Cours de Mathématique appliquées

I.E.T.C.PS

Année 2023-2024

Project IA prédictive COVID-19 pour les communes de Bruxelles

Lucie Delescaille

Etienne vallet

Rouyre Christophe

Alexandre Mihailofsky

# 1. Mise en Situation

Sciensano est un institut public national de recherche sur la Santé en Belgique opérant sous l'autorité fédérale du ministère de la Santé publique et du ministère de l'Agriculture. Le core business de Sciensano est l'avancée de la recherche dans des domaines tels que la santé publique, animale et alimentaire mais aussi de l'écologie, de la qualité des soins, des vaccins et la propagation de certaines maladies infectieuses.

<https://www.sciensano.be/fr>

Sciensano upload publiquement les résultats de leurs études sur leur site, vous pouvez en quelques clics obtenir des data de qualité se rapportant à la situation nationale de la grippe aviaire, influenza ou encore du coronavirus.

En tant qu'étudiants de deuxième et dernière années en bachelier d'informatique de gestion, c'est une opportunité de mettre en pratique nos compétences acquises tout au long de notre formation. Nous avons ainsi collaboré pour réaliser un projet ambitieux autour des données fournies par Sciensano. Notre objectif est de concevoir un outil de prédiction capable de modéliser l'évolution du taux de contamination et de propagation dans toutes les 19 communes de Bruxelles au plus fort de la pandémie.

Le but étant de parvenir à modéliser la propagation intercommunale bruxelloise du covid à l'aide de graphes probabilistes et ainsi avoir des pistes de raisonnement quant aux raisons ou autres événements ayant influencé cette propagation.

Retrouvez les même données que nous avons exploitées sur :

<https://epistat.wiv-isp.be/covid/>

# 2. Reprise

## 2.1 Travail année précédente

L'équipe précédente s'est engagée dans une démarche méthodique et rigoureuse d'exploitation des données de Sciensano. Ils ont d'abord téléchargé ces données qui étaient au format json, un format léger d'échange de données. Ils ont ensuite 'parsé' ou traité ces données, ce qui signifie qu'ils les ont converties en une structure plus utilisable pour le développement, c'est-à-dire des matrices de données Python.

La suite de leur travail a impliqué des manipulations complexes sur ces dites structures. Ils ont commencé par inverser ces matrices, une opération qui dans le contexte des mathématiques est l'équivalent de prendre l'inverse d'un nombre. L'inversion des matrices est une étape essentielle dans de nombreux algorithmes d'optimisation et de résolution de systèmes d'équations. Elle peut aider à transformer un problème complexe en un problème plus facile à résoudre.

Ils ont ensuite appliqué une topologie de graphe optimale. Cela suggère qu'ils ont représenté les données sous forme de graphe - une structure qui consiste en des points, appelés nœuds, connectés par des lignes, appelées arêtes. L'application d'une topologie optimale implique qu'ils ont organisé ces nœuds et ces arêtes de manière à maximiser le taux d'inversion, ce qui peut être interprété comme une tentative d'optimiser une certaine mesure de performance.

Leur objectif final était d'utiliser la régression linéaire pour prédire la valeur future de chaque élément de matrice en utilisant une projection LOGIT. La régression linéaire est une méthode statistique utilisée pour prédire une variable dépendante en fonction d'une ou plusieurs variables indépendantes. En utilisant une projection LOGIT, ils ont transformé ces prédictions en probabilités, ce qui est très utile pour la modélisation de situations où la variable de sortie est binaire, comme la présence ou l'absence d'une maladie.

3. Méthodologie choisie.

Dans ce cas nous avons choisi de découper notre travail en plusieurs étapes. La première étant de récupérer les données de Sciensano. De les traiter et de les nettoyer. En effet, nous avions remarquer que certaines données n’étaient pas correctes : certaines avait comme valeur <5.

Ces données ne peuvent pas être traitées par l’algorithme. Nous avons donc opté pour l’utilisation d’une valeur moyenne haute de 3. (L’utilisation de valeurs entières en lieu et place de valeurs réelles permettra de minimiser l’utilisation de ressources mémoires).