

Royaume de Belgique

Projet de spécialisation IA

**RGPD**

Par Harold Rajaonarison

Année Académique 2024 - 2025

**Table des matières**

[**Introduction** 1](#_Toc199694863)

[1. **Identification des données personnelles** 1](#_Toc199694864)

[2. **Finalité et limitation du traitement** 1](#_Toc199694865)

[3. **Pseudonymisation et sécurité** 1](#_Toc199694866)

[4. **Durée de conservation** 2](#_Toc199694867)

[5. **Encadrement des accès** 2](#_Toc199694868)

[6. **Transparence et documentation** 2](#_Toc199694869)

[**Conclusion** 2](#_Toc199694870)

**Introduction**

Dans le cadre de ce projet, l’objectif est de développer un modèle de segmentation comportementale des clients d’une marketplace fictive nommée "Amazing". Ce modèle repose sur l’exploitation de données utilisateurs afin de prédire leur profil d’achat. Dès lors, une attention particulière doit être portée à la protection des données personnelles et au respect des principes fondamentaux en matière de vie privée. Le présent document décrit la méthodologie mise en place pour garantir une conformité avec le Règlement Général sur la Protection des Données (RGPD) tout au long du projet.

1. **Identification des données personnelles**

Le jeu de données utilisé contient des informations comportementales sur les utilisateurs telles que :

* Nombre de vues de produits
* Nombre d’ajouts au panier
* Nombre d’achats
* Montant total dépensé
* Fréquence d’activité (jours actifs, récence, etc.)

Même si ces données ne contiennent pas de noms, prénoms ou adresses e-mail, elles peuvent potentiellement permettre d’identifier un individu de manière indirecte lorsqu’elles sont croisées avec d’autres sources. Elles sont donc considérées comme des données personnelles et doivent être traitées avec précaution.

1. **Finalité et limitation du traitement**

La finalité exclusive du traitement des données est académique. Il s’agit de mettre en œuvre une démarche de data science pour segmenter des utilisateurs selon leurs habitudes d’achat, à des fins de démonstration pédagogique. Aucune utilisation commerciale, publicitaire ou externe n’est prévue.  
Le principe de minimisation a été respecté : seules les données strictement nécessaires à l’objectif du projet ont été sélectionnées, sans collecte excessive d’informations.

1. **Pseudonymisation et sécurité**

Aucune information directement identifiable n’est présente dans les données. Afin de renforcer la protection des utilisateurs, toutes les données ont été anonymisées ou pseudonymisées dès leur importation.  
Par ailleurs, des mesures techniques ont été prises pour garantir la sécurité des fichiers :

* Utilisation d’un environnement isolé (virtualenv)
* Stockage local et sécurisé des fichiers CSV et des modèles
* Aucune exposition sur internet ni hébergement externe

1. **Durée de conservation**

Les données seront conservées uniquement pendant la durée du projet académique. À l’issue de la période d’évaluation, l’ensemble des fichiers contenant les données, les modèles dérivés et les traces de traitement seront supprimés définitivement du système.

1. **Encadrement des accès**

Le traitement a été réalisé dans un cadre strictement académique, limité à un étudiant développeur, sous la supervision pédagogique d’un enseignant. Aucun partage externe n’est autorisé. L’accès aux données est resté strictement local et contrôlé, sans circulation sur le web, sur des plateformes tierces ou entre plusieurs intervenants.

1. **Transparence et documentation**

Une documentation complète a été tenue tout au long du projet. Elle détaille l’origine des données, les étapes de transformation, l’entraînement du modèle ainsi que les mesures mises en œuvre pour garantir le respect de la vie privée. Cette documentation fait partie intégrante du livrable pédagogique et assure la traçabilité du traitement.

**Conclusion**

L’ensemble du processus de traitement a été conçu pour respecter les principes fondamentaux de protection des données personnelles. Bien que le projet soit fictif et à visée éducative, les bonnes pratiques du RGPD ont été appliquées pour garantir un niveau d’exemplarité, notamment en matière de sécurité, de limitation des accès, de conservation restreinte, et de transparence dans les traitements.