

Déroulement

Ce cas se divise en 3 parties.

- **Lecture du cas**

D'abord, ce document contient l'introduction au cas, la présentation de l'entreprise, une présentation des données, et les 2 questions auxquelles vous devrez répondre.

- **Lecture de la théorie**

Pour bien répondre à ces questions, il sera pertinent de consulter un ouvrage de référence. La 2e partie est donc la lecture de cet ouvrage, particulièrement les sections recommandées (présentées plus loin dans ce document).

- **Réponse individuelle et en petits groupes**

Avant la séance de cours, vous devrez soumettre votre réponse à la question 1 sur Zone-Cours. La réponse doit contenir (a) le ou les modèles que vous suggérez d'utiliser et (b) une justification étoffée de votre choix (entre 5 et 15 lignes). Lors de la séance de cours, vous serez d'abord placés en petits groupes pour discuter de vos solutions à la question 1. Ensemble, vous choisirez le ou les modèles les plus pertinents et préparerez une courte présentation sur les modèles sélectionnés. Certaines équipes choisies aléatoirement présenteront ensuite leurs solutions. Nous discuterons des solutions proposées et des solutions réellement mises en place par l'entreprise, puis de solutions possibles à la question 2.

Introduction

- Un système de recommandation!
- En es-tu certain?
- Ça aidera à personnaliser l'expérience utilisateur, et ça stimulera les ventes en ligne, renchérit Samuel, chef de l'équipe de science des données.
- Je te fais confiance! approuve Alexandre, chef des technologies chez Décathlon Canada.

Printemps 2018, Montréal. Samuel, chef de l'équipe de science des données chez Décathlon Canada, établit la liste de priorités du tout nouveau département sous sa gestion. Au-dessus de cette liste se trouve la personnalisation de l'expérience client, à travers l'utilisation de systèmes de recommandation.

Aux yeux de Samuel, cette priorité est un choix évident. De nombreuses études ont prouvé la pertinence de la recommandation, bénéfique à la fois pour l'augmentation des ventes et pour l'amélioration de l'expérience client.

Plusieurs domaines ont été profondément touchés par l'arrivée de systèmes de recommandation, notamment le commerce électronique et le divertissement numérique. «Already, 35 percent of what consumers purchase on Amazon and 75 percent of what they watch on Netflix come from product recommendations based on such algorithms. » [1]

Samuel cherche donc à mettre en place un outil de recommandation qui, sur les plateformes numériques de Décathlon, proposera une sélection personnalisée de produits intéressants à chaque utilisateur.

Comment s'y prendre?

Présentation de Décathlon

Décathlon, un géant de la vente d'articles sportifs

Avant tout, Décathlon est un magasin d'articles de sports généraliste. L'entreprise se veut accessible à tous et cherche à démocratiser la pratique de sports. Décathlon se démarque par son approche de bas prix tous les jours, rendue possible grâce à une grande intégration verticale. En effet, Décathlon oeuvre d'un extrême à l'autre de la chaîne de valeur: recherche et développement, conception, fabrication, logistique et vente au détail. L'offre de produits est gigantesque: une variété de plus de 65 sports, avec des articles pour néophytes et experts, pour un total d'environ 7500 produits offerts. Tous les produits vendus sont de marques *maison*, c'est-à-dire de marques privées possédées par Décathlon, comme Quechua pour la randonnée de montagne ou Kipsta pour les sports collectifs comme le soccer et le rugby.

Comme plusieurs géants du commerce, Décathlon a débuté dans la vente au détail, mais possède maintenant un grand écosystème numérique. Les consommateurs peuvent acheter des produits directement sur le site web ou l'application mobile de Décathlon, et peuvent rester en contact avec l'entreprise à travers son infolettre et ses plateformes de valorisation du sport.

Au total, Décathlon, c'est plus de 30 millions de transactions par mois, dont près de 100 000 au Canada [2]. Fondée en France en 1976, l'entreprise a depuis pris une grande expansion internationale : on compte maintenant près de 1700 points de vente dans près de 60 pays, qui servent plus de 110 millions de membres. Décathlon est une société privée détenue notamment par la famille du fondateur et par des employés, l'entreprise permettant aux salariés de devenir actionnaires. Dans son marché d'origine, la France, Décathlon domine : on y retrouve plus de 320 magasins, et l'entreprise fait partie intégrante de la culture. Décathlon s'est récemment implantée au Canada, et y compte déjà près de 300 000 membres et bientôt 8 magasins.

Des ressources informationnelles bien en place

Au fil des années, Décathlon a su développer une culture de la donnée. Plusieurs équipes sont en place pour permettre la collecte et le stockage de données de qualité. Cela permet à l'équipe de science des données – une équipe composée au Canada d'une dizaine d'experts – d'exploiter les données sans trop d'embûches logistiques. Le modèle d'affaires vertical de Décathlon rend l'entreprise apte à avoir une vue d'ensemble très complète de son offre : les données sont récoltées et utilisées à travers la chaîne de valeur, sans devoir composer avec des systèmes informatiques de fournisseurs tiers.

Bonifier l'offre grâce aux données : le système de recommandation

Avec toutes ces données prêtes à l'utilisation, Décathlon cherche maintenant de nouvelles façons d'améliorer l'expérience offerte aux clients. Une de ces améliorations passe par la personnalisation de l'expérience en ligne, notamment grâce aux systèmes de recommandation. Ces systèmes cherchent à détecter et à recommander des produits d'intérêt pour chaque consommateur. Cette recommandation agrémenterait l'expérience des internautes en leur permettant de trouver rapidement des produits intéressants parmi tout le catalogue, tout en stimulant les ventes. Les systèmes de recommandation ont d'ailleurs fait leurs preuves dans le monde numérique et y sont omniprésents.

Un des avantages de tels systèmes est de réduire le temps que les consommateurs passent à chercher et à considérer des produits. Les décisions d'achat de produits sportifs ne sont pas triviales et sont parfois difficiles à évaluer d'avance : la facture moyenne avoisine les 100\$ et les biens vendus sont souvent techniques. Comprendre quel produit parmi le catalogue entier convient à ses besoins peut s'avérer fastidieux; les systèmes de recommandation offrent aux internautes une présélection de produits considérés comme pertinents pour eux. Ces systèmes décuplent également les possibilités de vente croisée en proposant aux clients des produits complémentaires, à la manière d'Amazon et d'autres géants numériques avec une section « les clients qui ont acheté ce produit ont aussi acheté... », qui permet d'augmenter le nombre de produits vendus par client.

Décathlon envisage de présenter des recommandations dans quelques contextes différents : sur la page d'accueil du site web, sur la page d'accueil de l'application mobile et dans les courriels envoyés aux membres, notamment. Des exemples de recommandations sont disponibles en annexe A.

Les données à disposition : diverses et à grande échelle

Pour alimenter le système de recommandation, Décathlon possède une grande quantité de données. L'interaction entre chaque utilisateur et chaque produit est enregistrée (incluant l'identité de l'utilisateur). La base de données contient également des attributs sur les membres Décathlon, comme leurs sports pratiqués, et sur les produits, comme leurs caractéristiques physiques et les images et textes liés aux produits. Un exemple des jeux de données à disposition est disponible en annexe B.

La quantité d'utilisateurs des plateformes Décathlon – près de 300 000 visiteurs uniques chaque mois [2], couplés à la diversité des produits offerts – environ 7500 produits uniques, soulève un défi technique. Pour effectuer des recommandations pertinentes et personnalisées à un si grand nombre d'utilisateurs, énormément de données sont nécessaires. Cela peut donc compliquer et ralentir l'entraînement et la vitesse de prédiction de certains modèles, alors que ceux-ci devront être utilisables en temps réel dans l'écosystème numérique de l'entreprise.

Théorie sur les systèmes de recommandation

Pour en apprendre davantage sur les systèmes de recommandation, la référence suivante est suggérée :

Aggarwal, Charu C. *Recommender Systems: the Textbook*. Cham: Springer, 2016.

Disponible en version numérique à la bibliothèque de HEC Montréal à travers le lien suivant :

<https://hecmontreal.on.worldcat.org/oclc/946011635>

Suggestion de sections à lire

Chapter 1. Introduction

- 1.1 Introduction
- 1.2 Goals of Recommender Systems
- 1.3 Basic Models of Recommender Systems
 - 1.3.1 Collaborative Filtering Models (sauf 1.3.1.2, 1.3.1.3)
 - 1.3.2 Content-Based Recommender Systems
 - 1.3.6 Evaluation of Recommender Systems

Chapter 2. Neighborhood-Based Collaborative Filtering

- 2.1 Introduction
- 2.3.1 User-Based Neighborhood Models (sauf 2.3.1.1, 2.3.1.2, 2.3.1.3, 2.3.1.4)
- 2.3.2 Item-Based Neighborhood Models
- 2.3.4 Comparing User-Based and Item-Based Methods
- 2.3.5 Strengths and Weaknesses of Neighborhood-Based Methods

Chapter 3. Model-Based Collaborative Filtering

- 3.1 Introduction
- 3.6 Latent Factor Models (seulement 3.6.2, 3.6.3)

Chapter 4. Content-Based Recommender Systems

- 4.1 Introduction
- 4.2 Basic Components of Content-Based Systems
- 4.3 Preprocessing and Feature Extraction (sauf 4.3.4)
- 4.4 Learning User Profiles and Filtering (seulement 4.4.1, 4.4.5)
- 4.5 Content-Based Versus Collaborative Recommendations

Questions

Samuel explore les options qui s'offrent à son équipe. Tout est à faire : il souhaite donc entreprendre la tâche avec des systèmes moins complexes, pour ensuite tester les modèles les plus performants de l'industrie. Le système devra d'abord recommander des produits aux

membres actuels sur la page d'accueil. Éventuellement, Samuel aimerait étendre autant que possible les recommandations aux nouveaux membres.

Question 1

Pour le prototype, il n'y a pas de contraintes sur le temps de calcul pour entraîner le modèle ou pour servir les recommandations. En fonction des données disponibles et des objectifs de Décathlon, quel(s) modèle(s) de système de recommandation l'équipe de science des données devrait-elle choisir, et pourquoi?

Question 2

Une fois un premier modèle mis en place, quelles améliorations pourraient être proposées, ou quels modèles plus avancés devraient être testés en priorité?

Rappel : Avant la séance de cours, vous devrez soumettre votre réponse à la question 1 sur Zone-Cours. La réponse doit contenir (a) le ou les modèles que vous suggérez d'utiliser et (b) une justification étoffée de votre choix (entre 5 et 15 lignes). Lors de la séance de cours, vous serez d'abord placés en petits groupes pour discuter de vos solutions à la question 1. Ensemble, vous choisirez le ou les modèles les plus pertinents et préparerez une courte présentation sur les modèles sélectionnés. Certaines équipes choisies aléatoirement présenteront ensuite leurs solutions. Nous discuterons des solutions proposées et des solutions réellement mises en place par l'entreprise, puis de solutions possibles à la question 2.

Notes et références

1. MacKenzie, I., Meyer, C., and Noble, S. (2013, October 1). How Retailers Can Keep up with Consumers | McKinsey. Retrieved from <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>
2. Le nombre de visiteurs et d'achats a beaucoup fluctué depuis janvier 2020, étant donné la crise sanitaire mondiale. Les statistiques présentées donnent un ordre de grandeur, mais pourraient varier grandement d'un mois à l'autre.

Annexe A : Exemples de recommandation

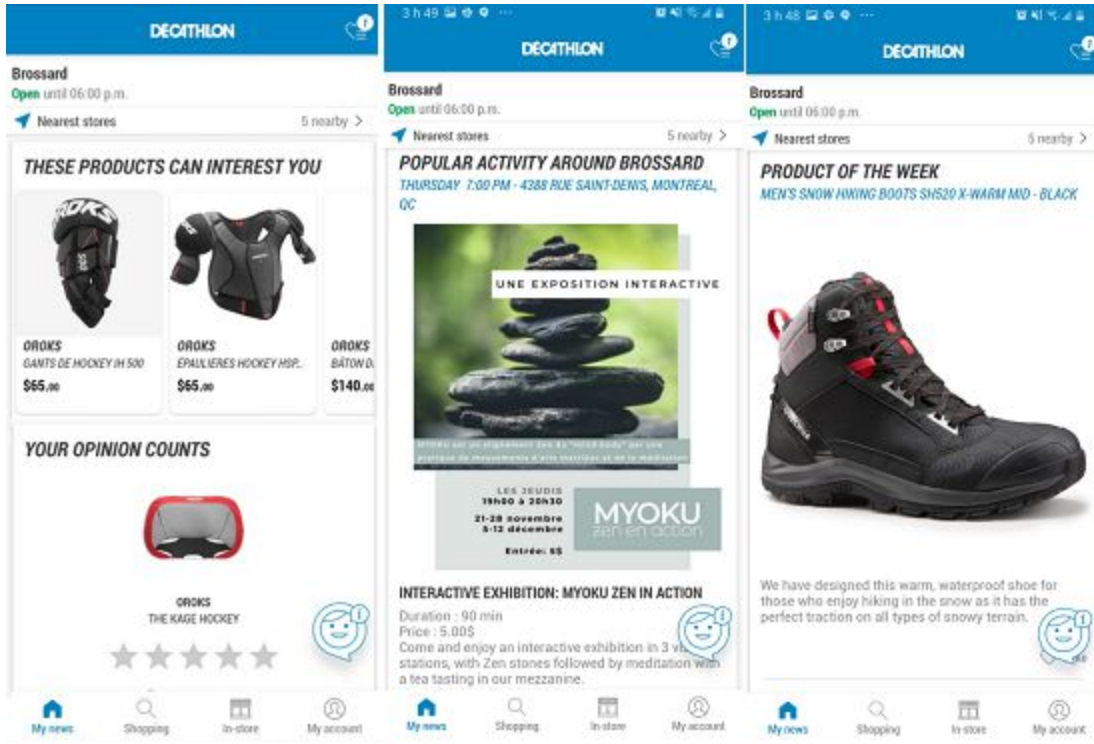
1. Site web

Exemple de page d'accueil du site web Décathlon, personnalisée pour un ou une membre intéressé(e) par l'haltérophilie.



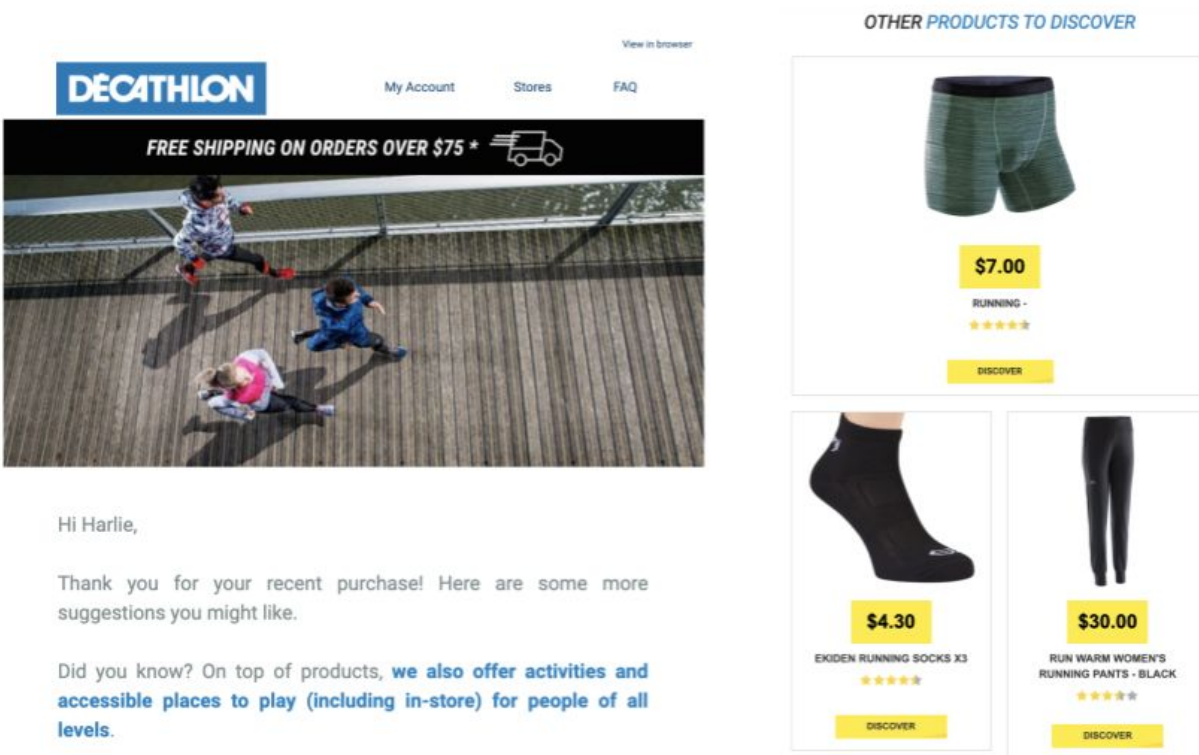
2. Application mobile

Exemple de page d'accueil de l'application mobile Décathlon, personnalisée pour un ou une membre intéressé(e) par le hockey.



3. Courriel post-achat

Exemple de courriel post-achat, personnalisé pour un ou une membre intéressé(e) par la course.



Hi Harlie,

Thank you for your recent purchase! Here are some more suggestions you might like.

Did you know? On top of products, we also offer activities and accessible places to play (including in-store) for people of all levels.

Annexe B : Jeux de données

Note : Les données ont été modifiées pour préserver la confidentialité de l'entreprise et des membres.

Tableau 1 - Interactions between customers and items

customer_id	item_id	timestamp	click	purchase
imbvblxwvtiywunh	3384934262863770	2018-01-01	0	1
nzhrkquelkgflone	8321263216904593	2018-01-01	1	0
rgqgpqysetgoxdmz	7400694405709236	2018-01-01	0	1
xbnkvygodmlyfete	1769704982990886	2018-01-02	0	1
eqeefiaxvjxsyfxk	5460080147661416	2018-01-02	1	0
...
mkfqblfqwgzivfga	8625784510427076	2019-12-30	1	0
rzwykqujaznzmth	4673617174927844	2019-12-30	1	0
evbdnuhdfcpiyotj	9384258747220968	2019-12-31	1	0
cgatomzvjiizvctb	2756920171861146	2019-12-31	1	0
cnspkotxubxnxtzk	5150255386059428	2019-12-31	0	1

Informations supplémentaires :

- Un `customer_id` est propre à chaque membre Décathlon. Les données sont uniquement pour des interactions faites par des membres.
- Un `item_id` est propre à chaque produit Décathlon.
- Un membre peut consulter la page d'un produit (`click`) ou l'acheter (`purchase`). Un membre peut avoir consulté la page avant d'acheter (à travers un achat en ligne, par exemple), ou pas (à travers un achat en magasin, par exemple).
- En moyenne, il y a environ 600 000 clics par mois, effectués par 31 000 membres différents. *Note: Ces statistiques varient beaucoup d'un mois à l'autre.*
- En moyenne, il y a environ 80 000 achats par mois, effectués par 20 000 membres différents. *Note: Ces statistiques varient beaucoup d'un mois à l'autre.*

Tableau 2 - User features

customer_id	is_male	is_female	declared_sports
imbvblxwvtiywunh	0	1	baseball;boxing
nzhrkquelkgflone	1	0	
rgqgpqysetgoxdmz	1	0	snowboard:golf;running
xbnkvygodmlyfete	0	0	
eqeefiaxvjsyfxk	0	1	ski;rugby
...
mkfqblfqwgzivfga	1	0	soccer
rzwykqujaznzmth	1	0	
evbdnuhdfcpiyotj	1	0	
cgatomzvjiizvctb	0	1	
cnspktxubxnxtzk	0	1	capoeira;beachtennis

Informations supplémentaires :

- Chaque membre est lié à une entrée dans la base de données `user features`. Cependant, un membre peut ne pas avoir déclaré son sexe ni ses sports préférés.
- Un membre peut avoir plusieurs `declared_sports` différents.
- Environ 20% des membres ont déclaré leurs sports préférés.
- Une très grande majorité des membres ont déclaré leur genre. *Note : Décathlon retirera progressivement les questions démographiques (âge et genre) de ses plateformes virtuelles. Ces données ne seront donc plus disponibles.*

Tableau 3 - Item features

item_id	is_male	is_female	is_junior	label	family	subdepartment	department	universe	linked_sports	image
3384934262863770	0	1	0	JACKET 920 SPACER HOOD W GREYNEPS	WOMAN PANT JACKET SWEAT	WOMAN APPAREL	PILATES, SOFT GYM	PILATES, SOFT GYM	muscultation	3384934262863770.jpg
8321263216904593	0	0	0	GLOVE RAIN M PAIR	GOLF BALLS, GLOVES, TEES	GOLF EQUIPMENT ALL LEVELS	GOLF	GOLF	golf	8321263216904593.jpg
7400694405709236	1	0	0	SUPPORT M SHOES RED2	REGULAR MEN JOGGING SHOES	JOGGING SHOES	JOGGING	JOGGING	running	7400694405709236.jpg
1769704982990886	0	0	0	TARO #1 HOLOGRAPHIC	SPOONS TROUT PERCHE	TROUT PERCH LURE FISHING	PREDATOR FISHING	FISHING	fishing	1769704982990886.jpg
5460080147661416	0	1	0	SH100 X-WARM W TROUSERS GREY	WOMEN SNOW HIKING WARM PANTS & FLEECEES	WOMEN SNOW HIKING APPAREL, SHOES	SNOW HIKING	SNOW HIKING	hiking	5460080147661416.jpg
...
8625784510427076	0	0	0	BAR 0,35M	FREE WEIGHTS AND EQUIPMENT	BODYBUILDING	BODYBUILDING	CROSS-TRAINING, BODYBUILDING	muscultation	8625784510427076.jpg
4673617174927844	0	0	1	GLOVE FH100 JUNIOR+ADULT BLACK	PROTECTION FIELDHOCKEY	FIELD HOCKEY	FIELD HOCKEY	FIELD HOCKEY	fieldhockey	4673617174927844.jpg
9384258747220968	1	0	0	SHORT B300 AD TARMAK BLACK	MAN BASKETBALL OUTFIT	MAN BASKETBALL OUTFIT	BASKETBALL	BASKET	basketball	9384258747220968.jpg
2756920171861146	0	0	0	TORCH LIGHT DYNAMO 50 WHITE COLLECTION	TORCH LAMPS HIKING/TREK	TREK BACKPACKS, POLES, HEADLAMPS, ACC	TREKKING SEVERAL DAYS	TREKKING SEVERAL DAYSniv	hiking	2756920171861146.jpg
5150255386059428	0	0	0	ARTENGO TR 560 LITE	ADULT BEGINNER/INTERMEDIATE RACKETS	TENNIS RACKETS/BALLS/ACCESSORIES	TENNIS	TENNIS	tennis	5150255386059428.jpg

Informations supplémentaires :

- Chaque produit est lié à une entrée dans la base de données `item features`.
- Un label appartient à une `family`, qui appartient à un `subdepartment`, qui appartient à un `department`, qui appartient à un `universe`. Cependant, plusieurs sous-catégories sont identiques. Par exemple, le `department` et l'`universe` de l'`item ARTENGO TR 560 LITE` sont le `TENNIS`.
- Chaque produit est lié à un ou plusieurs sports.
- Une image est disponible pour chaque produit.
- Un produit peut être spécialement pour hommes (`is_male = 1`), pour femmes (`is_female = 1`) ou pour enfants (`is_junior = 1`).