

Compétition Montréalaise en Sciences des Données

Montreal Data Science Competition

$HG(N, K, n)$ ϵ_i \geq $\%$ $+$ $\text{Poisson}(\lambda)$ $\text{Geom}(p)$ n σ^2 $X \sim N(0,3)$ s $\text{corr}(X, Y)$ β Md Π $E(X|Y)$ \wedge $U(0,1)$ H^0 $\text{gamma}(c, \lambda)$ $\text{Cov}(X, Y)$ $\text{Var}(X)$ μ a

Déroulement de la journée

- 9h00: Présentation de la Compétition
- 10h00: Début de la Compétition
- 12h00: Dîner et collations fournis
- 16h50: Fin de votre travail
- 17h00: Heure limite de la soumission de votre solution
- 17h00: Mot de la fin

Agenda

- 9:00: Competition presentation
- 10:00: Start of competition
- 12:00: Lunch and snacks provided
- 16:50: End of your work
- 17:00: Deadline for submission of your solution
- 17:00: Closing remarks

Chaire Co-operators en analyse des risques actuariels

Présentation de la Chaire

La Chaire de recherche Co-operators en analyse des risques actuariels (CARA) a été fondée en mai 2018. La mission et les objectifs de la Chaire sont multiples et touchent autant le développement de la formation de l'actuariat que le développement scientifique de l'actuariat dans le domaine de l'assurance générale.

Mathieu Pigeon, PhD (Directeur scientifique)

Co-operators Chair in Actuarial Risk Analysis

Chair Presentation

The Co-operators Research Chair in Actuarial Risk Analysis (CARA) was founded in May 2018. The Chair has multiple objectives, related to the development of actuarial education and of actuarial science in the field of general insurance.

Mathieu Pigeon, PhD (Scientific Director)

Co-operators

Présentation de Co-operators

Nous sommes une coopérative canadienne de services financiers de premier plan et nous avons à cœur de vous aider à bâtir et consolider votre sécurité financière. Nous offrons des solutions d'assurance multiproduit ainsi que des produits, services et conseils financiers adaptés à vos besoins uniques.

Amélie Beauregard (VP, tarification en assurances habitation et automobile)

Nicholas Caramagno (VPA, capital et modélisation)

Frédérick Guillot (VPA, recherche et innovation analytique)

Hugues Laquerre (VP, tarification en assurance entreprise et agricole)

Émilie Lavoie-Charland (VPA, intelligence d'affaires)



Chaire Co-operators en
analyse des risques actuariels

Co-operators

Presentation of Co-operators

We're a leading all-Canadian financial services co-operative deeply invested in helping Canadians build and grow their financial strength and security. We offer multi-line insurance and investment products, services, and advice that's personalized to your unique needs.

Amélie Beauregard (VP, Home and Auto Pricing)

Nicholas Caramagno (AVP, Capital and Modeling)

Frédérick Guillot (AVP, Research and Analytics Innovation)

Hugues Laquerre (VP, Commercial and Farm Pricing)

Émilie Lavoie-Charland (AVP, Business Intelligence)



UQÀM



Chaire Co-operators en
analyse des risques actuariels

Co-operators

Un emploi chez Co-operators vous intéresse?

- Stages pendant vos études
 - Consultez les portails de votre université (en septembre et janvier)
- Emploi permanent
 - Consultez le site web <https://www.cooperators.ca/fr-ca/about-us/careers>
- Questions?
 - Contactez caroline_morrisseau@cooperators.ca

Co-operators

Thinking about a career with Co-operators?

- Internships during your studies
 - Consult your institutions web portals (in September and January)
- Permanent position
 - Consult the website <https://www.cooperators.ca/en/about-us/careers>
- Questions?
 - Contact caroline_morrisseau@cooperators.ca

Problématique

Mise en contexte

- Vous êtes nouvellement employé·e chez Co-operators.
- On a besoin de vos talents pour améliorer le **processus de gestion des réclamations** en assurance automobile.
- À l'aide **d'algorithmes** et de **données** qui vous seront fournies, vous devrez proposer une solution afin de rendre ce processus plus efficace, ce qui permettra:
 - d'améliorer l'expérience client,
 - d'améliorer l'expérience des employé·e·s,
 - de sauver des coûts.



Problematic

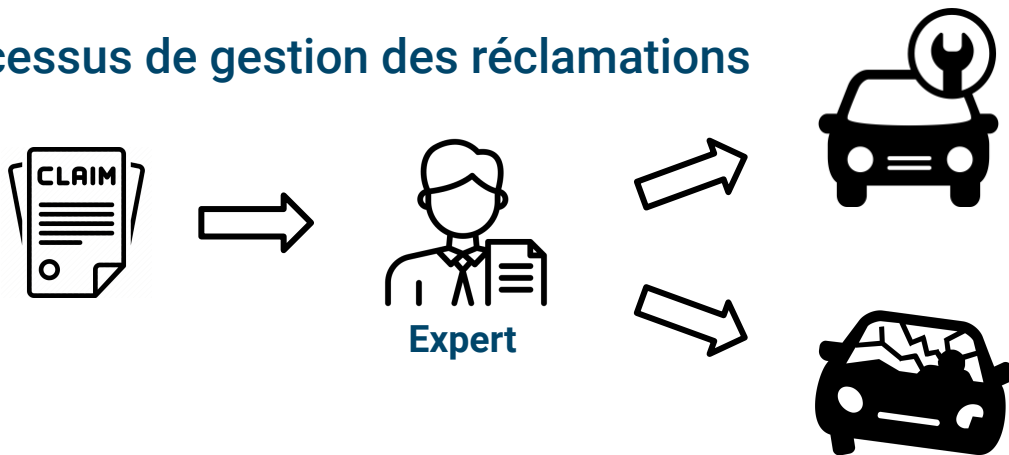
Context

- You are a new employee at Co-operators
- Your skills are needed to improve the auto insurance **claims management process**.
- Using algorithms and data provided to you, you will propose a solution to make this process more efficient, which will:
 - improve the customer experience,
 - improve the employees experience,
 - save money.



Problématique

Processus de gestion des réclamations



Véhicule réparable

- Co-operators rembourse les réparations.

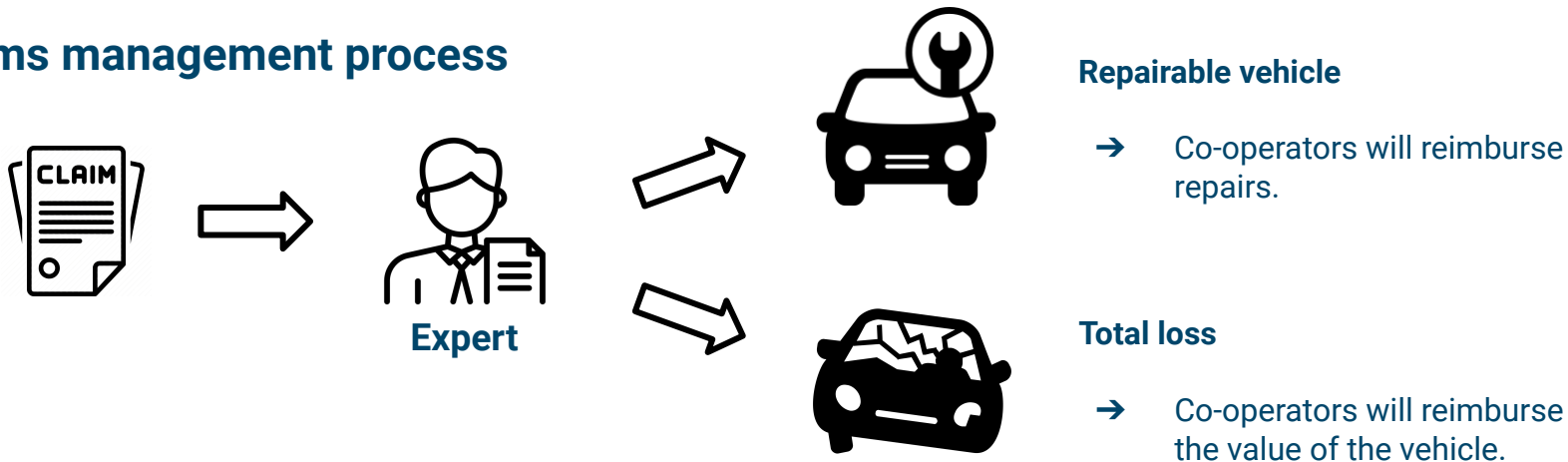
Perte totale

- Co-operators rembourse la valeur du véhicule.

- Le véhicule est dit « **réparable** » si le coût des réparations est **moins élevé** que la valeur du véhicule. Sinon, il est considéré « **perte totale** ».
- Chaque évaluation d'un expert coûte **1 500\$**.
- Un algorithme pourrait-il nous aider à classer certaines réclamations et ainsi éviter une partie des coûts liés à l'expert?
- Attention! Les erreurs de classification peuvent être coûteuses.

Problematic

Claims management process



- The vehicle is said to be “**repairable**” if the cost of repairs is **less than** the value of the vehicle. Otherwise, it is considered a “**total loss**”.
- Each expert evaluation costs **\$1,500**.
- Could an algorithm help us classify certain claims and thus avoid part of the costs related to the expert?
- Attention! Misclassifications can be costly.

Problématique

Actions possibles

Pour chaque réclamation du **jeu de données**, vous devez choisir l'une des 3 actions suivantes :

- Réparer le véhicule
- Rembourser la valeur totale véhicule
- Envoyer à l'expert



Problematic

Possible actions

For each claim in the **data set**, you must choose one of the following 3 actions:

- Repair the vehicle
- Reimburse total value of vehicle
- Send to the expert



Problématique

Évaluation de la solution

- Votre solution sera évaluée selon l'**économie réalisée** (en dollars) par rapport à l'évaluation de toutes les réclamations par l'expert.
- Exemple de solution:

Identifiant	Prediction	Coût réparations	Valeur véhicule
1	remplacer	6 000	8 000
2	réparer	1 000	4 000
3	expert	15 000	12 000

- Coût de la solution : $8\,000 + 1\,000 + (12\,000 + 1\,500) = 22\,500$
- Coût avec expert seulement : $(6\,000 + 1\,500) + (1\,000 + 1\,500) + (12\,000 + 1\,500) = 23\,500$
- Montant sauvé : $23\,500 - 22\,500 = 1\,000$

Problematic

Solution Evaluation

- Your solution will be evaluated based on the **savings** (in dollars) compared to the expert's evaluation of all claims.
- Solution Example:

ID	Prediction	Repair Cost	Vehicle Value
1	replace	6 000	8 000
2	repair	1 000	4 000
3	expert	15 000	12 000

- Cost of the solution: $8\,000 + 1\,000 + (12\,000 + 1\,500) = 22\,500$
- Cost by always using the expert: $(6\,000 + 1\,500) + (1\,000 + 1\,500) + (12\,000 + 1\,500) = 23\,500$
- Saved money: $23\,500 - 22\,500 = 1\,000$

Problématique

Jeux de données

- Chaque ligne correspond à une **réclamation**.
- 30 variables, dont la **valeur du véhicule** et le **coût des réparations**.
- Jeu d'entraînement :
 - 30k lignes
 - 30 colonnes
- Jeu test :
 - 100k lignes
 - 29 colonnes
- Le **coût des réparations** ne vous sera pas fourni dans le jeu test.



Problematic

Dataset

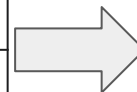
- Each row corresponds to one **claim**.
- 30 variables, including **vehicle value** and **cost of repairs**.
- Training set:
 - 30k rows
 - 30 columns
- Test set:
 - 100k rows
 - 29 columns
- The **cost of repairs** will not be provided in the test set.



Soumission de la solution

- Vous devez soumettre un **fichier CSV** contenant 2 colonnes et 100k lignes :
 - 1^{re} colonne : numéro d'identification de la réclamation (RECL_ID_CAT).
 - 2^e colonne : votre choix d'action pour chaque réclamation.

RECL_ID_CAT	PRED
a2b5-41fc264e2f89	1
9ed5-a8788e21fa5d	0
985e-c09497586fe8	2



Actions possibles:

- 0 : Réparer le véhicule
- 1 : Rembourser la valeur totale véhicule
- 2: Appel à l'expert

- Fichiers à envoyer (dans un fichier **Nom_de_lequipe.zip**):
 - **Nom_de_lequipe_Solution.csv**
 - Votre code
- Courriel : Compétition_Montrealaise_Sciences_Donnees@cooperators.ca

Solution submission

- You must submit a **CSV file** containing 2 columns and 100k rows :
 - 1st column : Claim ID (RECL_ID_CAT).
 - 2nd column : Action you chose for each claim.

RECL_ID_CAT	PRED
a2b5-41fc264e2f89	1
9ed5-a8788e21fa5d	0
985e-c09497586fe8	2



Possible actions:

- 0 : Repair the vehicle
- 1 : Reimburse total value of vehicle
- 2: Send to the expert

- Files to send (in a **Team_Name.zip** file):
 - **Team_Name_Solution.csv**
 - Your code
- E-mail: Competition_Montrealaise_Sciences_Donnees@cooperators.ca

Last things you need to know

- Wifi:
 - Réseau Eduroam
 - <https://servicesinformatiques.uqam.ca/services-offerts/reseau-sans-fil-visiteurs/>
- Documents
 - <https://github.com/francisduval/CMSD>

Question

- Avez vous des questions ?
- Do you have any questions?



GO!

- Documents
 - <https://github.com/francisduval/CMSD>

- Résoudre le problème en deux étapes :
 1. Construire un modèle en ignorant l'expert.
 2. Incorporer l'expert à partir de la prédiction de la première étape.

- Solve the problem in two steps:
 1. Build a model disregarding the expert.
 2. Incorporate the expert from the prediction in the first step.