

Sujet de stage « IMPACT » au Research LAB Vekia

Clustering d'articles par relation de cannibalisme ou d'effet « halo »

Présentation du sujet

Dans ce travail il s'agit d'étudier et de détecter les articles qui interagissent au niveau de leur comportement de vente.

Classiquement deux types d'interactions apparaissent :

- La cannibalisation : elle touche les articles qui sont en concurrence car ils répondent à un même besoin, par exemple deux paquets de spaghettis. Ainsi un article en promotion va très probablement faire baisser temporairement les ventes des articles rendant un service similaire.
- L'effet halo : cette interaction apparait entre articles répondant à un besoin complémentaire, par exemple un paquet de spaghettis et un pot de sauce bolognaise. Ces articles vont probablement être achetés en même temps, et une promotion de l'un d'eux provoquera probablement une montée des ventes de l'autre.

Une partie du métier de Vekia est la prévision des ventes. Dans ce contexte la prise en compte fine des effets d'interaction entre articles peut amener un gain de qualité.

Un préalable quasi incontournable à cette prise en compte dans des modèles dédiés est la formation de groupes d'articles sujets à la cannibalisation ou à l'effet halo. La nomenclature naturelle des articles, telle que fournie par nos clients, n'est pas assez fine et sans rapport direct avec le problème posé ici, d'où le besoin de créer automatiquement les clusters pertinents.

Outre leur utilisation au sein des moteurs de prévision, ces groupes peuvent également avoir un intérêt propre par l'éclairage qu'ils apportent sur les mécanismes de vente.

Travail et résultats attendus

L'étudiant(e) devra effectuer une étude bibliographique afin de dresser un panorama des techniques existantes ; il pourra également proposer ses propres idées. Un précédent rapport de stage abordant un sujet similaire pourra éventuellement servir de point de départ.

Idéalement le stagiaire mettra en œuvre un (ou plusieurs) prototype (sous technologie python/pandas/sklearn, ou R) sur des jeux de données réels, présentera et comparera les résultats obtenus.

Il s'agit d'un sujet plutôt orienté « recherche », où les initiatives personnelles et l'envie de défricher un sujet complexe seront déterminantes.