

## **Description des problèmes**

### **Problème no 1 (soumis par Air Canada)**

#### **Prévision des besoins en personnel des centres de services**

##### *Contexte*

Les centres de services à la clientèle d'Air Canada pour le fret s'occupent des requêtes de clients dont le fret part du Canada, y transite ou y arrive. Nous offrons des services téléphoniques et des services par courriel et gardons un œil sur d'autres canaux de communication comme le clavardage.

En plus des requêtes habituelles, les centres de services s'occupent des nouvelles réservations de fret et de réservations spéciales comme des « conteneurs actifs » et des animaux vivants (chevaux, chiens, chats, etc.). Les centres de services à Toronto (YYZ), Montréal (YUL) et Moncton (YQM) sont ouverts de 0600 HNE à 2200 HNE, sept jours par semaine, et offrent des services bilingues à notre clientèle.

##### *Problème*

Comme nous pouvons répondre à un appel d'offres d'horaires une fois par an seulement, nous devons prévoir les contacts et générer des horaires pour les trois centres de services qui soient liés entre eux. À l'heure actuelle nous utilisons des tableurs Excel et des cartes thermique pour déterminer les plus grandes quantités d'appels de la journée et l'horaire des employés. Toutefois ces outils sont seulement basés sur des données historiques et ne prennent en compte ni les exigences linguistiques (anglais / français) ni les estimations des quantités de fret futures.

##### *Objectif*

L'objectif est de concevoir un modèle estimant les besoins en personnel à partir des quantités de fret et tenant compte de nos exigences en matière de niveau de service et aussi du nombre total courant d'employés. Ce modèle devrait être basé sur une moyenne calculée chaque semaine.

##### *Données disponibles*

Nous fournirons au moins les données suivantes : les nombres courants d'employés incluant leurs compétences linguistiques ; les quantités d'appels téléphoniques et de courriels par périodes de 30 minutes, ainsi que des échantillons d'horaires ; l'historique des quantités d'appels et des temps d'attente ; l'historique des quantités de fret et les estimations des quantités de fret futures ; et l'horaire des vols.

## **Problème no 2 (soumis par Air Canada)**

### **Outil de simulation pour le facteur de modification du taux ponctuel**

#### *Contexte*

Air Canada Cargo est un fournisseur primé de services de transport de fret aérien. Nous sommes le plus grand fournisseur de fret aérien du Canada en termes de capacité de fret, avec une présence dans plus de 50 pays et des plaques tournantes avec auto-assistance à Montréal, Toronto, Vancouver, Chicago, Londres et Francfort. Nous desservons plus de 450 villes sur six continents avec des vols directs et offrons des centaines d'autres itinéraires grâce à nos partenariats inter-compagnies et de camionnage. Des investissements clés dans les données et l'analyse, les API, l'intelligence artificielle, et plus encore, nous permettent d'améliorer continuellement l'expérience de nos clients. L'un de ces investissements est un outil intuitif de recommandation de taux ponctuel qui apportera toutes les informations pertinentes à l'utilisateur, tout en fournissant automatiquement une recommandation de taux de façon explicable et cohérente.

#### *Problème*

La génération des taux ponctuels se fonde sur des règlements d'affaires et des données qui incluent des taux de base, des cibles pour les revenus, et une matrice de modification de taux prenant en compte (a) le nombre de jours avant le départ, (b) le facteur de charge courant, et (c) le facteur de charge prévu. La matrice des modifications engendrera un « facteur de modification » pour faire augmenter ou baisser le taux ponctuel calculé. À l'heure actuelle la détermination de ces facteurs est réalisée surtout manuellement.

#### *Objectif*

L'objectif est de concevoir un modèle de simulation qui nous permettra de tester efficacement des combinaisons variées de facteurs de modification et d'évaluer leur impact sur les revenus. Ce modèle de simulation pourrait nous aider à construire une matrice de facteurs quasi-optimale.

#### *Données disponibles*

Nous fournirons au moins les données suivantes : l'historique des données de réservations ; l'historique des Lettres de Transport Aérien (*AWB* en anglais) ; et l'historique de la demande des clients.

### **Problème no 3 (soumis par l’Autorité des marchés financiers)**

#### **Nettoyage des données et traçage des transactions**

##### *Nettoyage des données tracées manuellement*

Dans le domaine des cryptomonnaies et de la finance décentralisée (*DeFi* en anglais), les données de transactions sont souvent complexes et imbriquées. Elles impliquent des adresses internes et externes, des routeurs, des contrats intelligents et une variété d'opérations. Il n'est pas rare de constater qu'une transaction comprend jusqu'à 12 transactions internes. Généralement, ces dernières consistent en trois swaps (jeton, contrepartie, frais) avec pour chacun, quatre transferts entre le client, la cible, un routeur et le pool de liquidité. De plus, ces transactions peuvent emprunter de multiples chemins, comme les échanges, les conversions, les transferts multiples et directs. La première étape de ce projet vise donc à nettoyer ces données pour les rendre exploitables lors d'une analyse ou d'un futur apprentissage machine (voir projet 2). Cela implique la standardisation et la classification des transactions, ainsi que l'identification des transactions initiatrices. Pour ce faire, une approche globale peut être envisagée, en utilisant toutes les transactions d'un même jeton, en séparant les traçages fiables des traçages non fiables, puis en utilisant les données fiables pour compléter les données non fiables.

##### *Traçage des transactions*

Dans le cadre du premier projet, pour extraire une transaction dite complexe, il est nécessaire de reconstituer le prix et les transactions fiables autour des transactions jugées non fiables. Cette approche rend le traçage des transactions isolées très coûteux. Pour cette raison, l'approche de l'apprentissage machine est envisagée. Une fois les données nettoyées, il est possible d'entraîner un modèle pour tracer automatiquement les transactions. C'est ici que l'apprentissage machine entre en jeu. L'objectif est de développer une méthode d'apprentissage machine capable d'interpréter les transactions internes pour en extraire la véritable transaction. Pour cela, des groupes de données nettoyées préalablement seront utilisés pour entraîner un modèle dont les entrées seront les transactions internes et les métadonnées de la transaction, et dont la sortie sera les montants et noms des jetons échangés. L'architecture et l'approche du modèle restent à déterminer. Il serait par exemple possible de diviser et reconstituer les différents swaps internes avant de réaliser le traçage final. Il est important de noter que la fiabilité est un facteur clé de réussite. Un score de fiabilité de traçage pourrait donc représenter une valeur ajoutée très importante.

## **Problème no 4 (Environnement et Changement Climatique Canada) Estimation des niveaux d'eau extrêmes à des emplacements non jaugés le long de l'estuaire fluvial du Saint-Laurent**

Dans le cadre du Programme d'identification et de cartographie des aléas d'inondation (PICA), Environnement et Changement Climatique Canada (ECCC) a reçu le mandat de produire des simulations 2D des niveaux d'eau extrêmes dans l'estuaire fluvial du Saint-Laurent. Les élévations des niveaux d'eau dans ce système sont déclenchées par l'interaction complexe de processus hydrologiques, météorologiques et marégraphiques qui doivent être pris en compte pour simuler la dynamique fluviale et les inondations. Toutefois des contraintes liées aux ressources informatiques, au temps requis et au manque d'un arrière-plan de champs géophysiques limitent actuellement la production de simulations hydrodynamiques 2D à échelle fine à un ensemble limité d'évènements extrêmes relativement courts (environ 400 évènements d'une durée allant d'une heure à plusieurs semaines).

Par conséquent plusieurs outils de modélisation complémentaires ont été examinées afin d'étudier l'évolution des propriétés extrêmes du niveau d'eau. Des modèles statistiques multivariés et des outils d'apprentissage automatique (ML) ont démontré leur efficacité en reconstruisant des séries continues de niveau d'eau sur de longues périodes, ce qui est essentiel pour déterminer les distributions de probabilité extrêmes. Même si ces méthodologies sont prometteuses dans le cas des stations jaugées, il faut relever le défi de les appliquer à des emplacements non jaugés dans l'estuaire où seulement de courtes simulations 2D sont disponibles. Pendant l'atelier nous voulons tenter de répondre à deux questions spécifiques.

1. Comment peut-on estimer les caractéristiques extrêmes (par exemple récurrence, durée et saisonnalité) à des emplacements non jaugés en exploitant des simulations hydrodynamiques 2D courtes pour des évènements extrêmes et de longues reconstructions du niveau d'eau obtenues grâce à des outils statistiques ou de ML pour certains emplacements (par exemple des stations de jauge)?
2. Comment peut-on évaluer la fiabilité des reconstructions horaires et des estimations extrêmes à des emplacements non jaugés afin de déterminer la stratégie la plus convenable pour implantation dans le projet PICA ?

ECCC fournira des données pour la période 1970-2022 : niveau d'eau par heure pour chaque station d'un groupe de 15 stations; reconstruction du niveau d'eau par heure à 2 stations obtenue par une régression harmonique marégraphique non stationnaire; des simulations hydrodynamiques 2D correspondant à un sous-ensemble d'évènements extrêmes observés à des stations sélectionnées.

## **Problème no 5 (soumis par l'International Air Transport Association) Caractéristiques et prévision de la turbulence en transport aérien**

Les turbulences sont une cause majeure de blessures non mortelles dans l'industrie aéronautique et coûtent des millions de dollars chaque année. Pour aider les pilotes à éviter les turbulences, réduire les blessures et faire baisser les émissions, la International Air Transport Association (IATA) a créé la plus grande plateforme de données de l'industrie sur les turbulences en direct.

Dans le cadre des efforts continus de l'IATA pour améliorer la sécurité, nous souhaitons analyser les données que nous avons recueillies depuis 2019 et en tirer des enseignements précieux. La turbulence est un phénomène volatil. La première question que nous voulons aborder est celle à laquelle les météorologues tentent de répondre depuis longtemps. Combien de temps dure la turbulence ? Sa durée dépend-elle de la période de l'année ? Existe-t-il une relation entre la durée des turbulences et l'altitude ?

La réponse à la question précédente permet de répondre à la deuxième question que nous souhaitons aborder. Les pilotes peuvent être alertés si l'avion se dirige vers une zone de turbulence. Compte tenu de la position, du cap et de la vitesse de l'avion, nous voulons déterminer la probabilité de turbulence devant cet avion en nous basant sur des données de turbulence en temps réel et des données historiques. Ces informations peuvent ensuite être communiquées aux pilotes en temps réel afin qu'ils puissent prendre des décisions tactiques en connaissance de cause et éviter la turbulence.

Nous souhaitons répondre à une dernière question sur la turbulence thermique, qui provoque d'importantes secousses pendant la phase de descente de l'avion. Cette phase du vol est importante car les agents de bord sécurisent généralement la cabine pendant cette période. Nous voulons compter le nombre d'événements turbulents d'origine thermique et évaluer leur fréquence sur la base des données EDR et de la couverture nuageuse à basse altitude.

## **Problème no 6 (soumis par Revenu Québec)**

### **Identification de motifs dans une banque de données de transactions**

Les transactions d'actifs, ou séquences de transactions, qui surviennent entre différents intervenants, qu'ils soient des particuliers ou des entreprises, peuvent être accompagnées d'obligations fiscales qui varient en fonction du domaine d'application. Il arrive que certains intervenants se regroupent et développent des stratégies transactionnelles afin d'éluder leurs obligations fiscales. Afin de bien s'acquitter de sa mission d'assurer la conformité fiscale, Revenu Québec souhaite découvrir des comportements émergents dans le contexte de transactions d'actifs.

Dans cette optique, une approche d'identification de motifs (*patterns*) de transactions est une solution envisagée. Une fois identifiés, ces motifs pourraient faire l'objet d'analyses par des experts du domaine d'affaires afin de déterminer s'il s'agit réellement d'un comportement ayant été mis sur pied par un groupe d'acteurs dans le but d'éluder leurs obligations fiscales.

Pour identifier un réseau, une approche ascendante, où l'on part du détail pour complexifier progressivement, peut être utilisée. C'est-à-dire que l'identification de ce genre de comportements non-conformes peut débuter par la détection d'un intervenant ou d'une transaction suspecte. Par la suite, les informations relatives aux autres intervenants et transactions sont recherchées et une stratégie pour éluder les obligations fiscales peut être mise au jour.

Dans le présent contexte, il est souhaité d'aborder la problématique avec une approche descendante, c'est-à-dire que toutes les transactions, intervenants et liens seraient considérés et que l'on tenterait de dégager des motifs qui nous mèneraient à des groupes d'intervenants et de transactions méritant une analyse plus approfondie pour l'identification de comportements non-conformes.

Les données relatives au problème seraient des données anonymisées sur des transactions qui contiendraient au minimum les informations ci-dessous.

1. Intervenants : leur type, les liens entre eux et certaines caractéristiques pertinentes ;
2. Actifs transigés : leur type et certaines caractéristiques pertinentes ;
3. Transactions : leur type, leur date, les actifs et les intervenants impliqués.

À partir de telles données, il pourrait s'agir d'identifier des motifs fréquents ou des règles d'association qui permettraient d'identifier des stratégies pertinentes à analyser. Il pourrait également s'agir d'effectuer de la détection d'anomalies dans la

structure des relations ou bien d'utiliser une représentation en graphe pour détecter des sous-graphes intéressants. Idéalement, une recherche basée uniquement sur la structure des relations et qui n'est pas basée sur des réseaux de contrevenants déjà mis au jour serait préférée.

Cette approche est tout à fait nouvelle à Revenu Québec. Ainsi, les attentes et les résultats ne sont pas définis de façon précise. Entre autres, les concepts d'« anomalie », d'« intéressant » ou de « pertinent » ne sont pas formellement définis dans le contexte d'affaires. L'objectif premier de ce projet est d'explorer le potentiel de données de transactions décontextualisées pour la découverte automatique de réseaux suspectés d'avoir un comportement malveillant. Ainsi, il est attendu qu'à l'aide d'outils de science des données, certaines configurations soient identifiées comme anormales par rapport à l'ensemble des données fournies. De plus, le cheminement qui a mené à déterminer que ces configurations sont anormales devrait pouvoir être expliqué et reproduit.