

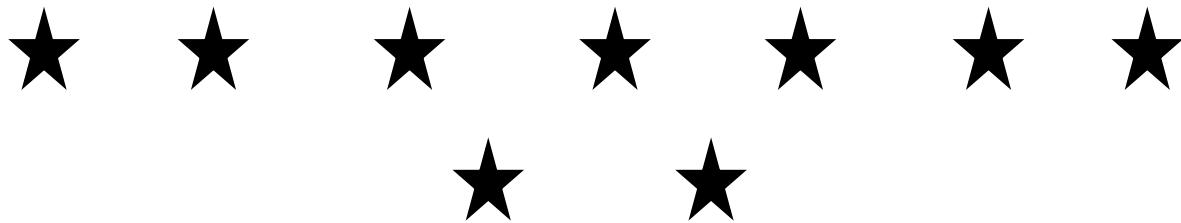


# Université Alioune Diop de Bambeley

## UFR Science Appliquée à la Technologie de l'Information et à la Communication (SATIC)

### Master Statistique et Informatique Décisionnelle (SID)

## Examen de Data Mining Approche Informatique



Présentés par :

**Samba KEBE - Habibatou DIALLO – Soriba  
KOMA – Ousmane SOW – Mouhamed DIENG –  
Aliou NIANG**

Année Universitaire  
2025 - 2026

## Introduction

Dans un contexte où les entreprises disposent de volumes de données de plus en plus importants, l'analyse et la valorisation des données clients deviennent un enjeu stratégique majeur. Les techniques de **Data Mining** et de **Business Intelligence** permettent aujourd'hui d'extraire des informations pertinentes afin d'améliorer la prise de décision, la connaissance client et la performance commerciale. Dans le cadre de cette étude, l'analyse repose sur un questionnaire réduit composé de 23 variables, sélectionnées à partir du questionnaire officiel du RGPH-5. Ce choix méthodologique vise à se concentrer sur les variables les plus pertinentes pour l'étude des caractéristiques sociodémographiques des individus et des conditions de vie des ménages, tout en facilitant le traitement et l'interprétation des données.

Dans un contexte où les entreprises disposent de volumes de données de plus en plus importants, l'analyse et la valorisation des données clients deviennent un enjeu stratégique majeur. Les techniques de **Data Mining** et de **Business Intelligence** permettent aujourd'hui d'extraire des informations pertinentes afin d'améliorer la prise de décision, la connaissance client et la performance commerciale.

L'objectif principal est de fournir un outil simple, visuel et interactif permettant d'identifier différents profils de clients, d'analyser leur comportement d'achat et d'aider à la prise de décisions marketing. L'application met également l'accent sur la clarté des résultats à travers des tableaux synthétiques, des indicateurs clés de performance (KPI) et des visualisations adaptées.

## I. Présentation et fonctionnement de l'application

### 1. Architecture générale de l'application

L'application est développée en Python à l'aide de **Streamlit**, permettant la création rapide d'interfaces web interactives. Elle est organisée autour d'un menu latéral offrant l'accès aux différents modules fonctionnels.

Les données sont chargées une seule fois et stockées dans la mémoire de session afin d'optimiser les performances.

### 2. Description des modules

#### ❖ Chargement des données

L'utilisateur importe un fichier CSV contenant les transactions clients. Des contrôles sont appliqués pour vérifier la structure des données et effectuer les conversions nécessaires (dates, montants, types numériques).

#### ❖ Analyse descriptive

Ce module fournit une vue globale de la base de données à travers des statistiques de base et des visualisations. Il permet d'identifier les produits les plus vendus et d'évaluer la performance globale.

#### ❖ Règles d'association (Apriori)

Un système de recommandation est intégré à partir des règles extraites. L'utilisateur peut saisir un panier de produits et obtenir des recommandations basées sur les associations fréquentes.

#### ❖ Segmentation clients

Deux méthodes de segmentation sont proposées :

- K-means pour une segmentation automatique
- RFM pour une segmentation marketing orientée comportement client

## II. Etapes de modélisations

#### ❖ Règles d'association – Algorithme Apriori

##### Étape 1 : Préparation des données

Les transactions sont transformées en une matrice binaire indiquant la présence ou l'absence de produits dans chaque panier.

## **Étape 2 : Extraction des itemsets fréquents**

L'algorithme Apriori identifie les ensembles d'articles dépassant un seuil minimal de support.

## **Étape 3 : Génération des règles**

À partir des itemsets fréquents, des règles d'association sont générées et évaluées selon le support, la confiance et le lift.

## **Étape 4 : Exploitation des règles**

Les règles les plus pertinentes sont utilisées pour recommander des produits aux clients.

### **❖ Segmentation par K-means**

## **Étape 1 : Sélection des variables**

Les variables quantitatives liées au comportement d'achat sont sélectionnées.

## **Étape 2 : Normalisation**

Les données sont normalisées afin d'éviter l'influence des échelles différentes.

## **Étape 3 : Clustering**

L'algorithme K-means est appliqué pour regrouper les clients en clusters homogènes.

## **Étape 4 : Interprétation**

Chaque cluster est interprété selon ses caractéristiques moyennes.

### **❖ Analyse RFM**

## **Étape 1 : Calcul des indicateurs**

Pour chaque client, les indicateurs Récence, Fréquence et Monétaire sont calculés à partir des transactions.

## **Étape 2 : Attribution des scores**

Chaque indicateur est découpé en classes afin d'attribuer des scores de 1 à 5.

## **Étape 3 : Segmentation marketing**

Les scores sont combinés pour classer les clients en segments marketing (fidèles, à risque, récents, etc.).

## **Étape 4 : Analyse des résultats**

Les segments sont analysés afin d'orienter des stratégies marketing adaptées.

## **Conclusion**

Ce projet démontre l'apport des techniques de Data Mining dans l'analyse et la segmentation des clients. La séparation entre le fonctionnement de l'application et les méthodes de modélisation permet une compréhension claire et structurée du travail réalisé. L'application constitue une base évolutive pouvant intégrer des modèles prédictifs ou des analyses avancées.