

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **14 mars 2024**

Nom de famille et prénom de l'auteur : **Monsieur VALLA Mathias**

Titre de la thèse : « *Dynamique temporelle dans les modèles par arbres et applications aux comportements de rachat en assurance vie.* »



Résumé

Cette thèse, *Dynamique temporelle dans les modèles par arbres et applications aux comportements de rachat en assurance vie.*, explore l'application de nouveaux modèles statistiques et de stratégies innovantes pour analyser le comportement de rachat en assurance vie. Ce travail est une combinaison de deux parties introductives et de trois articles de recherche, chacun offrant une perspective singulière et diverses contributions à la gestion du rachat ainsi qu'à l'inclusion d'une dimension temporelle dans les modèles de Machine Learning dits *par arbres*. L'ensemble de la thèse a été rédigée en anglais, ainsi nous proposons ce résumé en français, reprenant largement des éléments de l'abstract et de l'introduction du manuscrit complet.

La première partie de cette thèse est une introduction en plusieurs points. Tout d'abord, elle détaille le contexte dans lequel est né et s'est développé ce travail de thèse. Ensuite, il est détaillé, ce qui justifie ce sujet de recherche et comment il peut être bénéfique au domaine des sciences actuarielles, pourquoi nous utilisons l'intelligence artificielle (IA) pour résoudre des problèmes actuariels, pourquoi nous limitons l'analyse à des modèles basés sur des arbres et enfin, ce que nous entendons par la prise en compte de la dynamique temporelle dans la modélisation. Les récentes avancées dans le domaine de l'apprentissage statistique ont considérablement amélioré l'efficacité des modèles prédictifs, conduisant à de nouvelles utilisations dans divers secteurs. Néanmoins, le secteur de l'assurance a encore du mal à utiliser concrètement, et à grande échelle ces outils innovants. D'une part à cause du rôle fondamental de l'assureur dans la protection des individus contre de potentielles pertes économiques : tout processus de prise de décision dans le domaine de l'assurance affecte potentiellement de manière significative la vie des assurés. D'autre part, il s'agit également d'une conséquence des réglementations strictes mises en œuvre pour garantir le traitement équitable de tous les individus. Ces aspects soulignent la nécessité d'expliquer et d'interpréter les décisions en matière d'assurance. En outre, ces points soulignent la nécessité de stratégies centrées sur le client, d'abord pour répondre aux besoins et aux attentes des PH, ensuite

pour mieux identifier les comportements individuels et leurs conséquences sur la rentabilité de l'assureur. Cette partie explique que les problèmes de prédictions de rachat sont souvent abordés sous l'angle d'un problème de classification binaire (rachat/non rachat) et que ces approches gagnent, en pratique, à inclure des considérations temporelles.

Après avoir défini les termes du sujet et justifié les choix d'application à la prédiction du risque de rachat en assurance vie, cette partie s'achève par la présentation du plan du reste de la thèse : cette thèse comprend six parties principales, une introduction générale, une partie préliminaire qui présente des généralités théoriques sur les algorithmes *par arbres*, une troisième partie consacrée à la définition d'un cadre de gestion des rachats centré sur l'assuré pour les assureurs-vie, une quatrième partie suggérant que ce cadre peut être adapté à l'exploitation de variables explicatives variant dans le temps, une cinquième partie consacrée à l'exploration et au développement d'un nouvel algorithme *par arbres* qui peut utiliser des données longitudinales, et une dernière section qui constitue la conclusion générale de la thèse.

La seconde partie plonge dans le cadre théorique de notre recherche en présentant la méthodologie usuelle d'entraînement des modèles en apprentissage statistique, employée pour répondre à nos questions de recherche. Ensuite, elle propose une revue de littérature qui présente un historique des modèles *par arbres*, d'abord pour des problèmes de régression et de classification, puis pour l'analyse de survie, ainsi que les mesures d'évaluation qui sont utilisés pour les évaluer. Enfin, elle s'achève par l'énoncé de quelques résultats et références théoriques supplémentaires sur les modèles *par arbres*. Ce chapitre est pensé comme un guide méthodologique et théorique des techniques, algorithmes et processus utilisés dans les études et applications des parties suivantes.

La troisième partie de la thèse est basée sur un premier article, *Including individual Customer Lifetime Value and competing risks in tree-based lapse management strategies*, qui présente une stratégie de rétention qui va au-delà d'une simple prédiction du comportement assuré. Cette stratégie est basée sur un cadre de gestion des rachats intégrant les notions de valeur client et de rentabilité, en misant sur une individualisation des approches existantes. Ainsi, cette partie introduit d'abord la notion de valeur client, de manière générale en marketing, puis de façon plus ciblée pour le secteur de l'assurance. Il est ensuite expliqué comment la modélisation individualisée de la valeur client, reposant sur la prédiction des probabilités de rachat et de décès des assurés, peut être incluse dans une stratégie d'optimisation des stratégies de rétention. Un exemple d'application concret démontre comment les modèles de survie *par arbres* surpassent les approches paramétriques dans l'évaluation des profils de survie, contribuant ainsi à une optimisation des campagnes de rétention plus efficace et plus informée pour les assureurs vie. Sur les nombreux scénarios présentés dans l'application sur des données réelles, cette méthodologie génère en moyenne une augmentation de plus de 30% des profits espéré par l'assureur, comparativement à une méthodologie n'incluant pas la valeur client. Enfin, cette partie s'achève par une section moins appliquée mettant en avant les propriétés théoriques de la valeur client telle que nous proposons de la modéliser et plus globalement de la méthodologie proposée pour optimiser une stratégie de rétention.

La quatrième partie, s'appuyant largement sur les travaux de la partie précédente, suggère que la méthodologie d'optimisation de la stratégie de rétention peut être adapté à des covariables qui évoluent dans le temps. Dans cette partie, nous définissons la notion d'analyse longitudinale, c'est-à-

dire l'étude de sujets dont les caractéristiques et leurs évolutions sont observées et mesurées à plusieurs reprises dans le temps. Elle introduit les notations générales utiles à une telle analyse, présente les modèles longitudinaux les plus récents et examine comment ces modèles ont été et pourraient être utilisés pour répondre à des problématiques actuarielles. Il définit une méthodologie longitudinale pour l'optimisation de la gestion des comportements de rachat, prenant en compte la dimension temporelle d'une telle problématique. Cette nouvelle méthodologie, décrite dans l'article intitulé *A longitudinal ML framework for lapse management in life insurance* soumis dans *Annals of Actuarial Science* constitue une deuxième couche temporelle incluse dans le cadre méthodologique développé en partie 3. Ses points forts, ses faiblesses et les améliorations à y apporter sont examinés dans différents contextes. Elle propose également une application concrète de cette méthodologie, qui montre une nette amélioration en termes de précision dans la modélisation de la valeur client, comparée aux résultats obtenus sans prendre en compte l'aspect longitudinal des comportements. Lorsque l'historique longitudinal des comportements des assurés est disponible, s'en priver revient à occulter une richesse d'information importante, à fort pouvoir prédictif. Ce cadre permet d'affiner davantage la précision du ciblage des assurés à retenir en portefeuille, améliorant ainsi la compréhension globale des assureurs quant au risque qu'ils portent. Nous concluons qu'elle apporte une contribution précieuse au domaine de l'analyse des comportements de rachats pour les assureurs-vie et souligne l'importance de l'utilisation de l'historique complet des assurés, qui est souvent disponible dans les systèmes d'information des assureurs, mais rarement exploitée.

Le dernier article, *Time penalized tree (T_{pT}): a new tree-based data mining algorithm for time-varying covariates*, soumis dans *Annals of Mathematics and Artificial Intelligence*, présente un nouvel algorithme de construction d'arbre de décision acceptant des variables évoluant avec le temps, autrement dit, des données structurées longitudinalement. Ce travail contribue aux domaines de l'apprentissage automatique, de l'exploration de données et des sciences actuarielles en introduisant ce nouvel algorithme *par arbres* qui permet le partitionnement récursif de l'espace des covariables et du temps, conjointement, et en l'utilisant dans une application de *clustering* en assurance-vie. Dans le cadre de cette thèse, la partie V souligne les inconvénients potentiels des modèles *par arbres* capables d'accepter les données longitudinales et propose une nouvelle approche avec une conception originale. Si les principales contributions des parties III et IV résident dans la méthodologie et les exemples d'application qu'elles détaillent, la partie V contribue principalement au domaine de l'exploration de données en concevant un nouvel algorithme capable de détecter les temps auxquels une différence entre les variables explicatives de deux sujets observés aura le plus d'impact sur la variable à expliquer. Cette dernière idée est capitale : certains comportements auront une influence différente sur la variable cible, selon le moment où ils ont été mesurés. T_{pT} peut être utilisé pour savoir **quand** un changement dans la variable réponse est plus susceptible de se produire, ou **quand** une covariable influence fortement le résultat. Cette méthode innovante aide à capturer les tendances historiques importantes vis-à-vis d'un phénomène étudié, permettant une analyse plus précise et interprétable dans des environnements dynamiques. Une application de T_{pT} à l'analyse des comportements de rachat est présentée et les résultats sont discutés sur la base de visuels produits par l'algorithme.

En conclusion, cette thèse se concentre sur de nouveaux cadres méthodologiques et des algorithmes originaux pour analyser le comportement de rachat en l'assurance-vie. Elle vise à inclure la dynamique temporelle des données dans des approches de modélisation du rachat centrées sur

l'assuré. Elle souligne le potentiel de stratégies innovantes, notamment basé sur une analyse longitudinale, pour informer les décisions commerciales et stratégiques dans l'industrie de l'assurance tout en garantissant leur interprétabilité et leur explicabilité.