

Affectation de transaction à un store grâce au Scoring

Komlan Jean-Marie DANTODJI

Etudiant en M1 Big Data, Université Paris 8

Encadrante universitaire: Mme Rakia JAZIRI

Encadrant entreprise: Mr Thomas Moulin



Jun 17, 2020

Transaction
connect

Plan

2/18

- 1 Introduction
 - Présentation de l'entreprise
 - Solutions proposées
- 2 Problématique
- 3 Etat de l'art
 - Pattern Regrex
 - Store Locator
 - Alpha / Alpha City
- 4 Algorithme du Scoring
- 5 Apprentissage
- 6 Conclusion

Transaction Connect

3/18

- Start Up de B2B2C aux retailers
- Editeur de solution numerique basé sur la donnée de paiement

Solutions proposées

4/18

- Transaction Connect signe des contrats avec des foncières
- Fournit des services et outils de décisions aux clients foncières (Dashboard, enrichissement des données)
- Fidélisation de la carte bancaire
- Notification et Validation des Cashbacks aux clients acheteurs

Les modes d'intégrations des Clients Business (B)

5/18

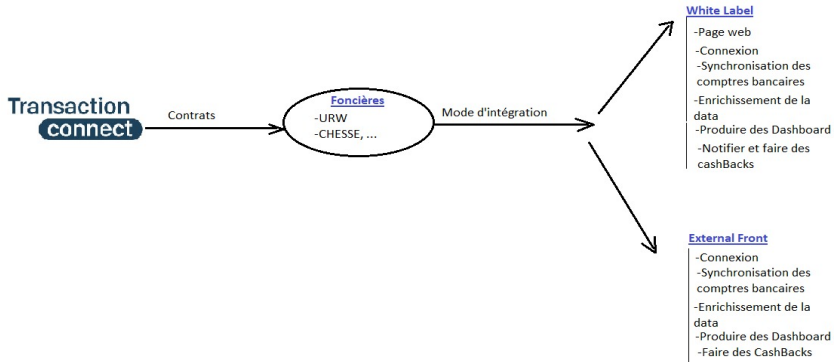


FIGURE – modes d'intégration B

Les modes d'intégration des Clients C

6/18

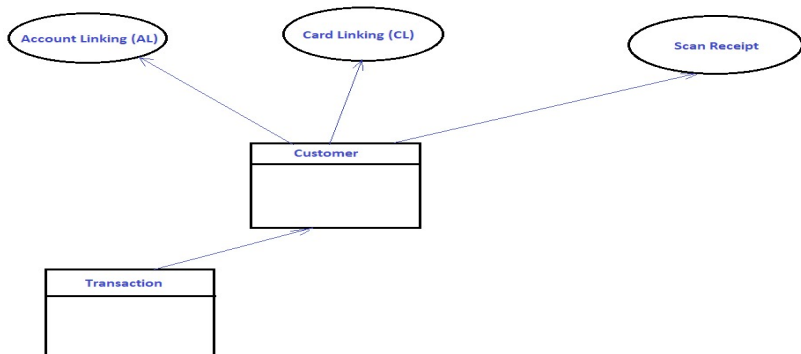


FIGURE – Mode d'intégration C

Problématique

7/18

- Les labels de transactions ne contiennent pas des informations nécessaires à l'affectation à un store
- Affectation des transactions à un store grace au scoring
- Automatisation des Modèles du scoring à un apprentissage mensuel

Etat de l'art : Pattern Regrex

8/18

- Patterns Regrex

Exemple Label : "CB MONOPRIX 0295 Paris 02/03/21 10€"

Retailer pattern : `'.*\mMONOPRIX\M.*'`

Store pattern : `'.*\mMONOPRIX 0295\M.*'`

- Une transaction ayant ce label est affectable parcequ'on peut identifier le retailer MONOPRIX et le store MONOPRIX 0295

Etat de l'art : Store Locator

9/18

- Store Locator

Exemple Label : "CB MONOPRIX Paris 02/03/21 10€"

Ville : Paris

Retailer pattern : '.*\mMONOPRIX\M.*'

- Si pour une transaction on connaît la ville et le Retailer, Store Locator peut affecter cette transaction à un store si cette dernière est unique dans la ville

Etat de l'art : Alpha / Alpha City

10/18

- Alpha / Alpha City (basé sur la détermination du Google Place Id d'une zone géographique)

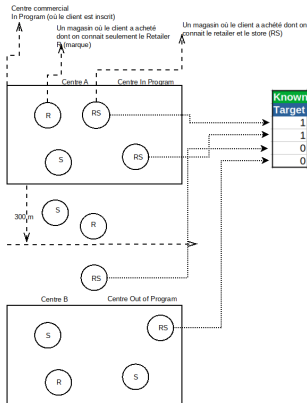
Algorithme du Scoring : Préalables

11/18

- Migration des requetes de calcule de Postgres vers Redshift Amazon
- Calcul des features nécessaires à l'apprentissage

Features Engineering

12/18



Known Stores and Retailers						Known Stores				Known Retailers			
Target	DayOfWeek	distance	distance_bin	Amount	Cannib	nb_transac	PM	SM	PNM	PC	PPM	PPNM	PPC
1	Samedi	5.1	(0,10]	10	0.166	10	1	0	1	4	1	1	3
1	Samedi	5.1	(0,10]	40	0.666	10	1	0	1	4	1	1	3
0	Samedi	5.1	(0,10]	5	0.083	10	0	0	0	4	0	0	3
0	Samedi	5.1	(0,10]	5	0.083	10	2	0	0	4	1	0	3

Pour les transactions identifiées (associées à un store et à un retailer)

Target:

Transaction In Program ou Out of Program

dayOfWeek:

jour de transaction extrait de la date de transaction

distance:

distance en (km) entre le centre commercial où le client est inscrit au store le plus proche en dehors de son centre

distance_bin:

classe de distance qu'appartient la distance précédente (a,b]

Amount:

montant dépensé dans le store

Cannib = cannibalisation:

rapport entre le montant dépensé chez le store sur le montant total dépensé dans la journée

nb_transac:

Nombre de transaction effectuées

PM = Purchases_InMail

Nombre de store dans le meme centre

SM = Services_InMail

Nombre de services utilisés (amazon locker, parking) à moins de 3km que le centre visité

PNM = Purchases_NextToTheMail

Nombre de stores visités en dehors du centre et à moins de 300m

PC = Purchases_InCity

Nombre de store visités dans la même ville que le centre visité

PPM = Purchases_Potential_InMail

Nombre de store de tous les retailers inconnu dans le meme centre

PPNM = Purchases_Potential_NextToTheMail

Nombre de stores de tous les retailers visités en dehors du centre et à moins de 300m

PPC = Purchases_Potential_InCity

Nombre de store de tous les retailers visités dans la même ville que le centre visité

Jeu de données

13/18

● Exemple de transactions du centre "Les 4 Temps" de Puteaux

	DayOfWeek	distance	distance_bin	amount	cannibalisation	nb_transac	Purchases_InMall	Services_InMall	Purchases_NextToTheMall	Purchases_InCity
0	Thursday	8.743127	(5.0, 10.0]	1.65	0.017518	3	0	0	0	0
1	Thursday	11.469140	(10.0, 20.0]	4.23	0.599103	3	0	0	0	0
2	Friday	11.526189	(10.0, 20.0]	15.50	0.333722	4	1	0	1	0
3	Friday	12.412205	(10.0, 20.0]	9.00	0.631520	4	1	0	1	0
4	Tuesday	6.868553	(5.0, 10.0]	67.40	0.322184	4	0	0	0	0
5	Saturday	4.337519	(-0.001, 5.0]	5.00	0.088437	5	0	0	0	0
6	Wednesday	4.282542	(-0.001, 5.0]	40.75	0.100635	11	0	0	0	0
7	Wednesday	4.349075	(-0.001, 5.0]	12.58	0.026281	11	0	0	0	0
8	Wednesday	8.743127	(5.0, 10.0]	4.99	0.020390	7	1	0	1	1
9	Wednesday	4.282542	(-0.001, 5.0]	48.94	0.097177	7	2	0	2	2

FIGURE – Transactions considérées

Jeu de données

14/18

	Services_InMall	Purchases_NextToTheMall	Purchases_InCity	Purchases_Potential_InMall	Purchases_Potential_NextToTheMall	Purchases_Potential_InCity	target
0	0	0	0	2.0	2.0	3.0	0
0	0	0	0	0.0	0.0	0.0	1
1	0	1	0	0.0	0.0	0.0	1
1	0	1	0	0.0	0.0	0.0	1
0	0	0	0	0.0	0.0	0.0	0
0	0	0	0	0.0	0.0	0.0	0
0	0	0	0	0.0	0.0	0.0	0
0	0	0	0	0.0	0.0	0.0	0
1	0	1	1	0.0	0.0	0.0	1
2	0	2	2	0.0	0.0	0.0	0

FIGURE – Transactions considérées

Informations données

15/18

```
Int64Index: 85095 entries, 0 to 85243
Data columns (total 14 columns):
DayOfWeek                85095 non-null object
distance                 85095 non-null float64
distance_bin             85095 non-null category
amount                  85095 non-null float64
cannibalisation          85095 non-null float64
nb_transac               85095 non-null int64
Purchases_InMall         85095 non-null int64
Services_InMall          85095 non-null int64
Purchases_NextToTheMall  85095 non-null int64
Purchases_InCity         85095 non-null int64
Purchases_Potential_InMall 85095 non-null float64
Purchases_Potential_NextToTheMall 85095 non-null float64
Purchases_Potential_InCity 85095 non-null float64
target                  85095 non-null int64
dtypes: category(1), float64(6), int64(6), object(1)
memory usage: 9.2+ MB
```

FIGURE – Les types de features

Apprentissage

16/18

- XGBoost
- Decision Tree
- Random Forest

Conclusion

17/18

- Prochainement, Metrics de choix de modèle
- Etendre le modele aux autres centres
- Créer un modele propre aux grands centres
- Automatiser l'apprentissage à chaque mois

Merci pour votre attention...