

Hackathon 2017

24-25 novembre au NUMA.



Le contexte

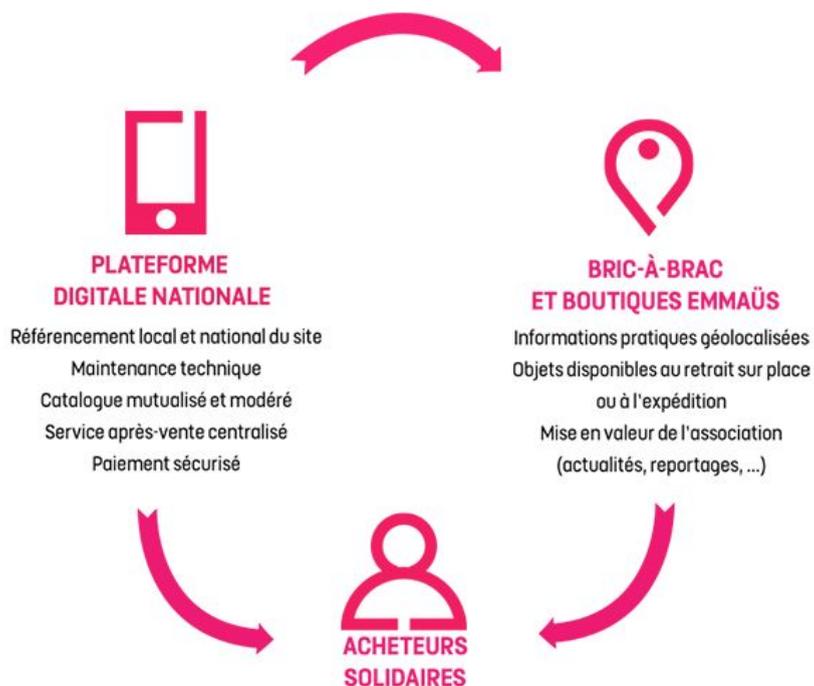
[Label Emmaüs](#) propose à la vente en ligne des objets rénovés ou créés par le mouvement [Emmaüs](#). Son catalogue d'objets est en croissance régulière. L'ajout d'un objet au catalogue depuis la réception de sa désignation (images, descriptif, entrepôt) n'est pas automatisé et prend de plus en plus d'importance.

La détermination d'un prix comme la rédaction d'une description ne sont pas toujours simples et il faut un peu de temps et d'expérience pour traiter rapidement un objet. L'ajout d'un produit prend aujourd'hui 40 minutes jusqu'à la mise en ligne. Beaucoup d'objets restent aussi très longtemps sur le site avant de trouver acquéreur et l'augmentation du catalogue nécessite plus d'espace de stockage qui n'est pas toujours disponible. Trois challenges ont été imaginés pour améliorer la qualité des données proposées aux futurs acheteurs et optimiser l'espace de stockage nécessaire.

Schéma de fonctionnement de la marketplace

label-emmaus.co est un site multi-vendeurs administré par une structure centralisée, Label Emmaüs SCIC-SAS, tiers de confiance entre vendeurs et acheteurs, qui sécurise la transaction, modère l'offre et assure un service après-vente centralisé.

Deux acteurs agissent sur le site : les marchands et Label Emmaüs. Les marchands sont des boutiques Emmaüs physiques qui vendent des produits en ligne par l'intermédiaire de la place de marché.



Les actions faites par les marchands

Le **choix des produits**, la **création des annonces**, la **gestion du stock**, le **suivi des commandes** et les **expéditions** sont effectués par les structures locales Emmaüs, et en priorité par les **compagnes**, **compagnons** et **personnes en insertion** via une **interface informatique de saisie d'annonces et de gestion de commande** à laquelle ils et elles sont formés. La création des annonces réalisées par les marchands inclut les tâches suivantes : prise de photos de chaque objet à vendre, estimation du prix, rédaction du titre et du descriptif, choix de la catégorie, saisie du poids, saisie des attributs standardisés (taille, dimension, couleur, état du produit, etc). Il y a également des opérations de gestion du stock comme ranger le produit en attente de vente dans un espace dédié à la vente en ligne, consulter ses commandes sur le back-office pour les confirmer (après vérification du stock), puis de les envoyer (faire le colis et le remettre au transporteur).

Les actions faites par Label Emmaüs

Les salarié(e)s en insertion et salarié(e)s qui composent l'équipe de Label Emmaüs modèrent les annonces avant leur mise en ligne pour garantir la cohérence nationale du catalogue de produits proposé aux acheteurs. Label Emmaüs prend également en charge le niveau de service client, la gestion des réclamations et des demandes de remboursement via un numéro unique.

vendeur 1 (Emmaüs Alternatives)



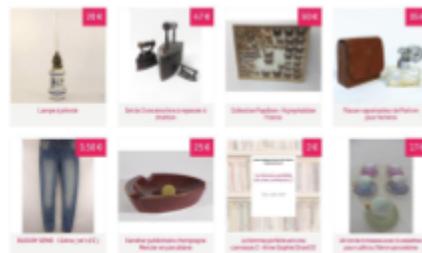
vendeur 2 (Emmaüs Lyon)



Chaque vendeur saisit son annonce dans son backoffice



Les produits de tous les vendeurs sont en vente ensemble sur label-emmaus.co



Pour en savoir plus :

<https://www.label-emmaus.co/fr/a-propos/qui-sommes-nous/>

Challenge 1 - deep learning

Nos compagnons passent 40' pour créer une annonce dont 30' sur l'identification et la qualification de l'objet. Tous n'ont pas les compétences pour faire ce travail d'identification et sont donc exclus de fait de l'accès à ce nouveau métier.

Objectif

Chaque objet est caractérisé par une ou plusieurs d'images. Il faut déterminer la ou les catégories auxquelles il appartient.

Données

200.000 images, 150.000 objets, 176 classes, la hiérarchie entre les classes. Un fichier avec le nom des images, les objets auxquels elles appartiennent et les classes associées.

Evaluation

Comme il y a plusieurs classes par objets, on s'intéresse principalement à prédire les bonnes classes sans regarder la confiance du modèle. C est l'ensemble des classes attendues, P la prédiction. Pour un objet, l'erreur est :

$$|C \cap P| / |C \cup P|$$

Les aspects auxquels le jury fera attention

- Comment Label Emmaüs pourra-t-il utiliser votre solution ? ROI ?
- Y-a-t-il des objets ou des classes plus difficiles que d'autres ?
- Comment votre modèle peut-il être utilisé pour proposer d'autres suggestions ?

Préparation des données - ce qui a été fait pour vous

Les images du site Label Emmaüs ont été scrapées, un fichier regroupe l'ensemble des noms de fichier avec la classe et l'objet à laquelle elles appartiennent.

Pour démarrer

[Search images with deep learning](http://www.xavierdupre.fr/app/mlinsights/helpsphinx/notebooks/search_images.html) : le notebook expose comment manipuler des images avec [keras](#) et comment utiliser le résultat des couches intermédiaires d'un réseau de neurones profond dans le but de recherche des images similaires.
http://www.xavierdupre.fr/app/mlinsights/helpsphinx/notebooks/search_images.html

Challenge 2 - machine learning

Certains groupes, vendent mieux que d'autre. Ceux qui vendent moins bien créent moins d'annonces : ils ont un espace de stockage qui tend à saturer, l'activité n'est pas perçue comme assez performante pour investir. A la question "pourquoi mes produits ne se vendent pas", Label Emmaüs ne sait pas répondre (sauf par du ressenti métier).

Objectif

Prédire la durée entre la mise en ligne et la vente d'un objet.

Données

10.000 objets, des données sur les objets (description, catégories), des données sur les acheteurs, les vendeurs, les données sur les paniers. La date de vente n'est connue que pour un sous-ensemble des produits (la base d'apprentissage). Il faudra prédire la durée pour les produits de la base de tests.

Evaluation

Les objets non vendus n'ont pas de durées définies. On leur attribue la valeur 180. Le modèle de prédiction doit répondre une durée entre 0 et 180 jours, ou 180 si l'objet se vend en plus de six mois ou s'il ne se vend pas. Soit $f(x)$ la prédiction en nombre de jours, l'erreur est :

$$\min(|f(x) - \min(x, 180)| / 180, 1)$$

Les aspects auxquels le jury fera attention

- Comment Label Emmaüs pourra-t-il utiliser votre solution ? Quel est son ROI ?
- Y-a-t-il des objets plus difficiles que d'autres ?
- Comment votre prédiction évolue en fonction de cinq variables clés (le prix, la catégorie, la distance entre l'acheteur et l'entrepôt, le nombre de produits dans les paniers, les mots-clés dans les descriptions) ?

Préparation des données - ce qui a été fait pour vous

Certains champs ont été supprimés ou anonymisés (mail utilisateur). Les produits ont été répartis en deux jeux : apprentissage et test. Les données panier et achat ont été supprimées pour les données de tests sur lequel les participants sont évalués. Quelques éléments de code :

http://www.xavierdupre.fr/app/ensae_projects/helpsphinx/projects/hackathon_2017.html#elements-de-code.

Challenge 3 - économie créative

La rotation du stock dormant est un enjeu important (coût et limites de stockage des produits, dynamique de ventes sur l'activité).

Objectif

Imaginer des fonctionnalités qui pourraient agir sur le stock dormant (plus de 3 mois ?) et qui respectent la [charte Emmaüs](#). Tout l'enjeu du sujet est d'inciter à l'achat tout en restant équitable.

Contraintes de la charte Emmaüs

Pas d'enchères ascendantes : le prix doit rester accessible à tous... Les ventes liées ou une réduction pour l'achat de deux objets. Chaque marchand doit être traité de façon équitable.

Les aspects auxquels le jury fera attention

- Est-ce que les données récoltées permettent de définir le caractère équitable de Label Emmaüs ?
- Comment les utilisateurs vont-ils modifier leur comportement ?
- Le prix est-il le principal facteur incitatif ?

Éléments de réflexion

Doit-on s'adresser aux acheteurs réguliers ? A ce qui ne le sont pas ? Doit-on les différencier ? Faut-il appliquer vos idées sur tous les objets ou certains objets, avec un prix élevé ou simplement encombrants, certaines saisons ? Pensez-vous à un système permanent, ponctuel ? Le prix est-il toujours bien choisi ? Pensez-vous à organiser une loterie pour certains acheteurs ? Des friperies ? Peut-on recommander un objet à un utilisateur, n'est-ce pas l'influencer ? Comment mesurer le caractère équitable du site de vente en ligne ? Il est difficile de quantifier ce qu'une idée peut apporter quand elle n'est pas encore mis en place. Ne peut-on pas tout de même faire un rapprochement entre deux objets similaire, un qui s'est vendu et un autre qui ne s'est pas vendu ?

Les enchères descendantes vérifient les contraintes imposées par Emmaüs, cela ne risque-t-il pas d'inciter les acheteurs à attendre que les enchères commencent ? Comment organiser une sorte de dégressivité équitable sans diminuer l'intérêt des acheteurs et des vendeurs pour le site ? Comment créer des bundles (achat groupé) avec plusieurs marchands sans favoriser l'un d'eux ? Faudrait-il mutualiser les risques et décider de la répartition du prix de la vente avant que celle-ci ait lieu ?

