L'IA pour la prédiction et la modélisation des risques en santé publique

Des performances prédictives à de meilleures décisions en matière de santé publique

Gemma Postill | *Population Health Analytics Lab* | *University of Toronto*

Méthodes d'IA en santé publique | *AI4PH* | Agence de la santé publique du Canada (ASPC)

Population Health Analytics Lab





Professeure
Laura Rosella

Sujets de la présentation d'aujourd'hui

01 Introduction à la prédiction et à la modélisation des risques

02 Modèles statistiques et d'apprentissage automatique pour la prédiction

03 Élaboration et évaluation d'un modèle de prédiction fondé sur l'apprentissage automatique

04 Utilisation responsable de la prédiction en santé publique

1. Introduction à la prédiction et à la modélisation des risques

Qu’essayons-nous de prédire, et pourquoi est-ce important?

Types de questions épidémiologiques

DESCRIPTIVE

À quoi ressemble la population atteinte de diabète?

p. ex. quelle est la prévalence du diabète dans notre communauté?

PRÉDICTIVE

Qui est susceptible de subir cette conséquence?

p. ex. quel est le risque de développer le diabète de type 2 sur une période de 10 ans?

CAUSALE

Que se passerait-il si nous intervenions?

p. ex. est-ce que le fait de modifier l'environnement alimentaire permettrait de réduire l'incidence du diabète?

Les approches descriptives, prédictives et causales posent des questions différentes

Types de questions épidémiologiques - 2

DESCRIPTIVE

À quoi ressemble la population atteinte de diabète?

p. ex. quelle est la prévalence du diabète dans notre communauté?

PRÉDICTIVE

Qui est susceptible de subir cette conséquence?

p. ex. quel est le risque de développer le diabète de type 2 sur une période de 10 ans?

CAUSALE

Que se passerait-il si nous intervenions?

p. ex. est-ce que le fait de modifier l'environnement alimentaire permettrait de réduire l'incidence du diabète?

Les approches descriptives, prédictives et causales posent des questions différentes; aujourd'hui, nous nous concentrons sur la prédiction

À quelles questions de santé publique les modèles prédictifs peuvent-ils apporter des réponses?

Prédiction du risque individuel

- Qui pourrait développer la maladie?
- À qui pourrait-on proposer des mesures de prévention?
- Comment le risque varie-t-il d'un groupe à l'autre?

Prévision à l'échelle de la population

- Combien de cas pourraient survenir?
- Où la demande de services pourrait-elle augmenter?
- Comment planifier les capacités?

Exemple : outil d'évaluation des risques liés au diabète dans la population (*DPoRT*)

La modélisation prédictive peut intervenir à différents niveaux et permettre de répondre à diverses questions en matière de santé publique.

Plan	Exemple de prédiction	Application en santé publique
Individuel	Qui risque de développer le diabète au cours des 10 prochaines années?	Prévention ciblée
Sous-groupe de population	Quels sont les groupes pour lesquels on prévoit la charge la plus élevée liée au diabète?	Identification des populations prioritaires
Géographique	Quelles sont les communautés qui devraient être le plus touchées par le diabète?	Répartition géographique des ressources
Temporel	Comment la charge future liée aux diagnostics ou aux complications évoluera-t-elle au cours des 20 prochaines années?	Surveillance et planification des capacités

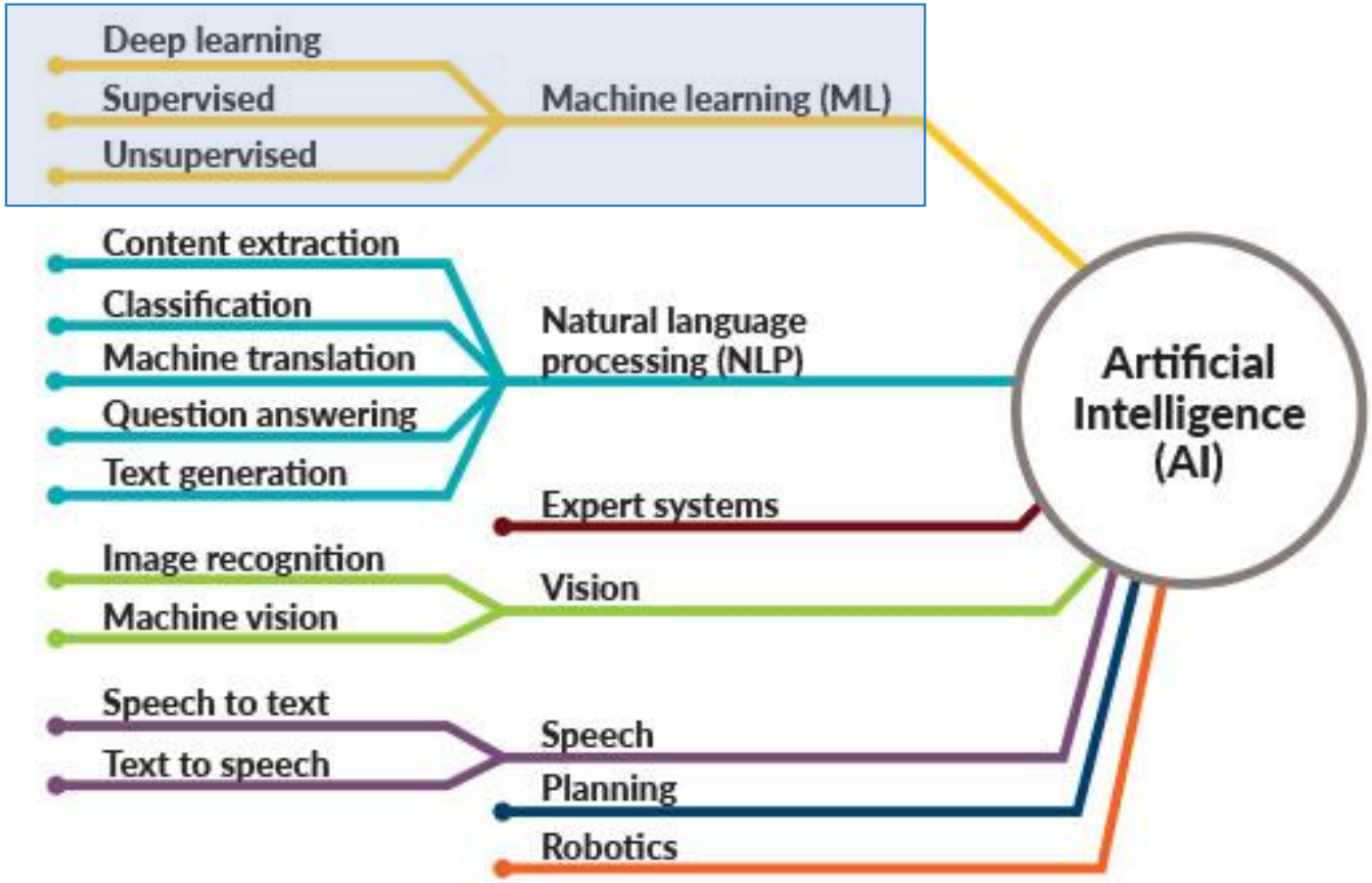
Traduction d'un exemple des travaux de Laura Rosella : L'outil d'évaluation des risques liés au diabète dans la population (DPoRT) évalue le risque futur de diabète de type 2 à partir des données d'enquêtes démographiques. Les estimations individuelles du risque au sein de la population peuvent être agrégées pour caractériser le risque au sein de la population et orienter la planification de mesures de prévention=.

Rosella et al. *J Epidemiol Community Health*. 2011 doi : 10.1136/jech.2009.102244.

2. Modèles statistiques et d’apprentissage automatique

Un algorithme plus complexe permet-il d’améliorer la décision?

Apprentissage automatique : une approche relevant de l'IA



*English only content / Contenu en anglais seulement

Image source : https://www.automation.com/en-us/articles/august-2022/ai-machine-learning-human-intelligent-systems

Approches de modélisation : la complexité est un choix et non un objectif

	Modèles traditionnels **p. ex. régression logistique, de Poisson, de Cox**	**Apprentissage automatique** **p. ex. algorithmes *Random Forests* et *XGBoost*, réseaux neuronaux**
Spécification du modèle	▪ Relations entre les variables prédictibles et les résultats **explicitement spécifiées**.	▪ Relations **apprises de manière plus flexible** à partir des données.
Non-linéarité et interactions	▪ **Spécification explicite** généralement nécessaire (p. ex. splines, termes d'interaction).	▪ Saisie possible de schémas complexes sans avoir à spécifier explicitement chaque relation.
Interprétabilité	▪ Interprétations statistiques familières pour les coefficients.	▪ Méthodes supplémentaires souvent nécessaires pour expliquer les prédictions.
Complexité	▪ Réglages et exigences de calcul souvent moins élevés.	▪ Réglages des hyperparamètres et de ressources de calcul impliqués davantage

Le terme « apprentissage automatique » ne désigne pas un modèle en particulier

Différentes approches d'apprentissage automatique peuvent être utilisées pour la prédiction des risques

Régression pénalisée
Lasso, crête, filet élastique

Utile lorsqu'il y a de nombreuses variables prédictives; réduit ou sélectionne les variables.

Méthodes arborescentes
Arbres décisionnels, *Random Forests*

Capable de saisir les interactions et la non-linéarité; souvent robuste et flexible.

Méthodes de « *boosting* »
XGBoost*, *LightGBM*, *CatBoost

Améliore la prédiction de façon séquentielle; offre souvent d'excellentes performances sur les données de santé sous forme de tableaux.

Séparateurs à vaste marge

Capable de classer des schémas complexes à l'aide de différents noyaux.

Réseaux neuronaux/apprentissage profond

Capable d'apprendre des relations très complexes; nécessite souvent des ensembles de données plus volumineux et plus de réglages.

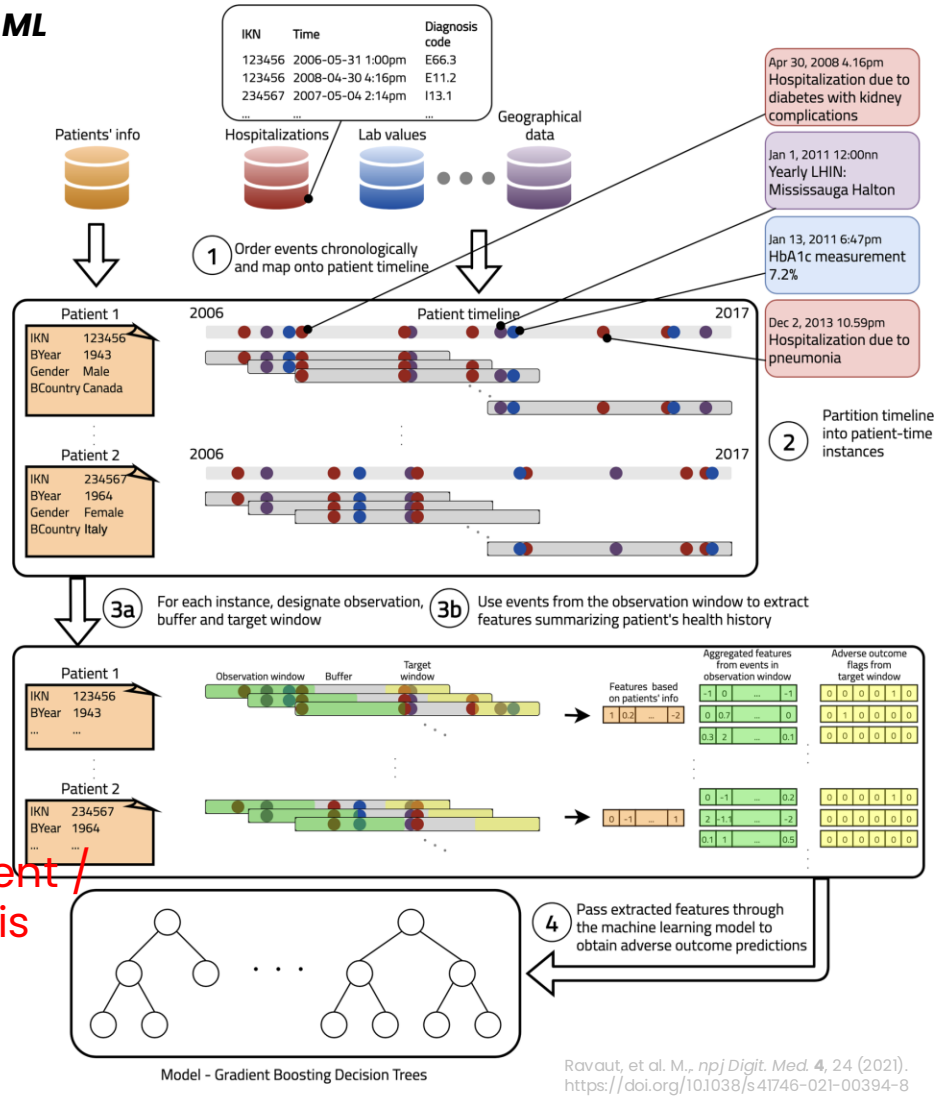
Apprentissage automatique pour la survie
Random Survival Forests*, *DeepSurv

Utile pour prédire le risque de délai jusqu'à l'événement.

La méthode appropriée dépend de la tâche de prédiction, des données et du besoin d'interprétabilité.

Exemple : prédiction du risque de diabète au sein de la population

Article sur *DPORT - ML*



Ravaut, et al. M., *npj Digit. Med.* **4**, 24 (2021). https://doi.org/10.1038/s41746-021-00394-8

***DPoRT* — outil d'évaluation des risques liés au diabète dans la population**

- Algorithme à l'échelle de la population permettant d'estimer le risque de développer le diabète de type 2 sur une période de 10 ans.
- Approche traditionnelle fondée sur la régression, à partir de facteurs de risque recueillis de façon systématique.

***DPoRT-ML* — approche fondée sur l'apprentissage automatique**

- Prédiction élargie du risque de diabète à l'échelle de la population grâce à l'apprentissage automatique.
- Exploration des relations plus flexibles, notamment la non-linéarité et les interactions entre les variables prédictives.
- Illustration des possibilités d'intégrer des données de santé plus riches et plus complexes.

La principale différence : l'apprentissage automatique permet de saisir des relations plus complexes dans les données, mais une plus grande complexité ne signifie pas nécessairement une meilleure prédiction.

*English only content / Contenu en anglais seulement

3. Élaboration et évaluation d’un modèle de prédiction fondé sur l’apprentissage automatique

Comment passer des données à des estimations de risque fiables?

Processus de la modélisation prédictive

Manuscript de *CMAJ*
Exemple de code

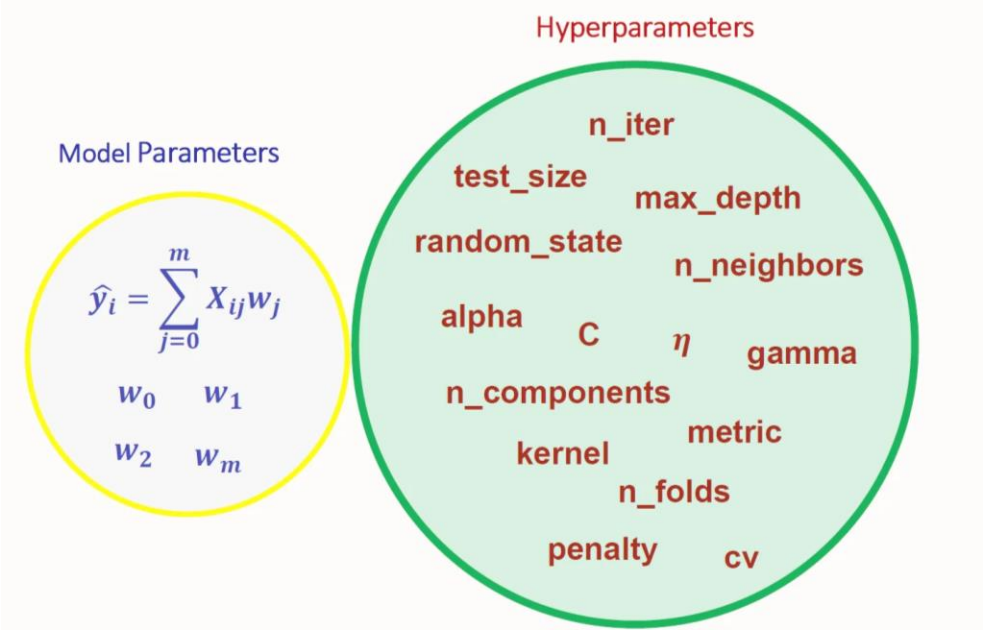
Définir le cas d'utilisation → **Constituer une cohorte** → **Choisir les variables prédictives** → **Fractionner et entraîner** → **Valider** → **Évaluer l'action**

- Les variables prédictives doivent être disponibles au moment de la prédiction
- Il faut séparer l'entraînement de la validation externe temporelle/géographique
- Il faut prévenir les fuites, évaluer les données manquantes et leur provenance, et préserver la population cible

En quoi l'entraînement des modèles d'apprentissage automatique est-il unique?

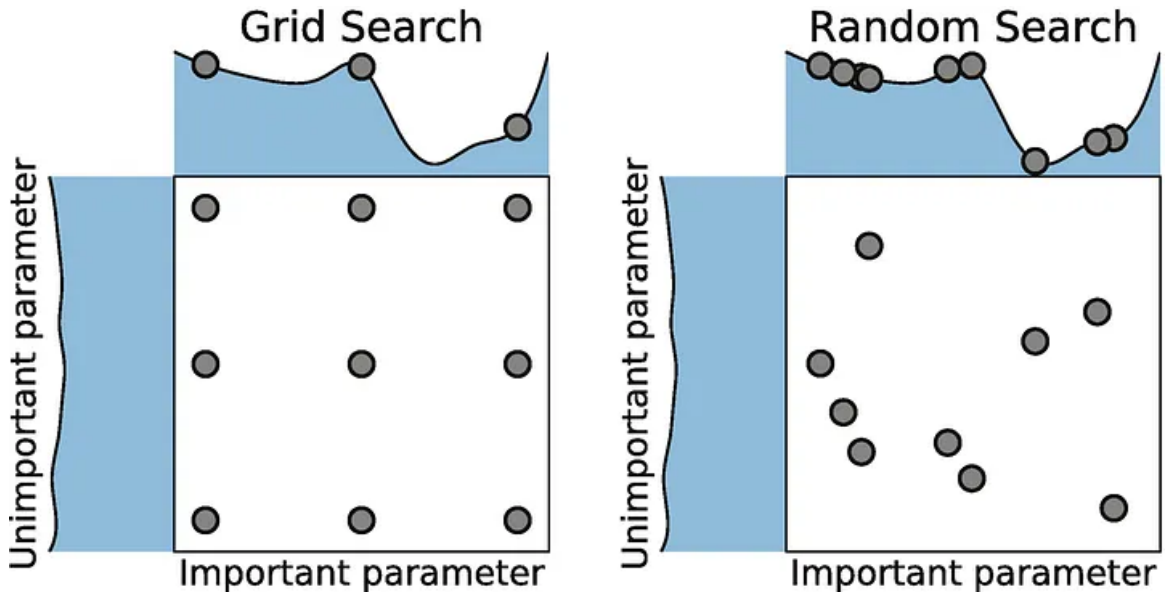
Manuscript de *CMAJ*
Exemple de code

Hyperparamètres du modèle



Source : https://towardsdatascience.com/model-parameters-and-hyperparameters-in-machine-learning-what-is-the-difference-702d30970f6

Approches pour choisir les hyperparamètres

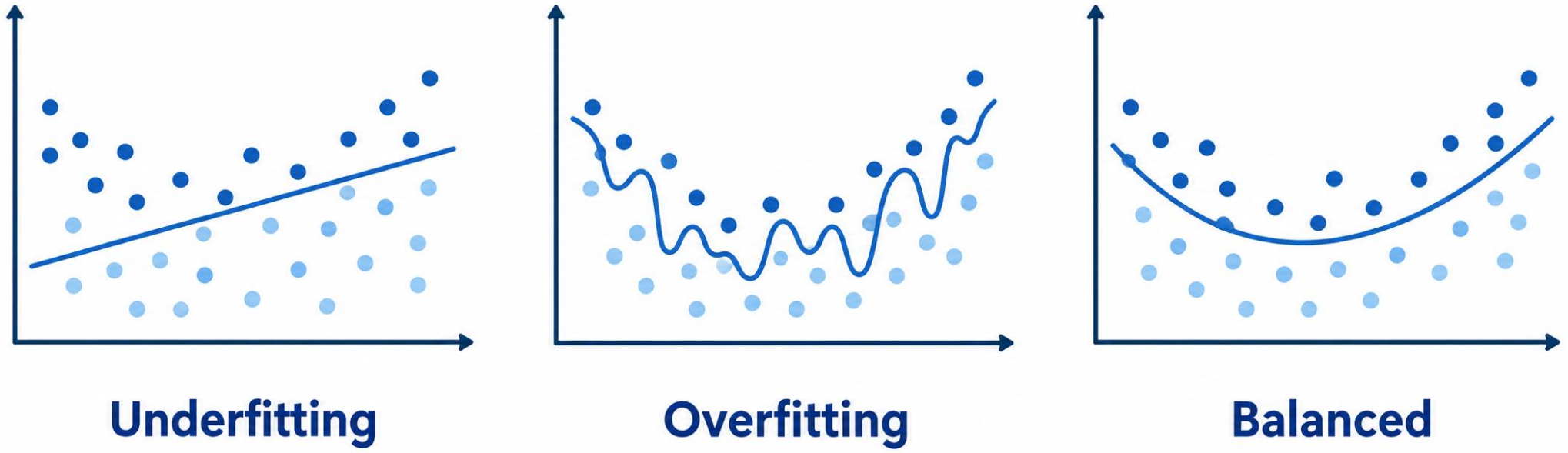


Source de l'image : Bergstra, J., Bengio, Y. : *Random search for hyper-parameter optimization*. Journal of Machine Learning Research 13, 281–305 (2012)

*English only content / Contenu en anglais seulement

Évaluation des modèles : une flexibilité accrue est-elle toujours synonyme d'amélioration?

Manuscript de *CMAJ*
Exemple de code



*English only content / Contenu en anglais seulement

Source de l'image : https://www.r-bloggers.com/2022/09/how-to-avoid-overfitting/

Évaluation des modèles : une flexibilité accrue est-elle toujours synonyme d'amélioration? - 2

Manuscript de *CMAJ*
Exemple de code

Discrimination

Le modèle peut-il classer les personnes à haut risque avant celles à faible risque?

Aire sous la courbe ROC / précision-rappel

Étalonnage

Les probabilités prévues correspondent-elles au risque observé?

Courbe d'étalonnage

Seuil de performance

Qui est identifié comme nécessitant une intervention à un seuil donné?

Sensibilité / spécificité / VPP

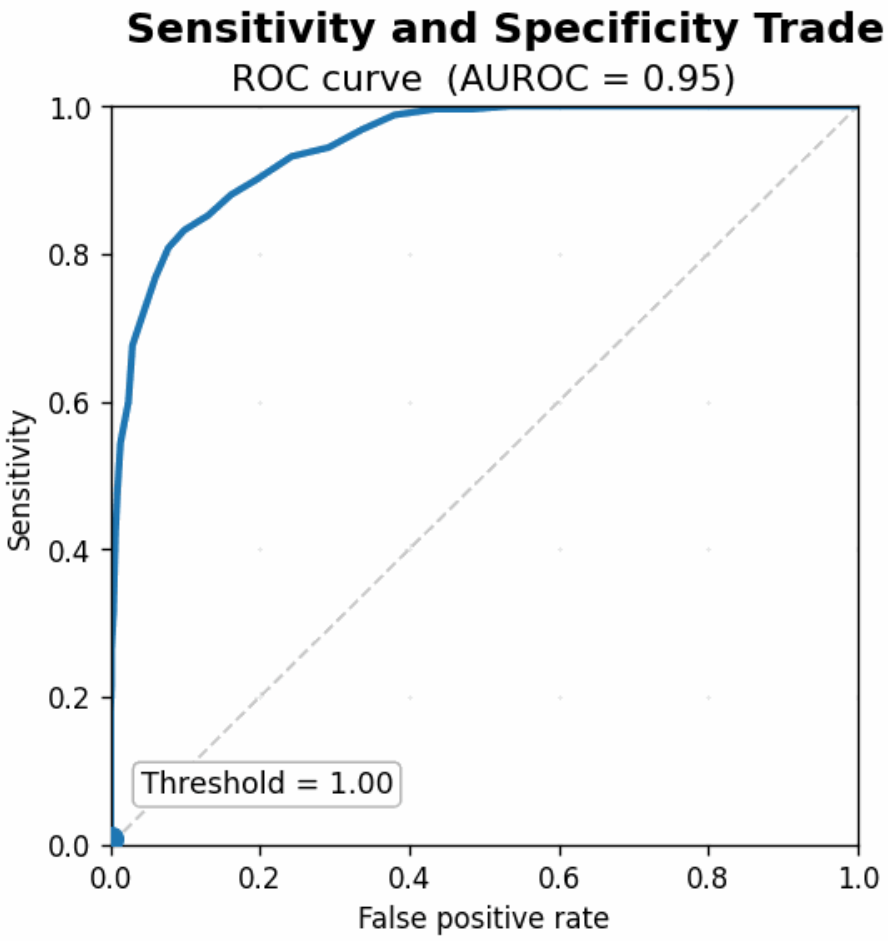
Utilité décisionnelle

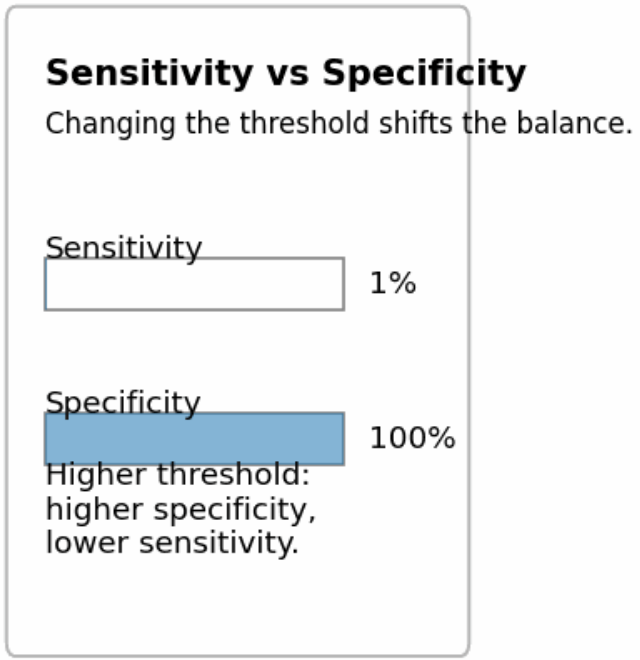
Le fait de s'appuyer sur le modèle permet-il de prendre de meilleures décisions?

Bénéfice net / évaluation d'incidence

Évaluation des modèles : discrimination

Manuscript de *CMAJ*
Exemple de code





La discrimination en matière de modèles d'apprentissage automatique est généralement évaluée à l'aide de l'aire sous la courbe ROC.

La valeur clinique de l'aire sous la courbe ROC dépend du seuil choisi, et par conséquent, du compromis entre la sensibilité et la spécificité

*English only content / Contenu en anglais seulement

Choisir les seuils de risque pour les interventions de santé publique

Manuscript de *CMAJ*
Exemple de code

Seuil inférieur

- Plus de personnes admissibles
- Sensibilité accrue (en général)
- Demande plus forte
- Plus de faux positifs

Seuil supérieur

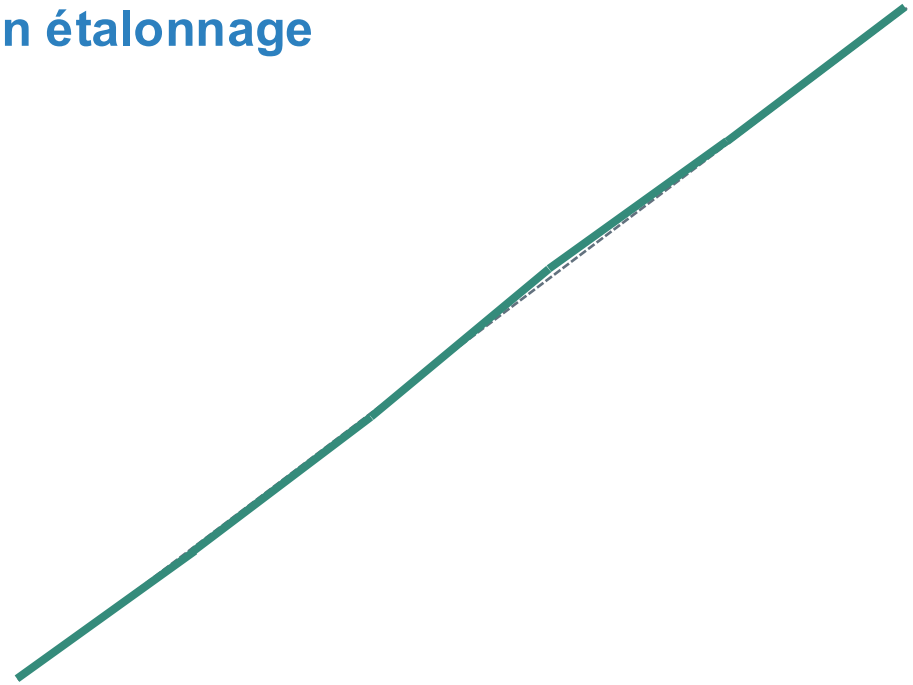
- Moins de personnes admissibles
- Spécificité accrue (en général)
- Demande moins forte
- Plus de cas non détectés

Le seuil optimal dépend des avantages de l'intervention, des capacités, des coûts et de l'équité.

Évaluation des modèles : discrimination ≠ étalonnage

Les modèles suivants pourraient présenter la même discrimination

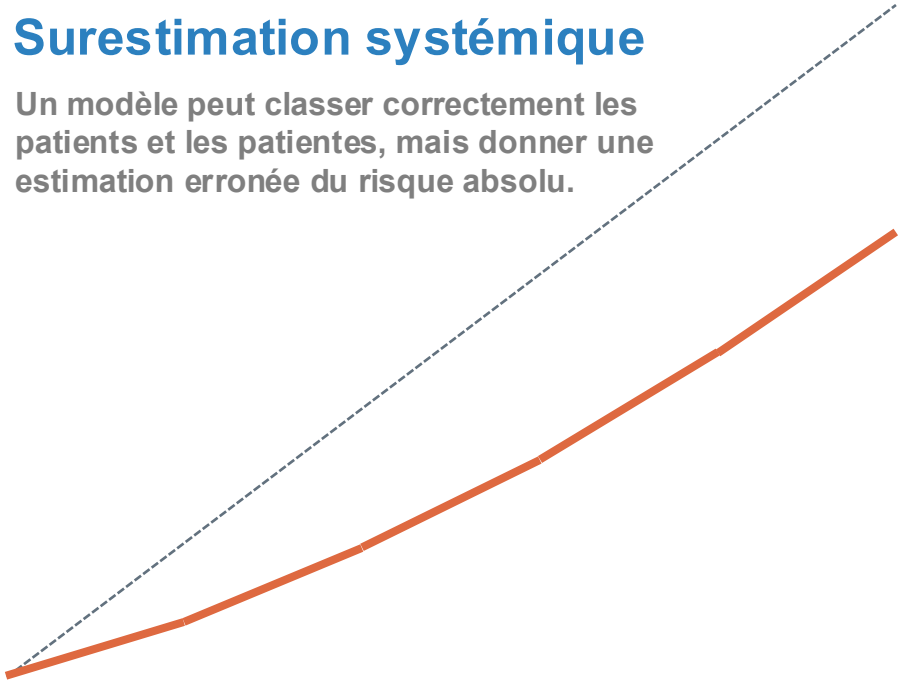
Bon étalonnage



Probabilité prévue →

Surestimation systémique

Un modèle peut classer correctement les patients et les patientes, mais donner une estimation erronée du risque absolu.



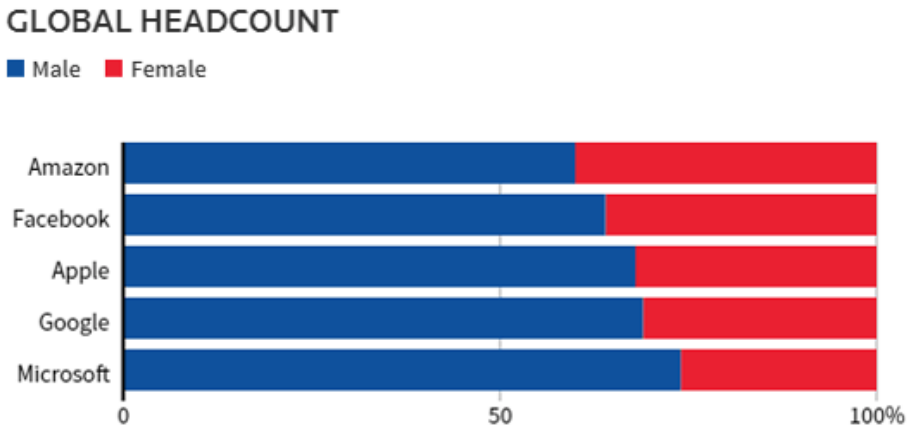
Probabilité prévue →

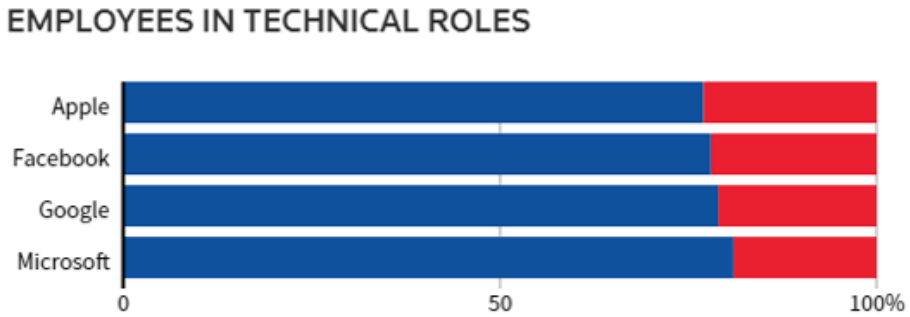
Courbes d'étalonnage à titre d'illustration (il ne s'agit pas des résultats d'une étude empirique).

Évaluation des modèles : évaluation des sous-groupes

En 2018, Amazon a abandonné son outil de recrutement secret basé sur l'IA, lequel a démontré un biais envers les femmes!

Le système pénalisait les CV qui comportaient le mot « *female* », comme dans « *President of Female Business Leadership Group Chapter* ».





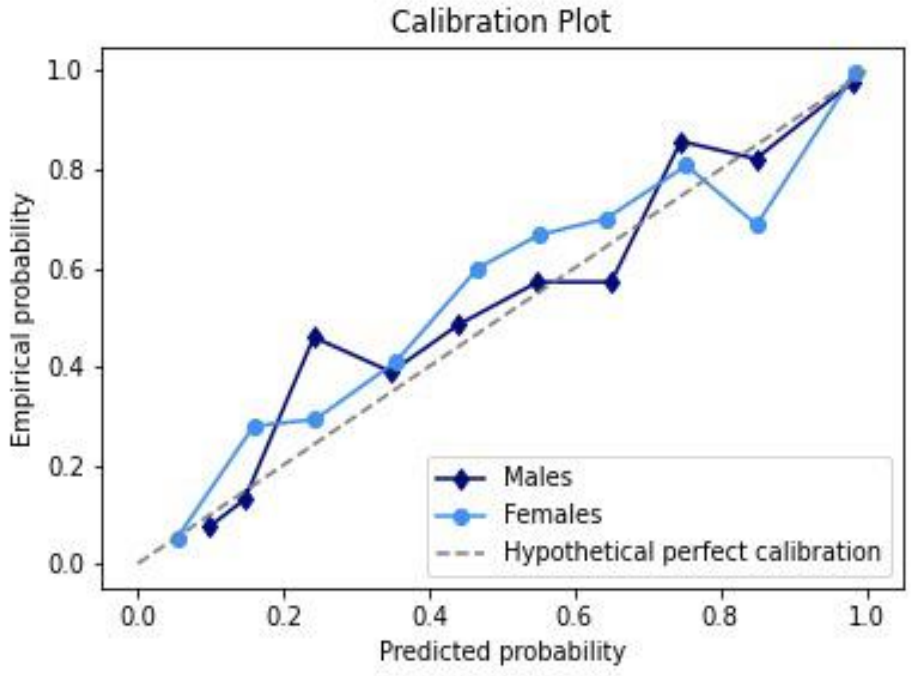
*English only content / Contenu en anglais seulement

Évaluation des modèles : évaluation des sous-groupes - 2

Manuscript de *CMAJ*
Exemple de code

L'évaluation des sous-groupes peut aider à détecter d'éventuels biais et favoriser l'élaboration de modèles équitables.

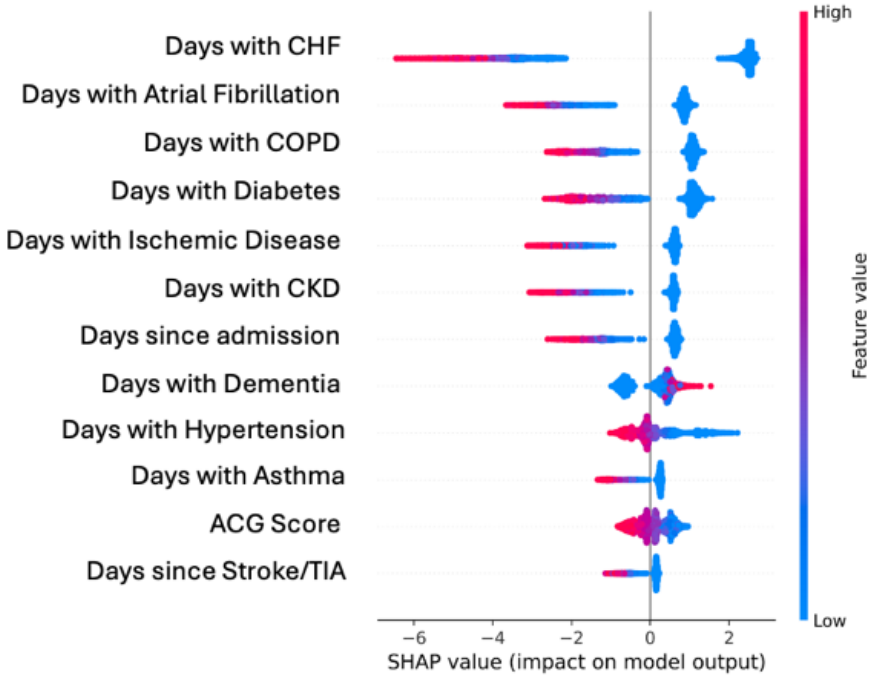
	Total (n = 1856)	Femmes (n = 898)	Hommes (n = 958)
Exactitude	0,89	0,90	0,88
Précision	0,89	0,91	0,88
Rappel	0,89	0,90	0,88
Score F1	0,89	0,90	0,88
ASC	0,89	0,91	0,88
Faux positifs, n (%)	57 (3,1)	24 (2,7)	33 (3,4)
Faux négatifs, n (%)	147 (7,9)	63 (7,0)	84 (8,8)



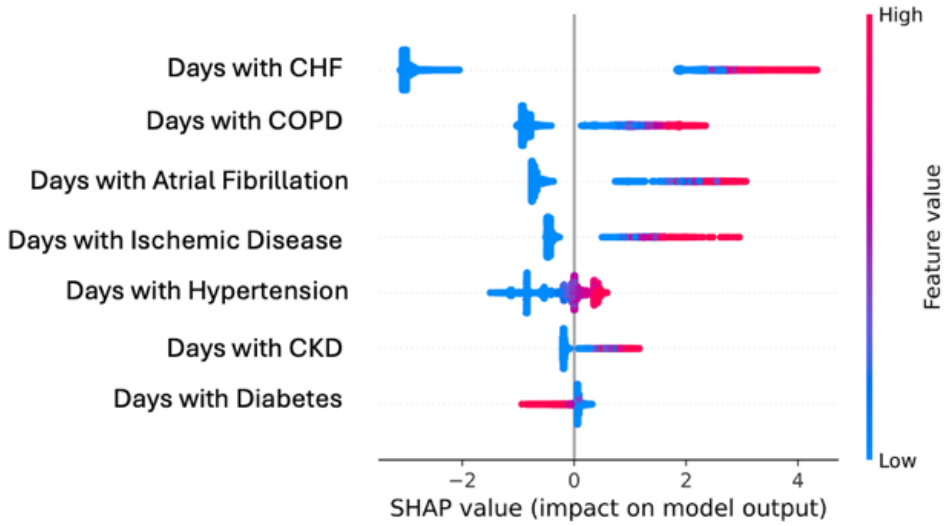
*English only content / Contenu en anglais seulement

Explicabilité des modèles – et notre mauvaise interprétation de l'explicabilité

(A) Feature importance of frail model – Cluster B
Measured as utility of features (and their values) to predict Cluster B (vs. Cluster A)



(B) Feature importance of *non*-frail model – Cluster C
Measured as utility of features (and their values) to predict Cluster C (vs. Cluster D or E)



*English only content / Contenu en anglais seulement

4. Mise en œuvre responsable

Quelles conditions doivent être remplies avant d’utiliser un outil de prévision?

Liste de vérification pour une utilisation responsable des modèles de prédiction fondés sur l'IA en santé publique

- Validation externe au sein des populations cibles, dans les contextes visés et pendant la période prévue
- Transparence dans la communication des résultats, gouvernance appropriée des données et mesures de protection de la vie privée
- Étalonnage, performance des sous-groupes et évaluation des conséquences inéquitables en aval
- Supervision humaine, processus décisionnels interprétables et responsabilité clairement définie
- Évaluation prospective des répercussions, suivi des écarts, critères de mise à jour et de retrait

Utilisation responsable = modèle valide + mesures justifiables + gouvernance continue.

Meilleurs indicateurs ≠ meilleurs résultats en matière de santé

Modèle

Meilleure discrimination et étalonnage adéquat

→

Décision

Changement des mesures, car le risque est présent

→

Répercussion

Prévention, meilleurs résultats, valeur équitable

Évaluer les répercussions potentielles, les coûts, les conséquences imprévues et les bénéficiaires.

Cinq messages à retenir si vous souhaitez créer un modèle de prédiction fondé sur l'apprentissage automatique

- Commencez par la décision de santé publique, pas l'algorithme
- Comparez l'apprentissage automatique à des références statistiques solides
- Évaluez l'étalonnage et l'utilité décisionnelle, pas uniquement la discrimination
- Ne confondez pas le risque prédit et le risque évitable
- Appuyez-vous sur la science de la mise en œuvre, l'équité et le suivi pour obtenir des avantages concrets

Merci!

Restons en contact

Courriel : gemma.postill@utoronto.ca
LinkedIn : https://www.linkedin.com/in/gemma-p-80481987/