

ROUGIER Félix and COPPERE Sylvain

Data-Scientists

KEYRUS

Keyrus Lyon
Campus Verrazzano
1 place Giovanni da Verrazzano
69009 Lyon

felix.rougier@keyrus.com
sylvain.coppere@keyrus.com

Dégun sait ce qui va se passer... Allez, vaï, industrialisons R !

Mots clefs : Prédiction, Industrialisation, Méthode Ensembliste, R

Dans le cadre d'un projet pour un grand aéroport français, Keyrus a élaboré des modèles de prédiction avec R et les a mis en production dans l'écosystème informatique du client.

Contexte :

L'enjeu principal était de réaliser une modélisation permettant de prédire grâce aux données des ventes, le chiffre d'affaires ainsi que le nombre de tickets de parking consommés sur le mois en cours et le mois suivant, en tenant compte des réservations qui ont pu être faites en amont.

Chaque jour, les données de la veille sont intégrées dans les systèmes, l'aéroport souhaitait donc que les prédictions soient recalculées automatiquement de manière quotidienne.

Cela sera un premier pas vers du yield management et permettra de mieux piloter la performance commerciale.

Modélisation :

Nous avons utilisé une combinaison de modèles adaptés aux différents stades de maturité des parkings et aux temps restants avant la fin du mois :

- Parking agé : Modèle de séries temporelles avec apprentissage sur 2 ans
- Parking agé : Modèle combinant les ventes déjà réalisés à l'instant t avec des probabilités (probabilités d'achats pour les prochains jours et les probabilités de que ces achats soient pour un usage dans le mois)
- Parking jeune : Modèle basé sur les pourcentages des achats réalisés en fonction de l'avancement dans le mois et des données historiques

Industrialisation :

Pour automatiser les prédictions avec un recalcul quotidien, nous avons réalisé les tâches suivantes :

- Création du Data-Mart contenant les données à fournir au process R, dans le datawarehouse existant (Oracle) et des tables de stockage des résultats
- Installation de R et des packages nécessaires sur les serveurs du client
- Interfaçage de R avec Oracle
- Paramétrisation des scripts R pour les rendre génériques
- Création de tests de recette avec R pour vérifier la qualité des données d'entrée et la conformité des données de sortie
- Création de modules R de suivi de performances pour contrôler la dérive du modèle
- Inclusion d'éléments permettant de gérer un versionning des modèles et des prévisions pour anticiper les futures évolutions
- Automatisation du lancement batch du process de prédiction vis SAP BODS