

Titre : Amélioration de la Gestion du Risque de Crédit par l'Utilisation du Machine Learning : Étude de Cas d'une Banque Internationale à Paris

Introduction

Dans le tumulte du monde financier moderne, la nécessité de personnaliser les services bancaires s'intensifie face à une clientèle diversifiée et exigeante. Confrontées à la dualité des attentes individuelles et à la prévention des risques de crédit, les banques modernes s'orientent vers la data science pour transformer les données en intelligence stratégique. Notre mission, aux confins de l'analyse de données et de la prévoyance, a pour objectifs de déchiffrer les mystères des défauts de paiement de cartes de crédit, d'anticiper les comportements à risque et de segmenter finement la clientèle pour une approche préventive et proactive.

Le Clustering : Définition et Principes

Le clustering, technique d'analyse de données non supervisée, permet de regrouper des clients en segments en fonction de leurs similitudes. Ce processus analytique, qui s'appuie sur des algorithmes avancés comme le K-means, organise les clients en groupes basés sur des similarités comportementales et financières. Il s'agit d'organiser le grand volume de données clients en sous-ensembles cohérents pour mieux comprendre et anticiper leurs comportements. Le clustering repose sur l'optimisation de la similarité intra-cluster et de la différence inter-cluster, mettant en lumière des caractéristiques communes souvent non évidentes. La force du clustering réside dans sa capacité à révéler des tendances et des patterns qui ne sont pas immédiatement évidents, fournissant ainsi un fondement pour des stratégies personnalisées.

Méthodologie :

Notre approche méthodologique débute par le nettoyage rigoureux et la normalisation des données financières pour neutraliser les écarts d'échelle entre les variables. Nous avons effectué un prétraitement sur les données pour les rendre appropriées pour le clustering. Nous avons procédé à la suppression de colonnes non pertinentes, le traitement des valeurs manquantes ou aberrantes, et la normalisation des caractéristiques.

Nous avons employé le K-means pour classer les clients en trois clusters, en nous basant sur les caractéristiques telles que l'historique de paiement et les montants de transactions. Parallèlement, un modèle prédictif de Random Forest a été calibré pour identifier les indicateurs précoces de défauts de paiement, offrant ainsi une possibilité de réaction avant que le risque ne devienne réalité.

➤ Phase 1: Exploration et Compréhension des Données

Outil utilisé: Google Colab

Utilisation de Google Colab pour charger, explorer et visualiser les données grâce aux bibliothèques intégrées comme Pandas pour la manipulation des données et Matplotlib/Seaborn pour la génération de graphiques exploratoires.

➤ Phase 2: Prétraitement des Données

Outil utilisé: Google Colab

Poursuite du travail dans Google Colab pour nettoyer les données, gérer les valeurs manquantes et standardiser les caractéristiques en utilisant StandardScaler de Scikit-learn.

➤ Phase 3: Modélisation Prédictive

Outil utilisé: Google Colab

Construction et entraînement d'un modèle de classification, spécifiquement le RandomForestClassifier de Scikit-learn, directement dans Google Colab.

Optimisation des hyperparamètres avec RandomizedSearchCV pour trouver les meilleurs paramètres pour le modèle au sein de Colab.

➤ Phase 4: Clustering et Segmentation des Clients

Outil utilisé: Google Colab

Nous avons effectué un prétraitement sur les données pour les rendre appropriées pour le clustering. Cela peut inclure la suppression de colonnes non pertinentes, le traitement des valeurs manquantes ou aberrantes, et la normalisation des caractéristiques si nécessaire.

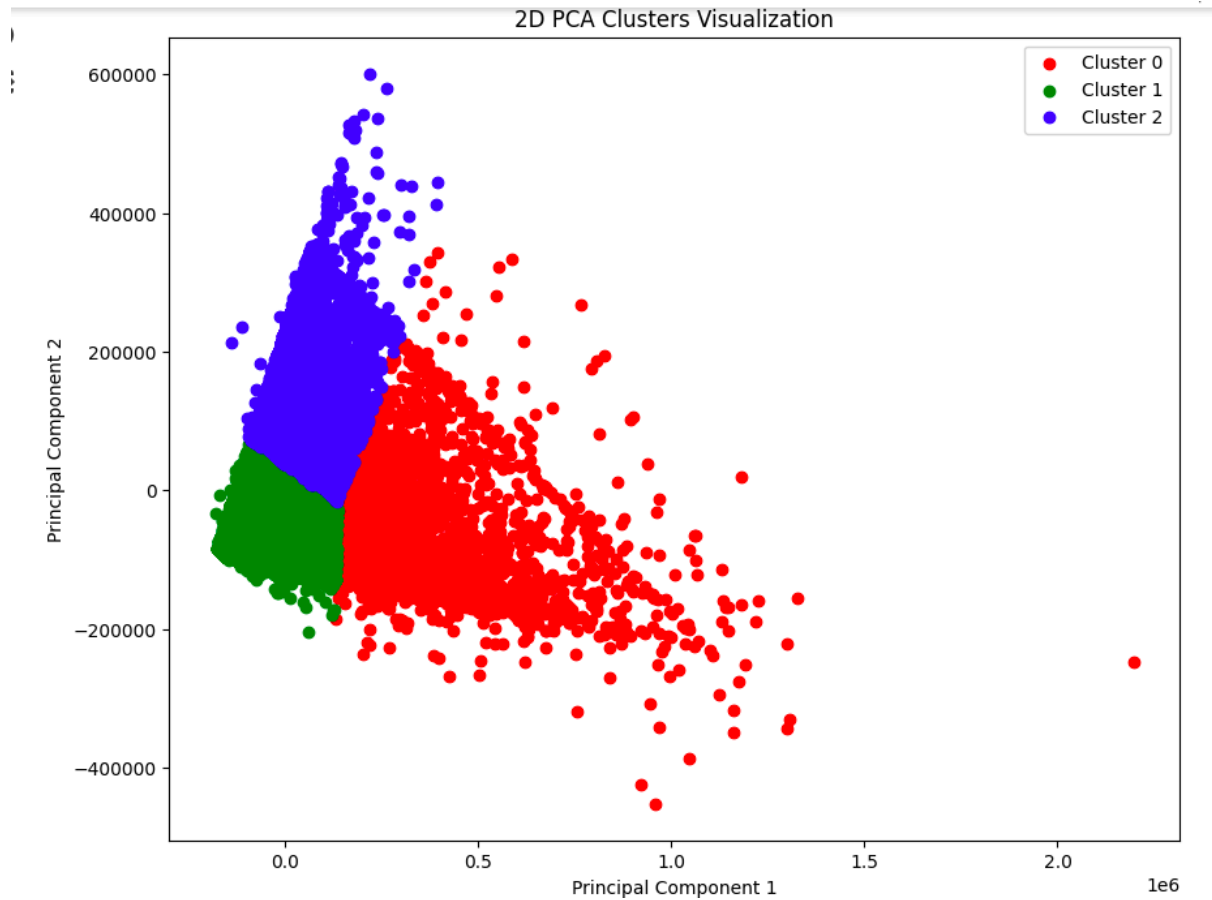
Application des techniques de clustering telles que K-Means directement dans Colab pour segmenter les clients en groupes significatifs basés sur leurs profils de consommation et autres caractéristiques.

➤ Phase 5: Développement d'Offres Commerciales Ciblées

Outil utilisé: Google Colab

Utilisation des insights tirés de l'analyse de clustering pour concevoir des stratégies de communication et des offres commerciales, tout cela orchestré depuis Google Colab en s'appuyant sur les résultats analytiques obtenus précédemment.

Le score silhouette moyen pour 3 clusters est de : 0.4495645349918882



Résultats de la Segmentation :

Segment	Description	Profil des Clients
Cluster 0 (Hauts Revenus)	Taille: 2951 clients	Limites de crédit moyennes: 272727.10 Montant moyen de la facture: 204302.79 Paiements moyens: 14309.08
Cluster 1 (Utilisateurs Prudents)	Taille: 14523 clients	Limites de crédit moyennes: 82944.43 Montant moyen de la facture: 34475.41 Paiements moyens: 3483.76

Cluster 2 (Jeunes Professionnels)	Taille: 6526 clients	Limites de crédit moyennes: 308585.66 Montant moyen de la facture: 19890.96 Paiements moyens: 6453.65
-----------------------------------	----------------------	---

Stratégie d'Offres Commerciales : Adaptation et Prévention

Segment	Offre Commerciale
Cluster 0	Programme Premium : Offre de services de gestion de patrimoine, conseils en investissement et taux d'intérêt préférentiels pour les prêts, pour fidéliser ce segment à haut revenu.

Cluster 1	Programme de fidélité : Bonus de cashback pour les paiements ponctuels, des taux d'intérêt réduits sur les prochaines échéances en cas de paiements anticipés, et des ateliers de gestion budgétaire pour encourager une utilisation prudente du crédit.
Cluster 2	Programme dynamique : Solutions de crédit flexible adaptées aux professionnels débutants, partenariats avec des réseaux professionnels pour offres spéciales, et points bonus pour des dépenses dans des catégories professionnelles pertinentes.

Conclusion :

L'intégration du clustering et des modèles prédictifs dans les opérations bancaires n'est pas seulement une réponse aux défis contemporains; c'est une marche vers un avenir où chaque client se sent unique et valorisé. Cette démarche stratégique n'offre pas seulement l'avantage de prévenir les risques de défaut, mais elle ouvre également la voie à des relations clientèles enrichies et à une fidélisation accrue. Les institutions financières qui adoptent cette approche basée sur les données sont celles qui se démarquent dans un paysage concurrentiel en perpétuelle évolution, où l'innovation et la personnalisation sont les clés du succès.