



## Mémoire de fin d'Etudes

Thème :

*Analyse de la rentabilité d'un portefeuille client par la Customer Lifetime Value au sein de la banque CPA.*

Présenté et soutenu

M. Saber SETTI

Encadré par

M. Neji BOUSLAMA

Etudiant parrainé par :

Crédit populaire d'Algérie

## Dédicaces

Je dédie ce travail à mes chers parents, dont l'amour, les sacrifices et le soutien indéfectible m'ont été le plus précieux des dons. Que Dieu les bénisse et les préserve. À mes frères et sœurs, pour leur encouragement et leur foi en moi. À tous ceux qui ont contribué de près ou de loin à ce voyage, votre aide a été un cadeau inestimable. À mes amis, pour leur compagnie et leur soutien constant, je vous suis profondément reconnaissant.

## **Résumé :**

Ce travail de recherche intitulé "Analyse de la rentabilité d'un portefeuille client par la Customer Lifetime Value au sein de la banque CPA" traite de l'importance croissante pour les banques et les institutions financières d'adopter une approche centrée sur le client. Le Crédit Populaire d'Algérie (CPA) a exprimé le besoin d'évaluer la rentabilité de ses clients, un domaine souvent négligé dans les agences bancaires où le Produit Net Bancaire (PNB) par client ou par portefeuille n'est généralement pas calculé. Le mémoire s'inspire des modèles d'analyse de survie, comme le DeepSurv, pour examiner la Customer Lifetime Value (CLV) comme outil stratégique pour la prise de décision en banque. Le travail est divisé en deux chapitres principaux : le premier se concentre sur l'état de l'art et les concepts clés tels que la rentabilité bancaire et la CLV, tandis que le second aborde la modélisation pratique de la CLV à partir des données de la banque. Les modèles de survie étudiés, tels que le modèle PH Cox et le modèle DeepSurv, ne calculent pas directement la CLV, mais estiment plutôt la probabilité qu'un client continue à faire affaire avec la banque. Le mémoire combine recherche documentaire, analyse de données, et entretiens avec des responsables de la CPA pour offrir un aperçu détaillé de l'utilisation de la CLV dans la gestion du portefeuille de clients de la banque.

**Mots clés :** Customer Lifetime Value, rentabilité client, Produit Net Bancaire, modèles d'analyse de survie, DeepSurv, CLV, modèle PH Cox, fidélisation client, stratégie bancaire, portefeuille client, Crédit Populaire d'Algérie.

## **Abstract :**

This research paper, titled "Analysis of Client Portfolio Profitability by Customer Lifetime Value within CPA Bank," addresses the growing importance for banks and financial institutions to adopt a client-centric approach. The Crédit Populaire d'Algérie (CPA) has identified a need to evaluate the profitability of its clients, an area often overlooked in bank branches where the Net Banking Income (NBI) per client or portfolio is not usually calculated. The thesis draws on survival analysis models, such as DeepSurv, to examine Customer Lifetime Value (CLV) as a strategic decision-making tool in banking. The work is divided into two main chapters: the first focuses on the state of the art and key concepts such as banking profitability and CLV, while the second addresses the practical modeling of CLV using the bank's data. The survival models studied, such as the PH Cox model and the DeepSurv model, do not directly calculate CLV, but rather estimate the probability that a client will continue to do business with the bank. The thesis combines documentary research, data analysis, and interviews with CPA officials to provide a detailed overview of the use of CLV in client portfolio management at the bank.

**Keywords :** Customer Lifetime Value, client profitability, Net Banking Income, survival analysis models, DeepSurv, CLV, PH Cox model, client retention, banking strategy, client portfolio, Crédit Populaire d'Algérie.

## **Remerciement**

Je tiens à exprimer ma profonde gratitude à Monsieur Neji BOUSLAMA pour son accompagnement et ses précieux conseils, qui ont été d'une aide inestimable tout au long de la rédaction de ce mémoire. Mes remerciements s'étendent également à Madame Hassam, divisionnaire RH, dont l'assistance et les orientations durant mon stage ont été déterminantes. Je suis également reconnaissant envers Madame Doumer mon tuteur de stage, pour son soutien et ses enseignements, ainsi qu'à Madame Radia pour sa collaboration chaleureuse et son aide précieuse.

Je ne saurais oublier tous les collaborateurs du Crédit Populaire d'Algérie avec qui j'ai eu le plaisir d'interagir durant mon stage. Leurs partages d'expériences et leur esprit d'équipe ont grandement enrichi mon immersion professionnelle.

Je souhaite aussi adresser une pensée particulière à toute l'équipe de l'IFID pour leur encadrement et leur dévouement tout au long de mon cursus académique.

Enfin, mes remerciements les plus sincères vont aux membres honorables du jury qui ont accepté d'examiner ce travail. Leur présence et leurs évaluations constituent pour moi un honneur et une source d'apprentissage exceptionnelle.

# Sommaire

<b>Introduction Générale :</b> .....	1
<b>Chapitre premier : La rentabilité client et la Customer life Time value.</b> .....	4
<b>Introduction</b> .....	4
Section 01 : Définition de la rentabilité client et des indicateurs pertinents pour l'évaluer .....	5
Section 02 : La customer Life time value définition et concepts .....	18
Section 03 : Indicateurs et modèles de mesure de la CLV dans la littérature : .....	24
<b>Chapitre deuxième : Proposition d'un modèle de calcul de la CLV pour la banque CPA</b> .....	33
<b>Introduction :</b> .....	34
Section 01 : Méthodologie .....	36
Section 02 : Résultats .....	45
Section 03 : Discussion et recommandations .....	65
Conclusion .....	67
<b>Conclusion générale</b> .....	68

## Liste des abréviations

ABC	Activity based costing
ANN	Artificiel Neural Network
ARC	Approche rentabilité client
C-INDEX	Concordance index
CPA	Crédit populaire d'Algérie
CLV	Customer life time value
CRM	Customer Relationship management
PNB	Produit net bancaire
PH	Proportionnel hazard
AUC	Aire Sous la Courbe
PSA	Probabilité de survie actualisée
KM	keplan meiler
TCI	Taux de cession interne
OPCVM	Organisme de Placement Collectif en Valeurs Mobilières
DF	Data frame
NBD	Negative Binomial Distribution
RNA	Réseaux de neurones artificiels
PLV	Publicité sur le Lieu de Vente.
RFM	Récence, Fréquence, Montant
ROA	Return on Assets
ROE	Return on Equity
SEB	SKANDINAVISKA Enskilda Banken
WACC	Weighted Average Cost of Capital

## Liste des figures

<b>Figure 1 :</b> Gestion de la rentabilité .....	6
<b>Figure 2 :</b> Les composantes du PNB .....	7
<b>Figure 4 :</b> Processus de la méthode ABC .....	16
<b>Figure 5 :</b> Architecture d'un réseau de neurones artificiels.....	30
<b>Figure 6 :</b> Boîtes à moustaches de l'Age par le Genre .....	39
<b>Figure 7:</b> Matrice de corrélation.....	40
<b>Figure 8 :</b> La courbe ROC du modèle de PH COX.....	46
<b>Figure 9:</b> La courbe ROC du modèle DeepSurv .....	47
<b>Figure 10 :</b> Exemples de courbes de survie selon les deux modèles .....	48
<b>Figure 11 :</b> Résidus de Schoenfeld pour la variable « salaire Estime ».....	53
<b>Figure 13 :</b> Fonction de risque de base $h_0(t)$ .....	56
<b>Figure 12 :</b> Importance des variables sur le risque de désabonnement.....	58

## Liste des tableaux

<b>Tableau 1</b> : Calcul Des Marges Sur Capitaux .....	9
<b>Tableau 2</b> : Test d'hypothèse des risques proportionnels .....	51
<b>Tableau 3</b> : Tests de proportionnalité des risques basés sur les résidus de Schoenfeld .....	51
<b>Tableau 4</b> : Résultats du modèle Ph-COX .....	54
<b>Tableau 5</b> : Calcul de la CLV pour le client A.....	63
<b>Tableau 6</b> : Calcul de la CLV pour le client B .....	63



## **Introduction Générale :**

Dans un environnement ultra compétitif boosté par les évolutions technologiques, les banques et les établissements financiers se trouvent dans l'obligation d'adopter une approche centrée client « Customer centric ». Permettant ainsi de mieux répondre aux attentes des clients, qui constituent une part importante dans le profit des banques de détails.

L'idée de ce sujet est née lorsque l'institution de parrainage le crédit populaire d'Algérie « CPA » a exprimé un besoin d'une étude sur « l'analyse de la rentabilité d'un portefeuille client ». Une telle étude permettra d'avoir une estimation de la rentabilité de chacun des clients constituant le portefeuille d'un chargé de clientèle. Dans de nombreuses agences, la rentabilité par client est complètement inconnue, tant de la part des directeurs d'agence que des chargés de clientèle. Le montant du PNB par client ou par portefeuille n'est pas calculé ni communiqué aux directeurs des relations-clients. En conséquence, ils doivent utiliser d'autres indicateurs pour évaluer la rentabilité globale de leur portefeuille.

Cette étude aborde la valeur à vie du client (CLV) dans le secteur bancaire, s'inspirant des travaux de L. Katzman 2018 et ses collaborateurs sur DeepSurv, un modèle d'analyse de survie éprouvé en médecine. Parallèlement, on observe l'utilisation de ces modèles pour la fidélisation dans le service des télécommunications. En transposant ces concepts au domaine bancaire, nous cherchons à mesurer la CLV, reflétant non seulement la performance financière mais aussi la pérennité des relations clients-banque.

De ce fait, l'objectif principal de cette étude est d'examiner comment la CLV peut être utilisée comme un outil stratégique pour la prise de décision dans le domaine bancaire, en se focalisant sur le cas de la banque CPA. De cela découle notre problématique de recherche : **Comment la CLV peut-elle être analysée et utilisée pour évaluer la rentabilité d'un portefeuille client ?**

L'utilisation de la CLV dans le secteur bancaire algérien est une innovation marquante, permettant une allocation des ressources plus efficace et une meilleure segmentation de la clientèle. Elle offre également la possibilité d'élaborer des stratégies de communication et de promotion ciblées, renforçant la fidélisation et l'engagement des clients. En intégrant cette approche, le CPA s'aligne sur les banques internationales leaders qui exploitent déjà la CLV

pour renforcer leur compétitivité sur le marché mondial.

Afin de mener à bien cette étude, nous structurerons notre analyse en deux chapitres distincts. Chacun ayant un objectif spécifique et contribuant à une compréhension globale de la problématique. Le premier chapitre sera consacré à l'examen approfondi de l'état de l'art concernant la rentabilité clientèle et la Customer Lifetime Value (CLV) dans le secteur bancaire. Ce chapitre commencera par une exploration détaillée de la littérature existante, analysant comment ces concepts ont été précédemment définis, mesurés et appliqués dans des contextes similaires ou différents. Il s'attachera particulièrement à clarifier les concepts clés, en établissant des définitions précises et en délimitant leur portée et leur pertinence dans le cadre de notre étude.

Tandis que, le deuxième chapitre sera dédié à la partie pratique, où nous allons essayer de modéliser la valeur à vie du client ou CLV à partir des données de la banque. Dans cette optique, il est important de préciser que les modèles de survie que nous envisageons de comparer le modèle PH Cox et le modèle Deep Surv; ne calculent pas directement la CLV, mais plutôt la probabilité qu'un client continue à faire affaire avec la banque ou reste "en vie" à un moment donné. C'est cette probabilité qui sera ensuite utilisée pour estimer la CLV des clients de la banque CPA.

Ce travail repose sur une combinaison de recherche documentaire, d'analyse de données et d'entretiens avec des responsables du CPA, afin de fournir un aperçu complet et approfondi de l'utilisation de la CLV dans la gestion du portefeuille de clients de cette banque.

# Chapitre premier :

## La rentabilité client et la Customer life Time value.

*« Réduit à l'essentiel, l'objectif du marketing  
est d'attirer et de fidéliser des clients rentables »*

Dr. Philip Kotler

# **Chapitre premier : La rentabilité client et la Customer life Time value.**

## **Introduction**

Ce premier chapitre de notre étude est dédié à une investigation minutieuse de l'état de l'art concernant deux notions cruciales dans le secteur bancaire : la rentabilité clientèle et la Customer Lifetime Value (CLV). L'objectif principal de ce chapitre est de fournir un cadre théorique robuste et complet pour notre analyse, en explorant en profondeur la littérature existante sur ces concepts. Nous analyserons de quelle manière la rentabilité clientèle et la CLV ont été définies, mesurées et appliquées dans différents contextes bancaires, tout en soulignant les variations et les similitudes dans les approches adoptées à travers différents milieux.

Une attention particulière sera accordée à la clarification des concepts clés, en établissant des définitions précises et en explorant leur pertinence spécifique à notre contexte d'étude. Cette démarche implique une compréhension approfondie des différentes dimensions de la rentabilité clientèle et de la CLV, y compris les modèles mathématiques et statistiques utilisés pour leur calcul, ainsi que leur signification dans la prise de décision stratégique au sein des banques.

## **Section 01 : Définition de la rentabilité client et des indicateurs pertinents pour l'évaluer**

### **I. Définition de la rentabilité bancaire**

La rentabilité d'une banque représente sa capacité à dégager des revenus suffisamment importants de ses activités, après déduction des coûts nécessaires au maintien de ses activités sur le long terme, (Monnier 2022) [1].

Dans un environnement très concurrentiel, il est indispensable pour les banques d'améliorer leur connaissance et leur suivi de la rentabilité bancaire. Notamment en matière d'opérations de crédit.

#### **1. Une banque, c'est un commerce d'argent**

La principale fonction de la banque est de collecter des fonds et les redistribuer, elle joue un rôle crucial dans l'intermédiation financière entre les agents à besoin de financement et les agents à capacité de financement. La marge d'intermédiation représente la valeur ajoutée sur capitaux.

#### **2. Une banque, c'est un commerce de services :**

La deuxième fonction de la banque est de garantir un service clientèle de qualité. La banque génère des revenus sous forme de commissions pour ces services. Ces frais comprennent diverses charges et taxes facturées aux clients en contrepartie des services fournis. Certaines commissions sont régulières, tandis que d'autres sont occasionnelles (par exemple, les frais de dossier pour les prêts immobiliers). Ces commissions proviennent principalement des sources suivantes :

- L'épargne financière (frais de gestion, frais d'entrée...) ;
- Le crédit (frais de dossier, pénalités pour remboursement anticipé du client...) ;
- Les conventions de compte (package...) ;
- Les services (opérations de change, frais de débit non autorisé...) ;
- L'assurance ;
- Les voyages, les alarmes... lorsque ces produits sont commercialisés en agence ;

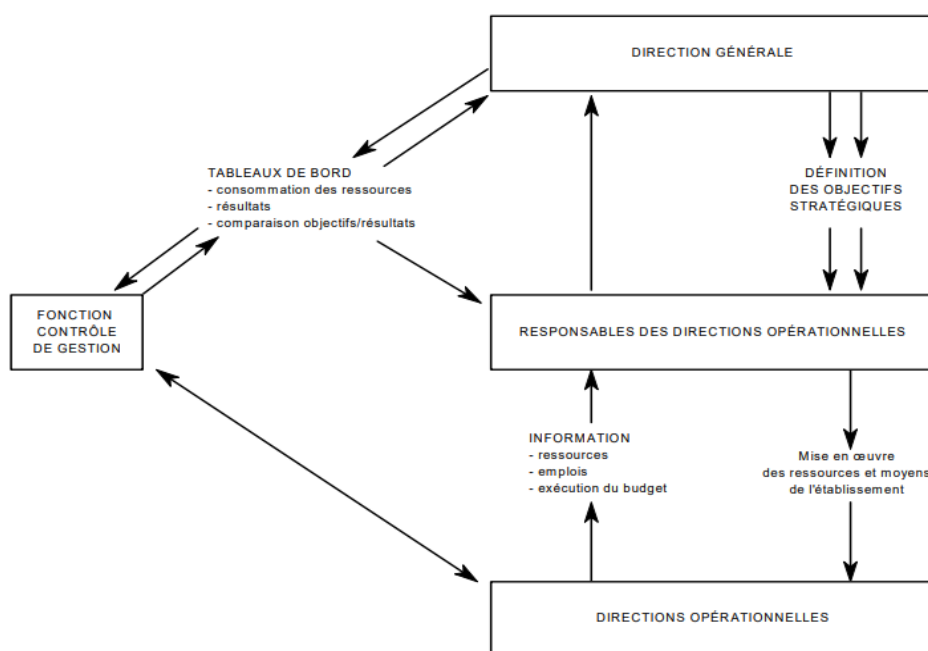
- La monétique (utilisation des distributeurs automatiques de billets et guichets automatiques bancaires, location des terminaux de paiement pour le marché des professionnels...).

## II. Le rôle de la mesure de la rentabilité dans les orientations stratégiques et budgétaire de la banque

Il est essentiel que les directions générales disposent d'informations précises et pertinentes pour orienter leur stratégie et leur politique commerciale de manière éclairée. Cette démarche analytique peut servir de base solide à ces orientations stratégiques. Cependant, il est important que ces décisions soient prises sous le contrôle des organes délibérants, assurant ainsi une gouvernance appropriée.

Des informations fiables et pertinentes concernant l'analyse de la rentabilité des produits, des services, des différents segments de clients nécessitent un système d'information performant et une approche d'analyse de la rentabilité efficace, qui fournit des informations permettant la prise de décision stratégique et budgétaire, favorisant ainsi les actions commerciales et marketing qui vise à améliorer la rentabilité du portefeuille client de la banque.

**Figure 1 : Gestion de la rentabilité**



Source : livre blanc sur la mesure de la rentabilité des activités bancaires

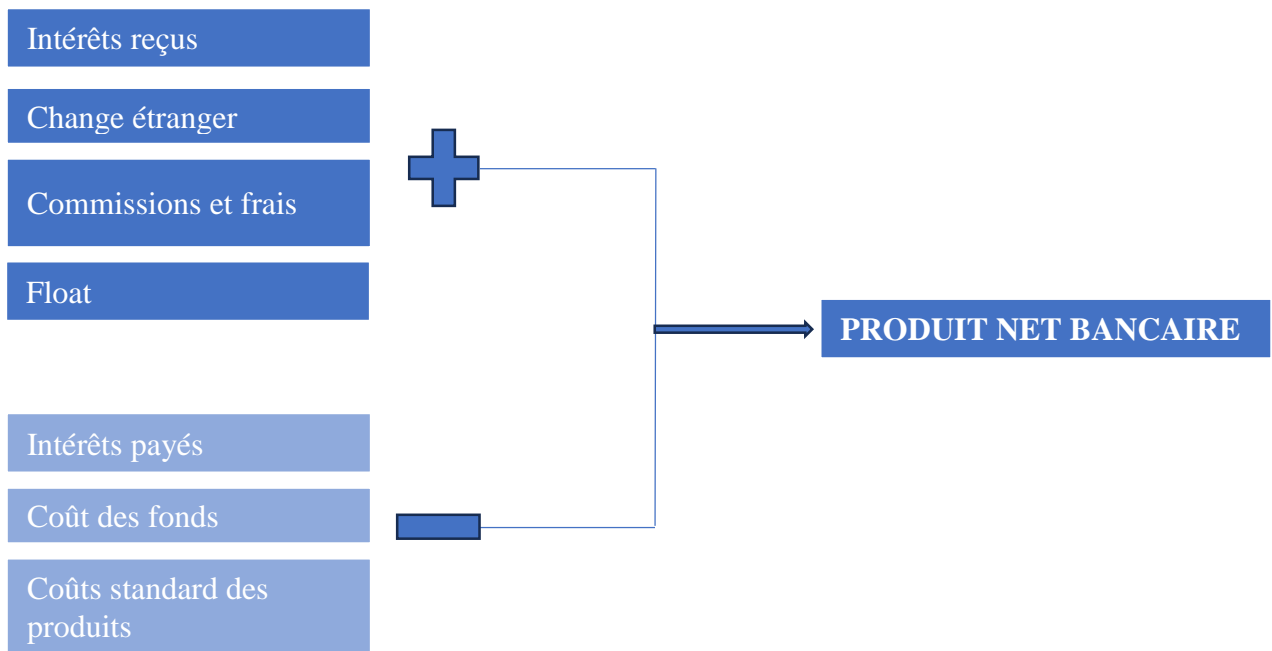
### III. Indicateurs clés de la rentabilité bancaire

Lorsqu'il s'agit de la rentabilité d'une institution bancaire, le premier indicateur qui nous vient à l'esprit est le produit net bancaire (PNB). Ceci est le point de départ pour calculer les indices de rentabilité. Les banques cherchent constamment à accroître leur PNB, lequel est maintenant fondé sur la recherche de commissions.

#### 1. Le Produit Net Bancaire (PNB) :

Le PNB est calculé par l'écart entre les produits bancaires d'exploitation et les charges bancaires d'exploitation (activité de prêt et d'emprunt, opérations sur titres, change, marchés dérivés...). Il mesure l'apport spécifique des banques à l'augmentation de la richesse nationale et peut donc être assimilé à la valeur ajoutée générée par les entreprises non financières.

**Figure 2** : Les composantes du PNB



*Source : Elaboré par nos propres soins*

## ❖ Décomposition du PNB

La mesure des revenus est basée sur le produit net bancaire, dans lequel la marge d'intermédiation est un élément clé.

La marge d'intermédiation clientèle est calculée comme l'écart entre [2]:

- Les intérêts perçus provenant des crédits octroyés aux clients,
- Et les intérêts déboursés pour les dépôts reçus des clients.

Majorée des produits ou charges de trésorerie liée à la position nette dépôts/crédits et au mode de refinancement ou de remplacement de cette position.

La marge globale d'intermédiation correspond à la somme de deux éléments : la marge d'intermédiation clientèle et la marge issue des opérations de trésorerie, autres que celles avec la clientèle.

Le PNB est généralement obtenu à partir de cette marge globale d'intermédiation en intégrant les produits ou charges qui ne sont pas directement associés à l'usage des capitaux, Cela inclut notamment :

- les commissions nettes bancaires, financières (OPCVM...), patrimoniales (assurances...),
- les revenus des activités de marché,
- les produits et charges nés des activités d'hors-bilan,
- les produits accessoires et divers.

Les diverses marges (comme la marge clientèle, la marge globale d'intermédiation, la marge bancaire globale, etc.) peuvent être calculées de manières variées, en fonction de l'approche adoptée. On distingue principalement deux approches : celle basée sur les aspects comptables ou notionnels (qui intègre les taux de cession interne) et celle basée sur les flux ou les stocks (analysant la production nouvelle). Les méthodes de calcul peuvent ainsi varier en fonction de ces différentes perspectives.



**Tableau 1 : Calcul Des Marges Sur Capitaux**

	<b>APPROCHE EN STOCKS</b>	<b>APPROCHE EN FLUX</b>
<b>APPROCHE COMPTABLE</b>	<p><b>Indicateur de taux de marge clientèle =</b> taux d'intérêt moyen des crédits - taux d'intérêt moyen des dépôts</p> <p><b>Marge d'intermédiation globale =</b> <math>\sum</math> intérêts courus à l'actif (crédits, trésorerie, portefeuille, immo...) / Cx employés (*) - <math>\sum</math> intérêts courus au passif (dépôts, trésorerie, fonds propres...) / Cx employés hors intérêts courus non échus sur créances douteuses</p>	<p><b>Indicateur de taux de marge clientèle sur la production nouvelle =</b> taux d'intérêt moyen de la production nouvelle de crédits de l'année - taux d'intérêt moyen de la collecte nouvelle de dépôts de l'année</p>
<b>APPROCHE NOTIONNELLE (utilisation des TCI)</b>	<p><b>Marge notionnelle d'intermédiation clientèle =</b> intérêts courus sur encours de crédits - intérêts de refinancement aux TCI des crédits - intérêts courus sur encours de dépôts + intérêts de remplacement aux TCI des dépôts</p> <p><b>Marge notionnelle d'intermédiation globale =</b> marge sur fonds propres non immobilisés + autres marges ventilables par métier par nature de capitaux (types de crédits, de dépôts, etc) par nature de taux par nature de produits</p>	<p><b>Rémunération immédiate =</b> commissions, etc + marge notionnelle d'intermédiation clientèle sur la première année</p> <p><b>Rémunération actualisée =</b> somme des marges notionnelles d'intermédiation clientèle actualisées sur la durée des produits</p>

(\*) En capitaux moyens équivalents crédit bilan et hors-bilan.

*Source : Livre blanc sur la mesure de la rentabilité des activités bancaires*

## 2. Le coefficient d'exploitation

La banque accorde une importance majeure à un autre paramètre crucial pour évaluer sa rentabilité : le coefficient d'exploitation. Cet indicateur est essentiel pour la gestion de la performance bancaire.

Dans son fonctionnement, l'établissement bancaire engage de nombreuses dépenses importantes. Les principales catégories de dépenses sont énumérées ci-dessous :

- ❖ Charges de personnel (masse salariale, formation...);
- ❖ Charges de logistique (entretien et rénovation de l'agence, achats de fournitures...);
- ❖ Charges informatiques (équipement, maintenance, achats logiciels...);
- ❖ Charges de production bancaire (affranchissement courriers...);

- ❖ Charges de marketing / communication (partenariats, publicité, prestations commerciales à payer au siège : PLV, mailing...).

L'ensemble de ces dépenses représente les coûts opérationnels de l'agence.

Le coefficient d'exploitation, calculé comme le ratio entre les frais généraux et le Produit Net Bancaire (PNB), est un indicateur supplémentaire crucial pour évaluer la rentabilité de l'agence. Il permet d'apprécier la proportion des gains générés absorbée par les coûts fixes. Une valeur plus faible de ce ratio indique une meilleure préparation de l'agence pour l'avenir. Cependant, l'acceptabilité de ce ratio varie en fonction de l'organisation spécifique du réseau d'agences. Néanmoins, il peut servir à comparer les agences au sein d'un même groupe (départemental, etc.).

### 3. Autres mesures de la rentabilité :

- ❖ **Retour sur capitaux propres (return on equity) :**

La banque évalue également la rentabilité de son capital propre. Le retour sur fonds propres (ROE) est calculé selon la formule suivante :

$ROE = \text{résultat net} / \text{fond propres}$

Pour les actionnaires de la banque, le rapport du bénéfice net et les fonds propres met en lumière le rendement de leur investissement.

- ❖ **Le retour sur actifs (return on assets : ROA)**

Il s'agit d'une expression de la rentabilité des actifs de la banque. Cet indicateur est calculé en divisant le résultat net par le total des actifs. Cet indicateur présente un inconvénient. Il met tous les actifs sur le même plan, alors que leurs risques sont différents. Or, les activités hors bilan se sont développées ces dernières années...

## **IV. La notion de la rentabilité client**

« La rentabilité de la clientèle au sein d'une banque désigne la mesure de la valeur économique générée par un client pour la banque. Elle englobe les revenus directs générés par le client, tels que les intérêts perçus sur les prêts, les frais de transaction et les frais de tenue de compte, ainsi

que les revenus indirects, tels que les recommandations de nouveaux clients ou la réduction des risques grâce à une relation stable. »

Il est essentiel de souligner que l'évaluation de la rentabilité du portefeuille est cruciale pour un chargé de clientèle. Cela va au-delà de considérer uniquement les performances commerciales (production bancaire). Il est également crucial d'évaluer le niveau de risque associé à son portefeuille.

### 1. Impact des dates de valeur

L'incidence des dates de valeur se concentre principalement sur les transactions avec les clients et comporte deux aspects distincts.

- a) Le décalage, qu'il soit en avance ou en retard, dans l'enregistrement des opérations sur les comptes des clients en appliquant les dates de valeur, permet à la banque de disposer d'une trésorerie gratuite, souvent appelée "float", en ne rémunérant pas les soldes des dépôts concernés, que ce soit sur les comptes courants ou les comptes d'épargne à régime spécial, conformément à la règle de la quinzaine.
- b) En ce qui concerne les mouvements sur les comptes courants, l'application des dates de valeur rend une partie des comptes débiteurs, occasionnant ainsi des intérêts débiteurs et des frais de transaction.

### 2. L'exploitation des résultats de rentabilité par le chargé de clientèle

L'évaluation de la rentabilité de la relation commerciale établie avec le client est un critère de segmentation essentiel pour la banque. Il peut être associé à d'autres critères, tels que la fidélité du client et la notoriété de la banque.

Dans notre quête de rentabilité au sein de la banque, notre démarche vise à cibler les clients qui ne sont pas rentables et qui ne manifestent pas de fidélité. Nous prévoyons d'engager une discussion avec ces clients pour les informer de leurs difficultés et leur proposer la possibilité de continuer la relation. Cela pourrait reposer sur des éléments clés tels que l'adoption d'un forfait, générateur de commissions régulières.

Les clients fidèles mais non rentables peuvent être encouragés à adopter des stratégies de gestion à distance visant à améliorer leur rentabilité. Des rencontres seront organisées avec ces

clients pour les persuader d'utiliser des services bancaires en ligne, réduisant ainsi le coût de la relation en effectuant certaines opérations à distance.

Pour les clients rentables et fidèles, des efforts seront déployés pour maintenir et renforcer leur rentabilité (réunions régulières pour suivre leurs besoins et offrir des solutions d'épargne complémentaires).

En ce qui concerne les clients rentables mais infidèles, nous chercherons à mieux les comprendre dans le but de les reconquérir. Nous leur présenterons des produits incitatifs visant à renforcer leur fidélité.

## **V. Méthodes d'analyse de la rentabilité d'un portefeuille client :**

L'analyse de la rentabilité d'un portefeuille client est essentielle pour évaluer la performance globale et surtout commerciale et prendre des décisions stratégiques éclairées. Voici les méthodes les plus couramment utilisées pour analyser la rentabilité d'un portefeuille client au sein de la banque :

### **1. L'approche rentabilité client (ARC) :**

Proposé pour la première fois par Cagnello-Charl (1999) [3]. L'ARC consiste à mesurer la rentabilité des portefeuilles de clients. Elle se rapproche, dans sa volonté de déterminer la rentabilité des clients, du contrôle de gestion qui, dans le domaine bancaire, cherche à évaluer la marge générée par la banque par client (ou segment de clientèle), et du marketing qui vise à segmenter les clients en fonction des niveaux de rentabilité, afin d'en tirer des conclusions comportementales. Elle s'inscrit également parfaitement dans la tendance actuelle de la réflexion sur le calcul des marges et des coûts issue de l'Activity-Based Costing (ABC), en utilisant cette même méthode pour calculer les coûts par client.

#### **❖ Formule de calcul de rentabilité client par l'ARC :**

Le calcul de la rentabilité client par l'ARC s'apparente à un calcul de marge classique. Le calcul de la marge par client revient à calculer un compte d'exploitation par client ou famille de clients. Il s'agit simplement d'attribuer à chaque client ou segment de client une évaluation de sa performance au moyen d'une métrique :

$$M_i^t = \sum_{j=1}^n CA_{ij}^t - \sum_{k=1}^m c_{ik}^t \quad [1]$$

$M_i^t$  : Marge générée par le client, durant la période T.

$CA_{ij}^t$  : Chiffre d'affaires obtenu par la consommation du produit ou du service (avec  $j=1, \dots, n$ ), par le client, durant la période t.

$c_{ik}^t$  : Coût du produit ou du service k avec ( $k=1, \dots, m$ ) consommé avec le client i, durant la période t.

Ainsi, tout calcul des marges des clients doit donc passer par trois étapes :

- 1) Identification préalable de tous les clients i, sur lesquels on souhaite focaliser l'analyse (par le biais d'une segmentation de la clientèle), une détermination des différentes ventes générées par la consommation des produits ou services j par les clients i, au cours d'une période t, et une affectation de ces ventes aux clients.
- 2) Une détermination des différents coûts générés par la consommation de produits ou services k par des clients i, et affectation des coûts aux clients.
- 3) Le calcul de la marge obtenue par la différence entre les ventes et les coûts.

#### ❖ **Objectif de l'approche ARC :**

L'objectif de l'approche de la rentabilité client est de se focaliser sur la marge créée par chaque client (ou segment de client), afin de mieux comprendre sa contribution aux bénéfices de la banque. C'est une approche qui se concentre davantage sur les clients que sur les produits ou services offerts. Cette approche repose sur le principe que ce sont les clients qui influencent directement les résultats de la banque, puisqu'ils sont les premiers responsables de la performance de la banque. De ce fait la banque doit porter un intérêt particulier à ses clients, plutôt qu'à n'importe quelle autre entité.

#### ❖ **Les principes de l'approche :**

Selon Cargnello-Charl (1999) [3], les principes sur lesquels repose l'approche ARC se

présentent ainsi :

- **1<sup>er</sup> principe** : un raisonnement sur un horizon assez long (5 à 10 ans), ensuite actualiser et probabiliser les événements futurs concernant les demandes de produits et services de la banque. Cette analyse du comportement du client conditionne les résultats attendus de ce dernier.
- **2<sup>ème</sup> principe** : le calcul des coûts associés à chaque client ou segment de client se fait à l'aide de la méthode ABC (Activity-based costing). Uniquement les charges relatives au client ou segment de clientèle leur seront affectées.
- **3<sup>ème</sup> principe** : mettre en œuvre les moyens susceptibles d'être utiles à une amélioration de la rentabilité clients, ceci nécessite l'identification des activités à valeur ajoutée et à non-valeur ajoutée (chaîne de valeur). Ainsi la politique de la banque s'oriente vers la recherche de l'optimum entre la valeur créée et les coûts.

Sert l'ARC est une approche utile pour évaluer la rentabilité client dans les banques. Mais elle présente certaines limites qui peuvent freiner son application au sein de la banque. Parmi ces limites :

- ✚ La complexité de l'affectation des coûts : Il n'est pas toujours facile d'affecter des coûts à un client spécifique. Certaines charges, telles que les frais généraux d'exploitation de la banque, sont réparties entre tous les clients, ce qui complique l'imputation des coûts spécifiques à chacun d'entre eux.
- ✚ Données incomplètes : Pour pouvoir calculer la rentabilité d'un client, il est nécessaire de disposer de données détaillées sur les revenus et les coûts associés à chacun des clients. Ce qui n'est pas toujours facile à obtenir notamment pour les clients qui utilisent plusieurs produits ou services de la banque.
- ✚ Horizon temporel : Contrairement au premier principe de l'ARC annoncé précédemment selon lequel la méthode stipule une analyse sur un horizon assez long, on ne retrouve pas réellement cet aspect dans la formule de calcul de la rentabilité. Et la méthode se concentre généralement sur une période de temps relativement courte, par exemple, une année. Cela peut ne pas donner une image complète de la rentabilité à long terme du client pour la banque.

## 2. L'approche ABC :

L'approche ABC (Activity Based Costing) permet d'évaluer le coût des produits, des services ou des clients en associant les coûts des centres de responsabilité aux activités ou aux processus utilisés pour la fourniture du produit ou service au client. Voici une synthèse du processus d'une démarche ABC :

Tout d'abord, cette méthode établit un lien analytique entre les coûts engagés par les centres de responsabilité et les activités ou les procédés employés pour la fourniture du produit ou du service au client.

Une activité est définie comme un ensemble de tâches ayant un objectif spécifique, tel que l'ouverture d'un compte bancaire ou la demande d'un prêt immobilier. En règle générale, plusieurs centres de responsabilité sont engagés dans l'exécution de ces activités.

Ensuite, la méthode consiste à déterminer les coûts associés à chaque activité en évaluant le temps alloué par les différents types de personnel et les dépenses connexes dans chacun des centres de responsabilité participant à l'activité.

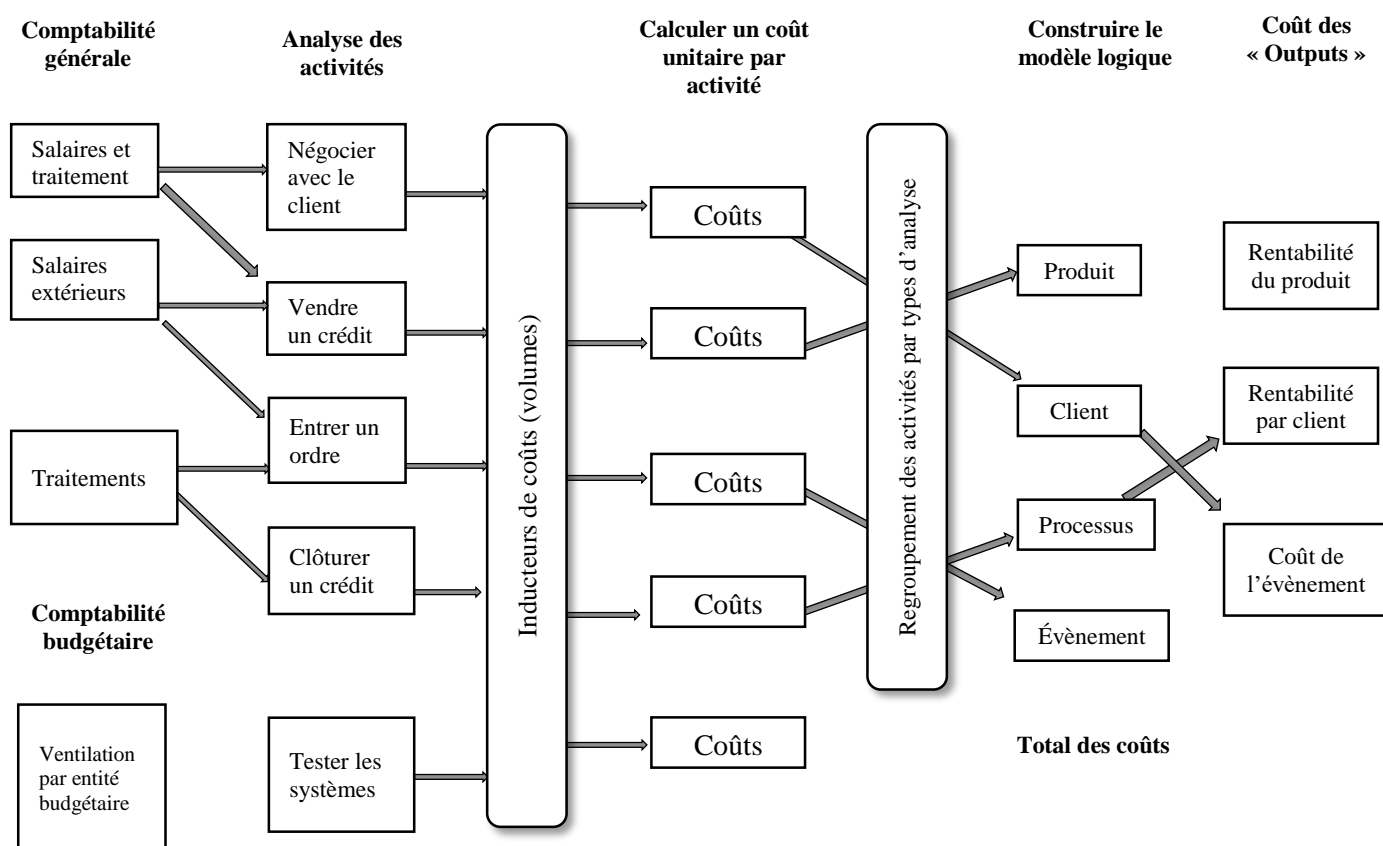
L'approche ABC évalue les coûts attribuables à chaque secteur d'activité, en suivant une répartition définie par chaque établissement. Ainsi, l'établissement doit préciser les activités liées aux produits ou aux clients.

Par la suite, la collecte des données peut être effectuée soit par des mesures systématiques, soit par des enquêtes menées auprès d'un échantillon représentatif de personnels travaillant dans les agences et les back offices. Cette collecte de données au sein des agences et des back offices vise à évaluer le coût par type d'opération qui contribue à l'activité. Ces évaluations servent de base pour le calcul des coûts standards par activité.

Les coûts sont alors imputés aux produits ou aux clients qui ont généré l'activité, sur la base de la consommation réelle.

Cette méthode présente l'avantage d'une grande flexibilité dans son champ d'application, allant de l'analyse d'un seul produit ou marché à celle de l'ensemble de l'entreprise.

**Figure 3 : Processus de la méthode ABC**



Source : Livre blanc sur la mesure de la rentabilité des activités bancaires

L'avantage de cette méthode est que son champ d'application est flexible, allant de l'analyse d'un produit, d'un marché ou de l'ensemble de l'entreprise.

La troisième méthode couramment utilisée dans le calcul de la rentabilité client concerne la méthode de la « customer lifetime value » ou la « durée à vie du client ». Cette méthode fera l'objet de la section 3.

## **VI. Importance de la gestion efficace d'un portefeuille client :**

La gestion efficace d'un portefeuille client revêt une importance cruciale au sein de la banque pour plusieurs raisons, tant d'un point de vue économique que stratégique. Voici quelques-unes des principales raisons, soutenues par des références scientifiques:



1. Maximisation des bénéfices :

Une gestion efficace du portefeuille de clients permet à une banque de repérer les clients les plus rentables et de concentrer ses efforts sur eux.

2. Fidélisation des clients :

Une approche proactive dans la gestion du portefeuille client permet de mieux comprendre les attentes et les besoins des clients, ce qui favorise la fidélisation. Des recherches telles que « Customer Behavior and Loyalty in Banking: Global Edition 2023 » [4] ont montré que la fidélité des clients peut être un facteur déterminant pour la performance globale d'une entreprise.

3. Réduction du risque de crédit :

Une bonne gestion du portefeuille client implique une évaluation rigoureuse de la solvabilité des emprunteurs potentiels. Des recherches comme « Gestion des risques et performance des institutions de microfinance » (Tchuigoua 2012)[5] ont mis en évidence l'importance de la segmentation des clients pour mieux évaluer le risque de crédit et réduire les pertes potentielles.

4. Amélioration de la prise de décision stratégique :

En analysant les données du portefeuille client, une banque peut obtenir des informations précieuses pour prendre des décisions stratégiques éclairées. De nombreuses recherches ont souligné le rôle essentiel des analyses de données pour améliorer la prise de décision dans le secteur bancaire.

5. Personnalisation de l'expérience client :

Une gestion efficace du portefeuille client permet d'identifier les préférences et les besoins individuels des clients, ce qui facilite la personnalisation des offres et des services, chose qui aura pour conséquence l'augmentation de l'adoption des services bancaires et donc la rentabilité du portefeuille.

En résumé, une gestion efficace du portefeuille client est essentielle pour maximiser la rentabilité, fidéliser les clients, réduire les risques et améliorer la prise de décision stratégique au sein d'une banque. Les références scientifiques fournies soutiennent l'importance de ces aspects pour le succès global de l'institution bancaire.

## **Section 02 : La Customer Life time value définition et concepts**

Cette section est consacrée à l'examen approfondi de la métrique de la Valeur Vie Client (CLV), qui constitue le cœur de notre étude. Nous explorerons la définition de cette mesure, son importance stratégique, ainsi que les différentes méthodologies appliquées à son calcul.

### **I. Définition de la Customer Lifetime Value (CLV)**

La valeur client, également connue sous l'acronyme CLV (Customer Lifetime Value), est une métrique prospective. Elle est calculée en prenant en considération la somme actualisée des marges générées tout au long de la durée de vie du client, en prenant en compte son historique comportemental et ses données sociodémographiques.

$$CLV = \sum \frac{(P_t - C_t) R_t}{(1+i)^t} - AC \quad [2]$$

- $P_t$  = prix payé par un client au moment  $t$ ,
- $C_t$  = coût direct du service au client à l'instant  $t$ ,
- $i$  = taux d'actualisation ou coût du capital pour l'entreprise,
- $R_t$  = probabilité qu'un client renouvelle son achat ou soit encore en vie au moment  $t$ ,
- $AC$  = coût d'acquisition,
- $T$  = horizon temporel pour l'estimation de la CLV

Pour une entreprise donnée, les variables  $T$  et  $i$  peuvent être considérées comme des constantes déterminées en interne, et la CLV peut donc être considérée comme une fonction des paramètres  $P_t$ ,  $C_t$  et  $R_t$  et de la variable  $AC$ . Par conséquent, l'objectif principal du calcul de la CLV est d'approximer et de déterminer ces processus et ces variables pour le client, le

groupe ou le segment spécifique.

En théorie, les modèles CLV devraient estimer la valeur d'un client sur l'ensemble de sa durée de vie. Cependant, dans la pratique, la plupart des chercheurs utilisent un horizon temporel limité à trois ou quatre ans (Donkers et al, 2007)[6]. Trois ou quatre ans est une bonne estimation de l'horizon au cours duquel l'environnement commercial actuel ne changerait pas de manière considérable. En outre, certaines recherches envisagent un horizon temporel encore plus court.

## **II. Le rôle et l'application de la valeur vie client dans les décisions marketing**

L'analyse de la valeur vie client propose que la valeur d'une relation avec un client peut être augmentée soit en augmentant le montant des bénéfices obtenus auprès du client, soit en prolongeant la durée de vie de la relation. Par conséquent différentes stratégies de marketing relationnel sont suggérées pour les clients à différentes étapes de leur cycle de vie. Pour cette raison, la valeur vie client est un outil important pour déterminer le budget des décisions marketing et aider les entreprises à prendre des décisions stratégiques ainsi que tactiques. (Berger 2002) [8] affirme que les entreprises doivent prendre quatre mesures pour allouer de manière optimale les ressources et prendre les actions marketing les plus appropriées pour acquérir et maintenir les relations avec leurs clients :

- 1- Créer une capacité de renseignement marketing guidée par une base de données pour le calcul de la valeur vie client ;
- 2- Segmenter leurs données en fonction des besoins des clients et de leur comportement d'achat, en prenant en compte des facteurs tels que le pouvoir d'achat, la régularité d'achat et les types de produits achetés ;
- 3- Prévoir la valeur vie client dans des scénarios alternatifs ;
- 4- Allouer des ressources pour maximiser la valeur de la base client.

Les actions mentionnées ci-dessus indiquent également que le rôle le plus important de la CLV dans les stratégies marketing est lié à son rôle dans l'élément stratégique central de la gestion marketing, à savoir la segmentation et le ciblage, tandis que l'allocation du budget marketing et la détermination des actions et tactiques marketing représentent l'impact de la CLV sur les aspects tactiques du marketing.

## 1. Segmentation et ciblage

La segmentation du marché est l'un des concepts clés du marketing car différents groupes de clients ont besoin de différents mix marketing. Smith (1956) [9] a d'abord défini la segmentation de marché comme « considérant un marché hétérogène comme un certain nombre de marchés plus homogènes, en réponse à des préférences différentes, attribuables aux désirs des consommateurs pour une satisfaction plus précise de leurs besoins ». Par conséquent, l'objectif principal de la segmentation des marchés est de satisfaire les besoins et les exigences des clients en atteignant chaque client et en allouant les ressources de l'entreprise aux zones les plus rentables.

Bien que de nombreuses tentatives aient été faites pour définir le concept de segmentation de marché depuis la fin des années 1950, la définition originale de Smith (1956) reste attrayante en raison de son idée que la dérivation des segments de marché devrait être motivée par une véritable hétérogénéité des besoins et des désirs des consommateurs (Alfansi, 2000) [10].

Traditionnellement, les entreprises segmentent leurs clients en utilisant quatre groupes principaux de critères de segmentation de marché : géographique, démographique, psychographique et comportement d'achat précédent. Malgré le fait que la segmentation démographique et géographique ont été les critères de segmentation traditionnels les plus courants, les critères psychographiques et comportementaux se sont avérés être de meilleurs indicateurs pour définir les besoins et les désirs, et donc le comportement d'achat des clients dans de nombreuses industries.

Enfin, la rentabilité des clients a été acceptée comme base importante pour la segmentation comportementale en raison de l'importance centrale des profits. (Mulhern, F. J. 1999) [11]

## 2. L'allocation des ressources pour les éléments du marketing mix et la formulation de la stratégie

Les décisions relatives au marketing mix impliquent l'allocation des budgets de marketing aux clients ou aux segments de marché. Cependant, la tâche principale du

département de marketing, qui consiste à augmenter les ventes à court terme et l'équité de la marque à long terme simultanément, rend difficile la mesure du succès d'un effort de marketing donné.

Auparavant, les critères les plus largement utilisés pour mesurer la performance marketing étaient la conscience du client, la satisfaction du client et la fidélité. Cependant, ces données ne prennent pas en compte l'impact financier. Dans un environnement hautement compétitif, les départements de marketing doivent considérer de manière approfondie l'efficacité de leurs gros budgets en termes de contribution à la rentabilité, à la croissance et à l'avantage concurrentiel à long terme de l'entreprise.

Les stratégies d'allocation de ressources doivent évaluer les clients en fonction de leur rentabilité future et recommander des initiatives marketing appropriées. Les études ont récemment porté sur le développement de modèles quantitatifs pour l'allocation optimale du budget marketing, plutôt que sur des discussions conceptuelles de l'allocation des ressources et des suggestions pour les stratégies de marketing connexes.

### **III. La CLV dans l'industrie bancaire de détail**

#### **1. Tendances actuelles**

La popularité de CLV dans l'industrie bancaire de détail a augmenté principalement en raison des progrès réalisés dans les technologies de l'information modernes et de la concurrence accrue. L'innovation a rendu possible la collecte et le stockage de vastes quantités de données client, ce qui a augmenté la possibilité pour les banques de calculer le CLV des clients individuels.

La demande croissante et l'intérêt pour certaines métriques commerciales peuvent être attribués aux changements dans l'environnement concurrentiel. Ces changements sont en partie dus aux avancées technologiques, notamment dans le domaine de l'information, qui ont mené à une standardisation des produits et services offerts aux clients dans plusieurs industries. Cela a réduit les différences entre concurrents, augmentant ainsi l'importance des stratégies de marketing. En outre, cela a intensifié la concurrence en favorisant l'émergence de nouveaux acteurs proposant des produits distincts.

## 2. les revenus par client :

Les gains  $pt$  générés par un client dans la banque de détail au temps  $t$  dépendent des produits qu'un client possède. De plus, la nature des produits dépend des finances générales du client. Les produits généraux sont les suivants:

- Compte chèque et épargne
- Hypothèques
- Cartes de crédit
- Prêts personnels
- Services de change et de transfert d'argent
- Assurances
- Services de gestion d'actifs
- Autres services bancaires en ligne

Toutefois, la banque reste un commerce très particulier dans la mesure où elle propose des multiproduits et multi-services, avec une large gamme de services de nature très différente, allant des services purs aux produits à forte intensité de capital. À cela s'ajoute le fait que ces services et produits sont souvent liés entre eux. La consommation d'un produit impliquant généralement celle d'un autre. Ces deux caractéristiques rendent souvent complexe la définition du **processus de production bancaire**.

## 3. Le désabonnement des clients

La caractéristique de l'industrie bancaire de détail est que les clients ont presque toujours besoin de services financiers tout au long de leur vie. Jusqu'à présent, les clients avaient également des relations presque monogames avec les banques, consolidant ces produits et services. Ces facteurs ont créé une industrie avec une valeur à vie client élevée en général et une faible probabilité de retour des clients une fois qu'ils ont résilié leur contrat. Par conséquent, le taux de résiliation des clients pourrait avoir un impact négatif important sur la valeur totale des clients, il est donc important de comprendre ce processus en détail dans ce contexte. De plus, la relation monogame avec les banques a commencé à changer et selon la banque suédoise SEB [12], il est de plus en plus courant que les clients répartissent leurs produits financiers sur différents acteurs du marché. Cela implique un besoin encore plus grand pour les

banques de comprendre les mécanismes et les effets sous-jacents du taux de résiliation des clients afin qu'elles puissent répondre de manière plus efficace aux effets de la nouvelle dynamique du marché.

#### **IV. La place de la CLV dans la stratégie marketing de la banque**

- La CLV n'est pas une solution miracle pour stimuler la croissance d'un marché mature, mais elle représente un levier essentiel pour optimiser les investissements et orienter la croissance des entreprises de manière efficace.
- Elle constitue une réponse appropriée face à la baisse anticipée de rentabilité, en permettant une allocation plus judicieuse des investissements commerciaux.
- La CLV est un outil puissant pour contrer les attaques de la concurrence, en dirigeant les investissements vers l'acquisition et la fidélisation de segments de clientèle spécifiques.
- La CLV est un moyen d'identifier les clients présentant un risque de volatilité ; ce qui permet de mettre en place des programmes de rétention adaptés à chaque profil.
- Sert à repérer les clients exposés à un risque de volatilité, facilitant ainsi la mise en place de programmes de rétention adaptés à chaque profil.
- Ni la gestion par la CLV ni l'adoption de nouvelles approches relationnelles ne constituent une solution complète. La menace d'une concurrence élargie, avec l'arrivée de nouveaux acteurs extérieurs au secteur bancaire, exige une réflexion prospective approfondie.

Le calcul de la valeur client au sein de la banque offre une vision plus précise du portefeuille client. L'orientation par la CLV jouera un rôle déterminant dans la bataille que les banques devront mener pour maintenir leurs parts de marché. La valeur client doit particulièrement être considérée comme un excellent moyen d'évaluer la rentabilité des clients et de déployer ensuite des stratégies marketing efficaces basées sur des données quantitatives.

### **Section 03 : Indicateurs et modèles de mesure de la CLV dans la littérature :**

Comme nous l'avons vu dans la section précédente, bien que la CLV soit un outil important et utile pour développer des stratégies de marketing efficaces et obtenir un avantage concurrentiel sur le marché, il n'existe pas de consensus sur les indicateurs ou les méthodes de mesure de CLV dans la littérature.

Par conséquent, avant de développer et de proposer un modèle de CLV, il est nécessaire d'examiner et de dresser une liste plus détaillée des indicateurs de CLV existants et de les comparer avec les autres indicateurs de CLV.

#### **I. Méthodes de calcul de la CLV**

Explorant les diverses méthodologies et perspectives qui ont émergé autour de cette métrique cruciale en marketing à travers la revue de la littérature ci-après.

Les prédicteurs de la CLV ont été analysés et répertoriés par divers chercheurs dans la littérature (Blattberg 1996 [13]; Ho et al., 2006; Ching et al., 2004; Gurau 2002[14]; Gupta et al, 2006 [15]; Haenlein et al., 2007[16]; Mulhern, 1999; Hartfeil, 1996[17]; Helgesen, 2006[18]).

Bien que certaines de ces études aient mis l'accent sur des prédicteurs ou des variables communs, il n'y a pas d'accord sur les facteurs restants analysés en raison de la diversité des industries étudiées. Les variables les plus largement utilisées ou discutées dans la littérature sont listées ci-dessous:

- Probabilité qu'un client soit en vie pendant une période donnée (par exemple, Kumar et al., 2004).
- Taux d'acquisition et de rétention des clients et les coûts y'afférent ( Blattberg et Deighton, 1996).
- Dépenses liées aux campagnes de marketing et au maintien de la satisfaction des clients (par exemple, Ho et al., 2006).
- Revenus attendus d'un client ou d'un groupe, dérivés des données de transaction précédentes (par exemple, Malthouse et Blattberg, 2005).
- Probabilités de transition entre les groupes ou les états (par exemple, Ching et al., 2004).
- Achats de clients, généralement modélisés comme des événements suivant la loi de Poisson, avec leur taux d'occurrence dépendant du degré de satisfaction de l'opération d'achat la plus récente (par exemple, Ho et al., 2006). La satisfaction peut être définie



par une distribution de Bernoulli (par exemple, Ho et al., 2006).

- Les valeurs de Récence-Fréquence-Montant (RFM) des achats précédents de chaque client (Gupta et al., 2006). La récence est la date du dernier achat du client, la fréquence correspond au nombre d'achats effectués par un client sur une période donnée et la valeur monétaire est le montant d'argent dépensé par le client pendant cette période.
- La probabilité de désabonnement est la probabilité qu'un client mette fin à sa relation avec l'entreprise sur une période donnée (par exemple, Ho et al., 2006).
- Le type et l'intensité de la possession du produit (par exemple, Haenlein 2007).
- Le niveau de réponse du client aux campagnes proposées montre si le client est actif et s'il a un potentiel d'augmentation de la CLV (par exemple, Mulhern 1999).
- L'âge, les caractéristiques démographiques et le mode de vie du client (par ex. Haenlein, 2007).
- La fidélité du client affecte la plupart des autres variables mentionnées dans cette étude et.

La liste des variables présentée ci-dessus indique que les variables les plus couramment prises en compte sont des critères orientés vers le profit ou les coûts. Bien qu'il soit très utile de mesurer la CLV à l'aide de ces types de critères et de données objectives disponibles dans les bases de données des entreprises, la prise en compte d'autres variables telles que la satisfaction du client, la fidélité du client et les variables liées à l'utilisation du produit est également nécessaire pour les entreprises qui souhaitent adopter une approche du marché axée sur le client.

## **II. Analyse de la littérature sur les modèles de mesure :**

Dans la littérature, le CLV est mesuré à l'aide de modèles déterministes ou stochastiques. Les techniques stochastiques telles que les chaînes de Markov sont largement utilisées dans la littérature pour modéliser la CLV afin de gérer les relations avec les clients, mais la principale difficulté réside dans la définition des états et la détermination des probabilités de transition entre les états (Dwyer, 1997 ; Haenlein et al., 2007 ; Morrison et al., 1982 ; Pfeifer et Carraway, 2000). Lorsque la définition des états change, les déterminations des probabilités de transition doivent également être modifiées.

Pour les modèles stochastiques, afin de calculer la CLV, il est nécessaire de prévoir si un individu continuera à être actif et quel sera son comportement d'achat (Jain et Singh, 2002). Le modèle de Pareto/NBD est largement utilisé dans les contextes non contractuels où les transactions peuvent avoir lieu à tout moment (Gupta et al., 2006). Singh (2003) a déclaré que les modèles de type Pareto/NBD proposés dans la littérature CLV présentent des limites en ce qui concerne les exigences en matière de données d'entrée. Ces modèles peuvent donner des résultats erronés si une série de données sur les transactions des clients s'étendant sur plus de deux ans est incluse dans le modèle.

### **III. Les modèles de durée de vie (Survival Analysis)**

Selon Lu 2002,[19] la survival analysis, ou analyse de survie, est une branche de la statistique qui étudie le temps jusqu'à l'occurrence d'un événement particulier. L'événement peut être la mort dans le cadre d'études médicales, ou, dans le contexte de la CLV, le désabonnement ou la fin de la relation client « Churn ».

La pertinence de l'analyse de survie pour la CLV découle de sa capacité à traiter les données censurées. Par exemple, dans une étude de CLV, certains clients sont toujours actifs et il n'est pas connu quand ils partiront ou si jamais ils partiront- ce sont des données censurées. Dans ce qui suit nous allons présenter les deux modèles les plus réputés dans l'analyse de survie.

#### **1. Le modèle Proportional Hazards de Cox (PH Cox)**

Le modèle Proportional Hazards de Cox est l'un des modèles les plus populaires en analyse de survie, en partie à cause de sa flexibilité (Katzma, 2018) [20]. Il ne fait pas d'hypothèses strictes sur la forme de la fonction de risque de base, permettant une analyse robuste dans divers contextes.

Les banques, avec leurs vastes ensembles de données sur les clients, peuvent bénéficier particulièrement de l'application du modèle PH Cox pour estimer la CLV. En utilisant ce modèle, les banques peuvent intégrer un large éventail de facteurs, tant démographiques que comportementaux, pour obtenir des estimations précises de la durée de vie du client.

- **Formule du Modèle de Cox**

La formule du modèle de risques proportionnels de Cox est donnée par :

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad [3]$$

où :

- $h(t, X)$  est la fonction de risque à un temps  $t$  pour un individu avec un ensemble de covariables  $X$ .
- $h_0(t)$  est la fonction de risque de base, qui est le risque de subir l'événement à un moment  $t$  pour un individu avec les valeurs de référence des covariables (c'est-à-dire, les covariables sont égales à 0). elle ne varie pas avec les covariables et sert de point de départ pour estimer l'effet multiplicatif des covariables sur le risque d'événement.
- $\exp(\beta_1 X_1 + \dots + \beta_p X_p)$  est le facteur de risque relatif, qui modifie la fonction de risque de base en fonction des valeurs des covariables de l'individu.
- $\beta_1, \beta_2, \dots, \beta_p$  sont les coefficients estimés par le modèle, qui représentent l'influence log-linéaire de chaque covariable sur le logarithme du risque de l'événement.

Ces coefficients sont typiquement estimés à l'aide de la méthode du maximum de vraisemblance partielle, qui est adaptée pour gérer les données censurées couramment rencontrées dans les études de survie.

- **Hypothèses du Modèle**

L'hypothèse de risques proportionnels est un fondement clé du modèle de Cox. Elle stipule que les effets des covariables (prédicteurs) sur le risque sont multiplicatifs et constants dans le temps. Autrement dit, le rapport des risques entre deux individus est supposé rester constant sur la durée de l'étude.

Pour un modèle de Cox ayant des covariables  $X$  et des coefficients  $\beta$ , l'hypothèse de risques proportionnels peut être formellement exprimée comme suit :

Si nous avons deux individus,  $i$  et  $j$ , avec des vecteurs de covariables  $X_i$  et  $X_j$ , l'hypothèse de risques proportionnels suggère que le rapport des fonctions de risque pour ces deux individus, qui est  $h(t, X_j) / h(t, X_i)$ , reste constant dans le temps, peu importe la valeur de  $t$ .

En termes de formule, cela serait :

$$\frac{h(t, X_j)}{h(t, X_i)} = \frac{h_0(t) \exp(\beta' X_j)}{h_0(t) \exp(\beta' X_i)} = \exp(\beta' (X_i - X_j)) \quad [4]$$

Ce rapport est indépendant de  $h_0(t)$ , la fonction de risque de base, et ne dépend que des covariables et de leurs coefficients. Cela signifie que l'effet relatif d'une covariable est le même quel que soit le moment où vous observez le risque. Si cette hypothèse n'est pas vérifiée, les estimations des coefficients et les interprétations des effets des covariables peuvent être biaisées.

Dans la pratique, il est important de tester l'hypothèse de risques proportionnels lors de l'utilisation du modèle de Cox. Des tests statistiques et des graphiques de vérification, tels que les graphiques de Schoenfeld résiduels, peuvent être utilisés pour évaluer si les risques proportionnels sont une hypothèse raisonnable pour les données à l'étude.

- **Coefficients du Modèle**

Les coefficients  $\beta$  du modèle de Cox, sont estimés à partir des données par la méthode du maximum de vraisemblance, ils quantifient l'impact de chaque variable sur le risque d'événement. Un  $\beta$  positif augmente le risque; un  $\beta$  négatif le diminue. Ils sont essentiels pour évaluer comment les attributs des clients affectent la durée de leur relation avec la banque.

- **La fonction de perte du modèle de Cox Proportionnel (CPH) :**

La fonction de vraisemblance partielle du modèle de Cox est donnée par :

$$L(\beta) = \prod_{i: \delta = 1} \frac{\exp(\beta' X_i)}{\sum_{j \in R(t_i)} \exp(\beta' X_j)} \quad [4]$$

où  $\delta_i$  est l'indicateur d'événement (1 si l'événement s'est produit, 0 pour la censure),  $\beta$  est le vecteur des coefficients,  $X_i$  est le vecteur des covariables pour l'individu  $i$ , et  $R(t_i)$  est l'ensemble de risque au temps  $t_i$ , c'est-à-dire l'ensemble des individus pour lesquels l'événement n'a pas encore eu lieu.

- **La relation entre la fonction de risque et la fonction de survie :** La fonction de survie et la fonction de risque sont les deux fonctions fondamentales de l'analyse de survie.

Elles sont étroitement liées, mais représentent des concepts différents dans l'analyse de survie :

- **La fonction de risque** (ou Hazard function), notée  $h(t)$ , décrit la probabilité instantanée qu'un événement se produise à un moment donné  $t$ , étant donné que l'individu a survécu (ou que l'événement ne s'est pas encore produit) jusqu'à ce moment. Pour le modèle de Cox, la fonction de risque est la fonction [3].
- **La fonction de survie**, notée  $S(t)$ , représente la probabilité qu'un individu survit (ou évite l'événement d'intérêt dans notre cas le désabonnement) au-delà d'un certain temps  $t$ . La relation entre la fonction de survie et la fonction de risque est donnée par :

$$S(t) = \exp\left(-\int h(u)du\right) \quad [6]$$

Pour résumer, la fonction de survie est une fonction cumulée qui décroît avec le temps, tandis que la fonction de risque est une fonction instantanée qui peut varier dans différentes directions selon le temps et les covariables.

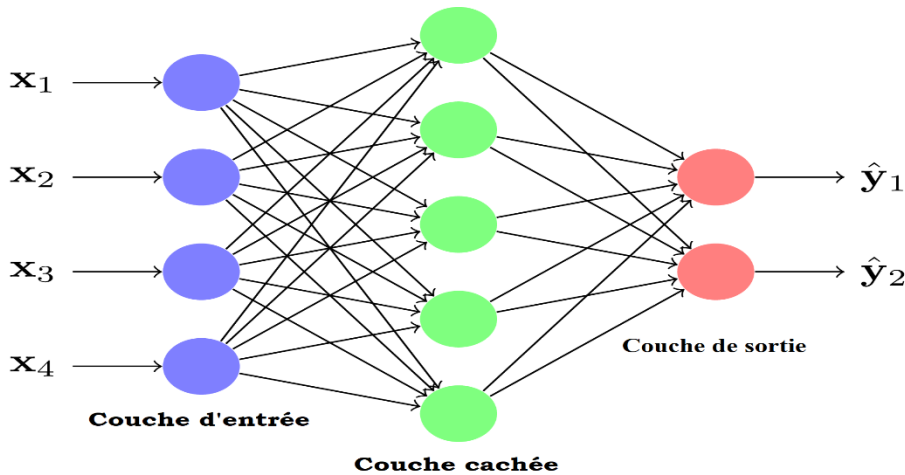
## 2. Présentation des Réseaux de Neurones Artificiel « RNA » et d'Analyse de Survie « Deep Surv »:

Les réseaux de neurones artificiels, souvent appelés Artificiel Neural Networks en anglais, sont conçus en s'inspirant grossièrement de la manière dont les neurones biologiques fonctionnent.

Ces réseaux, grâce à leur aptitude à apprendre, peuvent identifier des patterns complexes parmi de nombreuses variables sans nécessiter d'assistance extérieure. Ils sont idéalement employés pour des tâches telles que l'association, la classification, l'évaluation et la gestion de processus élaborés.

Ces systèmes sont formés de neurones, ou nœuds, interconnectés permettant le transfert de signaux. Ils sont structurés en plusieurs niveaux comprenant une couche d'entrée, une couche de sortie, et une couche intermédiaire, dite cachée. Chacune de ces couches peut être formée d'un ou de plusieurs neurones. Une illustration ci-après dépeint cette structure typique des réseaux de neurones :

**Figure 4 :** Architecture d'un réseau de neurones artificiels



Source : Oludare Isaac Abiodun 2018

L'analyse de survie, un domaine spécifique où les RNA trouvent leur application, bénéficie grandement de cette modélisation non linéaire. Le modèle DeepSurv, en particulier, tire parti de cette capacité pour prédire les risques d'événements en fonction des caractéristiques individuelles, offrant ainsi une approche plus affinée par rapport aux modèles traditionnels. DeepSurv, en tant que modèle de deep learning dédié à l'analyse de survie, représente une avancée significative, capable d'intégrer et de traiter des données complexes pour fournir des prédictions personnalisées et précises concernant la durée des relations client-banque.

- **Le Modèle DeepSurv**

DeepSurv est un modèle de survie qui étend le modèle de Cox à des relations non linéaires entre les covariables et le risque de terminaison de la relation client-banque. Il se sert d'une architecture de réseau de neurones pour capturer ces non-linéarités (Katzma, 2018)[\[20\]](#).

- **Modélisation Non Linéaire**

Contrairement à l'hypothèse linéaire du modèle de Cox, où le log-risque est une combinaison linéaire des covariables ( $\beta'X$ ), DeepSurv utilise un réseau de neurones pour apprendre une fonction de risque potentiellement non linéaire  $h\theta(X)$ .

- **Architecture et Fonction de Risque**

Le réseau est constitué de couches connectées, incluant des couches de "dropout" pour la

régularisation. La sortie du réseau est  $h\theta(X)$ , qui estime le log-risque individuel. Le modèle s'ajuste en minimisant la fonction de perte suivante (Katzma, 2018) [20], qui incorpore la vraisemblance partielle de Cox avec régularisation L2 :

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^N \left( h\theta(X_i) - \log \sum_{j \in R(t_i)} e^{h\theta(X_j)} \right) + \lambda \|\theta\|_2^2$$

où  $N$  est le nombre d'individus avec un événement observé,  $R(t_i)$  est l'ensemble de risque au temps  $t_i$ , et  $\lambda$  est le paramètre de régularisation.

- **Optimisation et Personnalisation**

Les poids du réseau, notés  $\theta$ , sont optimisés à travers la descente de gradient, avec des techniques modernes telles que l'Adam optimizer, pour apprendre les caractéristiques cachées des données. Cela permet à DeepSurv de fournir des prévisions personnalisées basées sur le profil de risque calculé.

En résumé, le modèle DeepSurv modernise l'approche de l'analyse de survie en offrant une meilleure flexibilité pour modéliser les effets complexes des covariables sur la durée de vie des relations client-banque, ce qui est crucial pour évaluer la Customer Lifetime Value et informer les stratégies de gestion de la relation client.

#### **IV. Conclusion**

Ce premier chapitre met l'accent sur l'importance de la rentabilité client et de la Customer Lifetime Value (CLV) pour les banques. Il souligne que les revenus bancaires sont principalement issus des intérêts et des commissions, tandis que la rentabilité est évaluée à travers des indicateurs tels que le produit net bancaire et le coefficient d'exploitation. La gestion stratégique du portefeuille client et la segmentation basée sur la rentabilité sont essentielles pour maximiser les profits et la fidélisation. La CLV est présentée comme un outil crucial pour élaborer des stratégies marketing précises et pour optimiser les investissements dans un marché concurrentiel.



Chapitre deuxième :  
Proposition d'un modèle de calcul de la CLV pour la  
banque CPA.

*"La connaissance est inutile sans l'action.  
C'est par l'application que les théories prennent vie,  
Transformant des idées en résultats tangibles."*  
- Thomas Edison

## **Introduction**

La partie empirique de notre mémoire s'attache à décomposer la complexité inhérente au calcul de la Valeur à Vie du Client (CLV) dans le secteur bancaire. Notre démarche s'ancre dans une application concrète de modèles de survie afin de prédire la durabilité des relations client-banque et d'en déduire la CLV correspondante.

Nous avons mobilisé deux modèles de survie complémentaires : le modèle Cox à risques proportionnels (CPH), reconnu pour son approche semi-paramétrique qui confère une flexibilité considérable dans la modélisation du risque sans postuler de distribution spécifique pour les temps de survie, et le modèle DeepSurv, une approche paramétrique qui, à travers des réseaux de neurones, tente de capturer des relations plus complexes et non linéaires entre les covariables et les temps de survie.

La section méthodologie détaille la préparation des données, le choix des variables explicatives, la mise en œuvre des modèles de survie, et l'évaluation de leur performance. Elle sert de fondement à notre exploration quantitative et guide le lecteur à travers les étapes de modélisation statistique.

La présentation des résultats met en lumière les conclusions tirées des modèles et examine la significativité des variables ainsi que leur capacité à prédire la CLV. Cette section est cruciale pour établir la validité de notre approche empirique.

La discussion qui s'ensuit vise à contextualiser nos résultats dans le cadre opérationnel de la banque, en mettant en exergue les implications stratégiques et opérationnelles des modèles de survie dans l'évaluation de la CLV.

Enfin, la section des recommandations propose des stratégies pratiques basées sur nos analyses pour optimiser la gestion de la relation client et augmenter la rentabilité à long terme. Nous insistons particulièrement sur l'importance de l'amélioration continue des systèmes d'information pour soutenir une prise de décision basée sur des données fiables et précises. Ensemble, ces sections constituent le cœur empirique de notre étude, transformant des concepts théoriques en outils pratiques pour une prise de décision éclairée au sein de l'institution bancaire.

Avant d'entamer tout cela nous vous proposons une brève présentation de la banque Crédit populaire d'Algérie « CPA ».

## Présentation du crédit populaire d'Algérie :

Le Crédit Populaire d'Algérie (CPA) est une institution bancaire établie en 1966, opérant comme l'une des principales banques commerciales du pays. En tant qu'entité détenue entièrement par l'État, le CPA présente un capital social de **48 milliards** de dinars algériens, reflétant sa stature et son importance dans le secteur financier algérien.

La banque s'articule autour de valeurs telles que la modernité, illustrée par son adoption de technologies bancaires avancées, et l'expertise, marquée par un savoir-faire développé au fil des années. L'engagement et la proximité sont également au cœur de ses principes, visant à maintenir une relation étroite et durable avec sa clientèle.

Avec un réseau étendu, le CPA dispose de **160 agences** dispersées à travers l'Algérie, assurant une présence significative sur le territoire. En complément, la banque a introduit 96 guichets dédiés spécifiquement à la finance islamique et a développé une vingtaine d'espaces digitaux pour faciliter l'accès aux services bancaires.

L'innovation dans le domaine numérique est une priorité pour le CPA, qui a lancé des produits de banque en ligne comme Mobile CPA et e-banking, ainsi que le service e-cp@, appuyés par des centres d'appels, y compris un dédié aux produits monétiques pour assurer une assistance continue.

Avec plus de **4451 collaborateurs**, le CPA s'engage à réaliser ses objectifs et à demeurer un partenaire de choix pour les entreprises de toutes tailles, des start-ups aux professionnels établis.

Sur le plan financier, le produit net bancaire de la banque a significativement progressé, avec un total de **70 148 million** de dinars algériens en 2022 et un résultat net de **37 500 million** de dinars algériens, témoignant de la solidité et de la croissance de ses activités.

Le CPA s'identifie également comme une banque citoyenne, engagée dans sa responsabilité sociétale. Elle participe activement en tant que partenaire, sponsor et mécène à diverses actions et événements qui visent le développement et la promotion de l'Algérie.

## Section 01 : Méthodologie

### I. Collecte de données

Dans le cadre de cette étude, les données utilisées ont été fournies par la banque et représentent une base de données comprenant des informations sur 10 000 clients. Ces données sont essentielles pour notre analyse de la durée de la relation client et son impact sur la rentabilité du portefeuille client. La base de données contient les variables suivantes :

- Score de crédit : Cette variable reflète le score de crédit attribué à chaque client, ce qui peut être un indicateur de la fiabilité financière du client.
- Nombre de produits : Il s'agit du nombre de produits financiers ou de services détenus par chaque client au sein de la banque, ce qui peut influencer leur relation avec la banque.
- Agence : Cette variable représente l'agence de la banque à laquelle chaque client est associé, ce qui peut être pertinent pour comprendre les variations régionales.
- Solde : Il s'agit du solde sur les comptes de chaque client, une variable financière clé qui peut avoir un impact sur leur relation avec la banque.
- Salaire estimé : Cette variable indique le salaire estimé de chaque client, ce qui peut être lié à leur capacité financière.
- Devise du compte : Cette variable précise la devise dans laquelle les comptes des clients sont libellés, ce qui peut être pertinent pour l'analyse des flux financiers internationaux.
- Âge : L'âge des clients est une variable importante, car elle peut influencer leur comportement financier et leur durée de relation avec la banque.
- Lieu de naissance : Cette information donne des détails sur le lieu de naissance de chaque client, ce qui peut être pertinent pour les analyses géographiques.
- Profession : La profession des clients peut avoir un impact sur leurs besoins financiers et leur engagement envers la banque.
- Genre : Cette variable indique le genre de chaque client, ce qui peut être un facteur dans l'analyse de la durée de la relation client.
- Actif ou non : Cette variable précise si le client est actif ou non, ce qui peut être lié à leur utilisation des services bancaires.

- Carte de crédit : Cette information indique si le client détient une carte de crédit de la banque.
- Capacité juridique : Il s'agit d'une variable qui peut être importante pour l'analyse légale et contractuelle de la relation client.
- Date d'ouverture du compte : Cette variable donne la date à laquelle chaque client a ouvert son compte auprès de la banque, ce qui peut être pertinent pour les analyses temporelles.
- Exited : Une variable binaire (souvent appelée variable de censure) indiquant si l'événement d'intérêt (le client quittant la banque) s'est produit (1) ou non (0).
- Durée de la relation : Une variable continue représentant le temps écoulé jusqu'à l'événement d'intérêt (dans notre cas combien de temps le client est resté avec la banque).

Parmi ces variables, nous avons sélectionné un sous-ensemble de 10 variables jugées essentielles pour notre analyse, notamment :

- Solde
- Salaire estimé
- Âge
- Genre
- Carte de crédit
- Nombre de produits
- Durée de la relation
- Exited
- Score de crédit

Ces variables ont été retenues en raison de leur pertinence potentielle dans l'analyse de la durée de la relation client et de leur impact sur la rentabilité du portefeuille client de la banque.

## **II. Prétraitement des données**

Le prétraitement des données est une étape cruciale dans toute analyse de données, visant à

garantir la qualité et la pertinence des données utilisées dans l'étude. Dans le cadre de cette étude, nous avons effectué un processus de prétraitement des données, comprenant les étapes suivantes :

Nettoyage des données : Nous avons d'abord procédé au nettoyage des données pour identifier et traiter toute donnée incohérente, erronée ou manquante. Toutes les données manquantes ont été gérées conformément aux besoins de chaque variable.

- Pour les variables numériques telles que 'Score de crédit', 'Âge', 'Solde', et 'Salaire Estimé', les valeurs manquantes ont été remplacées par leur moyenne respective. Cela permet de préserver l'intégrité des données en évitant la suppression de lignes entières.
- Pour la variable catégorielle 'Genre', nous avons utilisé les variables muettes (dummy) pour représenter les catégories masculines et féminines, ce qui a permis d'intégrer cette variable dans nos modèles.

Normalisation des données : Afin de faciliter l'entraînement des modèles statistiques, nous avons normalisé les variables numériques continues telles que 'Score de crédit', 'Âge', 'Solde', et 'Salaire Estimé'. La normalisation des données vise à mettre toutes les variables sur une échelle commune, généralement entre 0 et 1, ce qui garantit que les modèles sont sensibles à toutes les variables de manière équilibrée.

En somme, le prétraitement des données a permis d'obtenir un ensemble de données propre, sans valeurs manquantes et avec des variables normalisées, prêt à être utilisé pour l'entraînement et la validation de nos modèles statistiques visant à comprendre la durée de la relation client et son impact sur la rentabilité du portefeuille client de la banque.

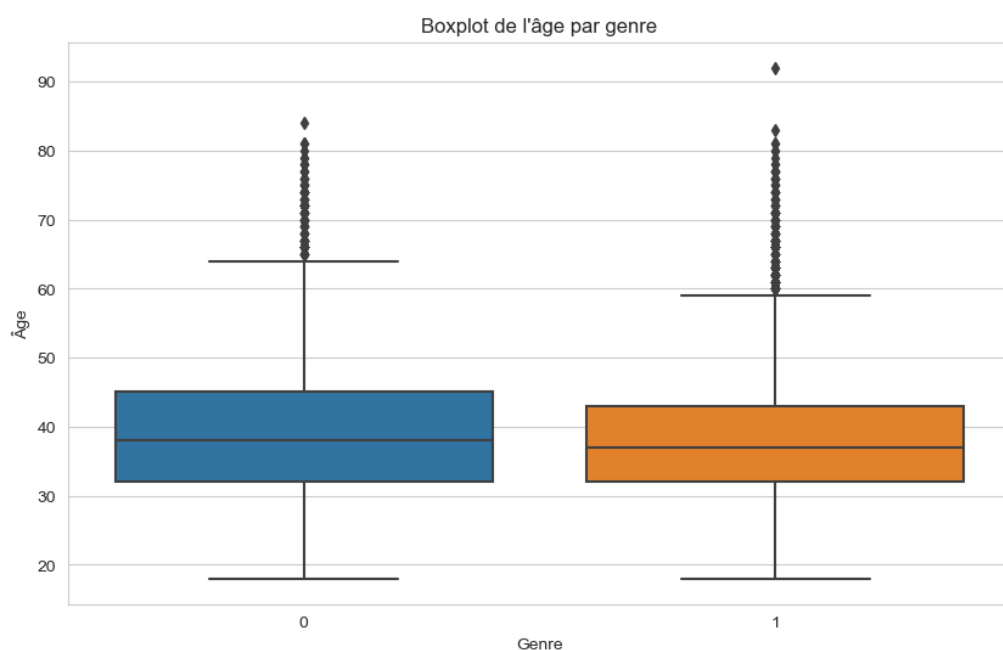
### **III. Exploration des données**

#### **1. Analyse descriptive**

A partir des graphiques nous allons essayer de visualiser la distribution de quelques variables construisant notre modèle.

#### **❖ La distribution de la variables Age par le genre**

**Figure 5 : Boîtes à moustaches de l'Age par le Genre**



*Source : élaboré par nos propres soins*

Le boxplot ci-dessus illustre la distribution de l'âge des individus en fonction de leur genre. Ou le signe "0" représente les femmes et "1" les hommes.

▪ Médiane :

Pour les femmes (bleu), la médiane est légèrement au-dessus de 40 ans. Cela signifie que la moitié des individus de ce genre ont moins de 40 ans et l'autre moitié a plus de 40 ans.

Pour les hommes (orange), la médiane se situe un peu en dessous de 40 ans.

▪ Quartiles :

Le fond de la boîte indique le 1er quartile (Q1) et le haut de la boîte indique le 3ème quartile (Q3). La hauteur de la boîte (l'intervalle interquartile, IQR) représente la répartition centrale de 50% des données.

Pour les femmes, Q1 est d'environ 30 ans et Q3 est proche de 50 ans.

Pour les hommes, Q1 est légèrement supérieur à 30 ans tandis que Q3 est un peu en dessous de 50 ans.

▪ Moustaches :

Les moustaches indiquent la dispersion des données en dehors des quartiles.

Les points au-dessus des moustaches sont considérés comme des "valeurs aberrantes". Cela ne signifie pas nécessairement qu'il s'agit d'erreurs, mais simplement qu'elles se situent en dehors de la distribution typique pour ces données.

- Valeurs aberrantes :

Il y a un nombre considérable de valeurs aberrantes pour les deux genres, principalement pour les âges supérieurs à 70 ans. Cela pourrait suggérer qu'il y a relativement peu d'individus dans ces tranches d'âge, mais ceux qui sont présents ont des âges bien au-dessus de la médiane.

- Comparaison entre les genres :

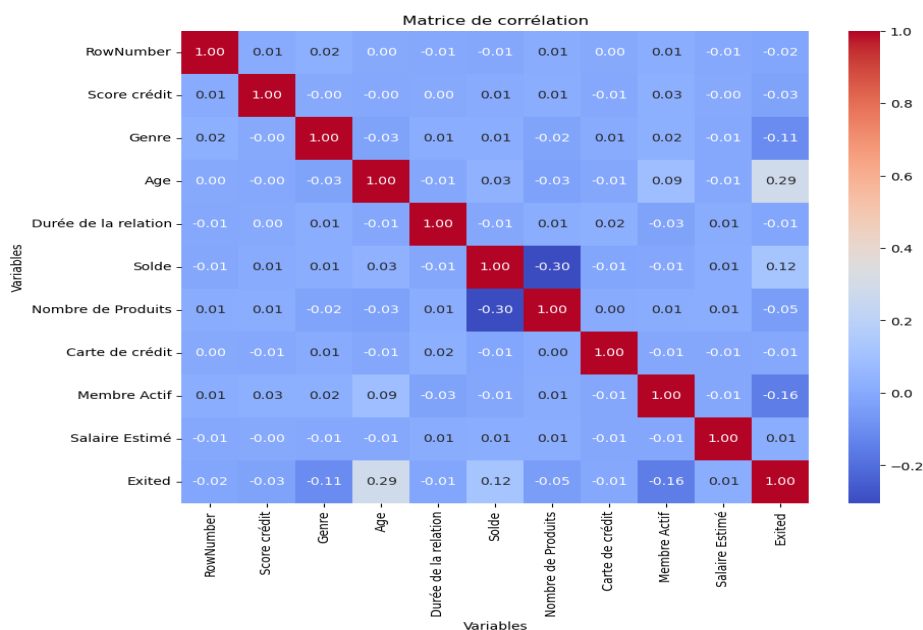
Globalement, la distribution de l'âge est assez similaire entre les deux genres, bien que les femmes semblent avoir une médiane légèrement plus élevée.

En conclusion, ce boxplot montre que la majorité des individus, indépendamment de leur genre, ont un âge compris entre 30 et 50 ans, avec quelques individus ayant plus de 70 ans. La distribution de l'âge est légèrement différente entre les deux genres, mais pas de manière significative.

## 2. Étude de la corrélation

Une matrice de corrélation a été calculée, afin d'identifier les éventuelles corrélations entre les variables. Ci-dessous une visualisation de cette matrice à l'aide d'une heatmap :

**Figure 6: Matrice de corrélation**



Source : python



Globalement, il n'y a pas de corrélations extrêmement fortes entre les variables, ce qui est généralement une bonne nouvelle car cela réduit le risque de multicollinéarité lors de la construction du modèle. Les corrélations les plus notables sont entre la variable « Age » et « Exited ». Cela suggère que l'âge peut avoir une influence sur la décision de quitter la banque. et entre « Membre Actif » et le fait de quitter la banque « Exited » suggérant que les membres actifs sont moins susceptibles de quitter la banque.

#### **IV. Division des données**

Comme cité précédemment, notre étude s'appuie sur les données de 10 000 clients, que nous avons gardés anonymes. La base de données a été divisée : 80% comme données d'entraînement pour créer notre modèle et 20% comme données de test pour vérifier si notre modèle fonctionne bien.

##### 1. Modélisation de la probabilité de survie

###### **❖ Mise en place du modèle de DeepSurv à l'ensemble d'entraînement**

Nous utilisons le modèle de Cox dans le but d'analyser les facteurs influençant la durée de la relation client avec la banque

Dans la première étape, un « *Data Frame* » d'entraînement « *train\_df* » a été créé en combinant les attributs des clients de « *X\_train* » avec les deux variables cibles : la 'Durée de la relation', qui représente la durée pendant laquelle le client est resté avec la banque, et 'Exited', qui indique si le client a quitté ou non la banque pendant cette période.

Après avoir préparé les données, une instance du modèle de Cox a été créée en utilisant la bibliothèque « *lifelines* » du langage python. Ce modèle a été ajusté aux données d'entraînement en spécifiant les colonnes représentant la durée (Durée de la relation) et l'événement « *Exited* ». Une fois le modèle ajusté, un résumé de ses performances et de ses coefficients est affiché à l'aide de la commande « *print\_summary()* ». Ce résumé offre des insights précieux sur l'importance et la significativité des différentes variables dans le modèle, ainsi que sur l'adéquation globale du modèle aux données. Nous verrons les résultats de ce modèle plus en détail dans la partie résultats.

En résumé, le modèle de Cox est une approche de modélisation qui combine des informations sur les covariables (les variables explicatives ou caractéristiques des clients) avec les données de temps jusqu'à l'événement (la durée de la relation et si le client a quitté ou non) pour estimer la fonction de survie et le risque de l'événement au fil du temps.

### ❖ **Mise en place du modèle de Deep Surv à l'ensemble d'entraînement**

DeepSurv est une bibliothèque d'apprentissage profond pour la recherche sur la survie. Il s'agit d'une adaptation du modèle de survie de Cox qui utilise un réseau de neurones pour prédire le risque d'un événement sur la base des caractéristiques d'un individu. En d'autres termes, DeepSurv est une mise à jour moderne du modèle de Cox, permettant d'apprendre des relations non linéaires entre les variables et le risque d'événement, ce qui est difficile à réaliser avec les modèles de survie traditionnels.

Le modèle DeepSurv que nous avons élaboré est structuré comme un perceptron multicouche. Il comprend une couche d'entrée dotée de 33 neurones qui utilisent la fonction d'activation ReLU, suivie d'une couche cachée de 8 neurones qui emploient la fonction d'activation SELU. Enfin, la couche de sortie utilise une activation linéaire. Pour optimiser la performance du modèle et éviter le surapprentissage,

L'optimisation est réalisée à l'aide de l'algorithme Adam, et un programme d'apprentissage est défini pour ajuster le taux d'apprentissage au fil des époques. De plus, un critère d'arrêt anticipé est utilisé pendant l'entraînement pour prévenir le surajustement et garantir que le modèle conserve la meilleure version de lui-même basée sur les performances de validation, Voir [annexe 5](#).

Il est à noter que DeepSurv prend les principes du modèle de Cox et les étend avec la puissance de l'apprentissage profond, permettant au modèle de capturer des relations non linéaires sans hypothèses préalables sur la forme de la relation entre les covariables et le risque d'événement.

## 2. Évaluation de la performance des deux modèles utilisés

Dans cette section, nous évaluerons les performances des deux modèles, le modèle de *Proportionnel Hazard de Cox (PH Cox)* et le modèle de réseaux de neurones *DeepSurv*, pour prédire la durée de la relation client. Nous utiliserons différentes méthodes d'évaluation, notamment des métriques quantitatives et des visualisations. Nous comparerons les modèles en utilisant des courbes ROC, le C-index, et des courbes de survie générées visuellement. Ces

évaluations nous aideront à choisir le modèle le plus approprié pour nos analyses futures.

- **Le C-index** : ou indice de concordance, est une mesure utilisée dans les études de survie pour évaluer la précision d'un modèle de prévision. Il reflète la capacité du modèle à discriminer correctement entre les sujets qui ont des événements (par exemple, la fin d'une relation client-banque) plus tôt et ceux qui les ont plus tard.

La formule du C index est la suivante :

$$C = \frac{\text{Nombre de paires concordantes}}{\text{nombre total de paire comparables}}$$

**Paires concordantes** : une paire est considérée comme concordante si, parmi deux sujets (clients), celui qui a l'événement en premier a aussi un risque prédit plus élevé.

**Paires comparables** : une paire de sujets est considérée comme comparable si l'un des sujets a eu l'événement et l'autre sujet soit a eu l'événement à un moment plus tard ou n'a pas encore eu l'événement (censuré).

Un C index de 1 signifierait une prédiction parfaite, tandis qu'un score de 0,5 indiquerait une performance de prédiction non meilleure qu'un tirage au sort. En général, un C index supérieur à 0,7 est considéré comme indiquant une bonne capacité prédictive du modèle.

- **La courbe ROC** est un outil graphique essentiel pour évaluer la capacité d'un modèle à discriminer entre des classes positives et négatives à différents seuils. L'aire sous la courbe (AUC) ROC est une mesure quantitative de cette capacité.

### 3. Calcul de la CLV :

La Customer Lifetime Value (CLV) est une métrique fondamentale pour évaluer la valeur future d'un client pour une entreprise. La formule de base de la CLV, que nous avons déjà introduite [1], se définit comme suit :

$$CLV = \sum \frac{(P_t - C_t) R_t}{(1+i)^t} - AC \quad [1]$$

Où :

- $P_t$  est le profit à la période  $t$ ,
- $C_t$  est le coût à la période  $t$ ,
- $R_t$  est le taux de rétention à la période  $t$ ,
- $i$  est le taux d'actualisation,
- $AC$  est le coût d'acquisition du client.

En raison de l'indisponibilité des données concernant le coût d'acquisition ainsi que la marge mensuelle par client, qui correspond à la différence entre les revenus ( $P_t$ ) et les coûts ( $C_t$ ), ou, pour employer le terme spécifique au secteur bancaire, le Produit Net Bancaire (PNB) par client, nous avons procédé à des ajustements méthodologiques.

Premièrement, nous avons pris la décision délibérée d'exclure le coût d'acquisition de la formule utilisée pour le Calcul de la Valeur à Vie du Client (CLV). Cette décision est fondée sur l'hypothèse suivante :

- Hypothèse 1 : Le coût d'acquisition ne sera pas incorporé dans le calcul de la CLV.

De plus, étant donné l'absence de données spécifiques au PNB par client au sein du Crédit Populaire d'Algérie, nous avons opté pour une approche simulée afin de pallier cette lacune. Cette démarche repose sur la seconde hypothèse :

- Hypothèse 2 : Une simulation du Produit Net Bancaire (PNB) par client sera mise en œuvre en raison de l'absence de cette métrique clé.

Néanmoins, il convient de souligner que l'ambition principale de notre étude demeure inchangée : elle vise à élaborer et présenter un modèle robuste capable d'évaluer la rentabilité d'un portefeuille client à travers l'analyse de la CLV.

Sur la base de ces deux hypothèses, la formule de la CLV devient :

$$CLV = \sum \frac{PNBc \times Rt}{(1+i/12)^t} \quad [8]$$

Où : PNBc = Produit Net Bancaire par client

Après avoir mené des entretiens approfondis avec les responsables du contrôle de gestion et de la comptabilité, il ressort clairement que la banque ne dispose pas actuellement d'un système de comptabilité analytique lui permettant de déterminer précisément les revenus et les coûts associés à chaque client, rendant ainsi difficile l'évaluation du Produit Net Bancaire (PNB) par client. Par ailleurs, il est également à noter que le service marketing ne dispose pas d'un outil de gestion de la relation client (CRM) avancé, capable de fournir des informations détaillées sur le coût d'acquisition des clients. Nous avons établi dans le premier chapitre que de telles analyses pourraient être facilitées par l'implémentation d'une approche de rentabilité client (ARC). Cependant, il est important de souligner que ce sujet ne constitue pas l'objet de notre étude actuelle.

## **Section 02 : Résultats**

### **I. Comparaison et évaluation de la performance des deux modèles :**

Cette section vise à comparer et évaluer les performances des modèles PH Cox et DeepSurv, en se basant sur trois critères clés : 1) l'indice de concordance (C-index), qui mesure la capacité prédictive du modèle, 2) la courbe ROC, qui évalue la spécificité et la sensibilité du modèle, et 3) la visualisation graphique des courbes de survie, qui illustre la probabilité de survie dans le temps. Ces outils d'analyse fourniront un aperçu approfondi de l'efficacité de chaque modèle dans le contexte de la prédiction des courbes de survie des clients.

#### **1. C-index :**

La première mesure qu'on va utiliser pour comparer la performance des deux modèles est le C-index.

- Données d'entraînement :

Sur les données d'entraînement, le modèle PH Cox a obtenu un C-index de 0.7138, tandis que le modèle DeepSurv ANN a enregistré un C-index de 0.6523. Cela indique que le modèle PH Cox a une meilleure capacité à prédire la survie des clients par rapport au modèle DeepSurv ANN<sup>1</sup> dans ce contexte spécifique. La différence de 0.0615 entre les deux modèles, bien que non énorme, est significative et suggère une meilleure performance du modèle PH Cox.

- Données de validation :

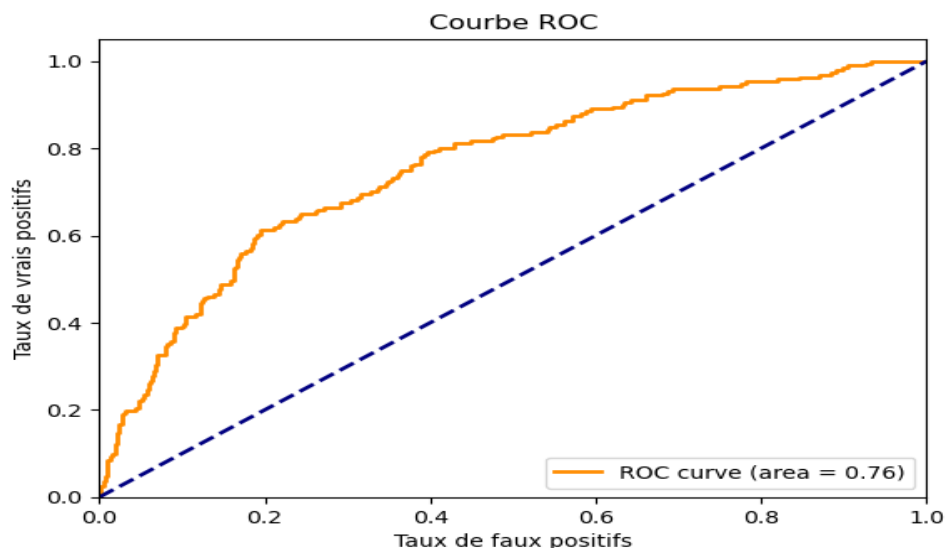
A l'image des données d'entraînement, le modèle PH Cox performe sur les données de validation avec un c-index de 0,7188 contre 0,66 pour le modèle de deep surv. Suggérant ainsi la validation du modèle de PH Cox.

Cependant, il est important de noter que bien que le modèle PH Cox surpasse DeepSurv ANN dans ce cas, cela ne signifie pas nécessairement qu'il est toujours supérieur. D'autres facteurs tels que la complexité du modèle, le temps de calcul, et la nature des données peuvent influencer le choix du modèle le plus approprié.

Dans la section suivante, nous poursuivrons la comparaison en examinant les courbes ROC des deux modèles.

2. Courbe ROC :

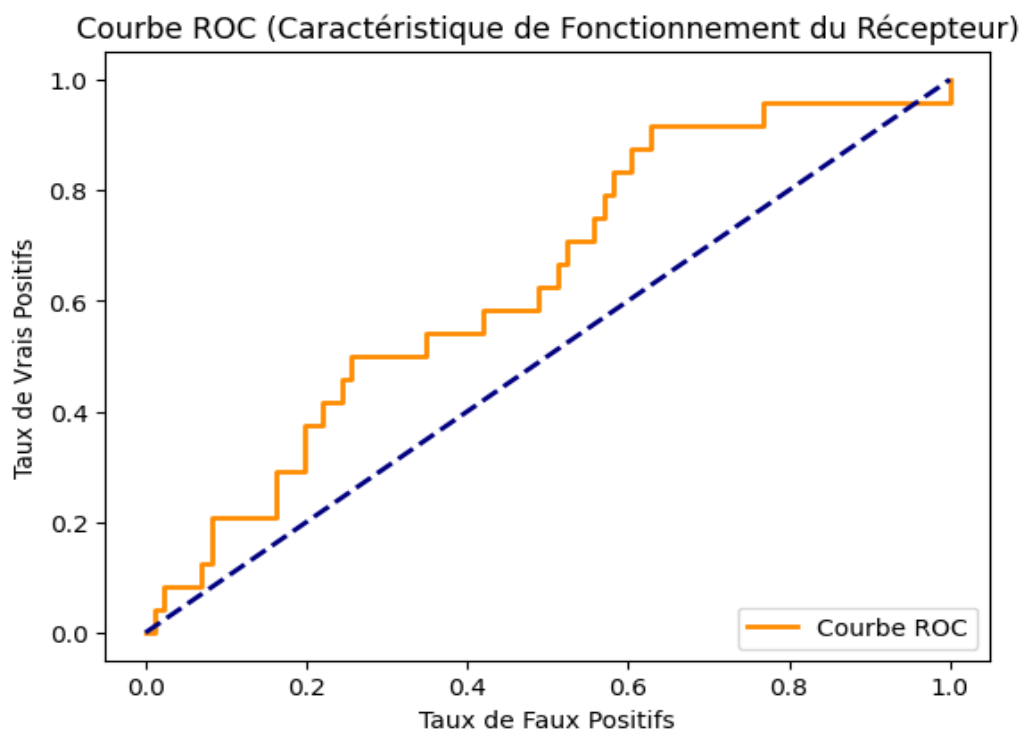
**Figure 7 :** La courbe ROC du modèle de PH COX



Source : python

<sup>1</sup> ANN : Artificial neural network

**Figure 8:** La courbe ROC du modèle DeepSurv



*Source : python*

Interprétation :

- **Modèle DeepSurv ANN :** La courbe ROC pour le modèle DeepSurv ANN présente une AUC de 0.6371. Cette valeur indique une capacité de discrimination modérée.
- **Modèle PH Cox :** La courbe ROC pour le modèle PH Cox affiche une AUC de 0.7576, démontrant une meilleure capacité de discrimination par rapport au modèle DeepSurv ANN.

En se basant sur l'AUC<sup>2</sup>, le modèle PH Cox démontre une nette supériorité en termes de capacité de discrimination par rapport au modèle DeepSurv ANN. Avec un AUC de 0.7576 pour le modèle PH Cox contre 0.6371 pour le DeepSurv ANN, il est clair que le modèle PH Cox est plus performant dans la discrimination des classes positives et négatives.

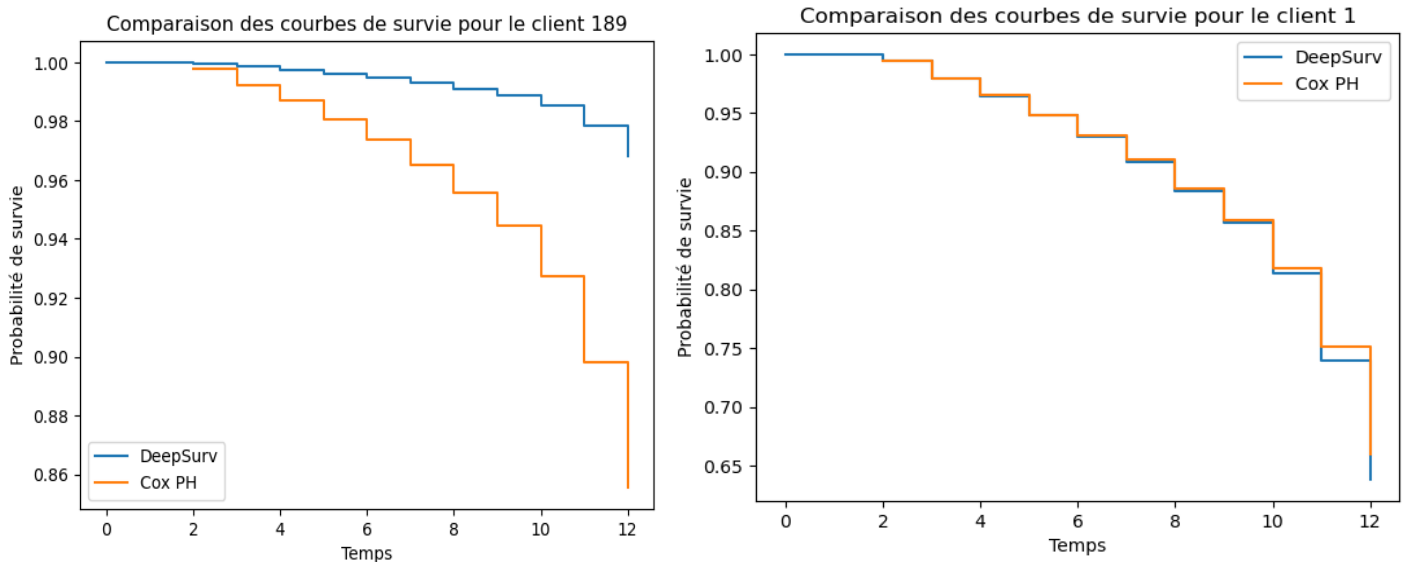
Il est intéressant de constater que le modèle PH Cox excelle à la fois en termes de C-index et d'AUC, renforçant l'idée qu'il est plus performant que le modèle DeepSurv ANN dans ce contexte d'étude.

---

<sup>2</sup> AIR UNDER THE CURVE

3. Comparons visuellement ces courbes de survie pour les deux modèles, DeepSurv et Cox PH.

**Figure 9 : Exemples de courbes de survie selon les deux modèles**



Source : python

**🚦 Client N° 189 :**

- DeepSurv : La courbe bleue représente le modèle DeepSurv pour le client 189. Elle montre une diminution plus progressive de la probabilité de survie au fil du temps.
- Cox PH : La courbe orange représente le modèle Cox PH pour le client 189. On observe une diminution plus nette de la probabilité de survie, avec des chutes plus marquées à certains moments.

Pour ce client, bien que les deux modèles montrent une diminution de la probabilité de survie avec le temps, le modèle Cox PH présente des chutes plus marquées, ce qui peut indiquer une prévision plus pessimiste ou simplement une réactivité plus grande à certains facteurs de risque.

**🚦 Client N° 01 :**

- DeepSurv : La courbe bleue représente le modèle DeepSurv pour le client 1. La courbe montre une diminution relativement linéaire de la probabilité de survie jusqu'à environ le temps 10, où il y a une chute notable.



- Cox PH : La courbe orange pour le client 1 montre une chute plus rapide de la probabilité de survie par rapport au modèle DeepSurv, en particulier autour du temps 6.

Pour le client N° 01, le modèle DeepSurv présente une diminution plus régulière et constante de la probabilité de survie jusqu'à une chute notable vers la fin. En revanche, le modèle Cox PH montre une chute plus rapide au milieu du temps observé, indiquant une estimation plus conservatrice de la survie pour ce client.

Dans l'ensemble, ces courbes visuelles permettent d'observer les différences dans les prévisions de survie des deux modèles. Chaque modèle peut interpréter les données d'une manière légèrement différente, ce qui se traduit par des variations dans les courbes de survie. Pour déterminer quel modèle est le plus approprié, il est important de considérer d'autres facteurs tels que la compréhensibilité du modèle, la facilité d'interprétation, et d'autres métriques de performance en plus des courbes de survie.

#### 4. Choix du modèle performant

Les différentes métriques utilisées pour évaluer les performances des deux modèles, DeepSurv et Cox PH, montrent des résultats variés :

##### ❖ C-index :

Le modèle Cox PH a un C-index de 0.7138, tandis que le modèle DeepSurv a un C-index de 0.6523. Un C-index plus élevé indique une meilleure concordance entre les prédictions du modèle et les événements observés. Ainsi, sur la base du C-index seul, le modèle Cox PH semble être le plus performant.

##### ❖ AUC (Aire Sous la Courbe) :

L'AUC pour le modèle DeepSurv est de 0.6371 et pour le modèle Cox PH, elle est de 0.7576. Une AUC plus élevée signifie que le modèle a une meilleure capacité à distinguer entre les clients qui vont présenter un événement (dans ce cas, possiblement quitter la banque ou cesser d'être rentable) et ceux qui ne le feront pas. Encore une fois, le modèle Cox PH est supérieur en termes de performance.

### ❖ Courbes de survie :

L'analyse visuelle des courbes de survie montre que, bien que les deux modèles prévoient une diminution de la probabilité de survie au fil du temps, leurs trajectoires et chutes sont différentes, reflétant des interprétations différentes des facteurs de risque et de survie.

### ii. Synthèse des résultats:

Bien que le modèle de réseaux de neurones DeepSurv En considérant les métriques et l'analyse visuelle, le modèle Cox PH semble être globalement plus performant que le modèle DeepSurv pour cette analyse particulière.

il est important de souligner que, bien que le modèle DeepSurv soit théoriquement plus performant en raison de sa capacité à détecter les relations non linéaires entre les covariables, le modèle Cox PH a montré un indice de concordance (c-index) supérieur. Cela suggère que, pour cette analyse spécifique, le modèle Cox PH est plus adapté. L'évaluation globale, en prenant en compte à la fois les métriques et l'analyse visuelle des résultats, indique que le modèle Cox PH excelle dans la prédiction des issues comparativement à DeepSurv.

### iii. Mise en place du modèle de Cox :

#### 1. Vérification de L'hypothèse du Modèle de Cox:

L'une des hypothèses fondamentales du modèle de Cox est celle des risques proportionnels. Cette hypothèse stipule que le rapport des risques entre deux individus demeure constant dans le temps. Pour vérifier cette hypothèse, nous utilisons les résidus de Schoenfeld qui permettent d'examiner si les résidus sont indépendants du temps. En pratique, une violation de cette hypothèse se traduit par une tendance notable dans les résidus par rapport au temps.

Le tableau ci-dessous résume les hypothèses testées pour chaque variable et les valeurs p associées aux tests de proportionnalité des risques basés sur les résidus de Schoenfeld.

**Tableau 2 :** Test d'hypothèse des risques proportionnels

		test_statistic	p	-log2(p)
Age	km	0.02	0.89	0.17
	rank	0.00	0.95	0.07
Carte de crédit	km	0.00	0.95	0.08
	rank	0.14	0.71	0.49
Genre	km	0.21	0.65	0.63
	rank	0.07	0.79	0.33
Membre Actif	km	1.54	0.21	2.22
	rank	0.97	0.32	1.63
Nombre de Produits	km	3.61	0.06	4.12
	rank	2.23	0.14	2.88
RowNumber	km	0.65	0.42	1.25
	rank	1.43	0.23	2.11
Salaire Estimé	km	4.00	0.05	4.46
	rank	4.85	0.03	5.18
Score crédit	km	0.04	0.84	0.25
	rank	0.03	0.88	0.22
Solde	km	3.68	0.06	4.18
	rank	2.80	0.09	3.41

Source : python

A partir de cet output nous allons essayer de trouver les variables qui présentent un risque proportionnel de celles qui présentent un risque non proportionnel. Dans le tableau si dessous nous récapitulons les résultats obtenus.

**Tableau 3 :** Tests de proportionnalité des risques basés sur les résidus de Schoenfeld

Variable	Hypothèse nulle (H0)	Hypothèse alternative (H1)	P-value (rank-transformed)	P-value (km-transformed)	Constat
<b>Âge</b>	Risques proportionnels avec 'Âge' dans le temps.	Risques non proportionnels avec 'Âge' dans le temps.	0,9544	0,8893	✓ On accepte H0
<b>Membre Actif</b>	Risques proportionnels avec 'Membre Actif' dans le	Risques non proportionnels avec 'Membre Actif' dans le	0,3241	0,2145	✓ On accepte H0

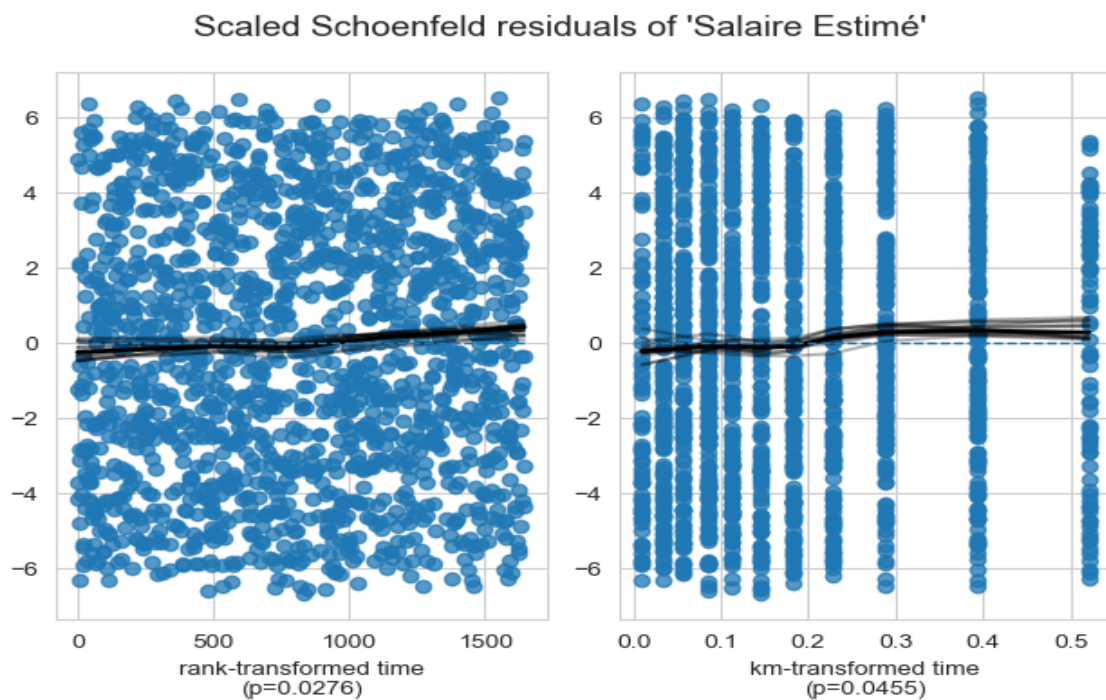
	temps.	temps.			
<b>Score crédit</b>	Risques proportionnels avec 'Score crédit' dans le temps.	Risques non proportionnels avec 'Score crédit' dans le temps.	0,8604	0,8433	✓ On accepte H0
<b>Genre</b>	Risques proportionnels avec 'Genre' dans le temps.	Risques non proportionnels avec 'Genre' dans le temps.	0,7934	0,6462	✓ On accepte H0
<b>Solde</b>	Risques proportionnels avec 'Solde' dans le temps.	Risques non proportionnels avec 'Solde' dans le temps.	0,0941	0,0552	✓ On accepte H0
<b>Nombre de Produits</b>	Risques proportionnels avec 'Nombre de Produits' dans le temps.	Risques non proportionnels avec 'Nombre de Produits' dans le temps.	0,1356	0.0573	✓ On accepte H0
<b>Carte de crédit</b>	Risques proportionnels avec 'Carte de crédit' dans le temps.	Risques non proportionnels avec 'Carte de crédit' dans le temps.	0.7114	0.9473	✓ On accepte H0
<b>Salaire Estimé</b>	Risques proportionnels avec 'Salaire Estimé' dans le temps.	Risques non proportionnels avec 'Salaire Estimé' dans le temps.	0.0276	0.0455	✗ H0 est refusée

Source : Elabore par nos propres soins

➤ **Interprétation :**

Les résultats du test obtenus démontrent que toutes les variables ont un risque proportionnel (Tableau 3), du fait que la P-value est supérieure au seuil de 0.05. Sauf la variable « Salaire Estimé » dont la p-value est inférieure au seuil de 5% nous devons donc confirmer le rejet de  $H_0$  à l'aide du graphique des résidus de Schoenfeld.

**Figure 10 :** Résidus de Schoenfeld pour la variable « salaire Estimé »



*Source : python*

Distribution des résidus : Les points semblent être répartis de manière aléatoire autour de la ligne horizontale à 0, ce qui est généralement une bonne indication que l'hypothèse des risques proportionnels est respectée. Sauf que la valeur p pour le temps transformé en km (keplan meiler) est de 0,0455, ce qui est également inférieur à 0,05. Cela renforce l'idée que l'hypothèse des risques proportionnels pourrait ne pas être respectée pour la variable salaire estimé.

Bien que la distribution des résidus pour la variable "Salaire Estimé" laisse présager une adéquation avec l'hypothèse des risques proportionnels, les valeurs p indiquent une possible violation de cette hypothèse. Toutefois, il convient de souligner que le non-respect de cette hypothèse pour une variable donnée ne remet pas nécessairement en cause l'intégrité du modèle dans son ensemble.

En dépit des observations faites sur la variable "Salaire Estimé", nous choisissons d'adopter le modèle. La raison étant que la validité globale du modèle peut demeurer intacte même si une seule variable ne respecte pas l'hypothèse des risques proportionnels, à condition que les autres variables la respectent. Par conséquent, il est judicieux de poursuivre l'utilisation de ce modèle, tout en étant conscient des éventuelles contraintes liées à "Salaire Estimé".

## 2. Significativité et Interprétation des coefficients :

Le modèle nous a fourni des coefficients  $\beta$ , estimés à partir des données par la méthode du maximum de vraisemblance. Rappelons qu'un coefficient positif augmente le risque alors qu'un  $\beta$  négatif le diminue.

Voyons aussi l'interprétation de chacun des coefficients et son impact sur le risque, ainsi que leur significativité au seuil de 5%. Pour évaluer la significativité statistique des coefficients, nous testons l'hypothèse nulle  $H_0$  selon laquelle chaque coefficient est égal à zéro, ce qui impliquerait qu'il n'a pas d'effet sur le risque. Si la valeur-p est inférieure à 0.05, nous rejetons  $H_0$  et concluons que la variable associée est significative. Cela signifie que nous avons des preuves suffisantes pour affirmer que la variable a un effet sur le risque, soit en l'augmentant soit en le diminuant, selon le signe de  $\beta$ .

**Tableau 4 : résultats du modèle Ph-COX**

model	lifelines.CoxPHFitter											
duration col	'Durée de la relation'											
event col	'Exited'											
baseline estimation	breslow											
number of observations	8000											
number of events observed	1644											
partial log-likelihood	-12966.94											
time fit was run	2023-11-10 17:07:04 UTC											
	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)	
Score crédit	-0.04	0.96	0.02	-0.09	0.00	0.91	1.00	0.00	-1.77	0.08	3.71	
Genre	-0.40	0.67	0.05	-0.49	-0.30	0.61	0.74	0.00	-7.94	<0.005	48.85	
Age	0.51	1.67	0.02	0.47	0.55	1.60	1.74	0.00	24.35	<0.005	432.80	
Solde	0.22	1.24	0.03	0.16	0.27	1.18	1.31	0.00	8.08	<0.005	50.49	
Nombre de Produits	-0.04	0.96	0.04	-0.13	0.04	0.88	1.04	0.00	-1.05	0.29	1.78	
Carte de crédit	-0.01	0.99	0.05	-0.11	0.10	0.89	1.10	0.00	-0.17	0.87	0.20	
Membre Actif	-0.76	0.47	0.05	-0.87	-0.66	0.42	0.52	0.00	-14.34	<0.005	152.57	
Salaire Estimé	-0.01	0.99	0.02	-0.06	0.04	0.94	1.04	0.00	-0.44	0.66	0.61	
Concordance	0.71											
Partial AIC	25949.87											
log-likelihood ratio test	853.61 on 8 df											
-log2(p) of ll-ratio test	592.11											

Source : python

A partir de ces résultats nous concluons que pour les variables "Genre", "Âge", "Solde", et "Membre Actif", les valeurs-p sont inférieures à 0.05, indiquant que ces variables sont statistiquement significatives dans ce modèle.

En revanche, les autres variables ont une p-value supérieure au seuil de significativité donc elles sont considérées comme des variables n'ayant pas un effet statistiquement significatif sur le risque. Cependant, cela ne signifie pas automatiquement que la variable est sans effet ; il se pourrait que les données ne montrent pas d'effet en raison de la taille de l'échantillon, de la variabilité des données, ou d'autres facteurs. Nous considérons dans ce cas la significativité réelle, c'est-à-dire ayant une importance dans le contexte de l'étude. Ces variables méritent souvent une étude plus approfondie avant de tirer des conclusions définitives sur leur impact. Dans ce qui suit nous allons interpréter l'effet de chaque variable sur le risque de désabonnement du client.

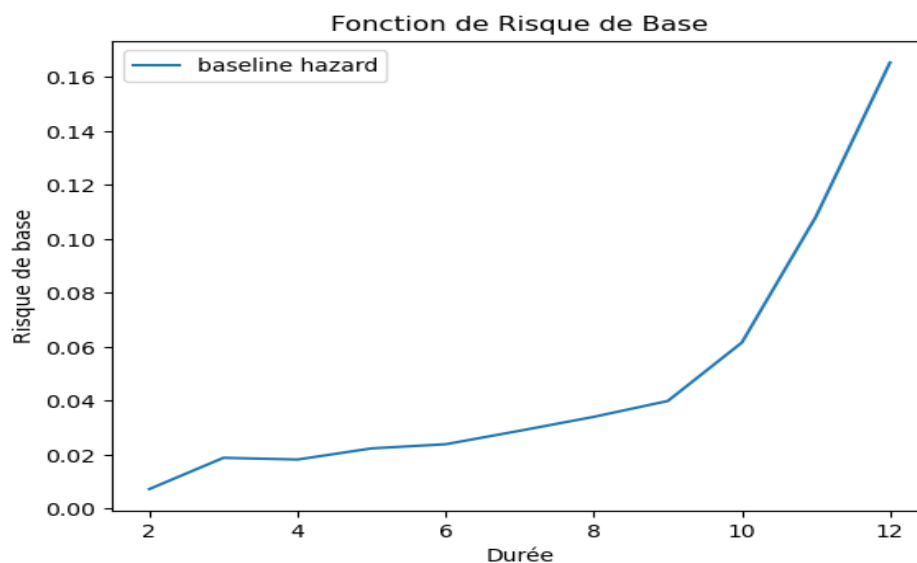
- **Score crédit** : Avec un coefficient de -0,04, cela suggère que pour chaque augmentation d'une unité du score de crédit, le risque diminue, ce qui est cohérent avec l'idée que des scores de crédit plus élevés sont généralement associés à des clients plus fiables.
- **Genre** : Coefficient (coef) : -0.39 : Le signe négatif de ce coefficient suggère qu'il y a une relation inverse entre le genre et le risque de désengagement. cela indique que les hommes (1) ont un risque de désengagement de la banque inférieur par rapport aux femmes (0). L'hazard ratio ( $\exp(\text{coef})$ ) de 0.67 soutient cette interprétation, signifiant que le risque de désengagement pour les hommes est 33% moins élevé que pour les femmes.
- **Âge** : Coefficient : 0.51. L'âge positif indique que, à mesure que l'âge augmente, le risque de désengagement augmente également. Cela pourrait suggérer que les clients plus âgés sont moins susceptibles de maintenir une relation à long terme avec la banque, peut-être en raison d'un changement de priorités ou de besoins financiers.
- **Solde** : Coefficient : 0.22. Un coefficient positif pour le solde suggère que les clients avec des soldes plus élevés ont un risque accru de désengagement. Ceci pourrait être contre-intuitif, mais cela pourrait indiquer que les clients avec des soldes élevés ont plus d'options bancaires disponibles ou sont plus sensibles aux fluctuations du marché.
- **Nombre de Produits** : Coefficient : -0.04. Le coefficient légèrement négatif pour "Nombre de Produits" suggère que, à mesure que le nombre de produits détenus par un client augmente, le risque de désengagement diminue légèrement. Cela peut indiquer que les clients qui détiennent plusieurs produits auprès de la banque ont une relation plus solide et sont moins susceptibles de se désengager.

- **Carte de crédit** : Carte de crédit : Le coefficient de -0,01 indique un effet très faible sur le risque. Étant donné que ce coefficient est presque nul, la possession d'une carte de crédit n'a pratiquement pas d'impact sur le risque de désengagement. La valeur-p est élevée (0,87), ce qui suggère que cet effet n'est pas statistiquement significatif.
- **Membre Actif** : Coefficient : -0.76. Le coefficient nettement négatif pour "Membre Actif" indique que les clients actifs ont un risque significativement réduit de désabonnement. Être actif pourrait impliquer une utilisation régulière des services bancaires, ce qui renforce la relation client-banque. La valeur-p est inférieure à 0,005, indiquant que cet effet est fortement significatif.
- **Salaire Estimé** : Avec un coefficient de -0.01, l'effet du salaire estimé sur le risque de désengagement est très faible, presque négligeable. La valeur-p est de 0,66, indiquant que l'association entre le salaire estimé et le risque de désengagement n'est pas statistiquement significative.

### 3. Fonction de risque (Hazard function)

Pour construire la fonction de risque  $h(t)$  en utilisant le modèle des risques proportionnels de Cox avec les coefficients fournis, nous allons remplacer chaque  $\beta_i$  par sa valeur correspondante donnée dans la **Tableau 4**. Les covariables (notées  $X_i$ ) resteront telles quelles car elles sont spécifiques à chaque individu dans l'ensemble de données. Le risque de base  $h_0(t)$  qui est commune pour tous les clients est représenté par la courbe dans la Figure 13.

**Figure 11** : Fonction de risque de base  $h_0(t)$



Source : python



La figure ci-dessus illustre l'estimation de la fonction de risque de base obtenue à partir de notre ensemble de données. Comme le montre le graphique,  $h_0(t)$  augmente avec le temps, ce qui indique que le risque de l'événement d'intérêt qui est dans notre cas, le désengagement du client - s'accroît au fur et à mesure que la durée augmente. Plus précisément, on observe une accélération dans la croissance du risque à partir de la période du dixième mois.

❖ **L'équation de la fonction de risque :**

L'équation complète avec toutes les variables s'écrit comme suit :

$$H(t) = h_0(t) \exp(-0.20 * \text{Score de crédit} - 0.4 * \text{Genre} + 0.51 * \text{Age} + 0.22 * \text{Solde} - 0.04 * \text{Nombre de Produits} - 0.01 * \text{Carte de crédit} - 0.76 * \text{Membre Actif} - 0.01 * \text{Salaire Estimé})$$

Où :

- $h_0(t)$  est le risque de base à l'instant  $t$ ,

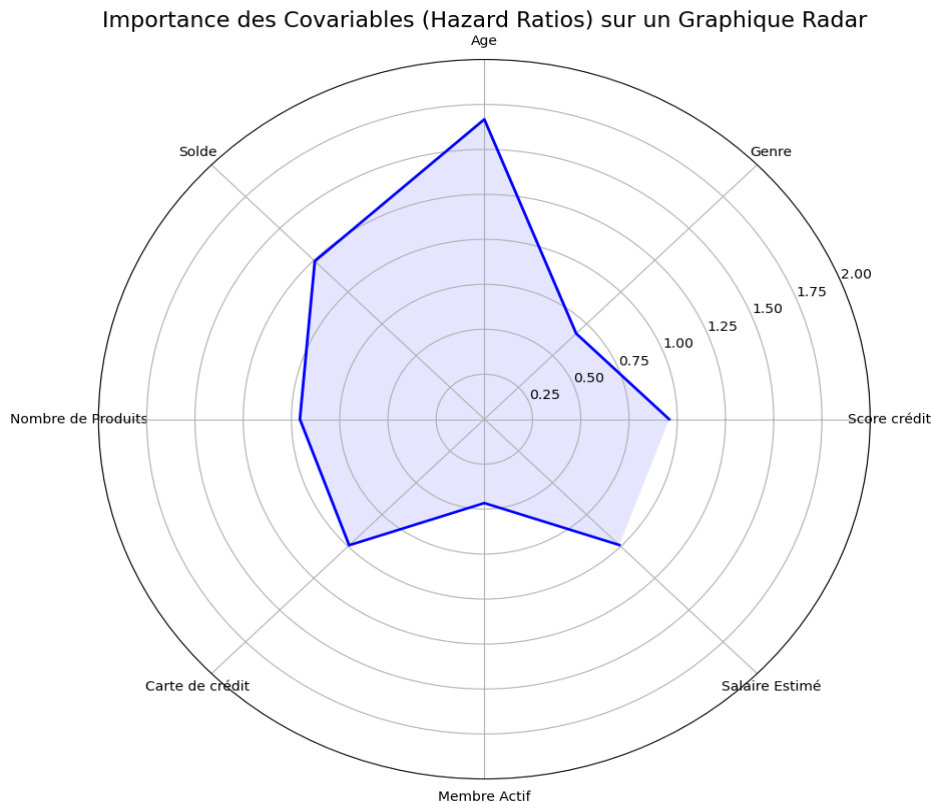
Toutefois une version épurée de cette équation qui ne comprend que les variables statistiquement significatives, se présente ainsi :

$H(t) = h_0(t) \exp(-0.4 * \text{Genre} + 0.51 * \text{Age} + 0.22 * \text{Solde} - 0.76 * \text{Membre Actif})$
--

4. **Interprétation des autres paramètres des modèles :**

- **exp(coef):** Ceci est l'exponentielle du coefficient, souvent appelée "hazard ratio". Un hazard ratio supérieur à 1 indique que la variable est associée à un risque accru de l'événement. Un hazard ratio inférieur à 1 indique un risque réduit. Par exemple, pour le "Genre", l'exp(coef) est de 0.67, ce qui signifie que le risque est réduit de 33% par rapport à la catégorie de référence. De même pour la variable « membre actif » avec un coef de -0.76, le hazard ratio est 0.47, indiquant que les membres actifs ont un risque de désengagement qui est 53% moindre que les membres inactifs (1 - 0.47). la figure 12 illustre l'importance de chaque variable sur le risque de désabonnement.

**Figure 12 : Importance des variables sur le risque de désabonnement**



*Source : python*

**Se (coef):** C'est l'erreur standard du coefficient, qui donne une idée de la variabilité ou de l'incertitude autour de l'estimation du coefficient.

- **coef lower 95% et coef upper 95%:** Ces colonnes fournissent l'intervalle de confiance à 95% pour le coefficient. Si cet intervalle ne contient pas 0, cela renforce l'idée que la variable est significativement associée au risque d'événement.
- **exp(coef) lower 95% et exp(coef) upper 95%:** C'est l'intervalle de confiance à 95% pour le hazard ratio. De même, si cet intervalle ne contient pas 1, cela renforce l'idée que la variable est significativement associée au risque d'événement.
- **z:** C'est la valeur de la statistique de test Z. Elle est calculée comme le coefficient divisé par son erreur standard. Une valeur absolue élevée de z (loin de 0) indique que le coefficient est statistiquement différent de 0.

- **p (-log<sub>2</sub>(p))**: C'est la valeur p transformée. Une petite valeur p indique une association significative entre la variable et le risque d'événement. En général, une valeur p inférieure à 0.05 est considérée comme statistiquement significative, bien que ce seuil puisse varier selon le contexte.

Ces mesures permettent de comprendre l'importance relative et la fiabilité de chaque variable explicative dans la prédiction de la durée de la relation client.

En résumé, en se basant sur le seuil de significativité de 0.05, les variables dont la p-value est inférieure à cette valeur sont considérées comme significatives dans le modèle. Dans les résultats que nous avons obtenus, les variables suivantes ont une p-value inférieure à 0.05 et sont donc statistiquement significatives :

- Solde
- Genre
- Age
- Membre Actif

Il est crucial de souligner que ces variables ont montré une association significative avec la durée de la relation client dans notre modèle de Cox, ces variables peuvent donc avoir un impact important sur la rentabilité d'un portefeuille client au sein de la banque via la CLV.

Les autres variables, bien qu'elles aient des coefficients, n'ont pas montré de significativité statistique à ce seuil, ce qui signifie que nous ne pouvons pas être aussi confiants quant à leur association avec la durée de la relation client par rapport aux variables significatives. Cependant, il convient de noter que le choix du seuil de 0.05 est conventionnel et que la pertinence des variables peut varier en fonction du contexte métier et des objectifs de l'analyse.

## **Synthèse**

L'application de la régression de Cox nous a permis d'identifier les variables clés qui influencent la durée de la relation client avec la banque. Ces insights sont essentiels pour comprendre et optimiser la rentabilité du portefeuille client au sein de la banque à travers la CLV.

#### IV. Calcul du coût du capital WACC

Dans le but de calculer la CLV, nous aurons besoin d'un taux d'actualisation afin d'actualiser qui est généralement le coût du capital de la firme. En fonction des données fournies par la banque nous allons essayer d'estimer le coût moyen pondéré du capital (CMPC ou Wacc) de la banque.

La formule générale du WACC est la suivante :

$$WACC = Re \times \left(\frac{E}{V}\right) + Rd \times (1 - Tc) \times \left(\frac{D}{V}\right)$$

Où :

- $E$  est la valeur de marché des capitaux propres,
- $D$  est la valeur de marché de la dette,
- $V=E+D$  est la valeur totale de l'entreprise,
- $Re$  est le coût du capital propre,
- $Rd$  est le coût de la dette,
- $Tc$  est le taux d'impôt sur les sociétés.

Nous disposons des données suivantes tirées du bilan de la banque :

##### ⇒ **Capitaux propres**

- Capital : 48 000 millions DA
- Réserves : 158 060 millions DA
- Écart de réévaluation : 15 921 millions DA
- Résultat de l'exercice : 30 441 millions DA

Le total des capitaux propres serait donc la somme de ces éléments.

##### ⇒ **Dettes**

- Dettes envers les Institutions Financières : 862 637 millions DA
- Dettes envers la clientèle : 1 812 898 millions DA
- Dettes représentées par un titre : 57 302 millions DA

Le total des dettes serait la somme de ces éléments.

Avec ces informations, nous pouvons calculer la valeur de marché totale de la banque, qui est la somme des capitaux propres et des dettes.

Ensuite, pour calculer le WACC, nous disposons des informations suivantes sur les coûts des dettes

- Coût de la ressource client (1,53%)
- Coût de la ressource sur le marché interbancaire (3%).

Nous utiliserons ces taux pour estimer le coût moyen de la dette.

Le coût du capital propre pourrait être approximé par le ROE qui nous a été donné par la direction du contrôle de gestion de la banque ROE = 12,63%.

Sur la base de ces données, nous allons calculer la structure du capital de la banque et le WACC :

⇒ **Total des capitaux propres**

Total capitaux propres = capital + réserves + écart de réévaluation + résultat exercice

Total capitaux propres = 252 422 millions DA.

⇒ **Total des dettes**

Total dettes = dettes institutions financières + dettes clientèle + dettes représentées par titre

Total dettes = 2 732 837 millions DA.

⇒ **Valeur totale de la banque:**

Valeur totale entreprise = total capitaux propres + total dettes = 2 985 259 millions DA.

⇒ **Structure du capital en pourcentage**

Pourcentage capitaux propres = (total capitaux propres / valeur totale entreprise) \* 100

Pourcentage dettes = (total dettes / valeur totale entreprise) \* 100

En résumé, la valeur totale des capitaux propres est de 252 422 millions DA, et la valeur totale des dettes est de 2 732 837 millions DA. Cela donne une valeur totale de l'entreprise de 2 985 259 millions DA.

La structure du capital en pourcentage est la suivante :

- **Capitaux propres : environ 8,46%**
- **Dettes : environ 91,54%**

Maintenant, pour calculer le WACC, nous aurons besoin de connaître le coût du capital propre et le coût de la dette. Nous prenons le ROE en tant que coût du capital propre, qui est de 12,63%. Pour le coût de la dette, nous allons devoir calculer une moyenne pondérée des différentes sources de dettes en utilisant les taux fournis pour la ressource client et la ressource sur le marché interbancaire. Ensuite, nous ajusterons pour les impôts en utilisant le taux d'impôt sur les sociétés

IBS de 25%.

Le coût de la dette pondéré pourrait être calculé comme suit :

$$Rd \text{ pondere} = \left( \frac{\text{dettes clientele}}{\text{Total dettes}} \times \text{cout ressource client} \right) + \left( \frac{\text{dettes institutions financieres}}{\text{Total dettes}} \times \text{cout ressource interbancaire} \right)$$

Où le coût de la ressource client est de 1,53% et le coût de la ressource sur le marché interbancaire est de 3%.

Calculons le coût de la dette pondéré et ensuite le WACC.

Le coût de la dette pondéré, après avoir pris en compte le coût des ressources clients et interbancaires, est de 1,962%.

$$WACC = 0,1263 \times \left( \frac{252\,422}{2\,985\,259} \right) + 0,01962 \times (1 - 0,25) \times \left( \frac{2\,732\,837}{2\,985\,259} \right)$$

$$WACC = 0,02415$$

En utilisant ce coût de la dette et le coût du capital propre approximé par le ROE de 12,63%, le WACC de la banque, après ajustement pour les impôts, est de **2,415%**.

#### **v. Calcul de la CLV pour les clients de la banque:**

Nous arrivons enfin à l'utilisation du modèle de survie qu'on a retenu dans cette étude qui est le modèle de PH COX, pour calculer la CLV des clients de la banque. Nous allons prendre quelques clients pour monter la formule de calcul de la CLV. Avec un taux d'actualisation qui correspond au coût du capital de la banque de 2,415%.

**Tableau 5 : Calcul de la CLV pour le client A**

	PNBc	Probabilité de survie	Facteur d'actualisation	PNBc actualisé	Probabilité de survie actualisée (PSA)	PNBc actualisé x (PSA)
1	200	1	0,997991542	199,59831	0,997991542	199,1974236
2	198,56	0,997852	0,995987118	197,7632	0,993847738	196,546511
3	199,5	0,992308	0,99398672	198,30035	0,986340974	195,5917609
4	201,5	0,98698	0,991990339	199,88605	0,979074625	195,7033627
5	200,5	0,980485	0,989997968	198,49459	0,970678158	192,6743655
6	202	0,97359	0,988009599	199,57794	0,961916265	191,9772657
7	197	0,965319	0,986025223	194,24697	0,951828882	184,8898754
8	199,5	0,955656	0,984044833	196,31694	0,940408349	184,6180933
9	198	0,944448	0,98206842	194,44955	0,927512555	180,3543964
10	201	0,927417	0,980095977	196,99929	0,908957671	179,064017
11	203	0,898255	0,978127495	198,55988	0,878607913	174,4562832
12	203	0,855562	0,976162967	198,16108	0,835167941	165,4977831
<b>CLV</b>						<b>2240,571138</b>

**Tableau 6: Calcul de la CLV pour le client B**

	PNBc	Probabilité de survie	Facteur d'actualisation	PNBc actualisé	Probabilité de survie actualisé (PSA)	PNBc actualisé x (PSA)
1	207,25	1	0,997991542	206,83375	0,9979915	206,41833
2	205,81	0,994276	0,995987118	204,98411	0,9902861	202,99291
3	206,75	0,979598	0,99398672	205,50675	0,9737074	200,10345
4	208,75	0,96562	0,991990339	207,07798	0,9578857	198,35704
5	207,75	0,94875	0,989997968	205,67208	0,9392606	193,17967
6	209,25	0,931045	0,988009599	206,74101	0,9198814	190,17721
7	204,25	0,910079	0,986025223	201,39565	0,8973608	180,72457
8	206,75	0,885964	0,984044833	203,45127	0,8718283	177,37457
9	205,25	0,858498	0,98206842	201,56954	0,8431038	169,94404
10	208,25	0,817789	0,980095977	204,10499	0,8015117	163,59254
11	210,25	0,750934	0,978127495	205,65131	0,7345092	151,05277
12	210,25	0,659399	0,976162967	205,23826	0,6436809	132,10795
<b>CLV</b>						<b>2166,0251</b>

Grâce à la valeur à vie du client, il est évident que le client A a une valeur plus élevée pour la banque que le client B, même si le client B semble plus rentable que le client A. En appliquant cette démarche à l'ensemble des clients qui construisent le portefeuille d'un charge de clientèle ou même de l'agence bancaire dans sa globalité, les dirigeants auront une vue plus claire sur la rentabilité de chacun des clients et donc du portefeuille en générale. Cela permet aussi de comparer et d'identifier les clients rentables de ceux qui sont moins rentables puis d'adopter les stratégies marketing adéquates. Comme nous l'avons vu dans le premier chapitre concernant « l'exploitation des résultats de rentabilité par le chargé de clientèle ».

### 1. Implications pour la banque

L'implication de ces résultats pour l'analyse de la rentabilité d'un portefeuille client, se concrétise par les éléments suivants :

#### ❖ Prévisions précises :

Un modèle plus précis, comme le modèle Cox PH dans ce cas, peut aider la banque à identifier les clients qui sont susceptibles de ne plus être rentables à l'avenir. Cela permet à la banque de prendre des mesures préventives pour retenir ces clients ou ajuster sa stratégie de portefeuille.

#### ❖ Segmentation des clients :

Les prévisions du modèle peuvent être utilisées pour segmenter les clients en différentes catégories de risque. Par exemple, ceux qui sont à haut risque de ne plus être rentables peuvent être ciblés avec des offres spéciales ou des incitations pour augmenter leur fidélité.

#### ❖ Stratégies d'investissement :

En comprenant le risque associé à chaque client ou segment de clients, la banque peut mieux allouer ses ressources et capital, en investissant davantage dans les segments à faible risque et en mettant en œuvre des stratégies de mitigation pour les segments à risque élevé.

#### ❖ Interprétabilité :

Bien que le modèle Cox PH soit plus performant, il est aussi généralement plus facile à interpréter que les modèles de réseaux de neurones comme DeepSurv. Cette interprétabilité



peut être précieuse pour les décideurs qui souhaitent comprendre les facteurs sous-jacents influençant les prédictions du modèle.

En conclusion, l'utilisation d'un modèle performant et précis pour évaluer la rentabilité future d'un portefeuille client peut offrir des avantages significatifs en termes de planification stratégique, d'allocation des ressources et de prise de décision éclairée pour maximiser la rentabilité et la fidélité des clients.

## **Section 03 : Discussion et recommandations**

### **I. Discussion**

Cette section discute des implications des résultats obtenus par l'analyse comparative de deux modèles prédictifs, le modèle Cox à risques proportionnels (PH) et le modèle DeepSurv, dans le contexte de la prédiction de la durabilité de la relation client-banque et de la propension à la défection des clients. L'étude a exploité un échantillon de 10 000 clients, anonymisé et scindé en un ensemble d'entraînement (80%) et un ensemble de test (20%).

Le modèle PH Cox, employant une méthode semi-paramétrique, a servi à identifier les variables influençant la durée de la relation client avec l'institution bancaire. Grâce à l'utilisation de la bibliothèque 'lifelines' en Python, l'ajustement du modèle a permis une interprétation significative de l'impact de variables distinctes sur la probabilité de départ des clients.

DeepSurv, quant à lui, adapte le modèle de survie de Cox en y intégrant des techniques d'apprentissage profond pour prédire les risques d'événements basés sur des caractéristiques individuelles, permettant ainsi de capter des relations complexes et non linéaires entre les variables. Ce modèle a été structuré comme un perceptron multicouche et optimisé pour éviter le surajustement, ce qui soulève des questions quant à son applicabilité et interprétation dans un cadre opérationnel bancaire.

L'évaluation de la performance a révélé que le modèle Cox PH, avec un indice C de 0,7138, surpassait le modèle DeepSurv qui enregistrait un indice C de 0,6523, soulignant ainsi la capacité supérieure du premier modèle à prédire la fidélité des clients. Néanmoins, cette comparaison des performances doit être mise en perspective avec la nature et la qualité des données utilisées, soulignant l'importance d'une collecte de données fiable et détaillée pour l'entraînement des modèles prédictifs.

La recherche n'a pas révélé de corrélations extrêmement fortes entre les variables, ce qui est

généralement positif puisque cela minimise le risque de multicollinéarité dans la modélisation. Toutefois, cela peut aussi indiquer une complexité inhérente au comportement des clients qui n'est pas entièrement capturée par les variables étudiées.

L'analyse s'est ensuite focalisée sur le calcul de la Customer Lifetime Value (CLV), un indicateur clé de la valeur future d'un client. Face à l'absence de données spécifiques telles que les coûts d'acquisition des clients et le revenu net bancaire par client, la formule de la CLV a été adaptée, ce qui met en lumière la nécessité pour la banque de raffiner ses méthodes de comptabilité pour mieux appréhender la rentabilité individuelle de ses clients.

Les résultats ont mis en exergue l'importance de variables comme le 'Solde', le 'Genre', l' 'Âge' et le statut de 'Membre Actif' qui se sont révélées significatives dans le modèle et qui pourraient donc avoir une influence marquée sur la rentabilité d'un portefeuille de clients.

En conclusion, cette discussion soulève la question de l'utilité pratique de modèles prédictifs précis et robustes pour l'identification de clients potentiellement non rentables, permettant ainsi à la banque d'adopter des mesures préventives ou d'ajuster ses stratégies de portefeuille en conséquence. Bien que le modèle DeepSurv présente théoriquement des avantages en raison de sa capacité à détecter des relations non linéaires, le modèle Cox PH s'est avéré supérieur dans cette analyse spécifique, tant en termes de métriques que d'analyse visuelle.

## **II. Recommandation**

Afin d'améliorer la stratégie de gestion de la relation client et de maximiser l'efficacité des modèles prédictifs à l'image de celui présenté dans cette étude, plusieurs actions sont préconisées. Ces recommandations visent à renforcer l'analyse des données clients et à optimiser les prises de décision marketing et stratégiques :

- Intégrer un système de gestion de la relation client (CRM) pour centraliser et analyser les données clients.
- Mettre en place un programme de formation pour le personnel sur l'utilisation et l'interprétation des données à travers le CRM.
- Assurer un suivi et une mise à jour continus du CRM pour maintenir la précision des données et l'efficacité des modèles prédictifs.
- Actualiser le système d'information de la banque pour garantir l'obtention de données fiables et détaillées rapidement.
- Investir dans l'approche de rentabilité client (ARC) et d'autres formes de comptabilité analytique pour une analyse de rentabilité client plus précise.

## Conclusion

En conclusion de cette partie empirique, il convient de noter que l'application du modèle de Cox à risques proportionnels (PH Cox) pour l'estimation de la Customer Lifetime Value (CLV) a manifesté une efficacité notable au sein de l'environnement bancaire étudié. La pertinence de ce modèle s'est avérée plus convaincante que celle du modèle DeepSurv, tant sur le plan de la précision prédictive que de l'interprétabilité des résultats. Cette supériorité caractéristique conduit à des recommandations stratégiques de valeur pour la banque, en particulier concernant la rétention des clients présentant une rentabilité limitée et l'ajustement des stratégies de gestion de portefeuille.

## Conclusion générale

En conclusion, cette étude a exploré l'importance stratégique de la Customer Lifetime Value (CLV) dans la gestion des relations clients au sein du secteur bancaire, en mettant l'accent sur la banque CPA. À travers une analyse comparative de deux modèles de survie, le modèle PH Cox et le modèle Deep Surv, nous avons mis en lumière la capacité de ces méthodologies à estimer la probabilité qu'un client continue son engagement avec la banque à un moment donné. Cette probabilité, qui ne constitue pas une mesure directe de la CLV, est néanmoins une composante essentielle dans le calcul de celle-ci, car elle permet d'évaluer la durée et la profondeur de la relation client.

L'application de ces modèles sur les données de la banque CPA a révélé des insights précieux sur la dynamique de la fidélité et de la rentabilité client. Les résultats obtenus offrent aux décideurs de la banque des leviers d'action pour optimiser les stratégies de marketing, de vente et de service, en se fondant sur une compréhension plus nuancée de la valeur que chaque client apporte sur le long terme.

En outre, ce travail a mis en évidence l'importance de disposer d'une approche analytique robuste pour éclairer les décisions stratégiques en matière de gestion de clientèle. Il souligne également le rôle clé des technologies de l'information et des compétences analytiques dans l'exploitation efficace des grandes quantités de données clients disponibles pour les institutions financières.

Alors que les outils et techniques de modélisation continuent d'évoluer, la banque CPA et ses semblables peuvent s'attendre à affiner encore davantage leurs prévisions et à renforcer leur compétitivité sur le marché. Cette recherche contribue à la littérature existante en fournissant un exemple concret de l'application de la théorie de la survie dans un contexte bancaire, tout en ouvrant la voie à des études futures pouvant élargir et approfondir la compréhension de la CLV dans le secteur des services financiers.

En fin de compte, l'adoption d'une orientation client centrée sur la CLV peut transformer la manière dont les banques interagissent avec leurs clients, passant d'une perspective transactionnelle à une relation plus durable et mutuellement bénéfique. Cela signifie que les institutions financières qui maîtrisent le calcul et l'interprétation de la CLV sont mieux

préparées à prospérer dans un paysage économique où la fidélité et la satisfaction des clients sont des monnaies de plus en plus précieuses.

## Bibliographie

- [1] Philippe Monnier, Sandrine Mahier-Lefrançois, Techniques bancaires 2022
- [2] Livre blanc sur la mesure de la rentabilité des activités bancaires, mesure des revenus et définition des marges, p14.
- [3] Emmanuelle Cagnello-Charl, ARC Décisions Marketing, Jan.-Avr. 1999, No. 16 (Jan.-Avr. 1999), pp. 25-37
- [4][https://www.bain.com/globalassets/noindex/2023/bain\\_report\\_customer\\_behavior\\_and\\_loyalty\\_in\\_banking\\_global\\_edition\\_2023.pdf](https://www.bain.com/globalassets/noindex/2023/bain_report_customer_behavior_and_loyalty_in_banking_global_edition_2023.pdf)
- [5] Hubert Tchakoute Tchuigoua et Mehdi Nekhili. Gestion des risques et performance des institutions de microfinance, Revue d'économie industrielle 2012
- [6] Donkers et al, 2007 ; Rust et al, 2000, Beniot et Poel, 2009
- [7] (Hwang et al, 2004)
- [8] Marketing Actions and the Value of Customer Assets: A Framework for Customer Asset Management. Paul D. Berger, Ruth N. Bolton, Douglas Bowman, Elten Briggs, V. Kumar, A. Parasuraman, and Creed Terry.
- [9] Wendell R. Smith (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies
- [10] Alfansi et Sargeant, 2000. Market segmentation in the Indonesian banking sector: the relationship between demographics and desired customer benefits
- [11] Mulhern, 1999; Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending. Journal of Managerial Issues.
- [12] Max völcker carl stenfelt, 2018, Modelling Customer Lifetime Value in the Retail Banking Industry,
- [13] Blattberg, Robert C., and John Deighton (1996), "Manage Marketing by the Customer Equity Test, 'Harvard Business Review'"
- [14] Gurau et Ranchhod, 2002; How to Calculate the Value of a Customer.
- Deng et al., 2019, Customer Portfolio Management and Firm Performance: a configurational perspective.
- Berger et al. (2002)
- [15] Gupta, S., Lehmann, (2003). CUSTOMERS AS ASSETS

[16] Haenlein et al., 2007; A Model to Determine Customer Lifetime Value in a Retail Banking Context, European Management Journal.

[17] Hartfeil, G. (1996), Bank one measures profitability of customers, not just products. Journal of Retail Banking Services

[18] Helgesen, 2006; Relationship between Customer Satisfaction and Customer Loyalty in the Retail Banking Sector of Ghana

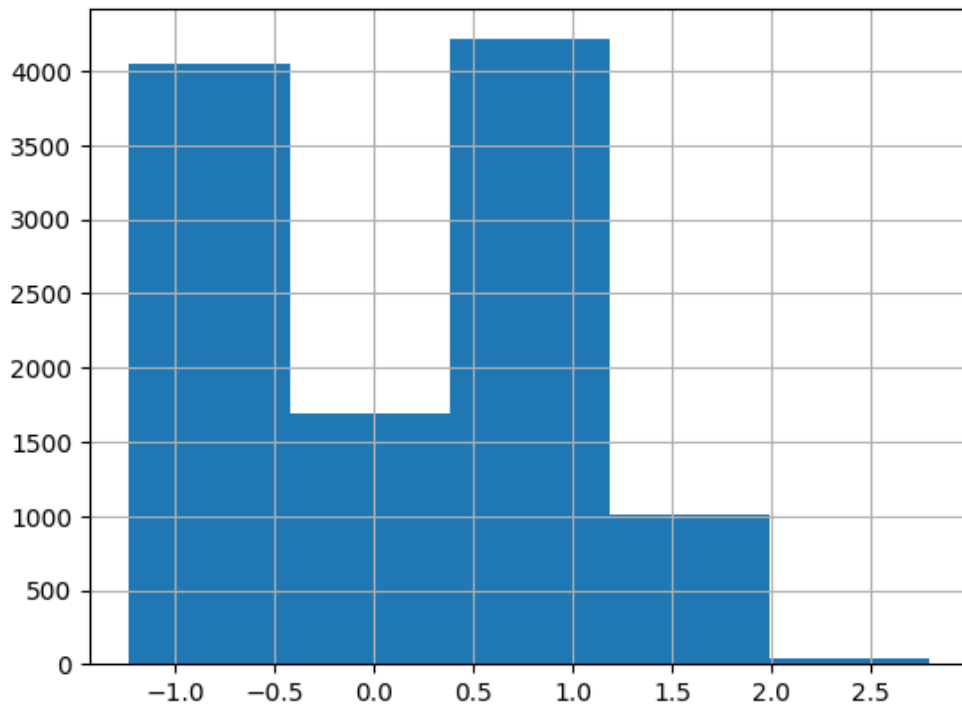
[19] Junxiang Lu, Ph.D. Overland Park, Kansas 2002 ; « Modeling Customer Lifetime Value Using Survival Analysis – An Application in the Telecommunications Industry| »

[20] Jared L. Katzman<sup>1</sup>, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang<sup>6</sup> and Yuval Kluger 2018. « DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. »

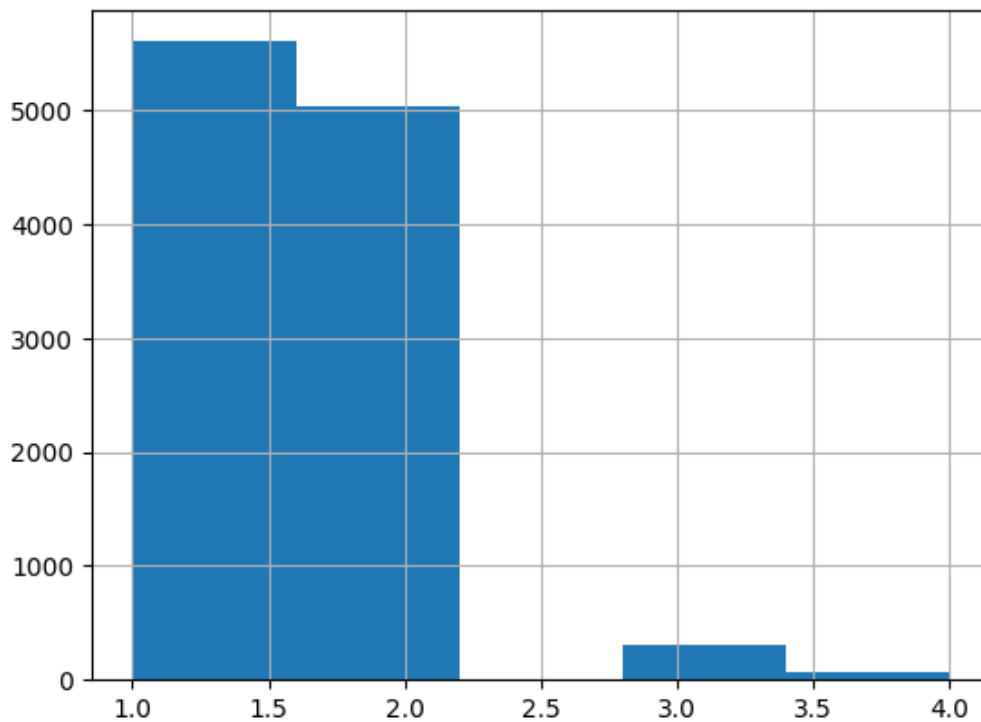
Oludare Isaac Abiodun a,b, et al.2018. State-of-the-art in artificial neural network applications: A survey

## Annexes

### Annexe 1: distribution de la variable SOLDE

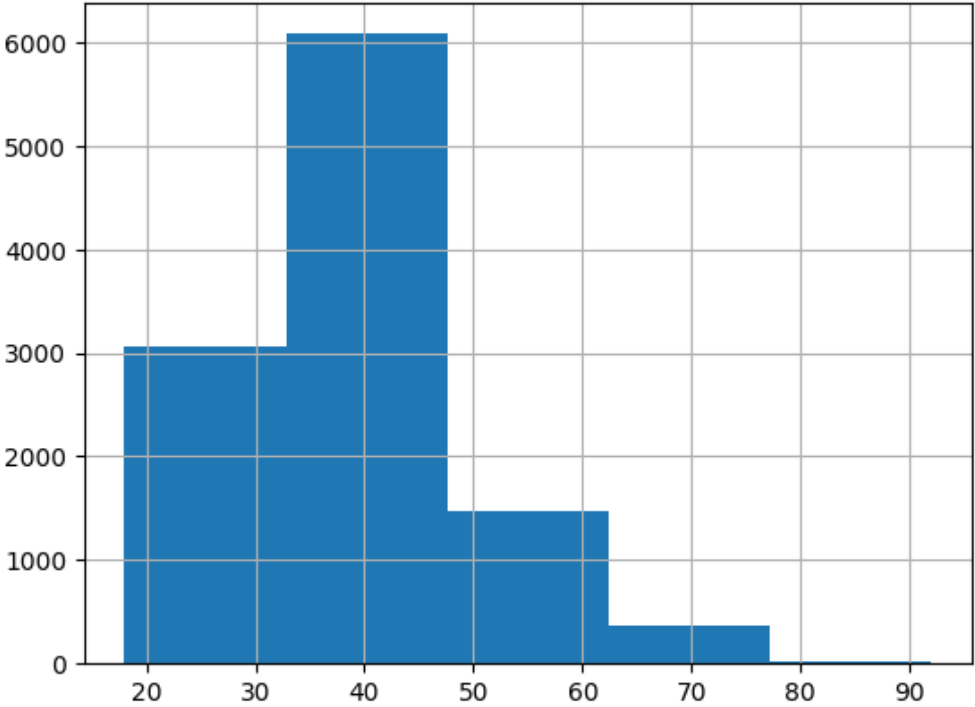


### Annexe 2: Distribution de la variable Nombre De Produits



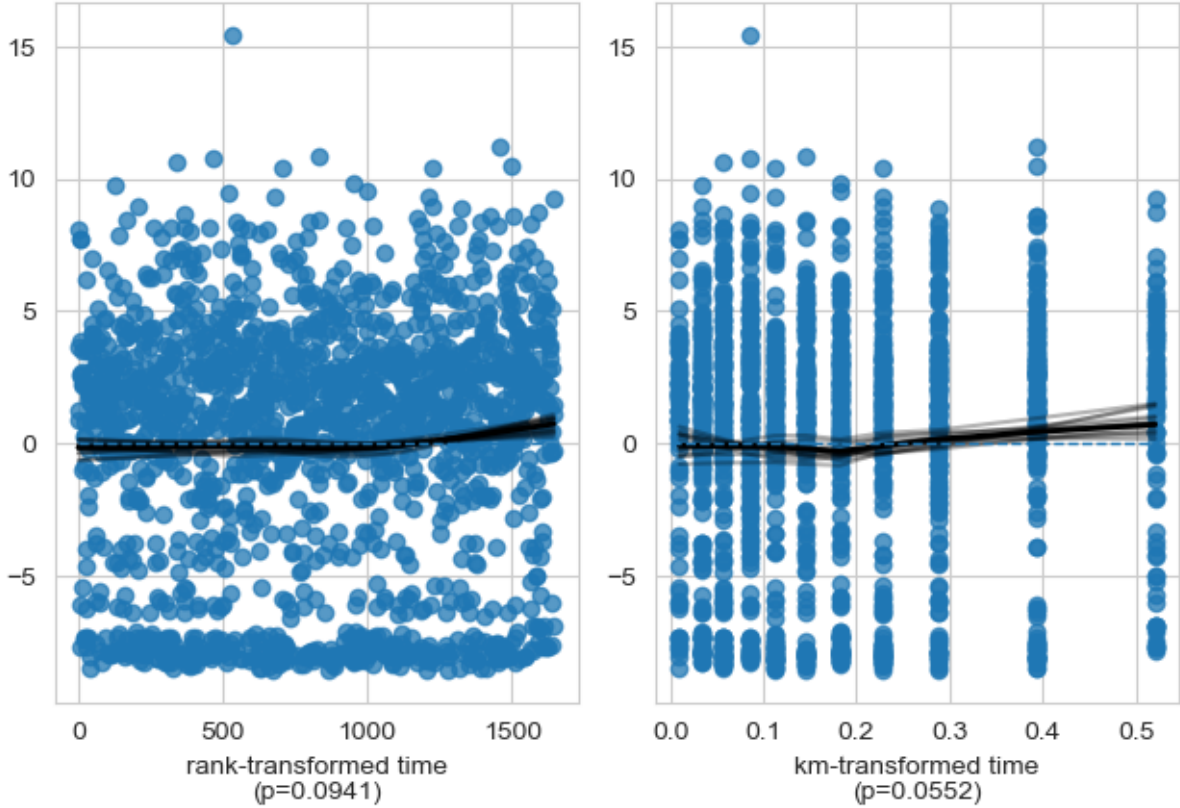


**Annexe 3:** Distribution de la variable Age.

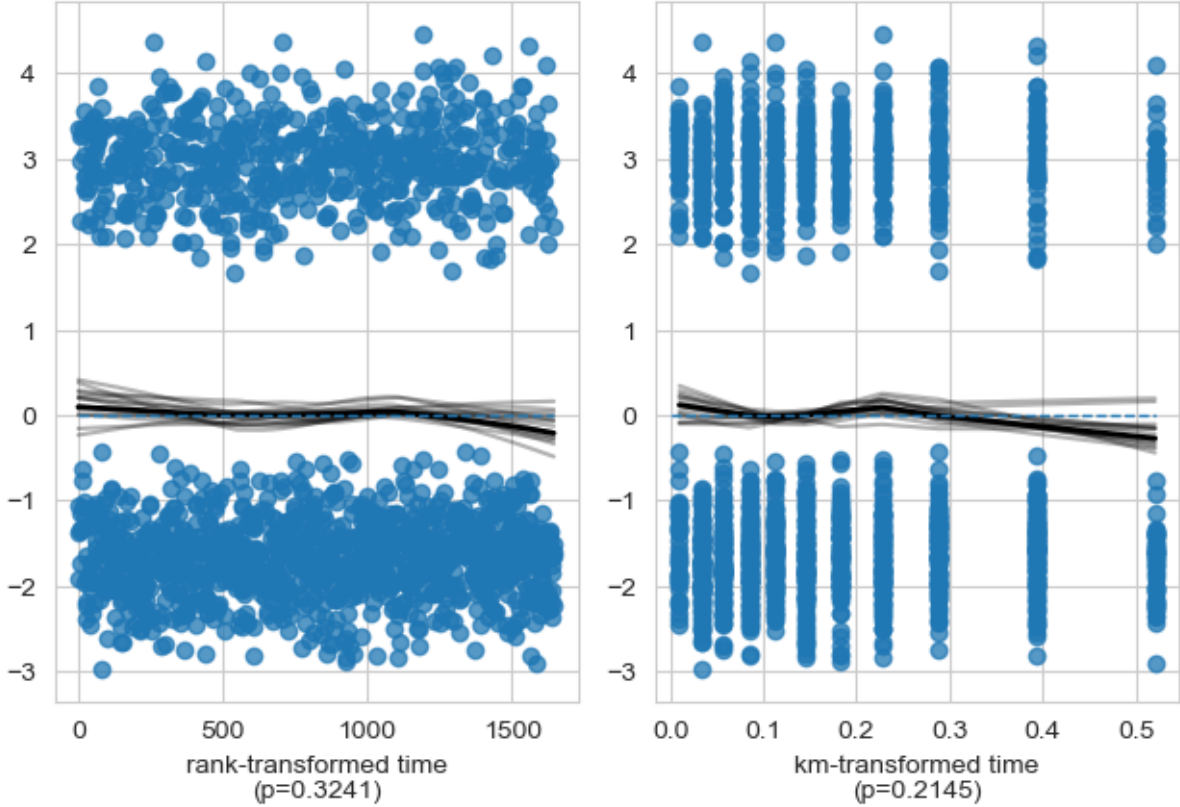


Annexe 4 : Visualisation des résidus de schoenfeld pour les différentes variables

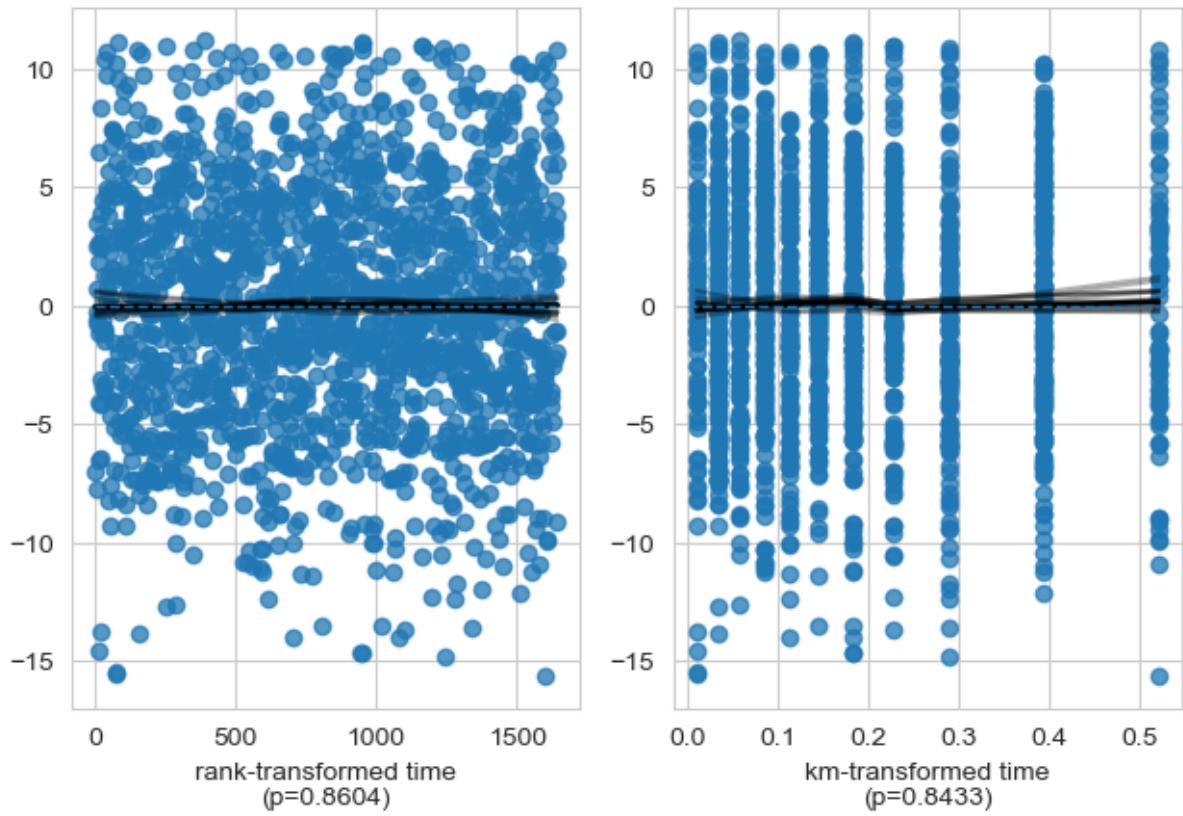
Scaled Schoenfeld residuals of 'Solde'



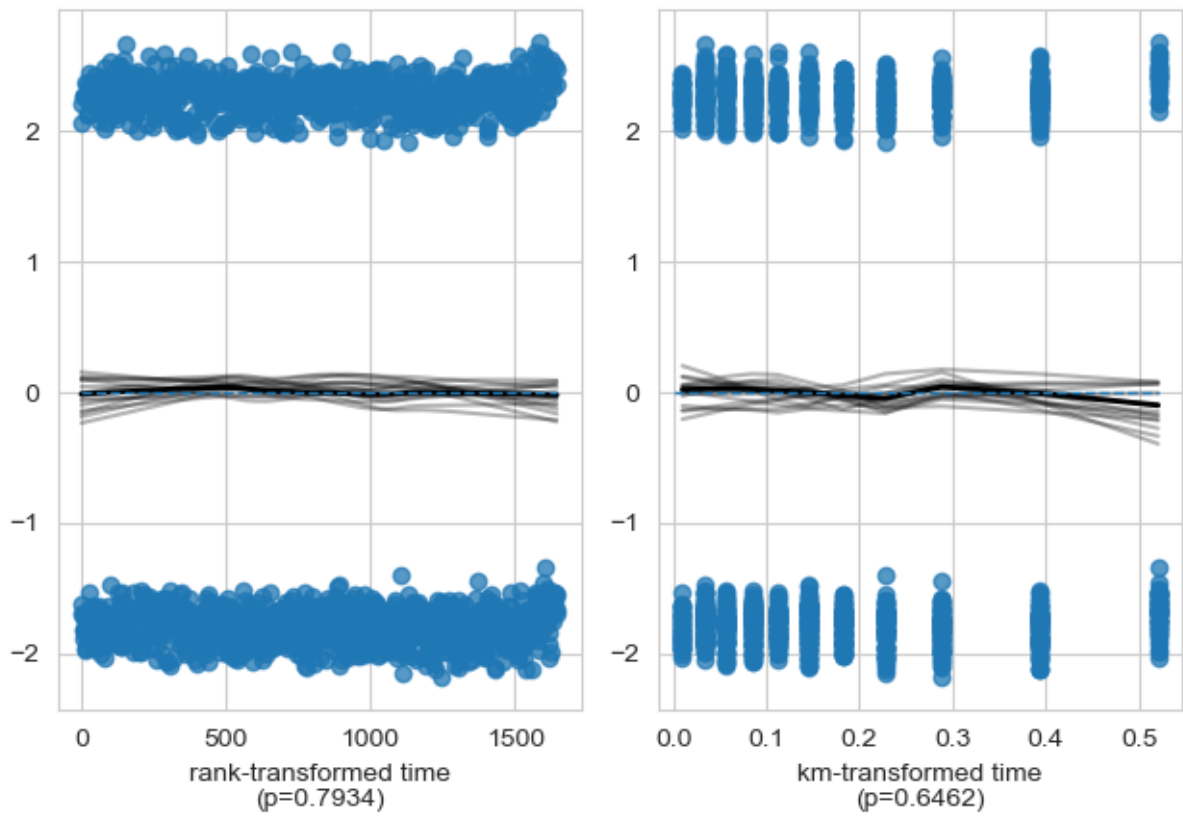
Scaled Schoenfeld residuals of 'Membre Actif'



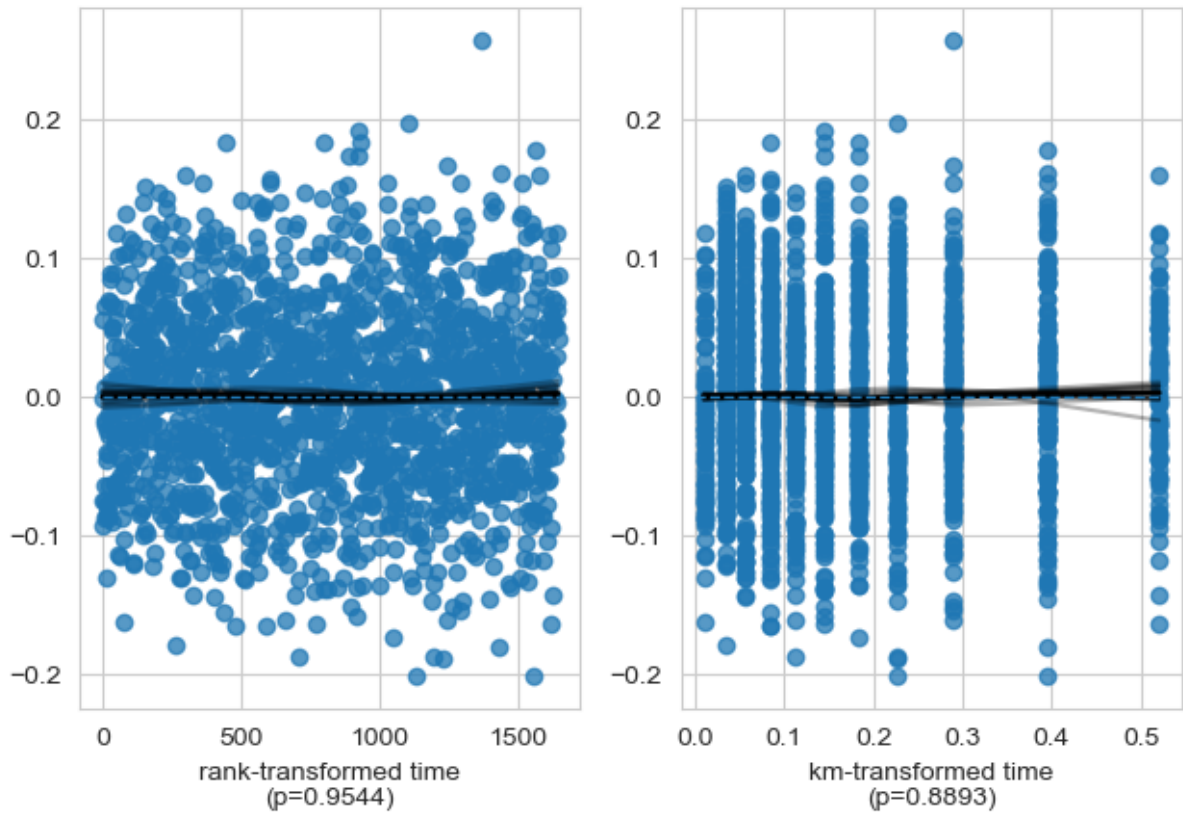
Scaled Schoenfeld residuals of 'Score crédit'



Scaled Schoenfeld residuals of 'Genre'



### Scaled Schoenfeld residuals of 'Age'



**Annexe 4** : Le bilan de la banque exercice 2021

## ACTIF

	U=Millions DA
Caisse, Banque Centrale, Trésor Public, Centre des Chèques Postaux	456 152
Actifs financiers détenus à des fins de transaction	3 982
Actifs financiers disponibles à la vente	203 126
Prêts et créances sur les institutions financières	588 440
Prêts et créances sur la clientèle	1 376 701
Actifs financiers détenus jusqu'à l'échéance	435 694
Impôts courants – Actif	2 964
Impôts différés – Actif	3 901
Autres actifs	12 188
Comptes de régularisation	33
Participations dans les filiales, les co-entreprises ou les entités associées	15 046
Immeubles de placement	-
Immobilisations corporelles	17 153
Immobilisations incorporelles	421
Écart d'acquisition	-
<b>Total de l'actif</b>	<b>3 115 801</b>

## PASSIF

	U=Millions DA
Banque Centrale	-
Dettes envers les Institutions Financières	862 637
Dettes envers la clientèle	1 812 898
Dettes représentées par un titre	57 302
Impôts courants- Passif	10 549
Impôts différés – Passif	542
Autres passifs	17 344
Comptes de régularisation	14 485
Provisions pour risques et charges	50 141
Subventions d'équipement Autres subventions d'investissements	-
Fonds pour risques bancaires généraux	36 111
Dettes subordonnées	-
Capital	48 000
Primes liées au capital	-
Réserves	158 060
Écart d'évaluation	1 370
Écart de réévaluation	15 921
Report à nouveau	-
Résultat de l'exercice	30 441
<b>Total du passif</b>	<b>3 115 801</b>

### Annexe 5 : le modèle de réseau de neurones Deep Surv :

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, BatchNormalization
from tensorflow.keras.regularizers import l1_l2
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import LearningRateScheduler
from tensorflow.keras.callbacks import EarlyStopping

def build_deepsurv_model(input_dim):
    model = Sequential()
        model.add(Dense(33, input_dim=input_dim, activation='relu',
kernel_regularizer=l1_l2(l1=0.01, l2=0.01)))
        model.add(BatchNormalization())
        model.add(Dropout(0.5))
            model.add(Dense(8, activation='selu',
kernel_regularizer=l1_l2(l1=0.01, l2=0.01)))
        model.add(Dropout(0.5))
```

```

model.add(Dense(1, activation='linear'))

return model

input_dim = X_train.shape[1]
model = build_deepsurv_model(input_dim)

# Définir l'optimiseur
optimizer = Adam(learning_rate=0.0000009990572606240212, beta_1=0.9,
beta_2=0.999, epsilon=1e-07, amsgrad=False)

# Compiler le modèle
model.compile(optimizer=optimizer,
loss=negative_partial_log_likelihood(event_indicator=y_train["Exited"].
values, risk_set=True))

# Programmation du taux d'apprentissage
def lr_schedule(epoch):
    if epoch < 10:
        return 0.00001
    elif epoch < 20:
        return 0.000007
    else:
        return 0.0000009

schedule = LearningRateScheduler(lr_schedule)
# Définir le critère d'arrêt anticipé
early_stopping = EarlyStopping(monitor='val_loss', patience=10,
verbose=1, restore_best_weights=True)
# Entraîner le modèle en utilisant le critère d'arrêt anticipé comme
rappel (callback)
history = model.fit(X_train, y_train, batch_size=6, epochs=250,
validation_data=(X_test, y_test), callbacks=[early_stopping])

```

## **Glossaire :**

**Perceptron multicouche (MLP) :** Type de réseau de neurones artificiels composé d'au moins trois couches de nœuds en connexion directe. Il peut distinguer des données qui ne sont pas séparables linéairement.

**ReLU (Rectified Linear Unit) :** Une fonction d'activation largement utilisée dans les réseaux de neurones. Elle transforme l'entrée en sortie en prenant le maximum entre l'entrée et zéro.

**SELU (Scaled Exponential Linear Unit) :** Une autre fonction d'activation qui, lorsqu'elle est utilisée dans un certain type de réseau, peut conduire à une normalisation automatique des poids. Elle peut améliorer la performance des réseaux de neurones.

**Activation linéaire :** Une fonction d'activation qui permet au neurone de produire une sortie directement proportionnelle à son entrée.

**Régularisation L1 et L2 :** Techniques utilisées pour prévenir le surapprentissage en ajoutant une pénalité à la magnitude des coefficients du modèle. L1 pénalise la somme absolue des coefficients (lasso) tandis que L2 pénalise la somme des carrés des coefficients (ridge).

**Normalisation par lots (Batch Normalization) :** Une technique utilisée pour améliorer la vitesse, la performance et la stabilité des réseaux de neurones en normalisant les activations des neurones dans une couche pour chaque lot d'entraînement.

**Surapprentissage (Overfitting) :** Lorsqu'un modèle d'apprentissage automatique est trop complexe et apprend à trop bien s'adapter aux données d'entraînement, il peut perdre la capacité de bien généraliser sur de nouvelles données.

## Table des matières

<b>Sommaire</b> .....	<b>4</b>
<b>Liste des abréviations</b> .....	<b>5</b>
<b>Liste des figures</b> .....	<b>6</b>
<b>Liste des tableaux</b> .....	<b>7</b>
<b>Introduction Générale :</b> .....	<b>1</b>
<b>Chapitre premier : La rentabilité client et la Customer life Time value.</b> .....	<b>4</b>
<b>Introduction</b> .....	<b>4</b>
<b>Section 01 : Définition de la rentabilité client et des indicateurs pertinents pour l'évaluer</b> .....	<b>5</b>
I. Définition de la rentabilité bancaire .....	5
II. Le rôle de la mesure de la rentabilité dans les orientations stratégiques et budgétaire de la banque .....	6
III. Indicateurs clés de la rentabilité bancaire .....	7
IV. La notion de la rentabilité client .....	10
V. Méthodes d'analyse de la rentabilité d'un portefeuille client :.....	12
VI. Importance de la gestion efficace d'un portefeuille client : .....	16
<b>Section 02 : La customer Life time value définition et concepts</b> .....	<b>18</b>
I. Définition de la Customer Lifetime Value (CLV).....	18
II. Le rôle et l'application de la valeur vie client dans les décisions marketing 19	
III. La CLV dans l'industrie bancaire de détail.....	21
IV. La place de la CLV dans la stratégie marketing de la banque.....	23
<b>Section 03 : Indicateurs et modèles de mesure de la CLV dans la littérature :</b> .....	<b>24</b>
I. Méthodes de calcul de la CLV .....	24
II. Analyse de la littérature sur les modèles de mesure : .....	25
III. Les modèles de durée de vie (Survival Analysis).....	26
IV. Conclusion .....	32
<b>Chapitre deuxième : Proposition d'un modèle de calcul de la CLV pour la banque CPA.</b> .....	<b>33</b>
<b>Introduction</b> .....	<b>34</b>
<b>Présentation du crédit populaire d'Algérie :</b> .....	<b>35</b>
<b>Section 01 : Méthodologie</b> .....	<b>36</b>



I.	Collecte de données .....	36
II.	Prétraitement des données .....	37
III.	Exploration des données .....	38
IV.	Division des données .....	41
	<b>Section 02 : Résultats .....</b>	<b>45</b>
I.	Comparaison et évaluation de la performance des deux modèles :.....	45
II.	Synthèse des résultats: .....	50
III.	Mise en place du modèle de Cox :.....	50
	Synthèse.....	59
IV.	Calcul du coût du capital WACC .....	60
V.	Calcul de la CLV pour les clients de la banque:.....	62
	<b>Section 03 : Discussion et recommandations .....</b>	<b>65</b>
I.	Discussion .....	65
II.	Recommandation .....	66
	<b>Conclusion .....</b>	<b>67</b>
	<b>Conclusion générale .....</b>	<b>68</b>
	<b>Bibliographie.....</b>	<b>70</b>
	<b>Annexes.....</b>	<b>72</b>

