

Modélisation prédictive

Document 218081

Transformer
les mégadonnées en
grandes possibilités



Canadian
Institute of
Actuaries



Institut
canadien
des actuaires

Articles sur la modélisation prédictive : Transformer les mégadonnées en grandes possibilités

Au cours des dernières années, les données sont devenues un moteur clé de la croissance économique et l'assise sur laquelle les industries sont bâties. Pour beaucoup, savoir quoi faire avec cette matière brute complexe est maintenant un enjeu organisationnel de taille. C'est là que les actuaires interviennent. En s'appuyant sur la modélisation prédictive, ils peuvent aider à transformer les mégadonnées en grandes possibilités.

La modélisation prédictive analyse des ensembles de données pour cerner des interdépendances significatives. Elle a ensuite recours à ces interdépendances pour mieux prévoir les résultats et prendre plus rapidement des décisions avisées et déployables. Elle s'appuie sur de l'information historique pour décrire des interdépendances antérieures dans lesquels puiser des tendances pour l'avenir. Ces tendances peuvent s'appliquer à plusieurs aspects d'une entreprise, par exemple, le comportement du consommateur, du fournisseur et du distributeur.

La modélisation prédictive s'inspire de nombreuses disciplines, notamment la statistique, la modélisation, l'optimisation, l'analyse de grappes, l'étude de marché et la programmation informatique. Son application repose habituellement sur une puissance informatique considérable et chevauche des domaines comme l'auto-apprentissage et l'intelligence artificielle.



Pourquoi les actuaires?

La modélisation prédictive présente des similitudes avec la science actuarielle. Les actuaires examinent les interdépendances dans de grands ensembles de données et les relient à des problèmes opérationnels concrets, généralement dans le contexte d'une société d'assurance, d'un régime de retraite ou d'une

fonction de gestion du risque. Les modèles générés sont instaurés dans plusieurs fonctions opérationnelles clés qui ont une incidence sur les résultats de l'entreprise.

Les actuaires reçoivent une formation en mathématiques axée sur l'élaboration de modèles et la résolution de problèmes complexes ayant des conséquences financières. Avec les bonnes compétences en analyse, la rigueur d'un professionnel, une aptitude pour l'informatique et l'habileté de voir au-delà du concept mathématique complexe, les actuaires excellent dans l'art de mettre au point des solutions opérationnelles dans un contexte incertain et en évolution.



Outre les domaines traditionnels qu'ils connaissent bien, les actuaires sont sollicités par des organismes œuvrant dans les secteurs bancaire, de la gestion des placements, du commerce électronique, de la gestion des risques météorologiques, des transports, de l'énergie et des programmes sociaux.

Renseignez-vous sur la façon dont la modélisation prédictive et les actuaires peuvent aider votre entreprise.

La Commission sur la modélisation prédictive de l'Institut canadien des actuaires tient à remercier les bénévoles dévoués, les nombreux et précieux auteurs ainsi que le personnel du siège social de l'ICA pour leur contribution à ce projet.

Pourquoi maintenant?

La modélisation prédictive n'est pas un concept nouveau. En fait, le milieu des affaires utilise depuis des décennies des techniques de base comme la régression logistique. La modélisation prédictive gagne en popularité, et ce, pour bien des raisons :

- Les entreprises font face à une énorme pression concurrentielle et doivent se différencier, offrir au client une expérience plus agréable et mieux organiser les processus.
- Il y a une révolution au niveau des mégadonnées. Nous créons maintenant 2,5 trillions d'octets de données par jour¹, prêtes à être analysées.
- De nouveaux types de données, par exemple, les médias sociaux, le Web, les capteurs, l'audio et les images sont maintenant disponibles.
- Améliorations significatives de la puissance informatique avec traitement parallèle à haute vitesse et réseau distribué à peu de coûts.

- Diminution des coûts de stockage des données.
- Nouvelle technologie hautement extensible pour entreposer et gérer les données structurées et non structurées, p. ex., Hadoop.
- Stockage en nuage et informatique en nuage.
- Nouvelles innovations en matière d'auto-apprentissage, d'apprentissage profond et d'intelligence artificielle.
- Disponibilité d'ensembles de données de style « open source ».
- Logiciels libres et de style « open source » (R et Python).

Quels problèmes l'analyse prédictive permet-elle de régler?

Les pronostics tirés des modèles prédictifs peuvent s'appliquer à plusieurs aspects d'une entreprise, y compris au comportement du consommateur, du fournisseur et du distributeur. La modélisation prédictive peut améliorer bien des processus opérationnels, dont les suivants :

- **Ventes et marketing** : cerner les groupes de vente cibles, déterminer chaque caractéristique d'un client corrélée avec la décision d'achat, comprendre les comportements des acheteurs et recommander le bon produit, et jumeler les clients prospectifs aux préposés des ventes les plus appropriés.
- **Expérience du client** : offrir des services adaptés aux besoins des clients et leur donner de l'information pertinente.
- **Gestion des activités courantes** : déterminer et fidéliser les clients, offrir des produits supplémentaires aux clients existants et établir le profil des consommateurs.
- **Établissement des prix** : établir plus exactement les prix et projeter l'impact des écarts par rapport aux paramètres de l'établissement des prix.
- **Gestion des risques** : déterminer la gamme des résultats des principales mesures du rendement, modéliser le capital/les fonds propres.
- **Détection de la fraude** : repérer les activités potentiellement frauduleuses, intervenir rapidement en cas de soupçon de fraude et découvrir les modèles de fraude.
- **Analyse des RH** : jumeler les employés aux meilleures fonctions, améliorer le taux de maintien en poste des employés et évaluer l'incidence des politiques en ressources humaines sur le rendement.

Dans les sociétés d'assurance, outre les applications ci-dessus, la modélisation prédictive peut améliorer ce qui suit :

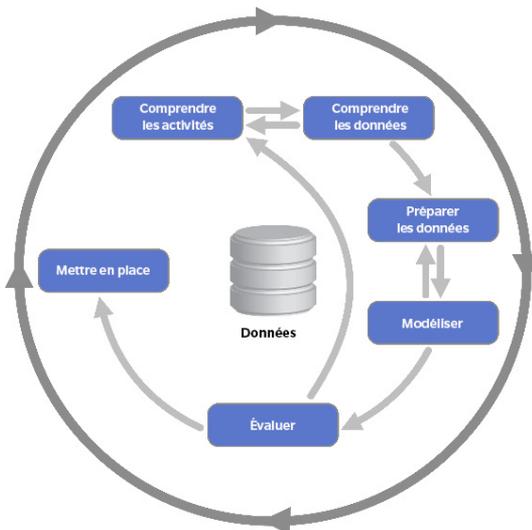
- **Souscription** : déterminer le meilleur profil de risque et prioriser les efforts d'acceptation, identifier les requérants qui doivent souscrire plus d'assurance, appuyer la souscription simplifiée.

- **Sinistres** : prévoir la fréquence et la gravité des sinistres, trier les sinistres, prioriser les ressources de gestion des sinistres.
- **Établissement des réserves** : établir des réserves plus exactes.
- **Analyse de l'expérience** : déterminer les moteurs derrière l'expérience, améliorer la modélisation des hypothèses de mortalité/déchéance.

Comment se déroule un processus type de modélisation prédictive?

Pour trouver les bonnes solutions aux problèmes opérationnels grâce à la modélisation prédictive, il faut collaborer de près avec les spécialistes en la matière du début à la fin. Voici les éléments centraux du processus :

1. Cerner un problème pour lequel les prévisions des résultats ou du comportement futur peuvent faire en sorte de prendre des décisions d'affaires plus exactes et efficaces.
2. Comprendre l'activité : connaître les produits, les besoins des intervenants, les ressources disponibles et les données dont on dispose, et déterminer les hypothèses et les contraintes ainsi que les moyens de mettre en place une solution de modélisation prédictive.
3. Définir clairement le résultat que doit prédire (la variable de réponse) le modèle prédictif.



Comprendre les données

Pour comprendre les données, la première étape consiste à déterminer les ressources nécessaires pour élaborer le modèle prédictif. À cette fin, il faut comprendre les sources et le flux de données afin de pouvoir recueillir les données. Une fois ces dernières regroupées, on procède à une exploration initiale pour bien les connaître, faire ressortir les pronostics, repérer les problèmes de qualité, puis rectifier la situation. Souvent, les données sont imparfaites et

il importe de bien comprendre et de documenter les limites ainsi que les solutions aux problèmes concernant les données.



La collecte et la préparation des données sont des étapes cruciales. Elles doivent être explorées, épurées, combinées, transformées et formatées. L'ensemble de

données qui en découle est ensuite réparti en au moins deux groupes, de sorte qu'une partie des données peut entraîner (calibrer) le modèle et l'autre, en évaluer la performance. Une structure de gouvernance des données aide à faciliter le processus et à éviter de perdre du temps à des itérations subséquentes. À cette fin, il faut habituellement faire appel à des processus automatisés de collecte des données provenant de sources internes et externes, traiter et transformer les données documentées, établir des règles sur leur seuil de qualité, puis les entreposer.

L'art de la science

La modélisation, c'est la rencontre entre l'art et la science. Le volet art, c'est l'intuition a priori et des hypothèses sur l'interdépendance entre chaque variable prédictive, c'est-à-dire les variables utilisées pour prédire le résultat, et la variable de réponse. Le volet science, c'est l'exécution de l'analyse quantitative des données et l'utilisation d'une grande collection de modèles. Pour choisir les modèles, il faut limiter le nombre de paramètres et, en bout de ligne, sélectionner un modèle que les intervenants peuvent comprendre et exécuter.

Le modèle doit être évalué en fonction de données qu'il n'a jamais vues. Un modèle prédictif doit être généralisable à des données invisibles, et non seulement décrire les données qui servent à calibrer le modèle. Il se peut que le modèle soit inapplicable, incorrect, instable ou mal utilisé. L'évaluation périodique d'un modèle aide à maintenir la confiance qu'il suscite et permet de l'améliorer en procédant à de nouvelles itérations du processus.

Mise en œuvre

Si le modèle fonctionne bien et que les intervenants opérationnels l'acceptent, la prochaine étape consiste à le mettre en œuvre. D'habitude, le service des technologies de l'information collabore avec l'équipe opérationnelle pour intégrer le modèle prédictif aux processus de l'organisation. La mise en œuvre comprend habituellement une période d'essai dans un environnement hors production avant le déploiement complet. Cette période d'essai permet aux modélisateurs d'identifier les écarts et de les corriger.

Les modèles de production devraient être examinés périodiquement en comparant les données réelles aux prédictions du modèle pour s'assurer que le rendement ne s'est pas considérablement détérioré. Il se peut que le modèle doive être rafraîchi ou recréé si la détérioration est évidente.

Quelles sont les techniques?

Les techniques de modélisation prédictive peuvent être classées de bien des façons. Le résumé qui suit, bien qu'il ne s'agisse pas d'une liste exhaustive de toutes les techniques disponibles, présente une classification générale de celles qui sont utilisées.

I. Catégorie des modèles d'apprentissage supervisé

Dans un problème d'apprentissage supervisé, le modélisateur a recours à un ensemble de données où la valeur des variables de réponse est connue. Par exemple, pour prédire si le titulaire résiliera sa police d'assurance, il pourrait utiliser une série de données historiques sur les polices dans lesquelles la décision sur la résiliation de chaque consommateur est connue et déjà codée dans les données. La plupart des applications courantes dans la modélisation prédictive pour le secteur de l'assurance se fondent sur les techniques d'apprentissage supervisé.

Dans le cadre de l'apprentissage supervisé, il y a deux sous-ensembles de modèle principaux :

i. Modèles de classification

Dans un modèle de classification, l'objectif consiste à prédire un résultat catégorique. Il pourrait s'agir d'une réponse binaire (c.-à-d., 1 si la police est résiliée et 0, si elle ne l'est pas) ou d'une



catégorisation générale/de multiples états (c.-à-d., 1 si le titulaire est en santé, 2, s'il est handicapé, 3, s'il est retraité et 4, s'il est décédé).

ii. Modèles de régression

Dans un modèle de régression, l'objectif consiste à prédire un résultat continu. Par exemple, la gravité d'une collision faisant l'objet d'une demande de règlement d'assurance automobile.

Il y a des techniques de modélisation qui peuvent régler des problèmes tant de classification que de régression. Cependant, certains modèles se prêtent mieux à l'un ou à l'autre. Par exemple, on peut utiliser avec efficacité un modèle de régression linéaire pour prédire une variable continue (un problème de régression), mais pas avec une aussi grande efficacité pour prédire une variable de réponse binaire (un problème de classification).

II. Catégorie des modèles d'apprentissage non supervisé

Dans un problème d'apprentissage non supervisé, le modélisateur ne tente pas de prédire un certain résultat, mais cherche plutôt à découvrir la structure ou les caractéristiques latentes des données. Par exemple, le modélisateur peut analyser la base de consommateurs d'une entreprise pour repérer ses principaux segments de clients.

Dans le cadre de l'apprentissage non supervisé, il y a deux sous-ensembles principaux de modèles :

i. Modèles d'analyse de grappes

Dans un problème d'analyse de grappes, l'objectif consiste à regrouper les données dans des catégories ou des grappes semblables. Étant donné que cela n'est pas supervisé, les algorithmes de regroupement tenteront de trouver dans les données sous-jacentes des modèles qui apporteront plus d'information au modélisateur. Parmi les exemples, il y a l'analyse en classification à K moyenne et l'algorithme DBSCAN (« density-based spatial clustering of applications with noise » ou *algorithme de regroupement d'applications spatial fondé sur la densité avec bruit*).

ii. Modèles de réduction de la dimensionnalité

Dans un problème de réduction de la dimensionnalité, l'objectif consiste à condenser le nombre de variables dont il faut tenir compte. À nouveau, étant donné que cela n'est pas supervisé, ces algorithmes tentent de trouver le nombre le moins élevé de variables qui fournissent la plus grande quantité d'information. Parmi les exemples, il y a l'analyse des composantes principales et l'analyse discriminante linéaire.

III. Apprentissage semi-supervisé

Dans un problème d'apprentissage semi-supervisé, le modélisateur est probablement confronté à un ensemble de données dans lequel seulement une partie des variables de réponse est connue. Dans ce scénario, une option consiste à utiliser la partie non supervisée des données pour améliorer un modèle supervisé. Ce pourrait, par exemple, être une technique pertinente si vous essayez de prédire l'efficacité d'une stratégie de ventes continues en place depuis cinq ans, mais que vous n'avez commencé que l'année précédente à rassembler des données sur les résultats des ventes.

Quelles sont les considérations relatives à la vie privée?²

Certaines applications des modèles prédictifs nécessitent des ensembles de données contenant de l'information granulaire sur les particuliers. Les données personnelles pouvant être recueillies comprennent notamment les informations personnelles identifiables (IPI), les données des médias sociaux, les données de navigation, les habitudes d'achat des consommateurs et le suivi de localisation.

Il faut évaluer les considérations relatives à la vie privée propres à chaque situation sous l'angle des objectifs visant à mettre au point le modèle prédictif



et à déterminer la façon de l'utiliser. Utiliser l'information médicale personnelle ou génétique d'un ensemble de données dans lequel on aura supprimé l'identité pour faire de la recherche afin d'élaborer des décisions socialement utiles pour le public, c'est très différent que de prédire le taux de grossesse chez les clientes pour faire de la publicité.

Dans toute application d'un modèle prédictif qui a recours à des renseignements potentiellement délicats, il faut tenir dûment compte de bien des aspects de la vie privée, notamment la connaissance et le consentement du consommateur, la transparence dans l'utilisation des données, l'utilisation éthique de l'analyse et la reddition de comptes.

Au Canada, les entreprises du secteur privé qui mettent au point des modèles prédictifs s'appuyant sur des données personnelles doivent veiller à ce que leurs pratiques soient conformes aux principes énoncés dans la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDÉ).

Qu'en est-il de l'innovation?

La modélisation prédictive est au centre des grandes innovations opérationnelles et technologiques.

Les activités ou les idées innovatrices cherchent à échanger des données sans intervention humaine pour alimenter un modèle prédictif qui déclenchera une décision opérationnelle.



La télématique intègre la collecte, la transmission et le stockage de données d'un objet à distance, comme un véhicule. Le secteur de l'assurance automobile a recours à des modèles prédictifs pour établir un lien entre les habitudes de conduite et les risques liés aux sinistres, et celui de l'expédition, pour gérer les coûts de la main-d'œuvre et de l'essence d'une flotte.

Les technologies portables sont des appareils électroniques intelligents qui peuvent être portés. L'industrie de l'assurance envisage le recours à cette technologie pour lier les habitudes de vie et les comportements à l'information concernant les sinistres. Vous ne le savez peut-être pas, mais votre téléphone intelligent fournit de l'information semblable à votre fournisseur de services et aux entreprises dont vous utilisez les applications.

Les marchés boursiers et les salles de négociation sont aussi touchés par la modélisation prédictive par le biais de la négociation par algorithme. Un modèle, par exemple, « lira » les gazouillis (*tweets*) pour y déceler de l'information au sujet d'une action et négociera en conséquence en quelques secondes.

Les modèles prédictifs peuvent utiliser les données recueillies auprès des centrales électriques pour anticiper le moment où il faudra peut-être procéder à des réparations, ce qui permet de faire un entretien préventif potentiellement moins dispendieux que des réparations d'urgence. Les modèles sont aussi utilisés dans d'autres domaines du secteur énergétique comme pour déterminer le résultat d'une éolienne. Vous recevez peut-être un relevé mensuel de votre consommation d'électricité : cette information et les suggestions pour diminuer votre consommation d'énergie sont générées par la modélisation prédictive.

Et la concurrence?

La concurrence sur le marché intervient beaucoup pour encourager les organisations à innover et à trouver des façons de continuer à réussir et à offrir un bon rendement sur les placements aux actionnaires. La modélisation prédictive peut aider à cette fin de diverses façons à l'échelle des fonctions.



Les entreprises qui le font correctement peuvent acquérir un avantage concurrentiel qui se traduit souvent par une diminution des coûts ou une augmentation des ventes et de la rentabilité. Les organisations qui ont recours à l'analyse prédictive pour sélectionner les risques peuvent mieux comprendre les risques et les coûts

et donc mieux les gérer en fonction de leur propension aux risques. Cette perspective permet aussi de mieux tarifer les produits pour les consommateurs en jumelant le risque à un prix proportionnel. Les entreprises dont les meilleurs prix correspondent au risque s'approprieront éventuellement une plus grande part du marché et rehausseront leur rentabilité.

Aujourd'hui, le consommateur réclame une meilleure expérience ainsi que des produits et un engagement plus adaptés à ses besoins. La modélisation prédictive donne l'occasion d'optimiser chaque interaction avec un consommateur et de prévoir le meilleur plan d'action personnalisé pour lui.

Si vos concurrents prennent des décisions plus éclairées grâce à la modélisation prédictive, vous devez leur emboîter le pas pour demeurer dans la course.

Pourquoi l'analyse prédictive est-elle si importante pour la direction et le conseil d'administration?

Compte tenu des possibilités de réduire les coûts et d'augmenter les ventes et la rentabilité, et étant donné que bien des entreprises ont des données dont elles peuvent tirer profit et qu'elles peuvent mieux utiliser, nul doute que les entreprises doivent investir temps et argent dans leurs capacités d'analyse prédictive.

Si elle est bien exécutée, l'analyse prédictive peut permettre de prendre des décisions plus éclairées, de réduire les risques et de mettre à jour des pronostics que l'entreprise peut utiliser pour acquérir un avantage concurrentiel. L'analyse peut fournir la base factuelle nécessaire pour améliorer le processus décisionnel et, avec le temps, les entreprises fonctionneront ainsi avec plus d'efficacité et d'efficacité. L'analyse peut réduire au minimum l'erreur humaine dans des situations où un modèle prédictif peut prendre une décision plus cohérente, exacte et objective, d'où la possibilité d'économiser et d'être plus rentable.



De quoi ai-je besoin pour utiliser la modélisation prédictive?

La modélisation prédictive se trouve à l'intersection des mathématiques, de la statistique et de l'informatique. Il faut aussi de l'expertise dans le domaine du problème à régler : actuariat, économie, ingénierie, médecine et ainsi de suite. Il est aussi nécessaire d'avoir un esprit critique pour faire preuve de jugement à l'égard des interdépendances établies par l'ordinateur et de trouver de l'intuition dans le modèle.

La complexité des modèles prédictifs varie et les intervenants doivent donc avoir des aptitudes marquées en communications pour avoir confiance dans le modèle et le comprendre. Il importe de pouvoir décrire en termes simples et de donner des détails au besoin et en fonction des demandes et des antécédents des membres de l'auditoire. Une aptitude marquée pour la communication permet de s'assurer que les limites du modèle sont comprises et que le modèle n'est pas mal utilisé.

Les données sont le matériau brut des modèles prédictifs. Il est essentiel de mettre en place un processus de gouvernance des données avec une solide infrastructure pour recueillir, entreposer, protéger, valider, examiner et réutiliser les données, et y avoir accès. Les organisations devraient s'efforcer de stocker toutes les données qu'elles recueillent même si elles n'en ont pas immédiatement besoin.

Pour exécuter la modélisation prédictive, il faut un logiciel spécial, une capacité de stockage et une puissance informatique. Heureusement, les récentes percées en informatique permettent une certaine souplesse dans le choix des solutions. Une grande organisation avec un budget imposant pourrait décider d'entreposer toutes ses données à l'interne et d'acheter des super ordinateurs pour exécuter les calculs. Une organisation plus petite pourrait décider d'utiliser des plateformes de programmation « open source » (R, Python), l'entreposage en nuage et l'informatique en nuage comme solution rentable avec des résultats comparables.

Peu de personnes possèdent toutes les compétences requises. Par conséquent, une organisation pourrait engager des spécialistes des données, des personnes connaissant les activités et des spécialistes de la technologie de l'information pour instaurer des solutions en analyse prédictive. Les actuaires contribuent grandement aux travaux d'une équipe de ce genre.

Sources :

¹ <https://cloudtweaks.com/2015/03/surprising-facts-and-stats-about-the-big-data-industry/>

² https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/recherche/consulter-les-travaux-de-recherche-sur-la-protection-de-la-vie-privee/2012/pa_201208/

https://en.wikipedia.org/wiki/Information_privacy_law

<https://www.digitaltrends.com/social-media/why-consumers-are-increasingly-willing-to-trade-data-for-personalization/>

Ce livret a été rédigé par la Commission sur la modélisation prédictive de l'Institut canadien des actuaires.

Membres :

Jean-Yves Rioux, FICA, FSA, CERA, président

Jeffrey Baer, FICA, FCAS

Patrick Duplessis, AICA, FSA, CERA

June Quah, FICA, FSA

Suzanna (Ying) Zhan, FICA, FSA

L'Institut canadien des actuaires (ICA) est l'organisme bilingue national et le porte-parole de la profession actuarielle au Canada. Ses membres se vouent à fournir des services et des conseils actuariels de la plus haute qualité. L'Institut fait passer l'intérêt public avant les besoins de la profession et de ses membres.

Les actuaires sont des experts en gestion du risque. Ils utilisent les mathématiques, la statistique et les probabilités pour assurer la sécurité financière des Canadiens. Les domaines d'exercice traditionnels sont notamment les assurances (assurance-vie et assurances IARD), les placements, les régimes de retraite, l'expertise devant les tribunaux et la gestion du risque d'entreprise.

Les opinions formulées dans le présent livret sont celles des auteurs.