

Stages Data Science et Data Analyse

Embarquez à bord de la direction Data

Et venez partager votre passion avec une équipe pluridisciplinaire de plus de 30 data scientists et analysts en travaillant sur de nouveaux gisements de valorisation de la donnée.

Les missions qui vous seront confiées

Rattachées au service Marketing/Commerce, les équipes Data contribuent à optimiser les performances d'un nombre grandissant d'équipes au sein de l'entreprise. Historiquement focalisés autour de la pertinence du moteur de recherche interne, les data scientists et data analysts abordent désormais des sujets tels que les systèmes de recommandation, l'acquisition de trafic ou encore le pricing dynamique. Pour cela, un large spectre de méthodes mathématiques sont appliquées aux centaines de téraoctets de données que l'infrastructure big data met à disposition, en forte interaction avec les services métiers qui pilotent l'activité et définissent les priorités business.

Vous travaillerez sur l'un des sujets ci-dessous, colorés Data Analyse (DA) ou Data Science (DS).

- Analyse des facteurs d'abandon du programme de fidélité (Cdiscount à Volonté) pour améliorer la rétention (DA) ;
- Application des algorithmes du traitement du langage naturel à l'état de l'art pour améliorer l'expérience utilisateur (DS) ;
- Optimisation du score du paiement en 4 fois par une segmentation à froid et l'ajout de variables liées aux parcours des clients (DS) ;
- Amélioration du remplissage des fiches produits grâce au machine learning (DS) ;
- Détection des zones à risque dans le code par apprentissage sur les bugs passés (DS) ;
- Construction de groupes de produits qui se cannibalisent en cas de mouvements de prix (DS) ;
- Construction d'une ontologie à partir des requêtes utilisateurs pour améliorer la pertinence du moteur de recherche (DS) ;
- Clustering des clients en utilisant les historiques de navigation pour mieux personnaliser (DS) ;
- Prédiction de la date de livraison, notamment pour les produits de notre marketplace (DS/DA) ;
- Entraînement et mise à l'épreuve un modèle d'embedding de requêtes utilisateurs spécifiques e-commerce via des techniques modernes de NLP (DS).

Vous trouverez ci-dessous une description plus détaillée des sujets proposés.

Rétention au programme de fidélité (Data Analyse)

Ce stage est proposé par l'équipe data-client, responsable de l'analyse et de l'algorithmie autour des sujets clients : satisfaction client, frais de port, programme de fidélité... Cdiscount à volonté est notre programme de fidélité, il regroupe plusieurs avantages de taille pour nos clients abonnés : la livraison gratuite, un accès à la presse ainsi que des promotions dédiées.

Les clients Cdiscount à volonté sont parmi nos clients les plus fidèles. Cependant, comme l'ensemble des programmes de fidélité, un nombre conséquent d'abonnés "churnent" (c'est-à-dire, ne se réabonnent pas à l'échéance du programme).

L'objectif de ce stage de mieux comprendre nos abonnés afin de guider les actions qui ont pour but d'augmenter leur satisfaction, et *in fine* augmenter la rétention au programme de fidélité. Pour cela l'ensemble des données de navigation mais également de nombreux retours explicites et critères socio-démo seront mis à disposition.

Ce stage sera réalisé en lien très étroit avec les équipes métiers, ce qui permettra de déclencher et tester rapidement les actions identifiées.

Améliorer l'expérience utilisateur par le traitement du langage naturel (Data Science)

Après la révolution de la "Computer Vision", le traitement naturel du langage ou "Natural Language Processing" (NLP) s'annonce comme le lieu de la prochaine évolution majeure en termes d'intelligence artificielle. Cette révolution est d'ores et déjà en marche avec les récentes publications d'algorithmes comme Transformer (2017) puis BERT (2018) qui voient tomber les uns après les autres les records des principaux challenges de NLP.

L'étape suivante est de mettre en œuvre ces nouveaux outils sur des cas d'usage concrets porteurs d'une réelle valeur ajoutée pour les utilisateurs. Cdiscount avec ses millions de visiteurs uniques chaque jour et ses nombreux échanges avec ses clients, via le chat en ligne ou par email, offre un terrain unique pour tester ces applications et révolutionner les usages.

Ainsi, le stage proposé au sein de l'équipe innovation de la direction data a pour objectif de s'approprier les outils du NLP à l'état de l'art pour explorer des cas d'usage concrets avec comme fils directeur l'amélioration de l'expérience utilisateur (UX). Vous vous appuyerez pour cela sur l'importante masse de données textuelles et sur les puissantes ressources de calculs mises à votre disposition.

Compte tenu de la nature exploratoire de ce sujet, il est attendu du candidat qu'il soit autonome, force de proposition et appétant aux innovations de la communauté data.

Optimisation du score de paiement en 4 fois (Data Science)

Ce stage est proposé par l'équipe scoring, responsable du score de solvabilité calculé en temps réel sur 120,000 pages panier par jour, qui détermine à quel client Cdiscount propose le paiement à crédit (étalé en quatre mensualités). L'objectif est donc de proposer des optimisations qui permettront de maximiser le taux d'affichage du paiement en quatre fois tout en maintenant le taux de défaut à un niveau acceptable. Deux pistes méritent selon nous d'être creusées.

On pourra d'abord chercher à construire une segmentation à froid des clients par des méthodes non supervisées, qui soit capable d'expliquer l'appétence au paiement en 4 fois, de décrire la solvabilité des segments et de fournir une indication métier sur la composition des différentes sous-populations. Une telle segmentation fournira d'une part un score de solvabilité qui sera comparé à l'approche supervisée en production grâce à un test AB sur le flux des transactions, d'autre part une cartographie descriptive utile au pilotage métier du service.

La seconde approche consistera à améliorer le score existant en y intégrant des variables liées aux caractéristiques des parcours de navigation des clients, aujourd'hui non prises en compte. On analysera donc les parcours clients pour proposer des variables comportementales pertinentes, et au passage chercher à expliquer les abandons de panier ou les anomalies de fonctionnement du service d'affichage du paiement en quatre fois.

Au cours de ce stage vous serez donc amené à appliquer des méthodes supervisées et non supervisées à de grandes volumétries constituées de données clients, de parcours de navigation et de transactions, et pourrez tester vos approches à grande échelle au moyen de tests AB.

Remplissage des données produits (Data Science)

La bonne qualification des produits du site a de nombreux impacts sur l'activité de Cdiscount : elle permet une bonne navigation à nos clients et des analyses pertinentes sur la demande, le marché et la complétude de la gamme proposée.

Avec environ 150 millions de fiches produits en base, réparties dans un arbre de catégorie définissant 8000 typologies de produits ayant chacune un ensemble de propriétés pour décrire finement les caractéristiques techniques, il s'agit d'un enjeu d'automatisation fort pour Cdiscount.

Le stage aura pour vocation d'attaquer un des multiples problèmes derrière la bonne qualification des données produits, au moyen de méthodes algorithmiques :

- Vérification de la catégorie du produit
- Vérification des images du produit (résolution, présence, fond blanc...)

- Vérification des valeurs de propriétés techniques
- Reconnaissance des produits offensants

Nous attendons la mise en place d'un formalisme d'évaluation rigoureux de l'approche développée pour résoudre ce problème, avec une comparaison de plusieurs techniques déployables. Si l'approche se révèle pertinente, elle pourra être ajoutée à notre panoplie d'algorithmes de traitement des données de produit.

Analyse des zones à risque dans le code (Data Science)

Le fonctionnel du site repose sur une large base de code modifiée et enrichie semaine après semaine et qui, comme toute base de code, contient ses zones à risques et son lot de bug. Cdiscount a un processus de release qui permet le test régulier en hors-prod des modifications à livrer.

L'idée de ce stage est d'étudier la possibilité de prédire, en utilisant l'historique des modifications et des tests fonctionnels qui ont échoués, un lien entre zone de code et fonctionnalité pour détecter les zones de code à risque.

Différentes méthodes peuvent être mise en place, du simple comptage ou modèle probabiliste à une analyse riche du code lui-même couplée aux données historiques pour essayer de repérer au mieux le lien entre instructions et fonctionnel.

Clustering de produits et cannibalisation (Data Science)

Le sujet de stage que nous vous proposons dans l'équipe Pricing concerne la cannibalisation des produits dans le cas de mouvements de prix. L'objectif du stage est d'identifier des clusters de produits se substituant les uns aux autres dans le cas de mouvements de prix en mettant au point un algorithme capable d'identifier les sous-groupes de produits homogènes dans notre catalogue contenant plusieurs centaines de milliers de produits. Ces sous-groupes seraient définis par leur "étanchéité", c'est-à-dire l'absence de cannibalisation / substitution inter-groupe. La démarche s'appuiera sur un apprentissage non-supervisé et utilisera des données relatives aux produits (données de caractéristiques techniques, prix, image etc.) et à la navigation des utilisateurs.

Construction d'une ontologie à partir des requêtes des utilisateurs (Data Science)

Au sein de l'équipe dédiée à la pertinence des résultats renvoyés par le moteur, l'objectif du stage sera de construire une ontologie des unités de besoin de nos clients à partir des requêtes des utilisateurs et du contenu des fiches produits. Cette ontologie aura pour vocation à affiner la pertinence des résultats du moteur, notamment sur les requêtes à faible occurrence.

Prédiction de la date de livraison réelle (Data Analyse/Science)

Ce stage est proposé par l'équipe data-client, responsable de l'analyse et de l'algorithmie autour des sujets clients : satisfaction client, frais de port, programme de fidélité...

Depuis quelques années, la rapidité de livraison est devenue l'un des principaux critères de satisfaction des clients. Par ailleurs, Cdiscount est une place de marché (une *marketplace*) de plus en plus conséquente : en d'autres termes, on donne l'opportunité à des vendeurs professionnels de vendre leurs produits via notre site web. Or, si une partie des leurs livraisons sont directement gérées par nos services logistiques, une autre partie reste livrée directement par nos vendeurs, qui ne disposent pas forcément de la même force de frappe logistique.

Ce stage a ainsi pour objectif la prévision de la date de livraison réelle des produits vendus sur notre marketplace. Cela permettra :

1. D'identifier les vendeurs qui ont des difficultés à livrer dans les temps : nous pourrions alors les accompagner dans cet exercice bien plus difficile qu'il n'y paraît ;
2. D'informer le client au plus vite d'un potentiel retard.

Pour ce faire, vous serez notamment amenés à appliquer des méthodes d'apprentissage supervisé aux données liées aux dizaines de milliers de vendeurs de notre site et aux millions de livraisons réalisées

chaque année. Ce stage se déroulera en lien très étroit avec les équipes métiers, ce qui permettra de déclencher et tester rapidement les actions identifiées.

Personnalisation, améliorations de nos algorithmes de clustering (Data Science)

Ce stage est proposé par l'équipe data-client, responsable de l'analyse et de l'algorithmie autour des sujets clients : satisfaction client, frais de port, programme de fidélité...

Nous cherchons à proposer une expérience de plus en plus personnalisée à nos clients. La plupart de nos algorithmes de clustering sont basés sur des données d'achats et sur les infos socio-démo que l'on possède sur eux.

L'objectif de ce stage est d'enrichir nos algorithmes avec l'ensemble des données de navigation sur le site et d'interactions avec nos différents canaux de communications : e-mails, push app, SMS... Cela permettra notamment d'améliorer nos ciblage client, la pertinence de nos recommandations produit et de notre moteur de recherche.

Advanced Word Embedding of E-commerce queries (Data Science)

L'équipe data trafic travaille à améliorer la position de Cdiscount dans les pages de résultats gratuits et payants des moteurs de recherche externes, notamment Google et Bing. Elle manipule donc au quotidien des millions de requêtes utilisateurs et doit faire en sorte que le site Cdiscount puisse y répondre au mieux : a-t-on une page pertinente qui répond à la requête ? Cette page est-elle bien placée dans la structure du site ? Afin de pouvoir gérer automatiquement cette volumétrie, il est fondamental de pouvoir déterminer de manière fiable la proximité sémantique entre plusieurs requêtes ainsi qu'entre une requête et le contenu d'une page. Il est clair que les méthodes simples basées sur la ressemblance entre chaînes de caractères, comme le comptage de mots en commun ou la distance de Levenshtein, ne répondent que très partiellement à ce besoin. On expérimentera donc les modèles de plongement lexical (*word embedding*) à l'état de l'art, qui seront entraînés et testés sur le vaste catalogue de données textuelles et de positionnement sur les moteurs de recherche externes dont nous disposons, afin de capturer au mieux les spécificités sémantiques liées notre activité de e-commerce.

Profil recherché

De formation supérieure de type Bac+4/5 ou équivalent en école d'ingénieur spécialisé en data science. Vous manifestez un intérêt pour le e-commerce et le web et vous êtes d'un naturel rigoureux, motivé et curieux.

Vous êtes à l'aise avec Python, SQL et avez potentiellement des notions en big data. Vous vous épanouissez dans les sujets analytiques et leur application à des cas concrets de l'entreprise. Vous avez une appétence certaine pour les sujets techniques et les problèmes complexes. Tout cela vous permet de transformer rapidement des idées théoriques en algorithme fonctionnel.

Vous cherchez une société où vous pourrez mettre en oeuvre votre créativité et votre expertise pour relever nos défis quotidiens.

Vous vous reconnaissez ? Ne cherchez plus, postulez et partez à la conquête du web !

Informations complémentaires

Type de contrat : Stage

Localisation : Bordeaux (33)

recrutement@cdiscout.com

Comment postuler

Vous souhaitez encore plus d'informations ? N'hésitez pas à visiter notre espace recrutement : <https://emploi.cdiscout.com/>

