

Fiche de lancement – Expérimentation SSP Lab-Indice des prix :

Collecte de données de prix par webscraping et estimation de modèles hédoniques en vue de l'amélioration de la qualité statistique de l'indice des prix à la consommation

**Sponsors** : Marie Leclair et Jean-Denis Zafar, division des prix à la consommation DPCEM/DSDS

**Sujet :**

L'indice des prix à la consommation (IPC) est l'indicateur mesurant l'inflation en France. Cet indice a un rôle très important dans l'économie : outre son impact sur la politique monétaire, il est utilisé pour l'indexation de contrats privés, de pensions alimentaires, pour la revalorisation du SMIC, des retraites et des loyers, dans des instruments financiers, etc.

L'IPC mesure, chaque mois, l'évolution « pure » des prix, à qualité constante. Ainsi, il s'agit de suivre un certain nombre de produits identiques dans le temps. Lorsque ceux-ci disparaissent, ils sont remplacés par des produits qui peuvent ne pas être équivalents. Il convient alors de distinguer, dans l'évolution des prix des remplaçants par rapport aux produits remplacés, un effet qualité (différence de prix pour un mois fixé), de l'effet inflation que l'on cherche à mesurer.

Certaines méthodes, dites « hédoniques », consistent à estimer cet effet qualité à partir de coefficients correspondant aux prix sous-jacents des différentes caractéristiques techniques du produit (par exemple la marque de l'ordinateur, la mémoire vive d'un ordinateur, le modèle et la fréquence du processeur, etc.). Cette estimation est typiquement réalisée par des modèles de régression. On recourt aujourd'hui à ce type de méthodes pour seulement quatre types de produits aujourd'hui dans l'IPC, mais la qualité statistique des estimations est faible, notamment à cause d'un nombre de prix relevés trop faible. Le recours au Webscraping des caractéristiques et des prix des produits offre l'opportunité d'augmenter significativement le nombre de relevés de prix et de caractéristiques des produits. Le but de l'expérimentation est donc de revoir les estimations et les modèles de prix hédoniques au vu des nouvelles données scrapées pour les produits pour lesquels les modèles hédoniques sont déjà utilisés et d'étendre le recours à ces méthodes à d'autres produits (en particulier des biens électroniques : smartphones, ordinateurs, tablettes, etc.). Les données collectées sur Internet par webscraping forment des ensembles d'estimation assez larges pour envisager des sélections de modèles et des modèles de régression pénalisée (de type Lasso) où le nombre de variables explicatives (caractéristiques) est très grand afin d'estimer l'effet qualité correspondant à chaque caractéristique. Les résultats obtenus permettront d'envisager une implémentation dans l'IPC, et ainsi une amélioration statistique de l'estimation de l'inflation en France.

Des travaux exploratoires seront également menés, sur la pertinence d'autres modèles de prédiction des prix : random forests, bagging, boosting, etc., à partir des caractéristiques techniques) machine learning, etc.

**Objectif** : Recourir au webscraping pour enrichir la base d'information sur les caractéristiques et les prix de certains produits dont on souhaite suivre l'évolution de la qualité afin de la maîtriser dans l'IPC, proposer et tester des modèles hédoniques exploitant ce large ensemble d'informations.

**Mode de travail** : Les partenaires consacreront en moyenne une journée par semaine à l'expérimentation.

Point d'avancement toutes les deux semaines avec les superviseurs.

**Groupe impliqué dans l'expérimentation**

Superviseurs : Marie Leclair (division des prix à la consommation) et Elise Coudin (SSP Lab)

Sponsor/Experts métiers : Jean-Denis Zafar, division des prix à la consommation  
Experts méthodologiques : Stéphanie Himpens et Mathilde Poulhes (SSP Lab, DMCSI)  
Conseil : Anna Simoni (CREST)

### **Rôles**

Les superviseurs définissent les objectifs de l'expérimentation, assistent aux démonstrations des avancées, jugent de la conformité des travaux à leurs attentes et redéfinissent les propriétés au besoin.

L'équipe de réalisation (sponsor/expert métiers et experts méthodologiques) définissent conjointement les étapes et les livrables.

### **Ressources**

Le SSP Lab met à disposition le lieu pour accueillir les expérimentateurs lors de session de travail en commun, du matériel pour accéder à la plateforme innovation le temps de l'expérimentation pour les agents impliqués qui n'en auraient pas. La DSI (Unissi) met à disposition les accès idoines à la plateforme d'innovation, et proposera son assistance pour le paramétrage des ressources correspondantes.

### **Calendrier et types de rendus**

mai 2018 : présentation conférence groupe Ottawa, note de faisabilité pour implémentation en production