

# INFO-F-311: Intelligence Artificielle

## Projet 3: Réseaux bayésiens

Pascal Tribel

Yannick Molinghen

Tom Lenaerts

### 1 Préambule

Ce projet vise à vous familiariser avec la notion de **réseaux bayésiens** et leur utilisation. Il s'agit d'analyser un ensemble de symptômes en lien avec la positivité de patient·e·s au COVID-19. L'analyse critique des résultats fait partie intégrante du travail.

### 2 Introduction

Le jeu de données <https://www.kaggle.com/datasets/saurabhshahane/brazilian-covid-symptomatic-patients-data> donne un ensemble de 3128 données de patient·e·s testé·e·s au COVID-19 au Brésil. Celui-ci offre une série de symptômes clairs pour une quantité relativement importante de patient·e·s.

### 3 Réseaux bayésiens

Certains des symptômes repris peuvent être dépendants, d'autres pas. Leur apparition est un phénomène aléatoire dont la dépendance conditionnelle peut être exprimée sous forme d'un **réseau bayésien**. Ce type de réseaux ne donne pas à *proprement parler* de relation de causalité, mais se permet de représenter correctement les relations d'indépendance conditionnelles. Ceci permet, par exemple, d'éliminer certaines observations de la liste de symptômes utile. En outre, un réseau bayésien permet, par le calcul d'*inférence*, d'étudier la probabilité conditionnelle *a posteriori* de certains symptômes, ou de la positivité à la maladie.

### 4 Consignes

Dans ce projet, vous vous baserez sur le jeu de données `both_test_balanced.csv` repris dans la source citée ci-dessus.

#### 4.1 Rapport

Produisez un rapport *scientifique*, qui suit la structure suivante :

1. Introduction
2. Cadre expérimental
3. Résultats
4. Analyse
5. Conclusion
6. Bibliographie (éventuelle)

et qui ne doit pas excéder 5 pages A4, en comptant les figures et la bibliographie. Veillez à obtenir un rendu similaire à Latex (mais vous pouvez utiliser un outil similaire, comme <http://typst.app>).

#### 4.2 Questions

On vous demande :

1. De présenter ce jeu de données, d'établir une courte analyse critique de son contenu, de sa(s) source(s), de la pertinence *a priori* des symptômes repris. Faites une hypothèse sur les relations causales possibles entre les grandeurs de ce jeu de donnée.
2. De créer un réseau de Bayes sur base de ce jeu de données, à l'aide du package pyAgrum (<https://agrum.gitlab.io>). Ce package contient la majeure partie des fonctions dont vous avez besoin pour

ce projet, explorez-le extensivement. Vous pouvez choisir si vous utilisez la totalité des symptômes présents dans le jeu de données, mais vous devez justifier si vous décidez d'en écarter. Choisissez un algorithme d'apprentissage parmi ceux disponibles, en justifiant votre choix.

3. De donner les probabilités conditionnelles pour les cas suivants :
  - Le sujet souffre de mal de tête
  - Le sujet est testé positif
  - D'autres grandeurs qui vous semblent pertinentes
4. De donner les probabilités, par inférence, des cas suivants :
  - Le sujet souffre de troubles de l'odorat
  - Le sujet souffre de troubles de l'odorat, sachant qu'il souffre de troubles du goût
  - Que le test soit positif, sachant que le sujet souffre de troubles de l'odorat, du goût, mais qu'il n'a pas de fièvre
  - D'autres grandeurs qui vous semblent pertinentes
5. De montrer le graphe d'inférence produit pour l'ensemble des symptômes dans le cas où le sujet souffre de troubles de l'odorat, du goût, et de fièvre. Analysez son contenu pour les relations n'apparaissant pas dans vos hypothèses.
6. Pour un des autres algorithmes d'apprentissage disponibles, de créer un deuxième réseau de Bayes pour le même jeu de données. Grâce à l'objet `GraphicalBNComparator`, établissez une représentation comparative entre vos deux réseaux. Écrivez une courte critique du graphe produit, en tentant d'analyser la notion de causalité entre les différentes grandeurs mesurées. Comparez cela avec vos hypothèses de la question 1.

#### 4.3 Utilisation de Chat-GPT, Copilot, ... et autres assistants

Dans la conclusion de votre rapport, détaillez votre utilisation éventuelle d'outils d'assistanat, que ce soit pour l'écriture du code ou du rapport. Toutes vos sources doivent être pertinentes et exister. L'utilisation de tels outils est autorisée. Ce travail est néanmoins *individuel*.

### Remise

- Un fichier `.zip` portant votre nom\_prenom (exemple: `tribel_pascal.zip`), contenant deux fichiers :
  - Un exécutable `nom_prenom.py`. Vous pouvez considérer qu'il sera exécuté dans un répertoire contenant aussi le fichier `both_test_balanced.csv`, mais n'incluez pas ce fichier `.csv` dans votre remise.
  - Votre rapport, au format `nom_prenom.pdf`

Le travail est **individuel** et doit être rendu sur l'Université Virtuelle pour le 19/11/2023 à 23:59. Vous pouvez envoyer vos questions à l'adresse **pascal.tribel@ulb.be**. Nous répondrons à celles-ci lors de la séance questions-réponse du **lundi 13 novembre**. Mentionnez l'intitulé du cours (F311) dans l'entête de votre mail.

*Bon travail !*