



Modélisation prédictive de l'attrition des consommateurs par apprentissage automatique dans la grande distribution à Kinshasa, République Démocratique du Congo

[Predictive modeling of consumer attrition through machine learning in the retail sector in Kinshasa, Democratic Republic of the Congo]

Erick Lumbala Tshipepele^{1,2}, Joel Ilunga Kabuya^{1,2} & Nathanael Mulenda Kasoro²

¹Centre de Recherche sur l'Enseignement de la Mathématique (CREM), Kinshasa, RD Congo

²Université de Kinshasa, Faculté des Sciences et Technologies, Kinshasa, RD Congo

Résumé

L'attrition des consommateurs est un phénomène qui demeure un enjeu stratégique majeur pour les décideurs dans pratiquement tous les secteurs d'activités. A Kinshasa (R.D. Congo), dans un contexte de concurrence accrue et de volatilité de la demande. La perte de clients impacte directement la rentabilité de l'entreprise vue les coûts d'acquisition, ce qui impose le passage d'une approche réactive à une approche prédictive de la gestion de la relation client. Cet article propose une méthodologie hybride d'analyse et de modélisation de ce phénomène d'attrition, fondée sur l'exploitation des données comportementales de clients d'un supermarché en RD Congo couvrant une période de quatre ans. Cette méthodologie est réalisée en deux phases : (i) phase analytique appliquant la technique analyse en composante principale (ACP) permettant la segmentation et l'identifier des profils homogènes en termes de valeur et de risque de churn, puis (ii) phase prédictive basé sur la classification supervisée mobilisant les techniques d'arbres de décision et de machine à vecteurs de support (SVM) pour la prédiction de l'attrition. Les modèles sont évalués à l'aide de matrices de confusion, d'indicateurs de classification et de courbes ROC-AUC issues des données du supermarché. Les résultats montrent que l'approche proposée permet de discriminer efficacement les clients à risque, les variables liées à l'intensité et à la régularité des achats étant particulièrement déterminantes. Cette étude met en évidence le potentiel de l'apprentissage automatique comme outil d'aide à la décision pour la mise en place de stratégies de fidélisation ciblées dans le secteur de la grande distribution en République Démocratique du Congo.


Mots-clés : Attrition des consommateurs ; marketing prédictif ; apprentissage automatique ; arbre de décision ; SVM.

Abstract

Consumer attrition remains a major strategic challenge for decision-makers across virtually all sectors. In Kinshasa (Democratic Republic of Congo), within a context of heightened competition and volatile demand, customer loss directly impacts company profitability due to acquisition costs, thereby necessitating a shift from a reactive to a predictive approach in customer relationship management. This article proposes a hybrid methodology for analyzing and modeling consumer attrition, based on behavioral data from supermarket customers in the DRC over a four-year period. The methodology is carried out in two phases: (i) an analytical phase applying Principal Component Analysis (PCA) to enable segmentation and identification of homogeneous profiles in terms of value and churn risk, and (ii) a predictive phase based on supervised classification techniques, specifically decision trees and Support Vector Machines (SVM), to forecast attrition. Models are evaluated using confusion matrices, classification indicators, and ROC-AUC curves derived from supermarket data. The results demonstrate that the proposed approach effectively discriminates at-risk customers, with variables related to purchase intensity and regularity proving particularly decisive. This study highlights the potential of machine learning as a decision-support tool for implementing targeted customer retention strategies in the retail sector in the Democratic Republic of Congo.

Keywords: Consumer attrition; predictive marketing; machine learning; decision tree; SVM forestière.

*Auteur correspondant: Erick Lumbala Tshipepele, (erick.tshipepele@unikin.ac.cd). Tél. : (+243) 818559879

 : <https://orcid.org/0009-0005-7300-2167>; Reçu le 18/03/2026 ; Révisé le 13/04/2026 ; Accepté le 05/05/2026

DOI: <https://doi.org/10.59228/rcst.026.v5.i2.272>

Copyright: ©2026 Tshipepele et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC-BY-NC-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Introduction

Les consommateurs constituent des acteurs indispensables pour l'entreprise, puisqu'ils représentent la source principale qui assure la pérennité et la rentabilité de son activité c'est-à-dire plus ils consomment, plus ils assurent la survie de l'entreprise (Kotler & Keller, 2016).

Cependant, celles-ci (les entreprises) qui s'appuient sur leurs consommateurs sont confrontées à ce défi majeur : la perte de clients (attrition ou churn), qu'il s'agisse d'un désengagement progressif ou d'un départ définitif. En RD Congo, le marché agroalimentaire et de la distribution se caractérise par une multiplication des acteurs, une diversification de l'offre : la concurrence s'intensifie et les consommateurs plus informés et deviennent de plus en plus volatils

Ce phénomène préoccupant qui se traduit par une baisse directe du chiffre d'affaires et par une augmentation des investissements nécessaires pour attirer de nouveaux clients, porte des conséquences directes sur la rentabilité, la croissance et l'image de marque. Le phénomène de l'attrition demeure au cœur des analyses et des débats stratégiques au sein des instances décisionnelles (Djerroud et al., 2016).

La littérature souligne que le coût d'acquisition d'un nouveau client est généralement supérieur au coût de la fidélisation d'un client existant, ce qui plaide pour une stratégie orientée vers la rétention et la valeur à long terme (Reichheld & Sasser, 1990). Dans cette perspective, le marketing prédictif, fondé sur l'exploitation des données massives (Big Data) et des techniques d'apprentissage automatique, permet d'anticiper les comportements futurs plutôt que de les subir. Appliqué à l'attrition, il s'agit d'estimer pour chaque client la probabilité de départ et de cibler les actions de fidélisation sur les individus à risque élevé (Verbeke et al., 2012).

Dans le contexte africain, et congolais en particulier, les études empiriques mobilisant l'apprentissage automatique pour la prédiction du churn en grande distribution demeurent rares, alors même que les enjeux de fidélisation sont importants. Le présent article vise à combler partiellement cette lacune en répondant aux questions suivantes : (i) quels profils de clients présentent une probabilité élevée d'attrition ou de fidélisation ? (ii) quels facteurs explicatifs sont les plus déterminants pour prédire ces comportements et orienter les stratégies de rétention ?

Pour y répondre, nous développons une approche hybride combinant une segmentation non supervisée

avec la méthode d'Analyse à Composante Principales (ACP) et des modèles supervisés avec les techniques des arbres de décision et Support Vector Machine). Ces techniques sont appliquées sur les données comportementales des consommateurs, couvrant deux années d'activités.

2. Matériel et méthodes

2.1. Présentation des données et environnement de travail

Pour appuyer notre recherche, nous avons extrait les données comportementales depuis la base de données client de l'un des grands supermarchés de la ville de Kinshasa dont le nom ne sera pas cité pour des raisons de confidentialité.

Ces données ont été extraites depuis la base de données SQL Serveur via deux fichiers CVS.

L'échantillon des données se présentent comme suit :

Tableau I. Répartition de Données initiales

Jeu de Donnée	Totale observations	Total caractéristiques
Client	4248	4
Ticket	246148	16

Ces données comportementales, comme dit précédemment, couvrent une période de deux années d'activités.

En ce qui concerne l'environnement de travail et d'analyse, nous nous avons opté pour le langage Python, celui-ci est un langage de programmation orienté objet, à la fois puissant, flexible et interactif. Il offre une grande richesse fonctionnelle grâce à ses modules, ses exceptions, sa syntaxe dynamique, ses classes et ses types de données évolués. Souvent comparé à des langages comme Perl, Ruby ou Java, Python se distingue par sa simplicité d'utilisation.

2.2. Techniques utilisées

2.2.1. Approche analytique : Méthode ACP

Cette première approche de notre méthodologie consiste à examiner les données historiques pour comprendre les tendances passées et expliquer les phénomènes observés.

L'Analyse en Composantes Principales (ACP) est une méthode statistique multivariée utilisée principalement pour réduire la dimensionnalité d'un jeu de données tout en conservant l'essentiel de l'information. Elle permet de transformer un ensemble de variables corrélées en un nouveau système de coordonnées orthogonales, appelées composantes principales, ordonnées selon la variance qu'elles expliquent (Jolliffe, 2002).

Son principe de fonctionnement passe par :

- Le centrage des données sur la moyenne

$$X_{\text{centrée}} = X - \mu \quad (1)$$

Où :

$X_{\text{centrée}}$ est l'ensemble de données centrées sur zéro

X est l'ensemble de données originales de dimension m observations et n variables

μ est le vecteur de la moyenne des données

- Le calcul de la matrice C de covariance

$$C = \frac{1}{m-1} X_{\text{centrée}}^T X_{\text{centrée}} \quad (2)$$

Avec C comme la matrice de covariance de dimension $n \times n$; et $X_{\text{centrée}}^T X_{\text{centrée}}$ est le produit de la matrice de données centrée sur la moyenne et de sa transposée.

- Le calcul des valeurs propres et de vecteurs propres

$$CV = \lambda V \quad (3)$$

Où :

V est un vecteur propre (la direction de la composante principale)

λ est la valeur propre correspondante (l'ampleur de la variance le long de cette composante)

Plus la valeur propre est importante, plus la variance est capturée par le vecteur propre correspondant.

- Sélection des composantes principales

Les valeurs propres sont triées par ordre décroissant. Les vecteurs propres associés aux plus grandes valeurs propres représentent les directions qui capturent le plus de variance dans les données. On choisit alors un sous-ensemble de ces vecteurs propres pour former une matrice de transformation P qui réduira la dimension des données.

- Projections des données sur les nouvelles composantes principales

Les données sont projetées dans l'espace des composantes principales en multipliant les données centrées par la matrice P des vecteurs propres sélectionnés :

$$X_{\text{projeté}} = X_{\text{centré}} P \quad (4)$$

Son application aux données comportementales a permis de réduire la dimension de données issues de la modélisation du comportement de client, qui comporte une matrice creuse afin de capturer beaucoup plus d'information dans les différents composants obtenu permettant de palier ainsi à ce problème de la matrice creuse.

2.2.2. Approche prédictive : Méthode des arbres de décision et Support Vector Machine

L'approche prédictive constitue la seconde composante de notre méthodologie. Elle vise à modéliser le comportement futur des clients.

a) Méthode des arbres de décision

L'arbre de décision est un modèle d'apprentissage supervisé utilisé pour des tâches de classification ou de régression. Il repose sur une structure arborescente où chaque nœud représente une condition sur une variable, et chaque branche correspond à une décision. L'objectif est de diviser l'espace des données en régions homogènes, en minimisant l'impureté à chaque étape de division (Breiman et al., 1984).

De manière générale, le fonctionnement de la construction d'un arbre de décision suit le processus suivant :

Choisir la meilleure caractéristique à diviser

Le choix de la meilleure caractéristique passe par le calcul de: l'entropie, l'indice de Gini, ou la variance pour chaque caractéristique.

- Calcul de l'entropie

Lorsqu'on divise les données en fonction d'une caractéristique X , le gain d'information (ou réduction de l'entropie) est défini comme la différence entre l'entropie avant et après la division:

$$\text{Gain}(X) = H(Y) - \sum_{v \in \text{valeurs}(X)} \frac{|D_v|}{|D|} H(Y|X = v) \quad (5)$$

Où D est l'ensemble des données et D_v est ensemble des données ayant la valeur v pour la caractéristique X .

- Calcul de l'indice de Gini

Le critère de Gini est une autre mesure de la pureté, utilisée couramment pour les arbres de décision. Pour une variable Y , l'indice de Gini est défini comme:

$$\text{Gini}(Y) = 1 - \sum_{i=1}^k p_i^2 \quad (6)$$

L'indice de Gini est faible (proche de 0) lorsque toutes les observations sont de la même classe, et élevé (proche de 0.5) lorsque les observations sont uniformément réparties entre toutes les classes.

Il y a lieu de signaler que l'approche que nous présentons dans ce travail est basée sur le problème de classification, c'est pourquoi nous avons jugé bon de ne pas présenter le critère de réduction de l'impureté qui essentiellement intervient dans un problème de régression.

Diviser l'ensemble de données :

À chaque nœud de l'arbre, on choisit la caractéristique X et la valeur v qui maximisent le gain d'information (en classification) ou minimisent la

réduction de la variance (en régression). Cela revient à trouver le critère de division qui conduit à la plus grande homogénéité possible des sous-ensembles

Répéter récursivement :

Appliquer le processus de division sur chaque sous-ensemble de données, en continuant à diviser jusqu'à ce qu'un critère d'arrêt soit atteint (par exemple, quand un nœud devient pur, ou qu'une profondeur maximale de l'arbre est atteinte).

Prédiction :

Pour prédire une nouvelle observation, on la fait passer dans l'arbre et on suit les décisions prises à chaque nœud jusqu'à ce qu'une feuille soit atteinte. La prédiction correspond à la classe majoritaire (en classification) ou à la valeur moyenne (en régression) des exemples de la feuille.

Dans cette phase, le jeu de données transformé par l'Analyse en Composantes Principales (ACP) est utilisé comme base d'entraînement pour le modèle d'arbre de décision. Cette démarche permet de combiner la réduction de dimensionnalité avec une méthode de classification supervisée, afin de prédire la fidélité des clients à partir de variables synthétiques mais informatives. Ainsi, l'entraînement du nouveau jeu de données obtenu à partir de l'analyse en composantes principales dans l'arbre de décision donne le résultat illustré dans la figure suivante:

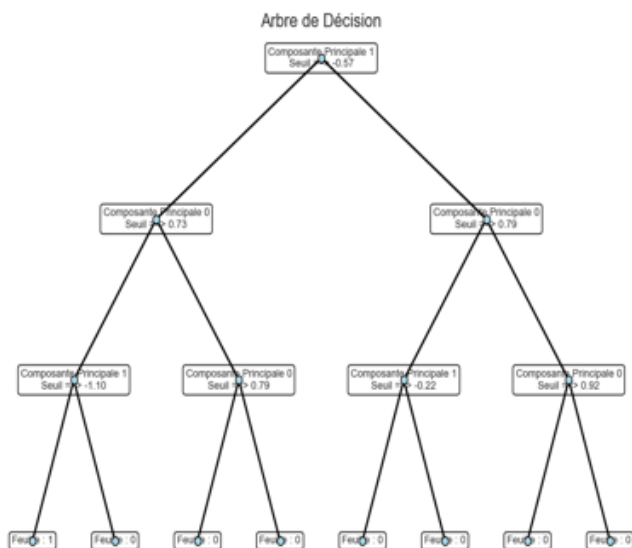


Figure 1 : Arbre de décision

b) Méthodes de Support Vector Machine

Les machines à vecteurs de support (SVM) sont des modèles d'apprentissage supervisé utilisés pour la classification (et la régression dans le cas des SVR, Support Vector Regression). L'idée principale derrière les SVM est de trouver un hyperplan qui sépare les

différentes classes de données avec la plus grande marge possible, tout en minimisant les erreurs de classification (Cortes & Vapnik, 1995).

Le fonctionnement de la méthode de SVM suit les étapes suivantes :

- Recherche de l'hyperplan optimal

L'objectif des SVM est de trouver un hyperplan H qui sépare les données en différentes classes, tout en maximisant la distance entre les points les plus proches de chaque classe et l'hyperplan (ces points sont appelés les vecteurs de support).

L'hyperplan dans un espace d-dimensionnel est défini par une équation linéaire de la forme:

$$w * x + b = 0 \quad (7)$$

Où: w est un vecteur normal à l'hyperplan, b est un biais qui détermine le décalage de l'hyperplan par rapport à l'origine, x est un point quelconque dans l'espace des caractéristiques.

- Maximisation de la marge

La marge est la distance entre l'hyperplan et les points les plus proches de chaque classe. Pour une classification binaire, les SVM cherchent à maximiser cette marge. L'idée est de trouver l'hyperplan optimal qui maximise la distance entre les vecteurs de support des deux classes, tout en séparant correctement les deux classes.

La distance de tout point x à l'hyperplan est donnée par la formule:

$$Distance = \frac{|w*x+b|}{\|w\|} \quad (8)$$

Les vecteurs de support sont les points pour lesquels cette distance est minimale, c'est-à-dire les points les plus proches de l'hyperplan, mais qui ne se trouvent pas de part et d'autre de la frontière.

- Problème d'optimisation

L'algorithme SVM consiste à résoudre le problème d'optimisation suivant:

$$\underset{w,b}{\text{minimiser}} \frac{1}{2} \|w\|^2 \quad (9)$$

sous la contrainte que tous les points de chaque classe soient séparés par une marge d'au moins 1, ce qui se traduit par :

$$y_i(w * x_i + b) \geq 1 \quad \forall i \quad (10)$$

Où $y_i \in \{-1,1\}$ est la classe du point X_i . Cette contrainte assure que tous les points de la classe 1 sont à droite de l'hyperplan et tous les points de la classe -1 sont à gauche.

Le terme $\frac{1}{2} \|w\|^2$ est l'objectif de minimisation, car il est équivalent à maximiser la marge $\frac{1}{\|w\|}$. Cela se fait par une approche d'optimisation quadratique.

- Support Vectors

Ce sont les points d'entraînement qui se trouvent sur les bords de la marge. Ce sont eux qui déterminent l'hyperplan optimal, et non les autres points de l'ensemble de données. Ces vecteurs de support sont cruciaux pour définir la solution du SVM.

Il y a lieu de signaler qu'il existe une autre manière d'aborder l'optimisation dans SVM, et c'est lorsque l'on fait face à un problème non linéaire, dont il s'agira de trouver le noyau (Kernel Trick en anglais), au lieu de chercher un hyperplan linéaire.

L'application de SVM au jeu de données issu de l'analyse en composantes principales présente le résultat final suivant:

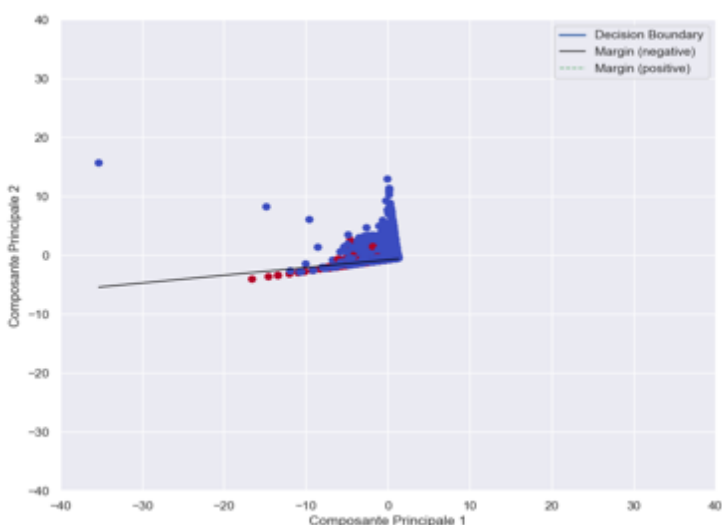


Figure 2 : Frontière de séparation de données finale

3. Résultats

Il nous est important de rappeler notre contribution réside sur un modèle de rétention des consommateurs basée sur la combinaison de deux approches d'apprentissage automatique notamment apprentissage supervisé et non supervisé, notre objectif est d'évaluer et de prédire le meilleur profile client susceptible de fidéliser. Dans ce qui suit, nous présentons les résultats de manière résumée de toutes nos expériences.

3.1. Evaluation des méthodes utilisées

La validation des modèles probabilistes utilisés et leurs résultats est basé sur les métriques d'évaluations suivants dont leurs performances s'avèrent être promettantes et stables dans le temps. Nous avons utilisé les algorithmes d'apprentissage automatique dans le point precedent particulièrement dans la construction des modèles prédictifs de rétention client. Ces métriques sont présentées et décrites de manière suivante:

3.1.1. Matrice de Confusion

La matrice de confusion est une table qui permet de visualiser la performance d'un modèle de classification en comparant les prédictions du modèle avec les vraies étiquettes. Elle donne une idée claire de combien de fois le modèle a fait des erreurs et de quel type elles sont (IBM, 2025).

a) Arbre de décision



Figure 3: Matrice de confusion de l'arbre de décision

b) Support Vector Machine



Figure 4. Matrice de confusion de SVM

3.1.2. Courbe ROC (Receiver Operating Characteristic)

a. Arbre de décision

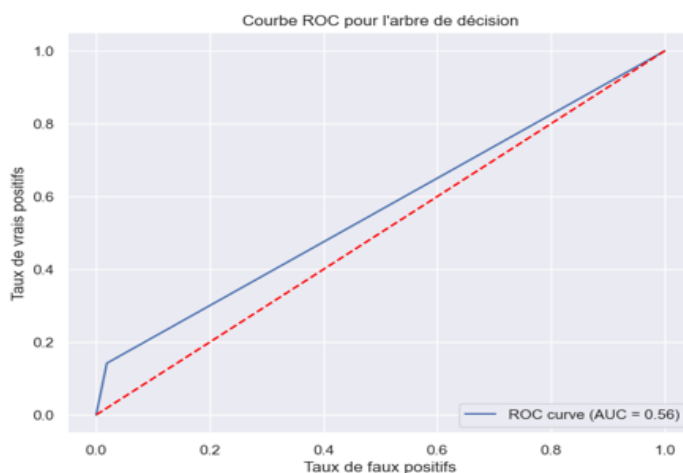


Figure 5 : Courbe ROC AUC pour l'Arbre de Décision

3.1.3. Rapport de classification

La classification report est une synthèse des principales métriques de performance pour un modèle de classification. Il est généralement présenté sous forme de

Tableau et inclut :

- **Précision (Precision)** : La capacité à ne pas classifier à tort des éléments comme positifs.
- **Rappel (Recall)** : La capacité à identifier tous les vrais positifs.
- **F1-score** : La moyenne harmonique de la précision et du rappel, qui équilibre ces deux métriques.

Support : Le nombre d'occurrences de chaque classe dans le jeu de données réel.

Tableau II : Rapport de Classification pour Arbre de décision

	Precision	Recall	F1-Score	Support
0 (Perdu)	0.88	1.00	0.94	1119
1 (Fidèle)	0.00	0.00	0.00	155
Accuracy			0.88	1274
Macro avg	0.44	0.50	0.47	1274
Weighted avg	0.77	0.88	0.82	1274

Tableau III. Rapport de Classification pour SVM

	Precision	Recall	F1-Score	Support
0 (Perdu)	0.88	1.00	0.94	1119
1 (Fidèle)	0.00	0.00	0.00	155
Accuracy			0.88	1274
Macro avg	0.44	0.50	0.47	1274
Weighted avg	0.77	0.88	0.82	1274

3.2. Présentation des résultats

Nous avons étudié le potentiel des modèles susmentionnés dans la classification binaire de la rétention client dans le marketing prédictif. L'objectif de cette classification est de construire un modèle capable de prédire si un parcours client mène à la rétention ou non de ce dernier.

Comme expliqué dans la section modélisation, nous disposons de la distribution statistique des données dans le tableau suivant :

Tableau IV. Répartition statistique des classes

Classe	Totale	Proportion (%)
1 (Fidèle)	515	12.1
0 (Non Fidèle)	3729	87.9

Pour rappel, l'objectif de ces expériences est de répondre aux questions de recherche posées ci-dessus.

A ce stade, peut-on dire que les modèles mis en œuvre sont capables de prédire la rétention d'un parcours client facilitant un marketing prédictif efficace ?

Le tableau suivant, présente les modèles et les métriques associés pour l'évaluation et la validation :

Tableau V. Évaluation de la performance du Modèle

Modèle	Accuracy (Précision)
Arbre de Décision	87,91 %
SVM (Support Vector Machine)	87.83 %

4. Discussion

Les résultats obtenus à travers les modèles d'apprentissage supervisé et non supervisé montrent une performance globalement satisfaisante dans la prédiction de la rétention client. Les taux de précision avoisinant les 88 % pour les deux modèles (Arbre de Décision et SVM) indiquent une bonne capacité à distinguer les clients fidèles des non fidèles. Toutefois, une analyse plus fine des rapports de classification révèle un déséquilibre important entre les classes, avec une faible capacité à identifier correctement les clients fidèles (classe minoritaire), notamment dans le cas du SVM où le rappel est nul.

Ce phénomène peut être attribué à la distribution déséquilibrée des données, où les clients non fidèles représentent près de 88 % de l'échantillon. Cela suggère que les modèles ont tendance à privilégier la classe majoritaire, ce qui limite leur efficacité dans une logique de fidélisation ciblée. Pour pallier cette limite, des techniques comme le rééchantillonnage (SMOTE), l'ajustement des seuils de décision, ou l'utilisation de métriques pondérées pourraient être envisagées.

Par ailleurs, l'intégration d'approches non supervisées dans la phase exploratoire a permis de mieux comprendre les profils comportementaux, ouvrant la voie à une segmentation plus fine et à des stratégies marketing personnalisées. À l'avenir, l'enrichissement du modèle par des données temporelles, des variables contextuelles ou des indicateurs d'engagement pourrait améliorer la précision des prédictions.

Enfin, ces résultats constituent une base solide pour le déploiement opérationnel d'un système de marketing prédictif, capable d'anticiper les risques

d'attrition et de proposer des actions ciblées en temps réel. L'approche hybride adoptée dans ce travail démontre la pertinence de combiner plusieurs types d'apprentissage pour répondre à des problématiques complexes de fidélisation client.

5. Conclusion

Ce travail s'inscrit dans la continuité des recherches sur la fidélisation client, en proposant une approche innovante fondée sur la combinaison de deux paradigmes d'apprentissage automatique : supervisé et non supervisé. En réponse aux limites des méthodes traditionnelles, nous avons développé un modèle hybride de rétention, intégrant à la fois des techniques de segmentation intelligente (ACP) et de prédiction comportementale (Arbre de décision et SVM).

Cette double approche permet non seulement d'identifier des profils clients pertinents, mais aussi de prédire leur probabilité de fidélisation avec une précision notable.

Les résultats obtenus confirment la robustesse et la stabilité du modèle, malgré la complexité et l'hétérogénéité des données du marché. L'originalité de notre contribution réside dans la capacité à croiser interprétabilité et performance, tout en offrant aux entreprises un outil exploitable pour anticiper les comportements et optimiser leurs stratégies marketing.

Ainsi, ce travail ouvre des perspectives concrètes pour le déploiement de solutions prédictives dans la gestion de la relation client, tout en posant les bases pour des recherches futures sur l'intégration de données temporelles, contextuelles et comportementales dans les modèles de fidélisation.

Remerciements

Les auteurs remercient M. Serge Matadi, IT Manager et son équipe avec qui nous avons travaillé particulièrement pour l'extraction et le traitement des données clients pour la réalisation de cette étude.

Financement

Cette recherche a été entièrement sur fond propre, couvrant toutes les activités de terrain, la collecte des données et les analyses statistiques nécessaires à la réalisation de l'étude.

Conflit d'Intérêt

Les auteurs déclarent ne présenter aucun conflit d'intérêt, qu'il soit financier, professionnel ou personnel, susceptible d'influencer la conduite, l'interprétation ou la publication des résultats. Toutes les informations rapportées reposent sur des données collectées et analysées de manière indépendante,

conformément aux standards éthiques de la recherche scientifique.

Considérations Ethiques

Cette étude a été menée dans le respect strict des principes éthiques de la recherche scientifique, incluant :

- L'intégrité et la rigueur scientifique tout au long de l'étude ;
- La confidentialité des informations recueillies auprès des participants ;
- L'obtention du consentement éclairé avant la collecte des données ;
- La transparence dans l'utilisation des sources et du matériel développé en collaboration, avec référence appropriée.

Les auteurs confirment que le travail présenté est issu de leur propre recherche et que les résultats peuvent servir de référence pour des travaux ultérieurs sous réserve d'une citation correcte

Contribution des Auteurs

T.L.E : Conception et supervision de l'étude, Analyse et interprétation des résultats, rédaction du manuscrit principal et validation finale.

K.I.J : Analyse et interprétation des résultats, Révision du questionnaire, interprétation des résultats, lecture critique du manuscrit et validation finale.

K.M.N : Validation des données, contribution à la discussion et approbation finale validation

ORCID des Auteurs

Tshipepele L.E : <https://orcid.org/0009-0005-7300-2167>

Kabuya I.J : <https://orcid.org/0009-0003-3356-9973>

Kasoro A.Y: <https://orcid.org/0009-0004-2228-6617>

Références bibliographiques

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.
- Cortes, C., & Vapnik, V. (1995). Djerroud, S., & Ferchouli, H. (2016). IBM. Qu'est-ce qu'une matrice de confusion? IBM Research, 2025. Disponible en ligne: IBM Knowledge
- Jolliffe, I. T. (2002). Kotler, P., & Keller, K. L. (2016). *L'impact de l'image de marque sur le comportement des consommateurs: Cas Activia de Danone*. [Mémoire de Master, Université Abderrahmane Mira de Béjaïa].
- Marketing management (15th ed.). Pearson Education.

-
- Principal Component Analysis (2nd ed.). Springer.
- Reichheld, F. F., & Sasser, W. E. (1990). Social network analysis for customer attrition prediction. *Applied Soft Computing*, 12(4), 190–197.
- Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Verbeke, W., Martens, D., & Baesens, B. (2012). Zero defections: Quality comes to services. *Harvard Business Review*, 68(5), 105–111.