

Améliorer un modèle de Machine Learning par la personnalisation : fonction de perte et enrichissement par les données

Mathias VALLA

Actuaire certifié

Université Lyon 1

Université KU Leuven

Pierrick PIETTE

Actuaire certifié

Fonction clé Actuariat – Seyna

Enseignant chercheur - ISFA

Sommaire

- A. Changement de fonction de perte et de métrique d'évaluation
 - 1. Présentation du problème: modéliser le comportement de rachat
 - 2. Approche statistique classique
 - 3. Approche avec une métrique d'évaluation personnalisée

- B. Enrichissement d'un modèle par les données
 - 4. Extensions du modèle par la personnalisation
 - 5. Extensions du modèle par la prise en compte de risques compétitifs
 - 6. Extensions du modèle avec des données longitudinales

- C. Conclusions et discussions
 - 7. Comparatif des approches
 - 8. Conclusions et pistes d'amélioration

A. Changement de fonction de perte et de métrique d'évaluation

1. Présentation du problème: modéliser le comportement de rachat

- Rachat dans l'assurance-vie

- Risque majeur pour la compagnie d'assurance vie
- Profits futurs sur les contrats
- Stratégie ALM

- Rachat et Churn

- Rachat : l'assuré met fin à son contrat d'assurance-vie avec la compagnie d'assurances
- Churn: le client met fin à son contrat avec la compagnie

A. Changement de fonction de perte et de métrique d'évaluation

1. Présentation du problème: modéliser le comportement de rachat

- Modèles pour la prediction de rachat

- GLM *Kim (2005)* ; CART *Milhaud et al. (2011)* ; GLMs Mixture *Milhaud (2012)*
- Boosting *Jamal (2017)* ; RF *Azzone (2021)*
- Détection des assurés qui vont racheter.

- Modèles pour la prediction du churn

- GLM and CART
- SVM, Bagging, Boosting, Neural Networks

- Buts en marketing

- Détection du churn ;
- Gestion de la retention proactive *Ascarza et al. (2018)*

A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique

- Division de la base de données \mathcal{D} en 10 sous-bases $\{\mathcal{D}_1, \dots, \mathcal{D}_{10}\}$
 - Division aléatoire et de taille égale
- Pour chaque $k \in \{1, \dots, 10\}$
 - \mathcal{D}_k la base d'entraînement, et $\mathcal{D} - \mathcal{D}_k$ en base de test.
 - pour obtenir les prédictions \hat{y} sur $\mathcal{D} - \mathcal{D}_k$ pour chaque modèle
- Mesure de validation
 - Définition d'une mesure de performance $\rho(y, \hat{y})$
 - pour comparer les modèles entre eux

A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique

- Prédiction d'une classification binaire : $(y_i, \hat{y}_i) \in \{0,1\}^2$
 - Vrai Positif (1,1) ; Vrai Négatif (0,0)
 - Faux Positif (0,1) ; Faux Négatif (1,0)
- Notation de la matrice de confusion

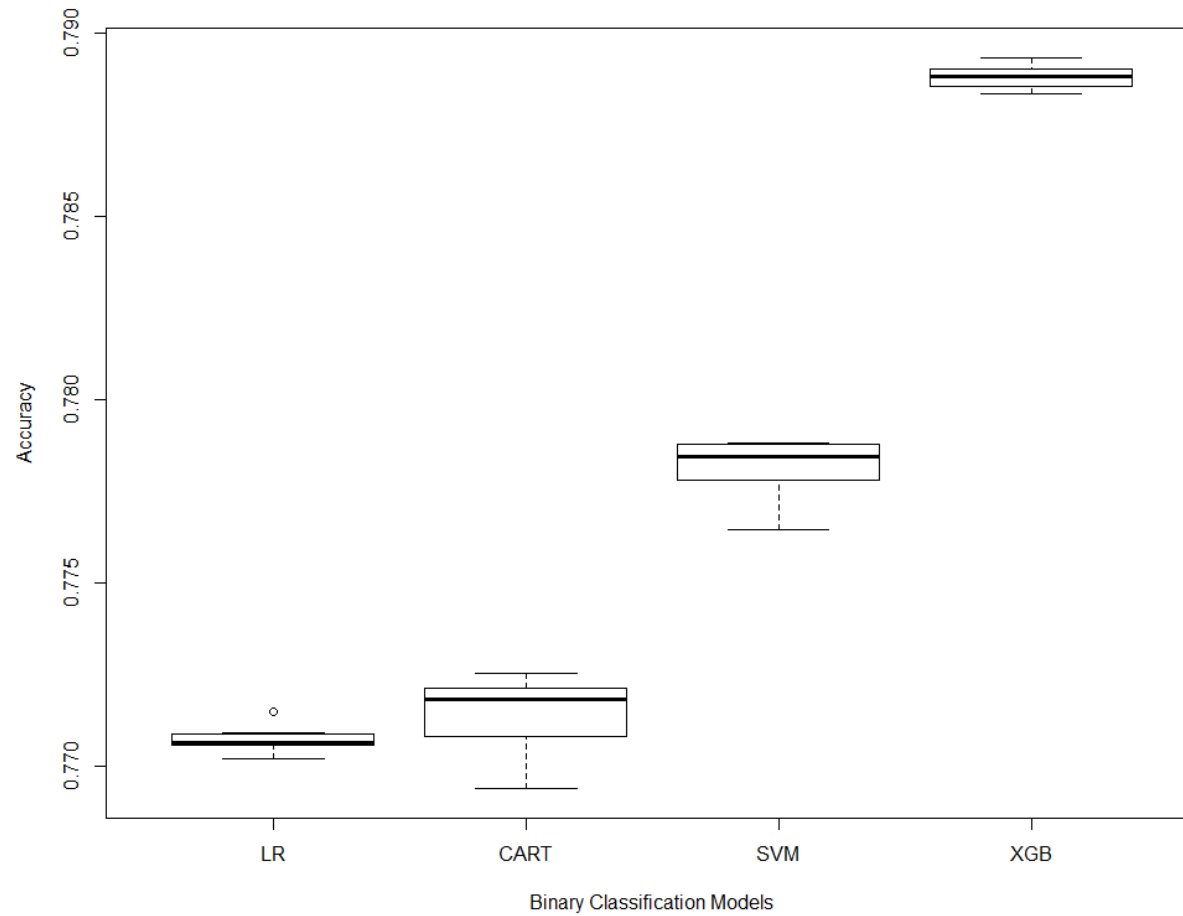
$$N(j, k) = \sum_i \mathbf{1}(y_i = j, \hat{y}_i = k)$$

- Beaucoup de mesures peuvent être alors définies, dont la plus connue

$$accuracy(y, \hat{y}) = \frac{N(1,1) + N(0,0)}{N} = 1 - error(y, \hat{y})$$

A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique



A. Changement de fonction de perte et de métrique d'évaluation
3. Approche avec une métrique d'évaluation personnalisée

	Reality: Predator Nearby	Reality: No Predator
Inference: Predator Nearby	Correct	False Positive
Inference: No Predator	False Negative	Correct

A. Changement de fonction de perte et de métrique d'évaluation

3. Approche avec une métrique d'évaluation personnalisée

- Profitabilité du point de vue de l'assureur *Neslin et al. (2006)*

$$CLV_i = \sum_{t=0}^{\Theta_i} \frac{p_{i,t} F_{i,t}}{(1 + d_t)^t}$$

- Simplifications

- Horizon de temps déterministe $T = 12$;
- Vecteur de rétention \mathbf{r}_i égale à \mathbf{r}_{stay} si pas de rachat observé et \mathbf{r}_{lapse} sinon;
- $p_{i,t} = p$;
- $F_{i,t} = F_i$ et $d_t = d$;

$$CLV_i = CLV(p, F_i, \mathbf{r}_i, d) = \sum_{t=0}^T \frac{p F_i r_t}{(1 + d)^t}$$

A. Changement de fonction de perte et de métrique d'évaluation

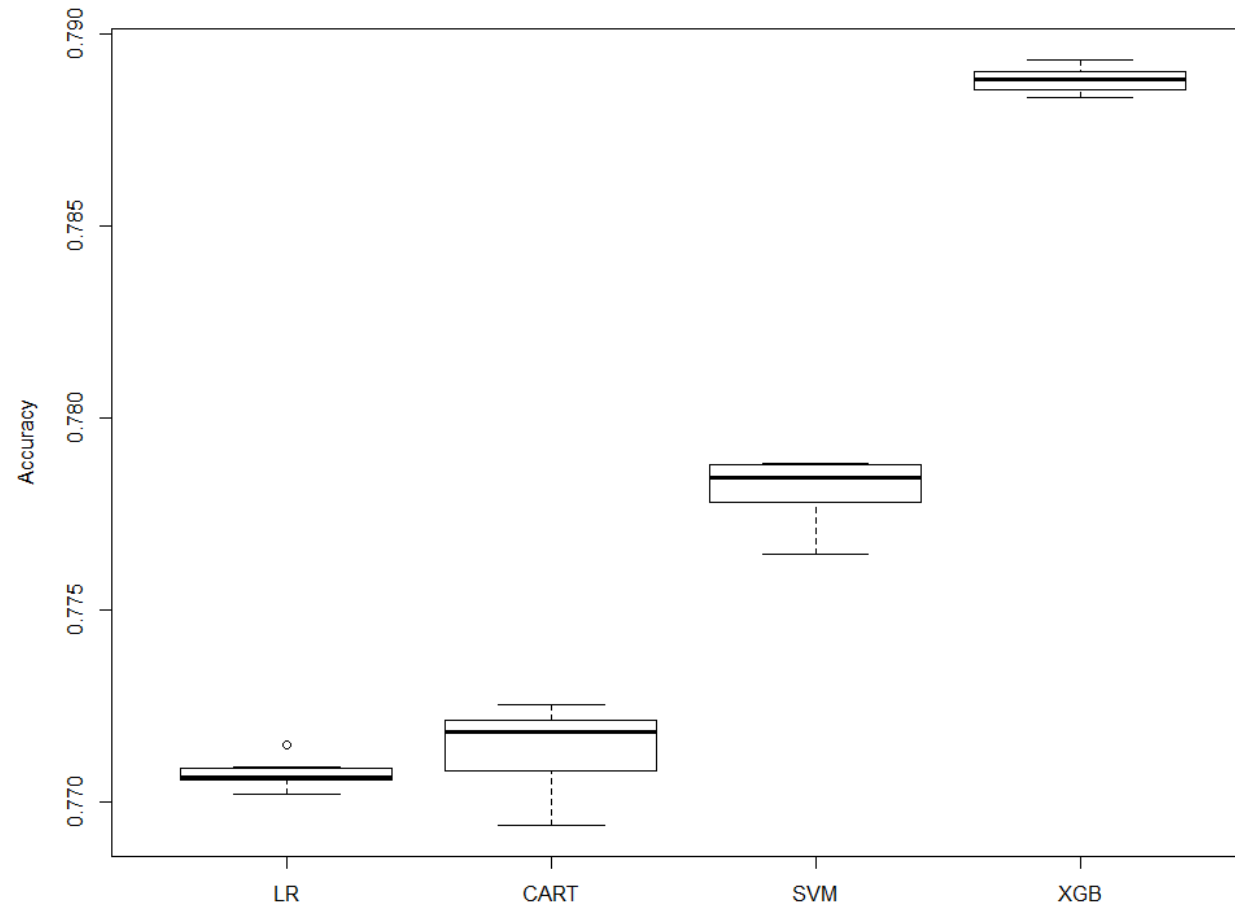
3. Approche avec une métrique d'évaluation personnalisée

- Stratégie de la rétention
 - Offre une incentive δ avec un cout de contact c
 - Acceptation avec probabilité
 - γ si l'assure allait racheter
 - 1 si non
 - Si un assure allant racheter accepte l'incentive sa probabilité de rétention change de r_{lapse} à r_{stay}
- Gain de la stratégie de rétention proactive: $RG(\delta, \gamma, c)$

$$\begin{aligned} & \gamma [CLV(\mathbf{p} - \delta, F(1,1), \mathbf{r}_{stay}, d) - CLV(\mathbf{p}, F(1,1), \mathbf{r}_{lapse}, d)] \\ & - CLV(\delta, F(0,1), \mathbf{r}_{stay}, d) - c(N(0,1) + N(1,1)) \end{aligned}$$

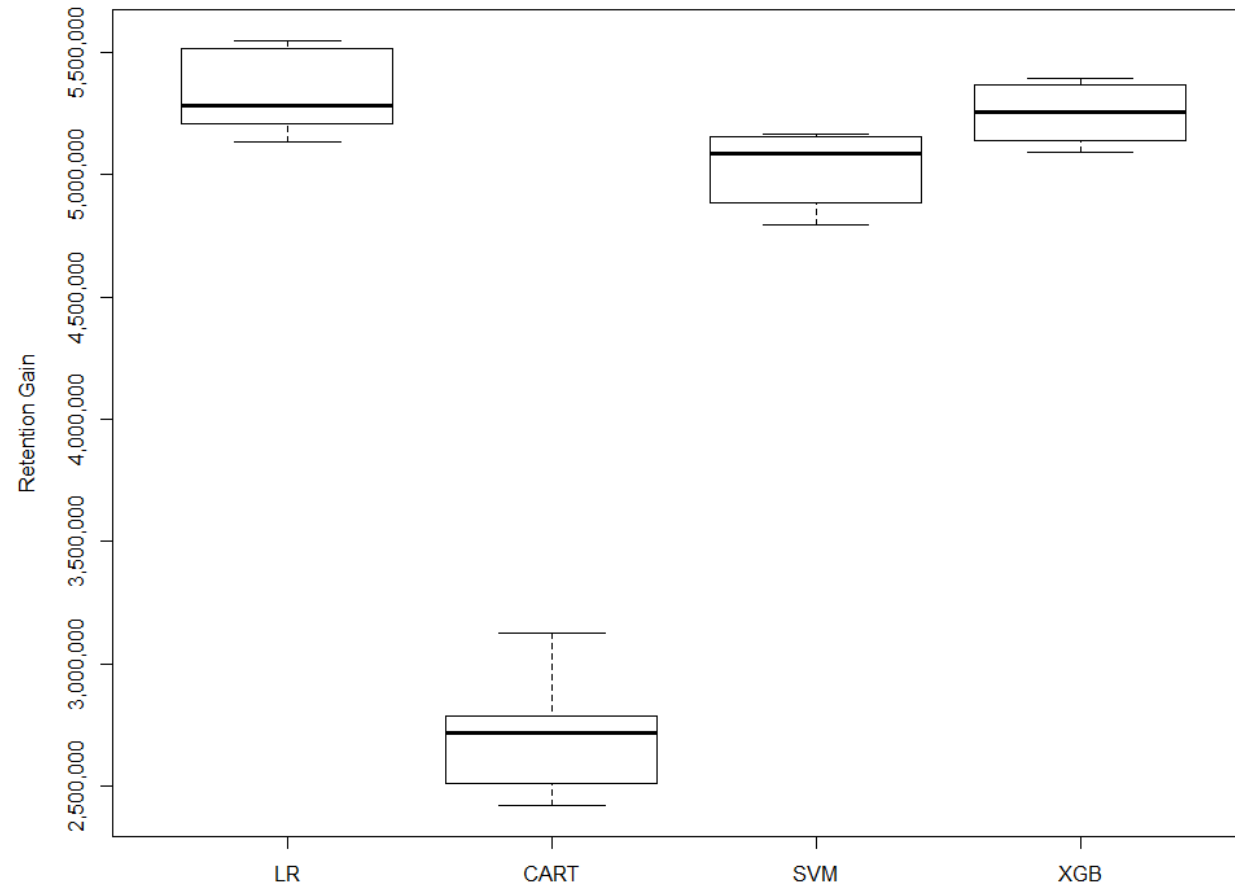
A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique



A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique



A. Changement de fonction de perte et de métrique d'évaluation

3. Approche avec une métrique d'évaluation personnalisée

- Les modèles standards de classification
 - mathématiquement créés pour savoir qui est à **risque de racheter** ;
 - mais pas pour détecter qui on **doit cibler** (*Lemmens and Gupta ; 2020*)
- Définition de la nouvelle variable d'intérêt

$$z_i^R = \begin{cases} -CLV(\delta, F_i, \mathbf{r}_{stay}, d) - c & \text{if } y_i = 0; \\ \gamma [CLV(\mathbf{p} - \delta, F_i, \mathbf{r}_{stay}, d) - CLV(\mathbf{p}, F_i, \mathbf{r}_{lapse}, d)] - c & \text{if } y_i = 1. \end{cases}$$

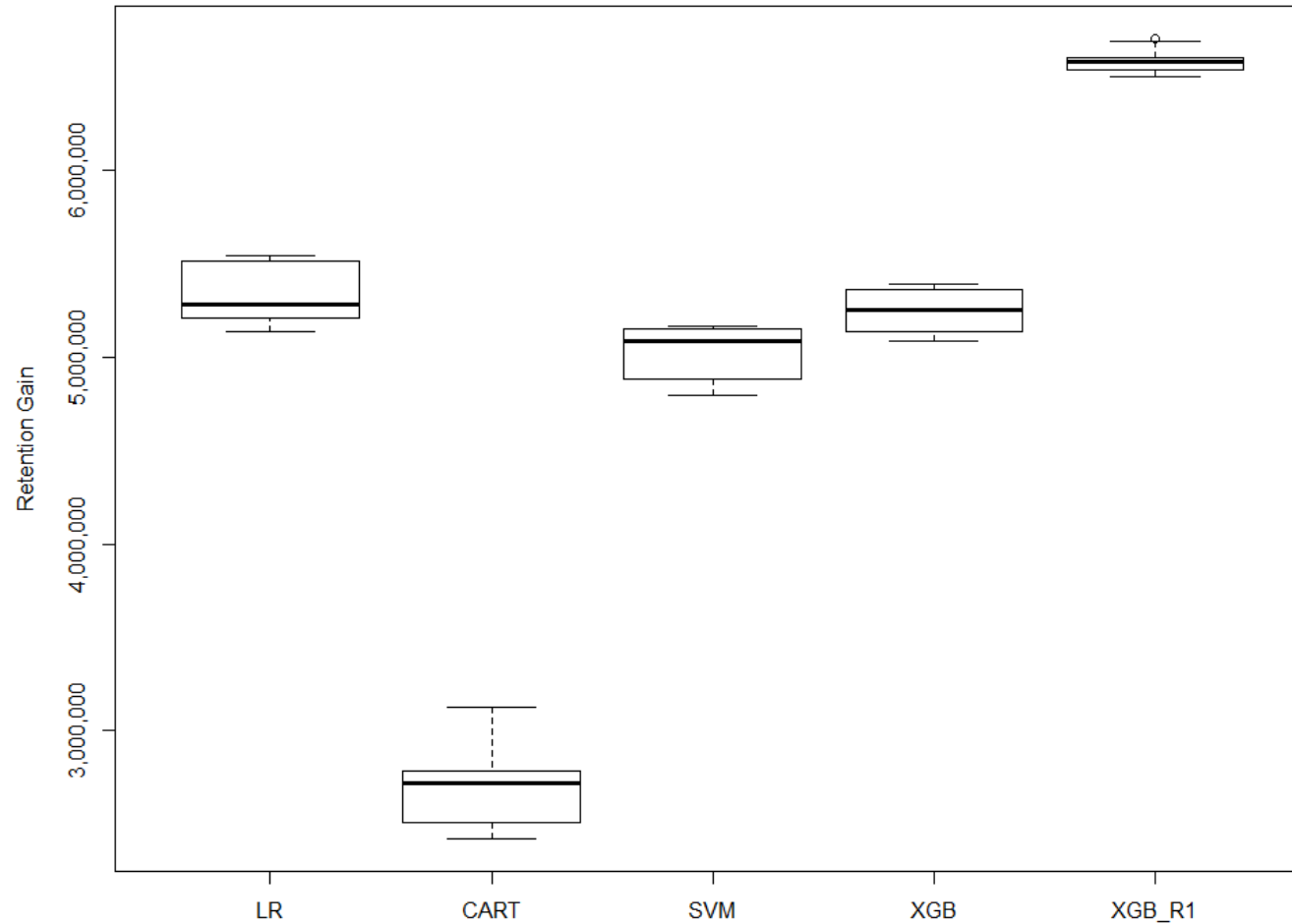
- XGBoost regression (XGB_R) avec une fonction de perte quadratique

$$L(z^R, \widehat{z^R}) = \frac{1}{N} \sum_{i=1}^N [z_i^R - \widehat{z^R}_i]^2$$

- Dans cette approche on offre l'incentive δ si $\widehat{z^R} \geq 0$

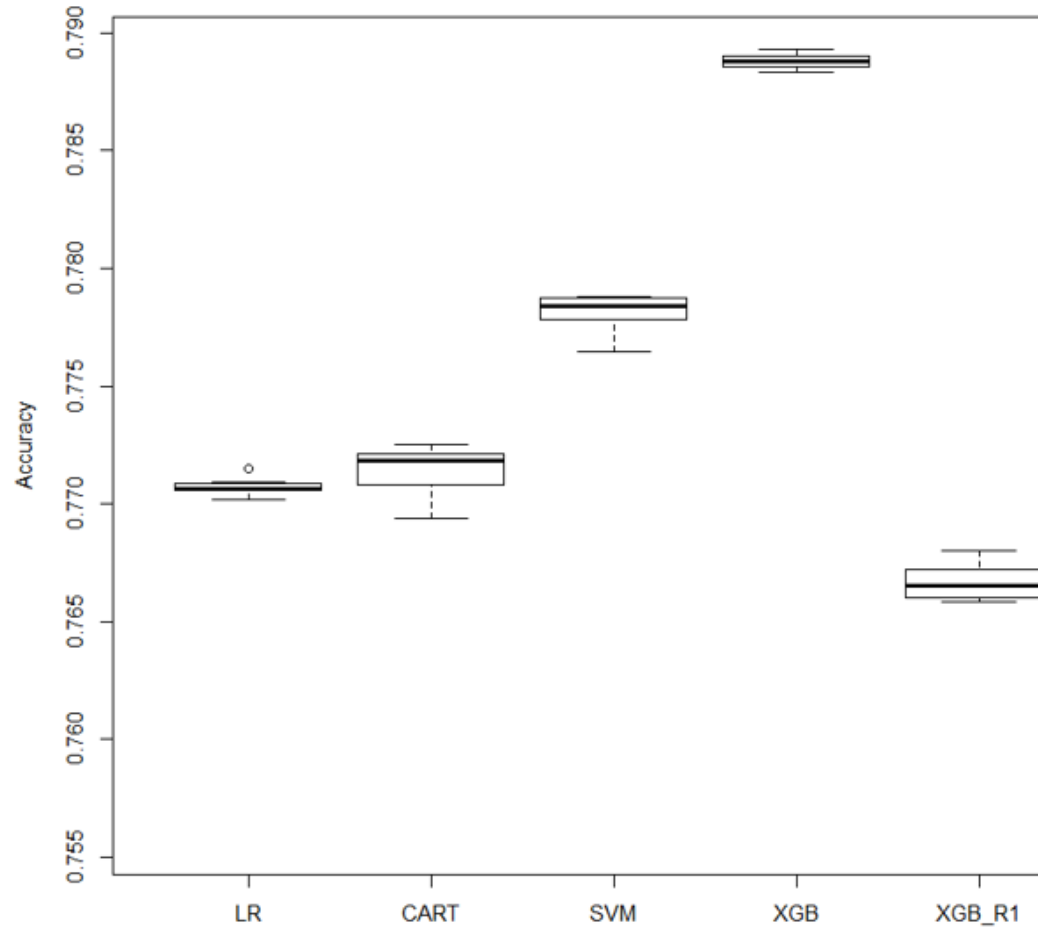
A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique



A. Changement de fonction de perte et de métrique d'évaluation

2. Approche statistique classique



B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation

- Modèle de survie à l'échelle individuelle pour r_{stay} et r_{lapse} :

Year	0	1	2	3	4	5	6	7	8	9	10	11	12
r_{lapse}	96%	87%	67%	37%	27%	21%	15%	12%	10%	8%	6%	5%	4%
r_{stay}	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Pistes d'extension:

- Individualiser le risque: [une ligne de risque par assuré](#)
- [Risque de décès](#): un assuré avec un profil de risque r_{stay} peut sortir du portefeuille ($\neq 100\%$)
- [Risque de décès](#): un assuré avec un profil de risque r_{lapse} peut décéder avant de racheter

B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation

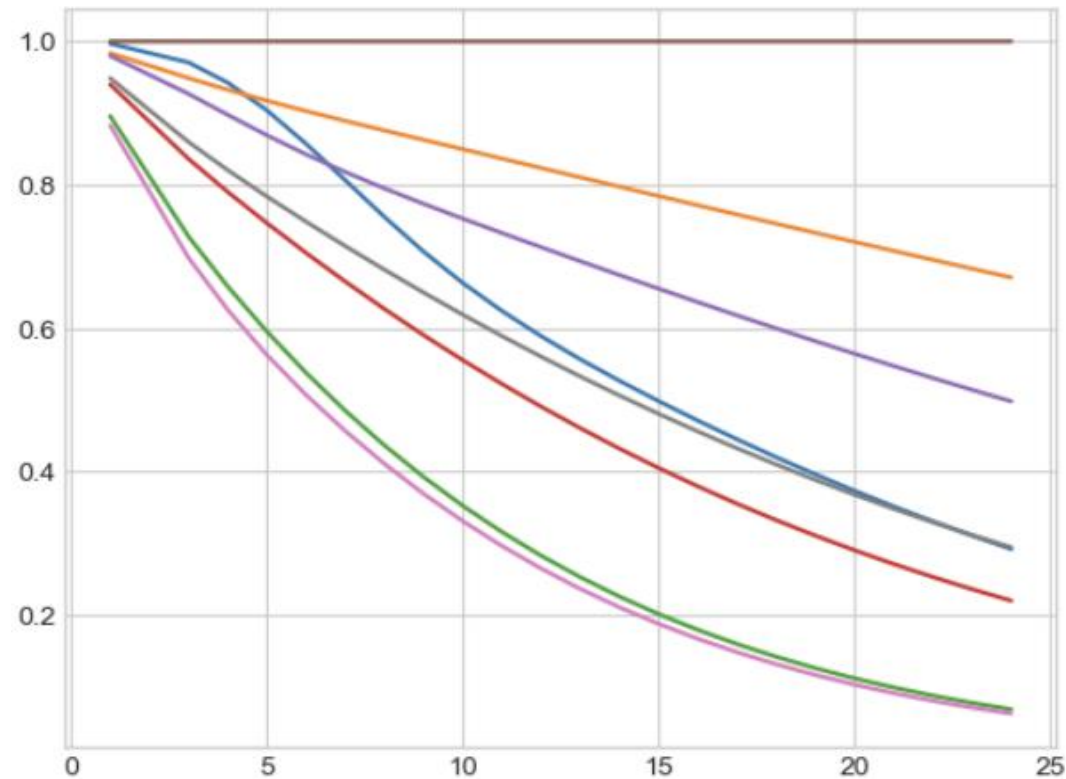
- Modèle de survie à l'échelle individuelle pour r_{stay} et r_{lapse} :

0.000001	1.000000	2.000000	3.000000	4.000000	5.000000	6.000000	7.000000	8.000000	9.000000	...	15.000000	16.000000	17.000000	18.000000	19.000000	20.000000	21.000000	22.000000	23.000000	24.000000
1.0	0.962656	0.923603	0.882695	0.834296	0.792650	0.752985	0.717424	0.687154	0.659541	...	0.429087	0.394909	0.362233	0.331141	0.301694	0.273933	0.247878	0.223536	0.200892	0.179921
1.0	0.974292	0.947069	0.918160	0.883404	0.852985	0.823540	0.796722	0.773563	0.752154	...	0.500308	0.467445	0.435541	0.404695	0.374993	0.346504	0.319287	0.293385	0.268826	0.245630
1.0	0.969314	0.936992	0.902867	0.862122	0.826717	0.792682	0.761892	0.735466	0.711173	...	0.300165	0.266739	0.235904	0.207635	0.181875	0.158542	0.137534	0.118729	0.101996	0.087193
1.0	0.950061	0.898549	0.845392	0.783597	0.731405	0.682570	0.639538	0.603484	0.571070	...	0.120919	0.098284	0.079218	0.063315	0.050178	0.039431	0.030723	0.023734	0.018178	0.013804
1.0	0.980942	0.960761	0.936036	0.911023	0.889213	0.866073	0.841448	0.811772	0.785733	...	0.689179	0.689179	0.453838	0.420304	0.387944	0.356872	0.327181	0.298943	0.272213	0.247027
...
1.0	0.962845	0.928776	0.910926	0.910926	0.910926	0.501328	0.437774	0.380021	0.327931	...	0.119351	0.098729	0.081172	0.066328	0.053866	0.043476	0.034874	0.027800	0.022023	0.017339
1.0	0.969231	0.936823	0.902613	0.861768	0.826282	0.792174	0.761320	0.734841	0.710504	...	0.350898	0.316636	0.284534	0.254620	0.226898	0.201344	0.177916	0.156549	0.137163	0.119666
1.0	0.987680	0.973072	0.957226	0.932298	0.906302	0.879032	0.845988	0.812974	0.784531	...	0.595828	0.572272	0.550725	0.539456	0.539456	0.539456	0.379899	0.347857	0.317294	0.288300
1.0	0.988185	0.975314	0.958749	0.939208	0.918138	0.885259	0.851330	0.816135	0.774043	...	0.542466	0.507623	0.476839	0.450986	0.427689	0.415643	0.415643	0.415643	0.194631	0.168215
1.0	0.967984	0.935608	0.902971	0.870174	0.837312	0.804480	0.771769	0.739267	0.707057	...	0.524059	0.495848	0.468426	0.441828	0.416088	0.391232	0.367282	0.344253	0.322158	0.301002

B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation

- Modèle de survie à l'échelle individuelle pour r_{stay} et r_{lapse} :



B. Enrichissement d'un modèle par les données

5. Extensions du modèle par la prise en compte de risques compétitifs

- Le risque de décès doit être pris en compte
- Un assuré décédé ne rachètera pas
- Une police rachetée ne peut pas se terminer par le décès de l'assuré

- Deux approches:
 - Cause-specific
 - Subdistribution

B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation
5. Extensions du modèle par la prise en compte de risques compétitifs

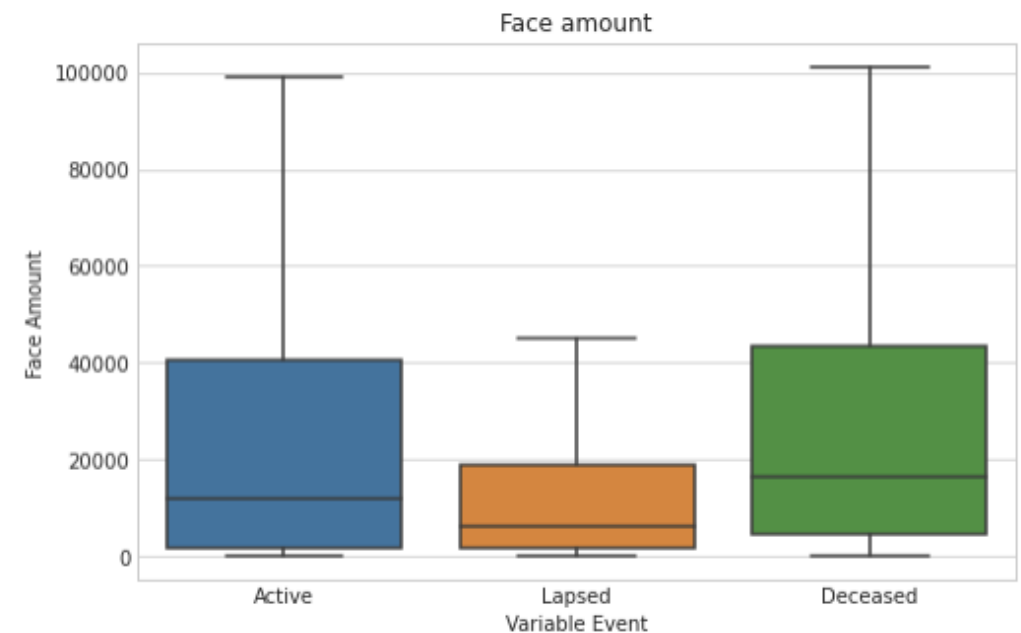
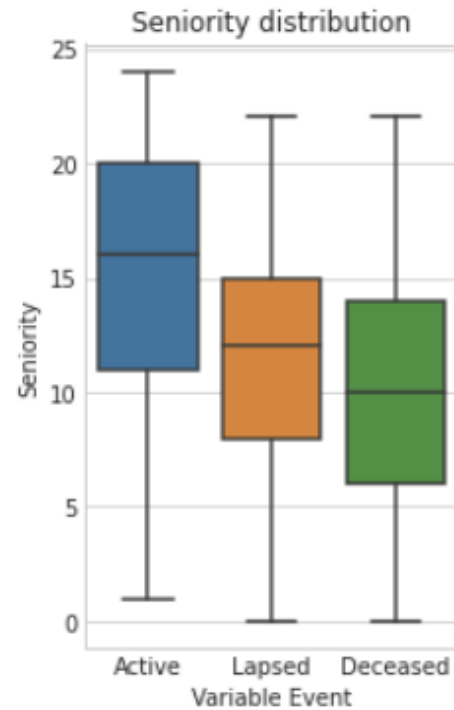
Disclaimer: pour des raisons de confidentialité:

- Toutes les données, statistiques et périmètres présentés dans ce document ont été **modifiés** ou **anonymisés**
- Toutes les analyses, résultats, discussions et conclusions restent **identiques**.

B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation
5. Extensions du modèle par la prise en compte de risques compétitifs

- Real-world french life insurance portfolio
- From 1998 to today (almost!)
- 249k unique policies
- 235k unique policyholders
- 43 covariates



B. Enrichissement d'un modèle par les données

4. Extensions du modèle par la personnalisation
5. Extensions du modèle par la prise en compte de risques compétitifs

L'extension du modèle s'opère sur la modélisation des profils de risque.

Nous considérons 3 modèles:

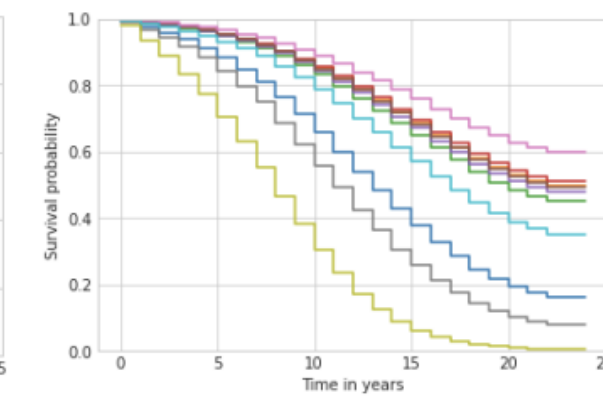
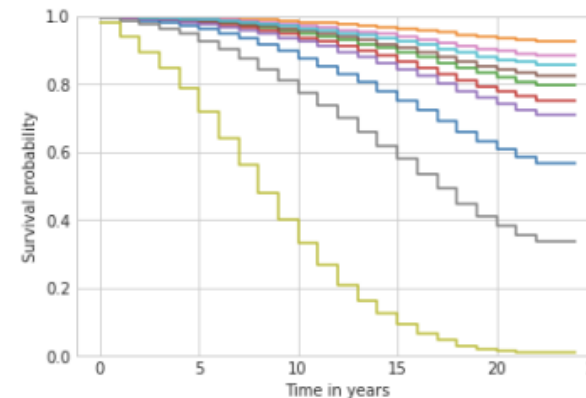
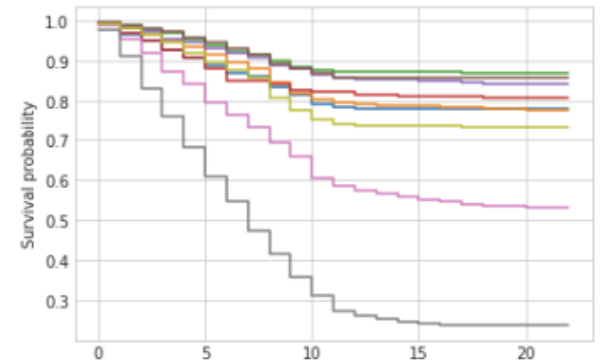
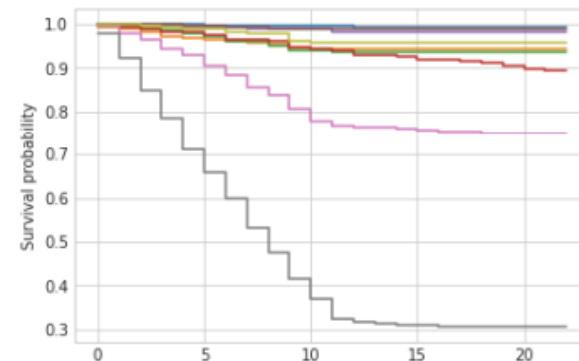
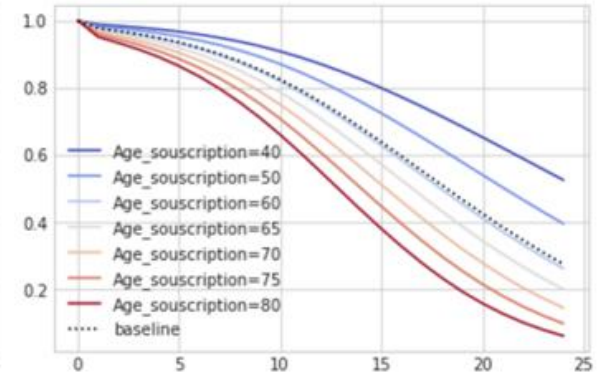
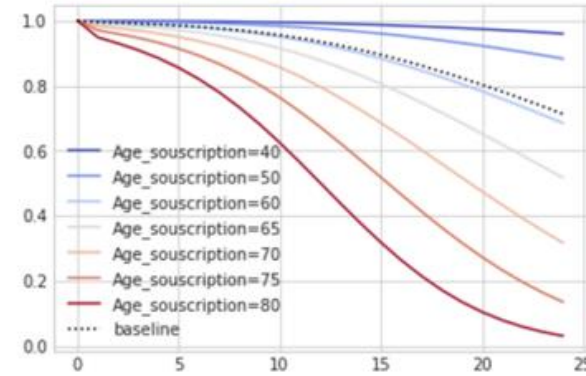
- Cox model
- Random survival forest
- Gradient boosting survival model

En prenant en compte le décès via une approche [cause-specific](#)

B. Enrichissement d'un modèle par les données

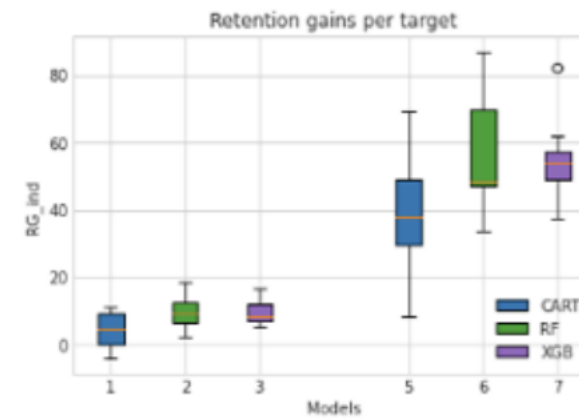
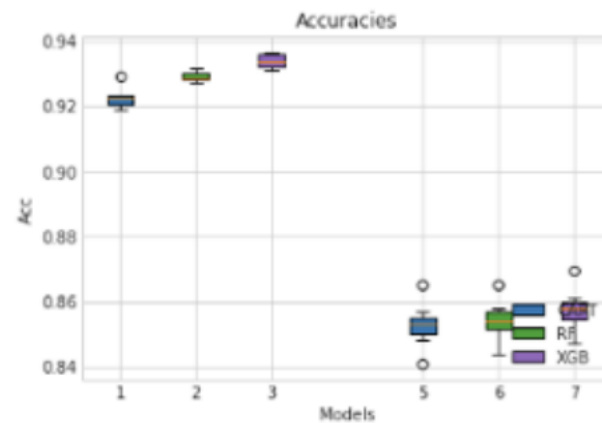
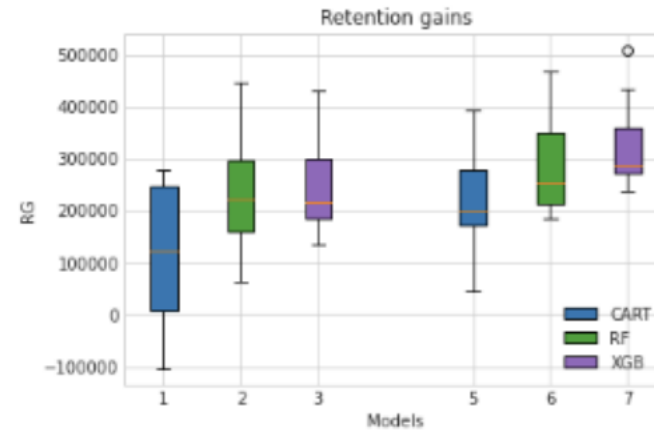
4. 5. Résultats

	Concordance index	
	r_{lapse}	r_{stay}
Cox model	69,5%	80,7%
RSF	71,6%	83,7%
GBSM	73%	84,1%



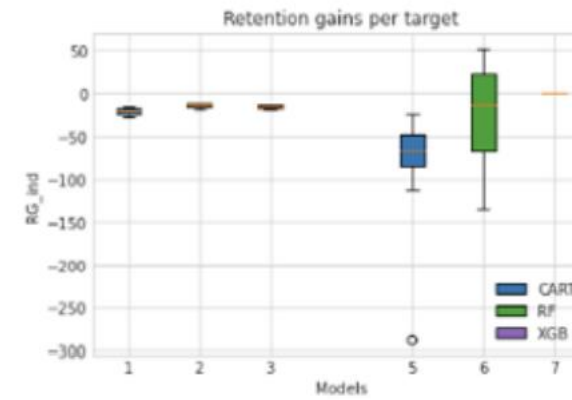
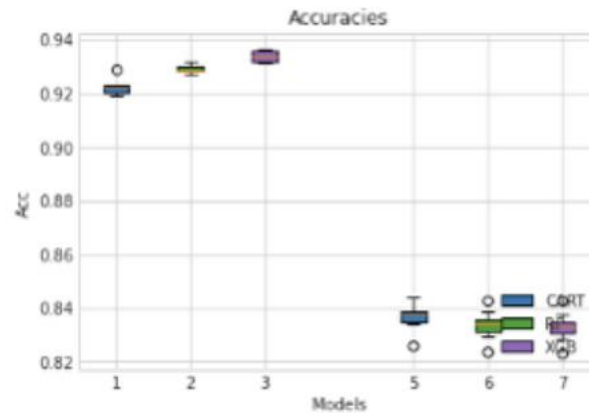
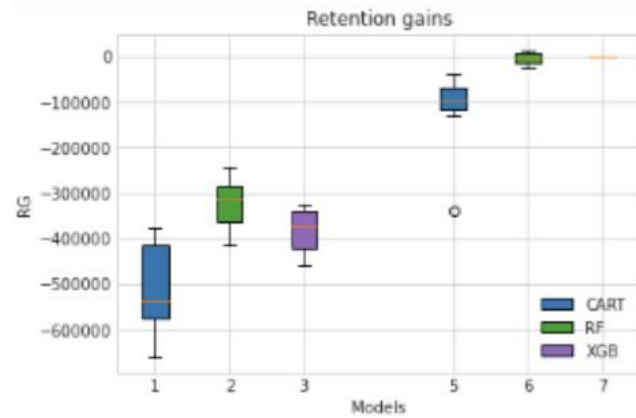
B. Enrichissement d'un modèle par les données

4. 5. Résultats



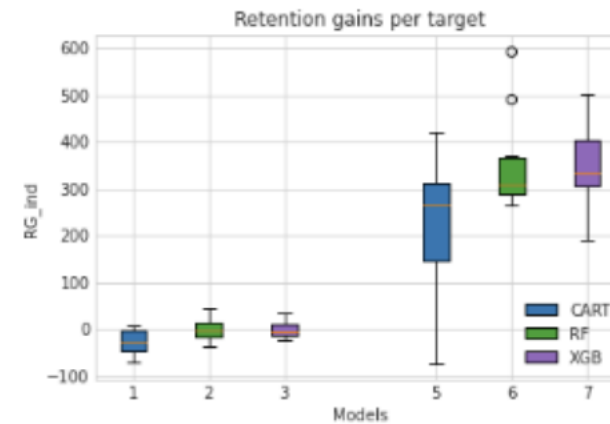
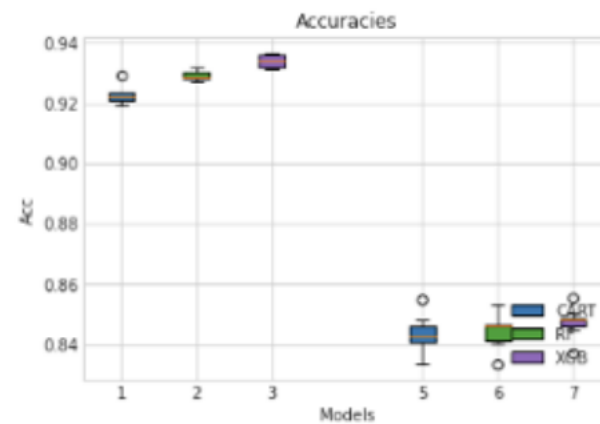
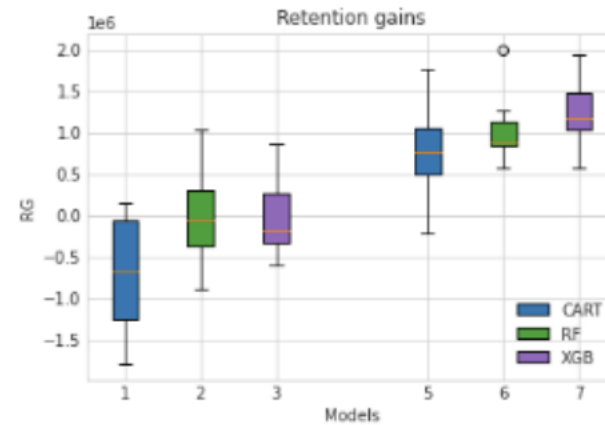
B. Enrichissement d'un modèle par les données

4. 5. Résultats



B. Enrichissement d'un modèle par les données

4. 5. Résultats



B. Enrichissement d'un modèle par les données

4. 5. Résultats

The average observed improvement of a CLV-augmented framework over the classical lapse one is 57,9%^a. If we weight these results by the expected RGs, the average improvement is still 31,7%.

^aUsing XGBoost

The modelization:

- showed to be loss-limiting (Strategy n°4)
- showed that the improvement of a lapse management strategy including CLV grows with the proportion of lapsers with a negative CLV

B. Enrichissement d'un modèle par les données

4. 5. Résultats

This framework can be used in several ways by the insurer. It can help:

- understand what differentiate a subject for which $y_i = 1$ and $\tilde{y}_i = 0$ from the others
- interpret the results at an individualized level
- compares future hypothetical lapse management strategies in order to chose the best one
- answer questions like:
 - For which incentive δ the retention strategy becomes profitable ?
 - For which acceptance probability γ the retention strategy becomes profitable ?
 - at which horizon T , the retention strategy become profitable ? In other words, when can the insurer expect a return on investement ?
- Measure the expected gain of a real retention campaign from the past, at various time horizons

B. Enrichissement d'un modèle par les données 4. 5. Résultats

This framework can be used in several ways by the insurer.

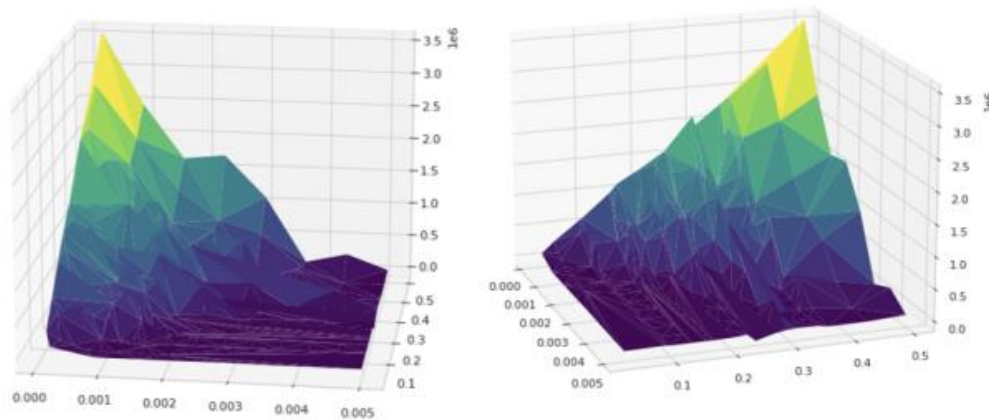


FIGURE: 3d plot (δ , γ , RG)

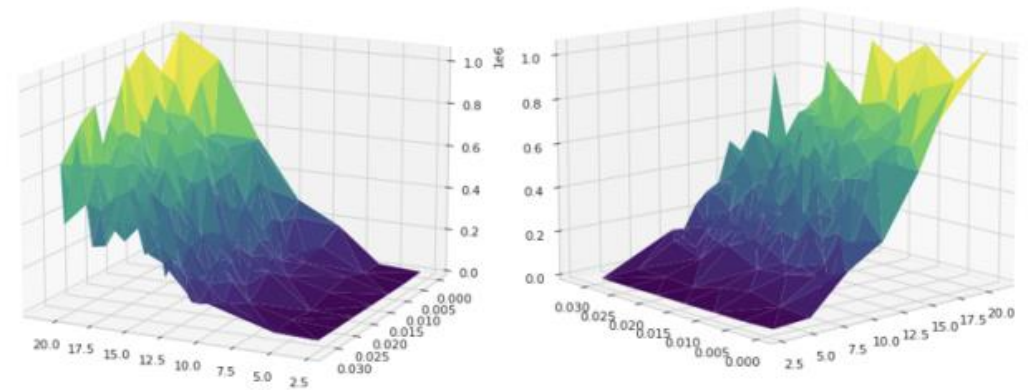


FIGURE: 3d plot (d, T, RG)

B. Enrichissement d'un modèle par les données

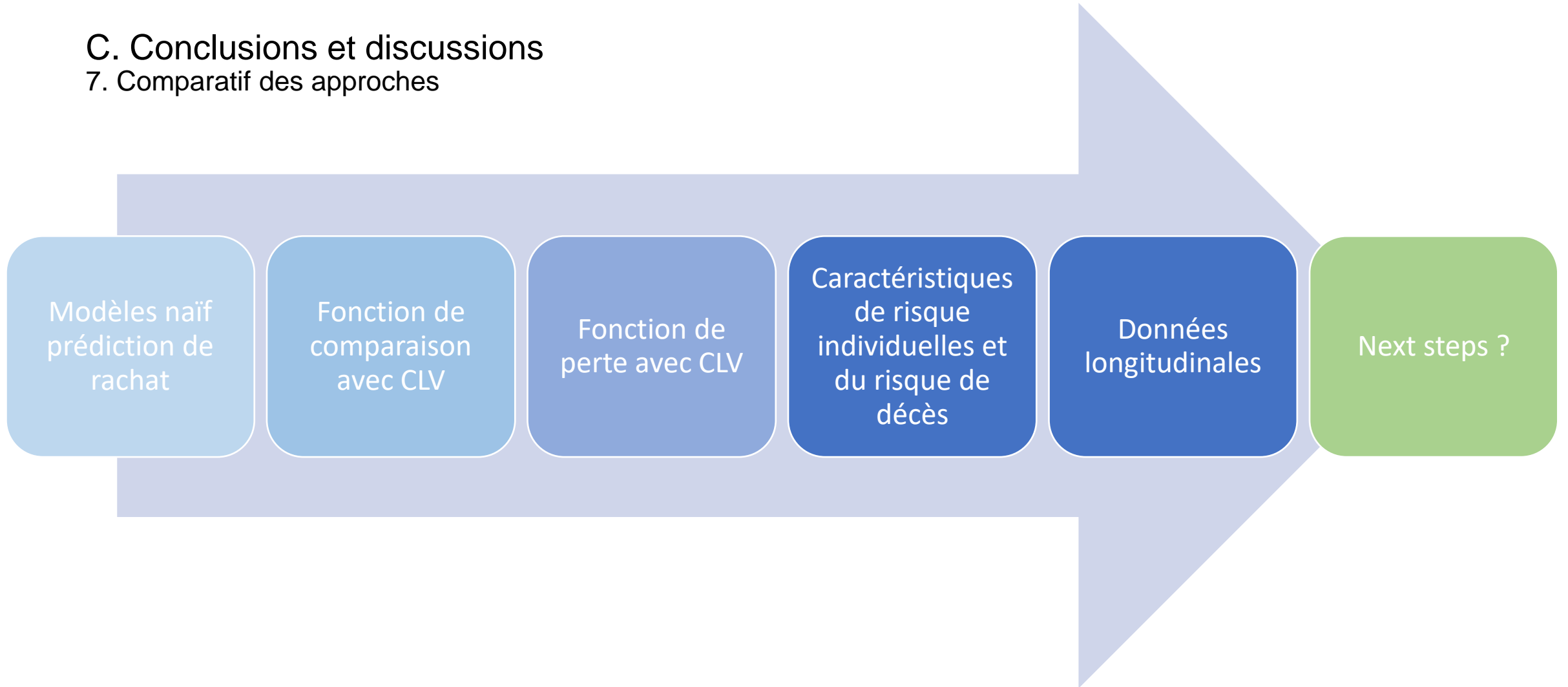
6. Extensions du modèle avec des données longitudinales

- Deux assurés avec le même encours, et les mêmes caractéristiques auraient une prédiction similaire
- Or, leur trajectoire passée respective peut-être complètement différente !
- Exploiter la richesse contenue dans l'historique des données de l'assuré
- En considérant non pas une seule observation par assuré, mais des observations régulières (tous les mois/ans par exemple)

- Travail en cours sur cette extension

C. Conclusions et discussions

7. Comparatif des approches



C. Conclusions et discussions

8. Conclusions et pistes d'amélioration

Conclusion en termes opérationnels

- Prédire si un assuré va racheté son contrat ou prédire si l'assureur devrait investir pour le retenir dans le portefeuille sont **deux problèmes différents**.
- La prise en compte de la valeur client dans la stratégie de rétention **aide à la prise de décision** de l'assureur quand aux politiques de rétention
- Les modèles de Machine Learning peuvent surpasser les approches paramétriques classiques, y compris en analyse de survie
- Ces modélisations permettent de détecter une mauvaise stratégie de rétention, qui serait vouée à générer une perte pour l'assureur

C. Conclusions et discussions

8. Conclusions et pistes d'amélioration

Quelques limites et pistes d'amélioration

- Une stratégie unique, figée pour tous les assurés n'est pas réaliste
- Le profil de risque d'un assuré peut évoluer au cours du temps : r_{stay} pourrait contenir un peu de risque de rachat
- Les paramètres γ et δ sont fortement corrélés, cela pourrait être étudié

Merci de votre attention !

Des questions ?

Annexe

Table 3a: Average Confusion Matrix of XGB

		Predicted	
		Stay	Lapse
Actual	Stay	309,111 54.6%	38,450 6.8%
	Lapse	81,177 14.3%	137,660 24.3%

Table 3b: Average Confusion Matrix of SVM

		Predicted	
		Stay	Lapse
Actual	Stay	310,258 54.8%	37,303 6.6%
	Lapse	88,339 15.6%	130,498 23.0%

Table 3c: Average Confusion Matrix of CART

		Predicted	
		Stay	Lapse
Actual	Stay	296,320 52.3%	51,241 9.0%
	Lapse	78,209 13.8%	140,628 24.8%

Table 3d: Average Confusion Matrix of LR

		Predicted	
		Stay	Lapse
Actual	Stay	315,184 55.6%	32,377 5.7%
	Lapse	97,486 17.2%	121,351 21.4%

Annexe

Table 3a: Average Confusion Matrix of XGB

		Predicted	
		Stay	Lapse
Actual	Stay	309,111 54.6%	38,450 6.8%
	Lapse	81,177 14.3%	137,660 24.3%

Table 8a: Average Confusion Matrix of XGB_R1

		Predicted	
		Stay	Lapse
Actual	Stay	329,357 58.1%	18,204 3.2%
	Lapse	113,948 20.1%	104,889 18.5%

Table 3c: Average Confusion Matrix of CART

		Predicted	
		Stay	Lapse
Actual	Stay	296,320 52.3%	51,241 9.0%
	Lapse	78,209 13.8%	140,628 24.8%

Table 3d: Average Confusion Matrix of LR

		Predicted	
		Stay	Lapse
Actual	Stay	315,184 55.6%	32,377 5.7%
	Lapse	97,486 17.2%	121,351 21.4%

Annexe

Table 11: Improvements of XGB_R over the Best of Binary Models

Profitability (p)	Incentive (2nd year)	Probability (gamma)	Cost (c)	Discount Rate (d)	Best of Binary Models	XGB_R	Increase	Percentage of Increase
0.30%	0.03%	20%	10	2.0%	-942,888	1,642,720	2,585,608	157%
1.00%	0.03%	20%	10	2.0%	23,250,096	23,634,878	384,782	2%
0.50%	0.05%	20%	10	2.0%	-546,631	3,213,095	3,759,725	117%
0.50%	0.01%	20%	10	2.0%	12,578,031	13,272,649	694,618	5%
0.50%	0.03%	25%	10	2.0%	8,693,534	9,477,878	784,344	8%
0.50%	0.03%	15%	10	2.0%	2,324,515	3,995,017	1,670,502	42%
0.50%	0.03%	20%	50	2.0%	-821,181	3,424,200	4,245,380	124%
0.50%	0.03%	20%	1	2.0%	6,828,897	7,782,795	953,898	12%
0.50%	0.03%	20%	10	5.0%	4,268,910	5,311,218	1,042,308	20%
0.50%	0.03%	20%	10	1.0%	5,741,453	7,104,433	1,362,981	19%