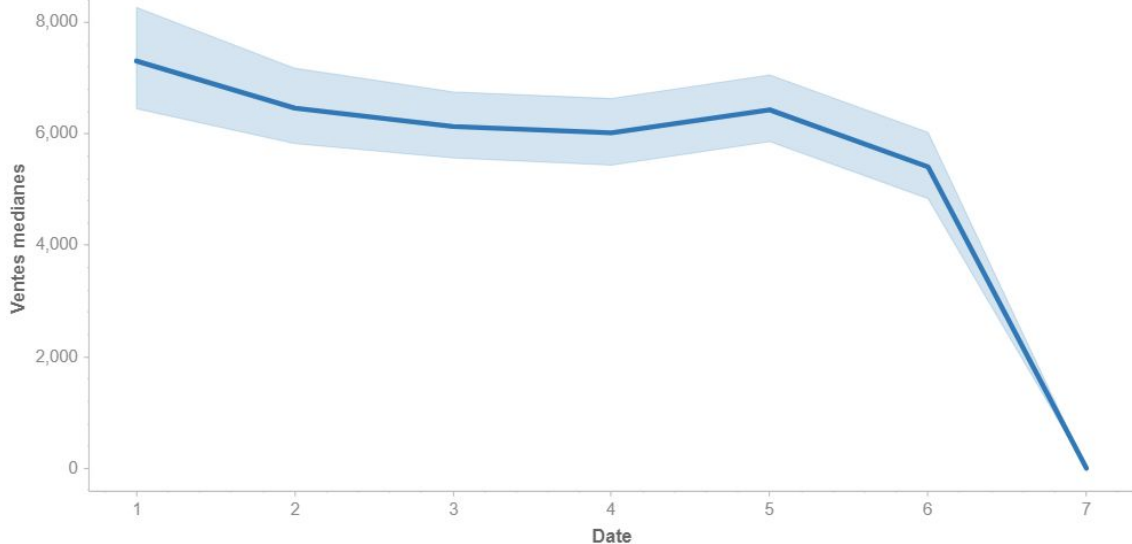
- Légère croissance globale
- Percentiles 0.4-0.6 éloignées
- Pics de ventes en fin d'année

Analyse temporelle

Tendance hebdomadaire

Tendance des ventes par jour de la semaine

Ventes journalières médianes



Observations

Différence de 20% entre lundi et samedi

Fermeture le dimanche

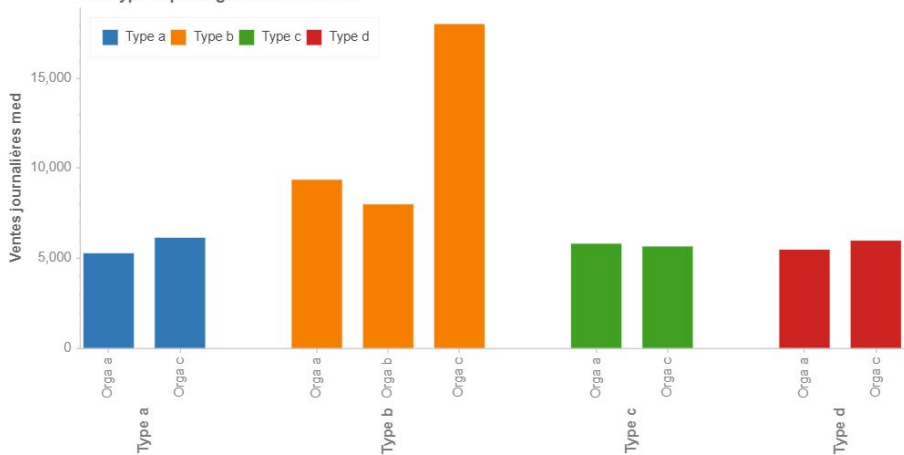
Analyse données intemporelles

- Catégorie de magasin (A-D)
- Organisation interne (A-C)

Chiffre d'affaire

Ventes journalière medianes des magasins

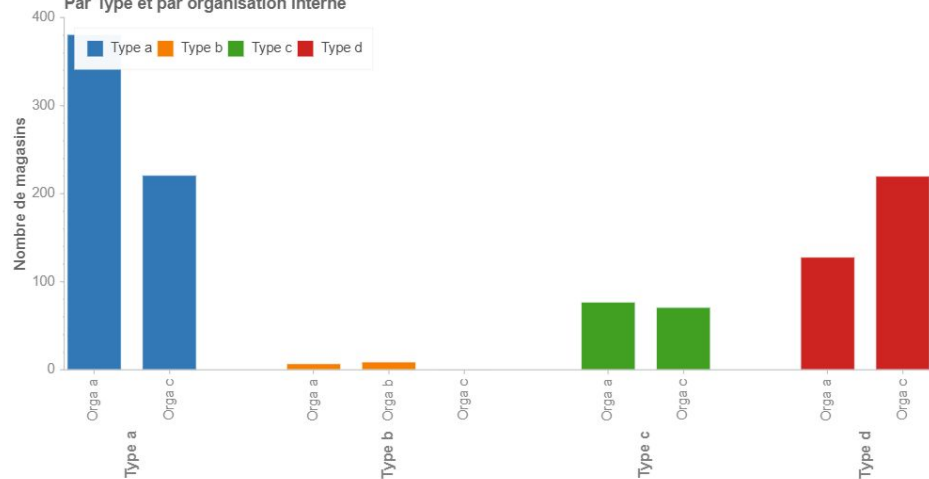
Par Type et par organisation interne



Distribution

Répartition des magasins

Par Type et par organisation interne



Transformations effectuées

- Competition
- Promotions
- Jours fériés
- Vacances scolaires



Dates



Nb jours avant / après

Ventes journée (instant T)



- Ventes médianes trimestre
- Ventes médianes semestre
- Ventes médianes semaine année précédente

Modélisation

Pour notre modèle, nous
montrons d'abord un modèle
'state of the art' nommé Prophet

Puis nous le comparerons à une
approche plus classique en
utilisant XGBoost

Prophet

Qu'est ce que Prophet ?

- Librairie de prédiction temporelle
- Développée par Facebook
- Apparue en 2017

Avantages

- Simple et intuitif
- Performante
- Rapide

Fonctionnement

- Détection de tendances
- Saisonnalité
- Vacances et jours fériés

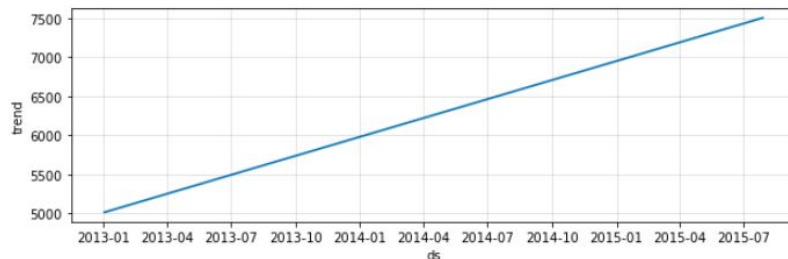
Limitations

- 1 seule entrée par date
- Pas de paramètres intemporels
- Peu de personnalisation

Analyse par Prophet

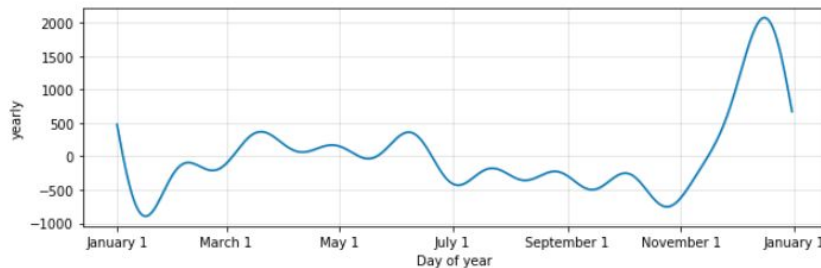
Tendance Globale

- Croissance linéaire



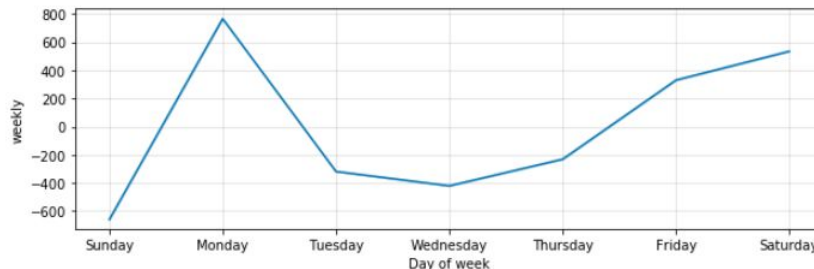
Tendance Annuelle

- Bon chiffres en fin d'année
- Baisse en février



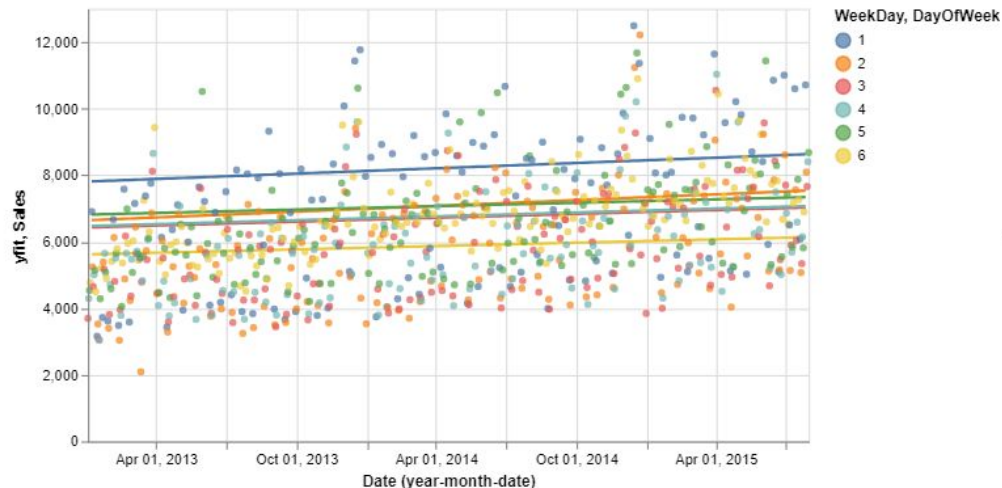
Tendance Hebdomadaire

- Creux en milieu de semaine



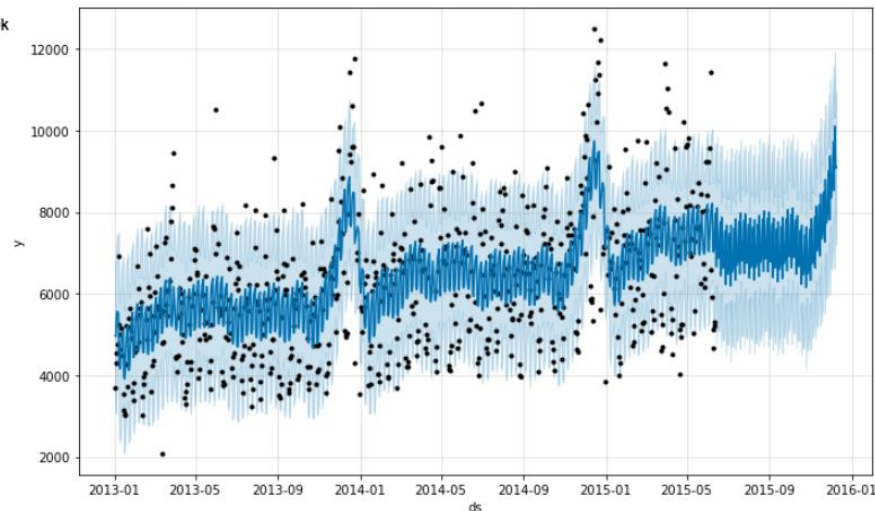
Prophet Modélisation

Ventes d'un magasin



- Variations hebdomadaires importantes

Prévisions Prophet



- Semblent réalistes
- Pics de fin d'année
- Interval d'erreur

Résultats

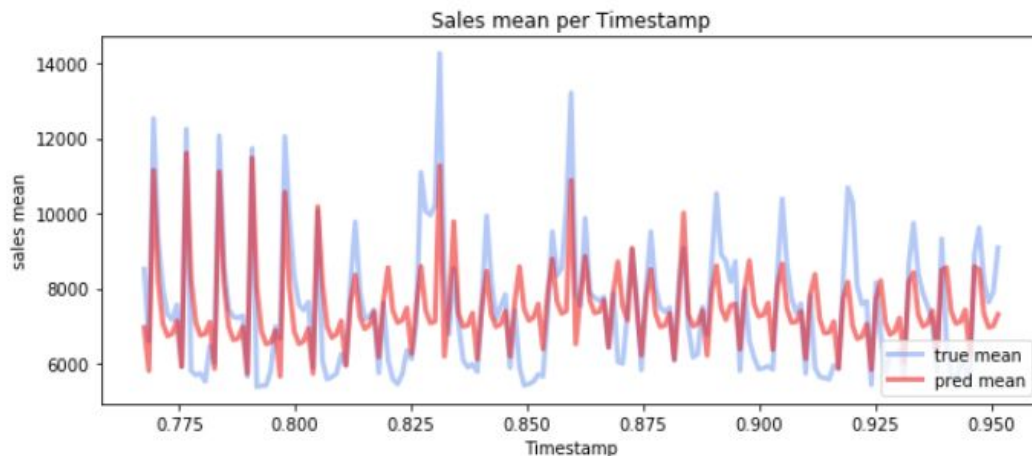
Métrique de score : RMSPE

Root Mean Square Percent Error

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2},$$

Résultats Prophet sur l'ensemble des magasins

RMSPE : **0.22**



XGBoost

Qu'est ce que XGBoost?

- eXtreme Gradient Boosting
- Modèle de classification / regression
- Evolution de Random forests
- Apparue en 2014

Avantages

- Versatile
- Très réputé
- Plus rapide que random forests

Fonctionnement

- Arbres séquentiels
- Gradient Descent

Limitations

- Potentiel overfitting
- Hyperparamètres
- Pensé pour la classification

XGBoost paramètres

subsample

- Ignorer entrées au hasard

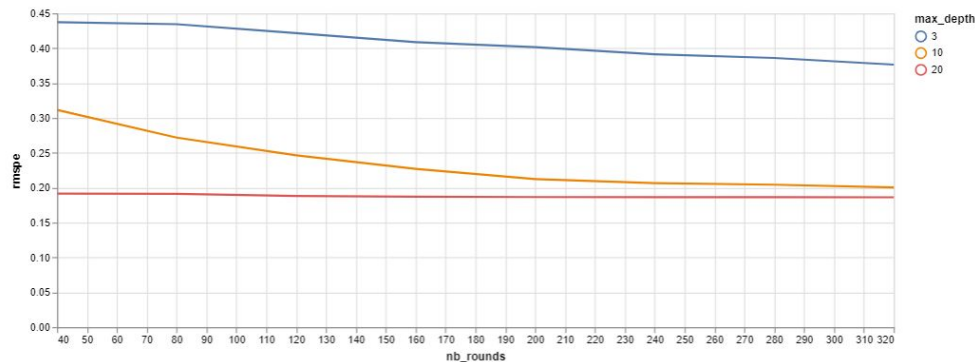
max_depth

- Limiter la profondeur

eta/learning_rate

- Equilibrer rapidité et précision

Impact de la profondeur maximum



XGBoost Résultats

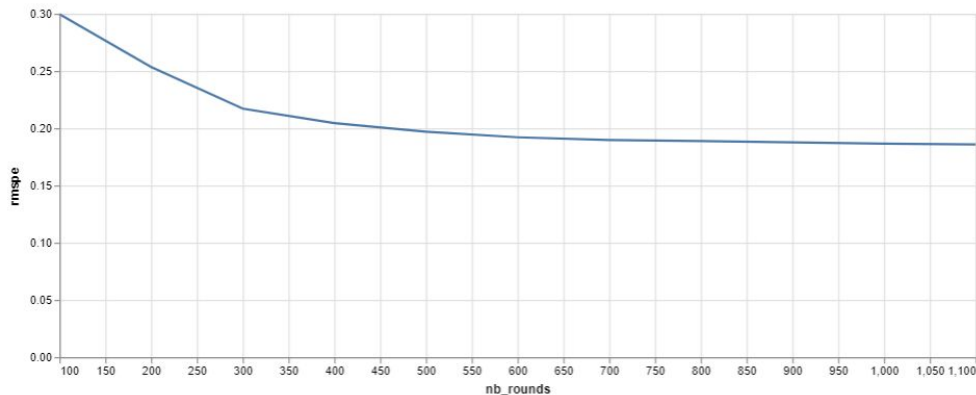
Métrique de score : RMSPE

Root Mean Square Percent Error

Résultats XGBoost sur l'ensemble des magasins

RMSPE : **0.18**

Entrainement final



Résultats & Conclusion

Regardons dans le détails les résultats obtenus.

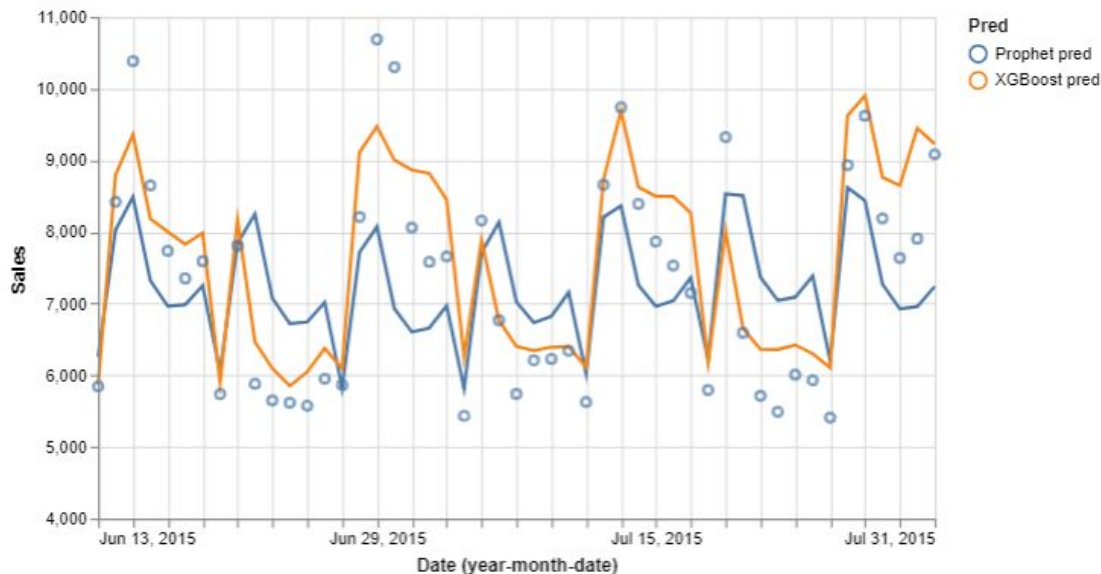
Puis une autre approche envisagée

Avant de conclure et exposer quelques améliorations possibles

Comparaison visuelle

Prédictions sur l'intervalle cible de 48 jours

XGBoost vs Prophet



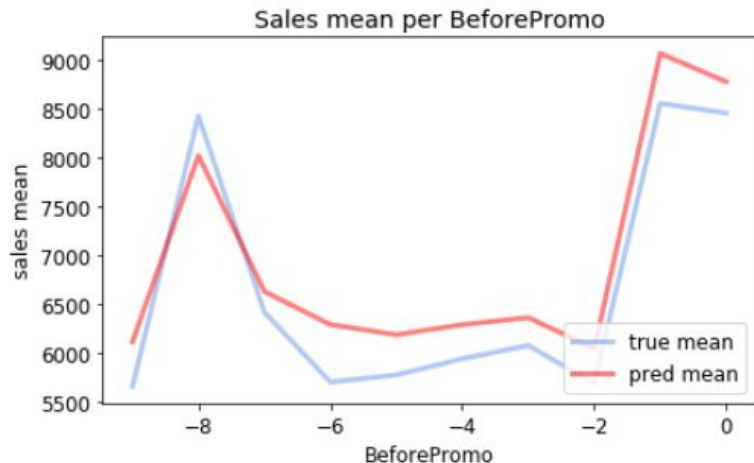
Observations

Prophet ne parvient pas à prédire la double tendance observée

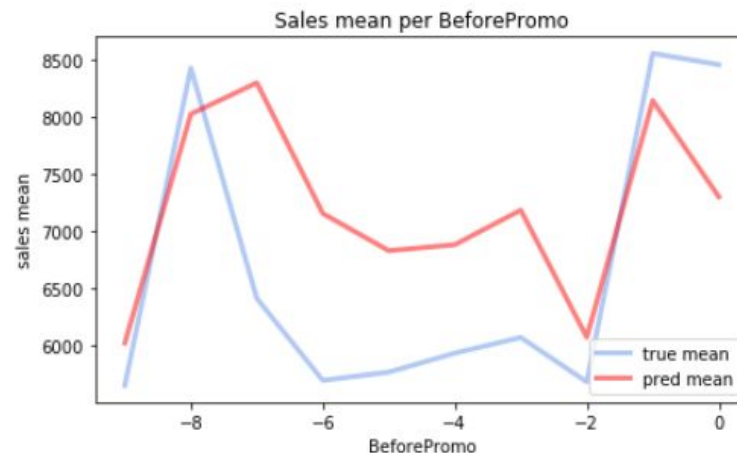
Comparaison visuelle

Comparaison sur la variable de promotion

XGBoost



Prophet



Une approche hybride

Combiner Prophet + XGBoost

Avantages

- Meilleur des deux mondes
- Erreur réduite (0.14)

Inconvénients

- Peu prévisible
- Effet Black Box
- Maintenance
- Temps d'entraînement

XGBoost classique plus

Conclusion

XGBoost toujours compétitif (Dans certaines conditions)

Prophet très simple d'utilisation mais cas particuliers peu complexes.

Améliorations possibles

Mieux diviser le modèle par magasin
Appliquer des poids XGBoost

Questions

