

# Documentation du Projet : Analyse Client et Modélisation Prédictive

## 1. Introduction

### 1.1 Contexte

L'entreprise **Chez Dj Irving** souhaite améliorer ses stratégies marketing en comprenant mieux le comportement de ses clients. Ce projet vise à analyser les données clients et à développer un modèle prédictif pour identifier les clients les plus susceptibles de répondre aux campagnes marketing.

### 1.2 Objectifs

- Segmenter les clients en fonction de leurs caractéristiques démographiques et de leur comportement d'achat.
- Construire un modèle prédictif pour estimer la probabilité de réponse des clients aux campagnes marketing.

## 2. Méthodologie

### 2.1 Collecte des Données

Les données utilisées dans ce projet comprennent :

- Démographiques : année de naissance, niveau d'éducation, état matrimonial, revenu.
- Comportementales : nombre d'enfants à domicile, fréquence d'achat, montant dépensé par produit.

### 2.2 Préparation des Données

- Nettoyage des données : gestion des valeurs manquantes et des doublons.
- Création de nouvelles variables : calcul de l'âge des clients à partir de l'année de naissance.

### 2.3 Analyse Exploratoire des Données (EDA)

- Visualisation des distributions des variables.
- Identification des corrélations entre les variables.

### 2.4 Segmentation des Clients

- Application de l'algorithme de clustering (K-means) pour segmenter les clients en groupes basés sur leurs caractéristiques.

- Analyse des caractéristiques de chaque cluster : démographie, comportement d'achat, réponse aux campagnes.

## 2.5 Modélisation Prédicative

- Sélection du modèle : régression logistique pour prédire la réponse des clients.
- Entraînement et validation : division des données en ensembles d'entraînement et de test.
- Évaluation des performances du modèle en utilisant des métriques telles que l'accuracy, la précision, le rappel, et le F1-score.

## 3. Résultats

### 3.1 Segmentation des Clients

- **Clusters Identifiés :**
  - Cluster 0 : Clients avec un revenu moyen faible et une faible probabilité de réponse.
  - Cluster 1 : Clients avec un revenu moyen élevé et une réponse modérée aux campagnes.
  - Cluster 2 : Clients avec un revenu très élevé et une forte réponse aux campagnes.

### 3.2 Performance du Modèle

- **Accuracy :** 87%
- **Précision :** 60%
- **Rappel :** 29%
- **F1-Score :** 0.41

### 3.3 Interprétation des Résultats

- Le modèle présente une bonne précision globale, mais des performances médiocres en termes de rappel, indiquant qu'il manque des réponses positives. Cela souligne la nécessité de réajuster le seuil de classification.

## 4. Conclusions et Recommandations

### 4.1 Conclusions

Le projet a permis d'identifier des segments de clients et d'estimer la probabilité de réponse aux campagnes. Bien que le modèle soit performant, il nécessite des améliorations pour mieux capturer les véritables réponses.

### 4.2 Recommandations

- **Amélioration des Données** : Collecter des données supplémentaires sur le comportement d'achat des clients pour affiner le modèle.
- **Affinage du Modèle** : Tester d'autres algorithmes (arbres de décision, forêts aléatoires) pour améliorer les performances.
- **Suivi et Itération** : Mettre en place un système de suivi pour évaluer continuellement l'efficacité des campagnes basées sur les résultats du modèle.