

REPUBLIQUE DU CAMEROUN

Paix-Travail-Patrie

Ministère de l'Enseignement Supérieur

Université de Douala

ECOLE SUPERIEURE DES SCIENCES DE LA
FINANCE DE L'ASSURANCE ET DES RISQUES

REPUBLIC OF CAMEROON

Peace-Work-Fatherland

Ministry of Higher Education

The University of Douala

HIGHER SCHOOL OF SCIENCE, FINANCE
INSURANCE AND RISKS



PROJET DE MACHINE LEARNING

Thème : Segmentation client avec l'algorithme DBSCAN

Rédigé par :

- AICHA OUSSEINI : M2 SBD
- GUEKAM FIRMIN Donatien : M2 ACT

Sous l'encadrement académique de : Dr Habib SIDI

ANNÉE ACADÉMIQUE : 2024 - 2025

PLAN DU PROJET

INTRODUCTION

MÉTHODOLOGIE

RÉSULTATS

CONCLUSION

RECOMMANDATIONS

INTRODUCTION

Ce projet a pour objectif de segmenter les clients d'un centre commercial en utilisant des techniques de Machine Learning. L'objectif est d'identifier des groupes de clients ayant des caractéristiques similaires afin de pouvoir adapter les stratégies marketing et améliorer l'expérience client.

Méthodologie

1. Chargement et Prétraitement des Données

- **Source des données** : Fichier mall_customers.csv.
- Les bibliothèques principales incluent sklearn, pandas, matplotlib, seaborn, et numpy.
- **Variables principales** :
 1. **Customer ID** : Numéro unique attribué à chaque client
 2. **Gender** : Sexe (Homme/Femme)
 3. **Age** : Âge du client.
 4. **Annual Income (k\$)** : Revenu annuel en milliers de dollars.
 5. **Spending Score (1-100)** : Indicateur des dépenses du client.
- Les données ont été inspectées pour identifier les valeurs manquantes et les types de variables.

Traitements effectués :

- Inspection des données pour détecter des valeurs manquantes : aucune absence de données.
- Résumé statistique (moyenne, médianes, etc.) pour identifier les tendances générales.
- Suppression des colonnes non pertinentes (Customer ID et Gender) pour une analyse focalisée sur des variables numériques.

2. Exploration des Données

- **Visualisation des données** :
 - Analyse des distributions des variables numériques stratifiées par sexe.
 - Diagrammes en secteurs montrant que 56% des clients sont des femmes.
 - Analyse des corrélations avec un graphique pairplot.

Observations :

- Les variables catégorielles comme le sexe et l'ID ne contribuent pas significativement à la segmentation.
- Les variables quantitatives restantes (**Annual Income**, **Spending Score**) sont jugées pertinentes pour la segmentation

3. Clustering avec l'algorithme DBSCAN:

- L'algorithme DBSCAN a été utilisé pour segmenter les clients en fonction de leur revenu annuel et de leur score de dépense.
- Le graphique des distances k-voisins a été utilisé pour déterminer une plage valeur optimale possible pour le paramètre **eps** de DBSCAN.
- **Le Silhouette Score** a été utilisé pour évaluer la qualité des clusters obtenus avec différentes valeurs de eps.

RÉSULTATS

L'analyse a révélé 5 groupes de clients distincts :

1. **Clients à faible revenu et forte propension à dépenser:** Ce groupe représente une opportunité pour des promotions et des offres attractives.
2. **Clients à revenu élevé et forte propension à dépenser:** Ce groupe est susceptible d'être intéressé par des produits haut de gamme et des expériences exclusives.
3. **Clients à revenu élevé et score de dépense modéré:** Ce groupe pourrait être ciblé avec des offres de fidélisation et des programmes de récompenses.
4. **Clients à revenu élevé et faible propension à dépenser:** Ce groupe nécessite une attention particulière pour comprendre les raisons de leur faible dépense et adapter les offres en conséquence.
5. **Clients à revenu moyen et propension moyenne à dépenser:** Ce groupe représente le segment de clientèle "moyen" et peut être ciblé avec des offres générales.

CONCLUSION

La segmentation des clients réalisée avec DBSCAN a permis d'identifier des groupes de clients distincts en fonction de leur comportement d'achat. Ces informations peuvent être utilisées pour :

- **Personnaliser les offres et les promotions.**
- **Développer des produits et services adaptés aux besoins de chaque groupe.**
- **Améliorer l'expérience client en proposant des communications et des interactions plus pertinentes.**
- **Optimiser les stratégies marketing et augmenter la rentabilité.**

RECOMMANDATIONS

- Il est recommandé de continuer à suivre l'évolution des clusters et d'adapter les stratégies marketing en conséquence.
- Il serait intéressant d'explorer d'autres algorithmes de clustering et de comparer les résultats obtenus.
- Des données supplémentaires, telles que les préférences d'achat ou les données démographiques, pourraient enrichir l'analyse et permettre une segmentation plus fine.