

Caractérisation de profils clients à partir de données transactionnelles partielles et anonymes

Application au secteur du commerce de proximité

Localisation : Site du Futuroscope, France

Date de début : début 2026 (flexible)

Durée : 6 mois

Encadrants : Loïc ADAM, Mickaël BARON, Brice CHARDIN et Stéphane JEAN (*l'encadrant dont le nom est souligné est l'encadrant référent : baron@ensma.fr*)

Mots clés : analyse de données, apprentissage automatique, personnalisation, IA prédictive, clustering, apprentissage semi-supervisé.

Contexte du stage

Bimedia, filiale du groupe Orisha, est une entreprise spécialisée dans les solutions destinées aux points de vente de proximité (bureaux de tabac, boulangeries, etc.). Son offre repose sur un ensemble de services comprenant du matériel (caisses enregistreuses à double écran : le premier pour le commerçant et le second pour le client final), un logiciel d'encaissement, des solutions dématérialisées (cartes prépayées, etc.), ainsi que des dispositifs d'animation commerciale via des publicités diffusées sur l'écran client des caisses.

Si le point de vente est identifié, le client final, lui, ne l'est pas. Pour pallier cette limite, une initiative récente baptisée BMyShop¹, proposée par Bimedia sous la forme d'une application mobile, permet d'associer le client final à un programme de fidélité. Les informations collectées se limitent aux données strictement nécessaires (prénom, nom, âge) en conformité avec la réglementation (RGPD). Toutefois, cette base de clients finaux identifiés reste restreinte par rapport au volume de transactions réellement effectuées dans les commerces par des clients finaux anonymes.

Bimedia souhaite mieux caractériser les typologies de clients finaux fréquentant les points de vente, et croiser ces informations avec celles relatives aux commerces eux-mêmes. L'objectif est de mieux qualifier l'audience présente en magasin afin de proposer aux annonceurs des espaces publicitaires plus pertinents.

Objectifs du stage

L'objectif de ce stage est d'explorer des approches d'analyse et d'enrichissement de données à partir des bases transactionnelles (l'historique d'achats) issues d'une part des clients identifiés via l'application BMyShop et, d'autre part, des clients anonymes. Le stage visera à

1. <https://bmyshop.com/>

définir des méthodologies permettant de caractériser et d'associer des profils clients à partir de données partielles. Plusieurs pistes de recherche pourront être explorées :

- Construction de profils clients sur la base BMyShop : élaborer des typologies de clients à partir des données de tickets de caisse et des informations disponibles (prénom, âge, etc.). L'utilisation de sources de données externes (ex. : distribution statistique des prénoms par génération, caractéristiques socio-démographiques de la zone) pourra être envisagée pour enrichir les profils.
- Projection des profils sur la base des clients anonymes : étudier la possibilité d'associer des comportements d'achat (panier moyen, typologie des produits consommés) à des profils-types établis précédemment afin d'inférer des caractéristiques pertinentes sur les clients anonymes.
- Évaluation de l'impact de nouvelles variables sur la qualité des profils : analyser l'effet de l'enrichissement des données BMyShop par l'ajout de caractéristiques complémentaires sur la précision des typologies générées. Cela impliquera la génération de données synthétiques adaptées aux caractéristiques complémentaires.

Réalisation attendue

Les objectifs du stage en adéquation avec les livrables sont les suivants.

- Étude bibliographique sur l'utilisation de techniques d'apprentissage automatique pour répondre aux besoins (apprentissage semi-supervisé).
- Conception et implémentation de prototypes logiciels pour attester de la faisabilité de la solution sur les données fournies par la société Bimedia.
- Rapport de synthèse décrivant l'étude, l'état de l'art, le prototype et les résultats obtenus.

À noter que les prototypes développés dans le cadre de ce stage seront des **démonstrateurs**. Ils ne visent pas être développés à des fins de mise en production. Les données (≈ 100 Go) et les besoins de l'étude seront fournis par Bimedia.

Profil du candidat

Le candidat doit être en Master 2 en Informatique ou en dernière année de préparation d'un diplôme d'ingénieur spécialité Informatique. Une bonne connaissance du langage de programmation Python et des bibliothèques usuelles d'apprentissage automatique est requise. Un **bon niveau** en français est fondamental.

Documents à fournir

Les deux documents suivants sont indispensables ; sans eux, la candidature sera rejetée.

- Curriculum Vitae ;
- Notes de Master ou équivalent (si possible avec le classement) ;
- Autres documents utiles pour appuyer la candidature (ex. : projets, GitHub).

