

Segmentation des Clients

Synthèse

Contexte :

Un magasin possède de nombreuses informations sur les achats de leurs clients. A partir de celui-ci, Datazon nous demande de mettre en place une segmentation de la clientèle afin de déterminer au plus tôt quel comportement d'achat aura un client et quel seront ses intérêts.

Problème :

Ce problème est un problème de classification multiclassées. A l'aide d'un algorithme de Clustering, les différents clients vont être regroupés par comportement. Sur ce résultat de clustering, un classifieur sera mis en place pour prédire le groupe auquel appartiendra un client le plus tôt possible basé sur son historique d'achat.

Données :

Le dataset fourni regroupe les articles achetés avec la date d'achat, la quantité, le prix unitaire et le numéro de facture. Une des feature fournit aussi un identifiant du client utile à la segmentation par la suite.

Approche :

Après un nettoyage des données (suppression des faux articles et des achats du magasin), un 1^{er} algorithme de Clustering va être mis en place sur les articles afin de les grouper dans des clusters. Par la suite, chaque article sera converti en son cluster respectif. Pour chaque facture, le prix total du panier, son ancienneté et sa répartition par cluster sera calculée. Par la suite le dataset sera regroupé par clients avec le nombre d'achats, ancienneté du premier et du dernier achat. Sur ce dataset un second clustering sera mis en place. Pour grouper les comportements dans des clusters. Ces clusters sont ensuite utilisés pour entraîner le classifieur.

Performances des modèles :

Les algorithmes de clustering n'ont pas des scores de silhouette très haut mais malgré cela, les paniers sont assez distincts. Le résultat du clustering sur les clients est ensuite meilleurs avec des différences visibles. Celui-ci permet notamment de trouver des clients particuliers auquel il faut apporter une attention particulière (VIP) et aussi de visualiser les différences particulières entre les différents types de clients. Quant au Classifieur, il apporte de bonnes performances et permet de classifier avec très peu d'erreur les clients même avec un petit historique.

Résultats :

Les différents modèles ont été sauvegardés et une API a été mise en place afin de pouvoir prédire, basé sur des historiques d'achats, dans quel cluster sera un client. Quant aux type de Clusters, une étude est disponible dans le notebook qui permettent d'analyser le comportement lié à un cluster particulier.