

# INFO-F311 - Project - Réseaux bayésiens

Luca Palmisano - 000516721

Novembre 2023

# 1 Introduction

## 1.1 Source

Les données ont été collectées par l’Agence de la Santé Publique de la ville de Campina Grande, au Brésil, sans inclure les détails personnels des patients. Seules les informations concernant les divers symptômes et les résultats des tests ont été enregistrées.

Il est souvent difficile d’attester à 100% de la fiabilité des données trouvées sur internet. Dans ce cas particulier, ces données ont été mises à disposition dans le cadre d’une étude réalisée par plusieurs professeurs brésiliens [1]. Cette étude ayant pour but de donner la priorité aux patients présentant des symptômes pour effectuer un test de dépistage du COVID-19.

## 1.2 Pertinence

Les symptômes recensés sont pertinents. Après avoir effectué des vérifications sur divers sites d’organismes de santé publique tels que le NHS[2], le CDC[3] et SPF Santé[4], il est à noter que tous citent les symptômes suivants : dyspnée, perte de goût, perte d’odorat, fièvre, mal de tête, mal de gorge, toux et coryza.

Il est cependant intéressant de souligner que certains symptômes sont considérés comme majeurs par rapport à d’autres[5], en particulier la dyspnée, la perte du goût et d’odorat, ainsi que la toux.

## 1.3 Hypothèses

Le COVID-19 fait partie de la famille des coronavirus, une catégorie de virus similaires aux rhinovirus[6]. Ces derniers sont principalement responsables des rhinopharyngites[7], présentant des symptômes similaires à ceux observés chez les coronavirus, tels que la toux, le mal de gorge, la rhinorrhée et la congestion nasale. En effet, bien que la majorité des rhinopharyngites soient attribuées aux rhinovirus, les coronavirus sont la cause dans environ 15% des cas[8]. Par conséquent, la distinction entre ces deux groupes de virus basée uniquement sur les symptômes mentionnés ci-dessus peut s’avérer difficile.

Parmi les autres symptômes, il est important de noter que la fièvre et les maux de tête peuvent également être présents lors d’une grippe[9] ou de maladies telles que la méningite[10].

# 2 Cadre expérimental

## 2.1 Chargement des données

La première étape consiste à charger les données à l’aide du package `pandas`[11]. Les colonnes non pertinentes sont éliminées et les valeurs sont inversées : dans

les données fournies, la présence d'un symptôme est indiquée par 0 au lieu de 1. La colonne "Class" est renommée en "Positive".

## 2.2 Création du réseaux Bayésiens

Le réseau bayésien est construit à l'aide du package `pyAgrum`[12], qui offre la capacité de créer un réseau basé sur un ensemble de données. Plusieurs algorithmes de recherche sont disponibles, et pour ce projet, l'algorithme K2 a été choisi. Contrairement à d'autres algorithmes, K2 permet de spécifier un ordre au préalable, permettant ainsi de réduire la complexité de l'espace de recherche[13].

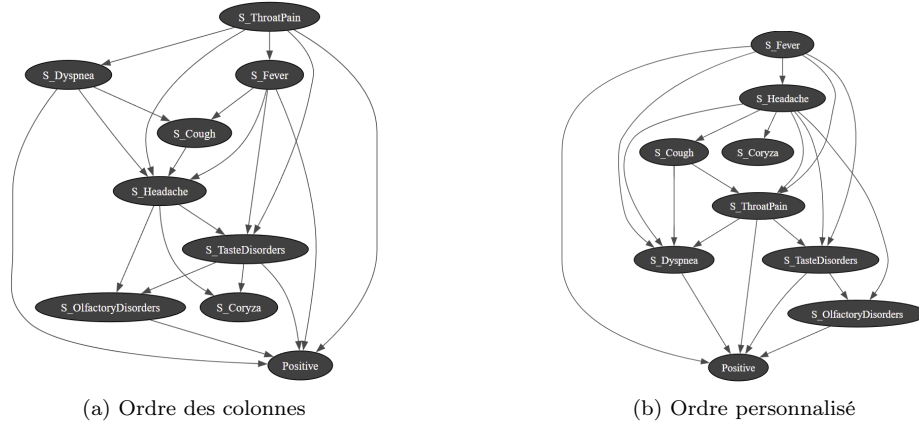


FIGURE 1 – Exemple d'ordres différents pour l'algorithme K2

## 3 Résultats et Analyse

### 3.1 Probabilités conditionnelles

En ce qui concerne le mal de tête, on observe que la probabilité d'avoir ce symptôme est plus élevée que la probabilité de ne pas l'avoir, mais seulement lorsque tous les autres symptômes sont absents. Cela va à l'encontre d'une des hypothèses présenté dans l'introduction dans laquelle les maux de têtes sont souvent associés à d'autres symptômes comme la fièvre.

Dans le second cas, on constate que les symptômes de perte du goût/odorat et de dyspnée semblent avoir un impact significatif sur la probabilité d'un test positif. Il est toutefois intéressant de noter que la probabilité d'un test positif diminue lorsque tous ces symptômes sont présents par rapport au cas où seulement deux des symptômes sont présents.

				S_Headache	
S_Fever	S_ThroatPain	S_Dyspnea	S_Cough	0	1
0	0	0	0	0.3813	0.6187
			1	0.8951	0.1049
		1	0	0.8269	0.1731
			1	0.7864	0.2136
	1	0	0	0.8515	0.1485
			1	0.7971	0.2029
		1	0	0.9437	0.0563
			1	0.7452	0.2548
1	0	0	0	0.8900	0.1100
			1	0.8559	0.1441
		1	0	0.6860	0.3140
			1	0.8945	0.1055
	1	0	0	0.8398	0.1602
			1	0.8089	0.1911
		1	0	0.6816	0.3184
			1	0.7032	0.2968

(a) Le sujet souffre de mal de tête

					Result	
S_Headache	S_Fever	S_ThroatPain	S_Dyspnea	S_Cough	0	1
0	0	0	0	0	0.9950	0.0050
			1	0	0.9919	0.0081
		1	0	0	0.9861	0.0139
			1	0	0.9787	0.0213
	1	0	0	0	0.9189	0.0811
			1	0	0.9171	0.0829
		1	0	0	0.8889	0.1111
			1	0	0.8788	0.1212
1	0	0	0	1	0.9875	0.0125
			1	1	0.9788	0.0212
		1	0	1	0.9404	0.0596
			1	1	0.9287	0.0713
	1	0	0	1	0.8875	0.1125
			1	1	0.8843	0.1157
		1	0	1	0.8233	0.1767
			1	1	0.8075	0.1925
2	0	0	0	2	0.9747	0.0253
			1	2	0.9713	0.0287
		1	0	2	0.9351	0.0649
			1	2	0.9270	0.0730
	1	0	0	2	0.8889	0.1111
			1	2	0.8800	0.1200
		1	0	2	0.8375	0.1625
			1	2	0.8288	0.1712

(b) Le sujet est testé positif

FIGURE 2 – Probabilités conditionnelles

### 3.2 Probabilités par inférence

En ce qui concerne les troubles de l'odorat, les deux premiers cas mettent en évidence une relation de causalité claire entre les troubles de l'odorat et du goût : la probabilité de souffrir de troubles de l'odorat est nettement plus élevée lorsque le sujet présente déjà des troubles du goût.

Le troisième cas ne fait que renforcer les observations issues des résultats des probabilités conditionnelles : les symptômes tels que les troubles de l'odorat et du goût ont un impact significatif sur le résultat positif du test, contrairement à des symptômes tels que la fièvre, par exemple.

S_OlfactoryDisorders	
0	1
0.8323	0.1677

(a) Troubles de l'odorat

S_OlfactoryDisorders	
0	1
0.2165	0.7835

(b) troubles de l'odorat, sachant troubles du goût

Positive	
0	1
0.3402	0.6598

(c) Test positif, sachant troubles de l'odorat, du goût, mais pas de fièvre

FIGURE 3 – Probabilités par inférence

### 3.3 Graphe d'inférence

Il est intéressant de noter que la relation entre des symptômes tels que la toux, le mal de gorge et la coryza est soit inexistante, soit indirecte, se manifestant à travers d'autres symptômes. De même, bien que la relation entre la fièvre et le mal de tête persiste, la fièvre ne semble pas avoir un impact significatif sur la présence d'un mal de tête. Par ailleurs, la plupart des symptômes semblent être des causes directes d'un résultat positif au test, à l'exception de la coryza, de la toux et du mal de tête.

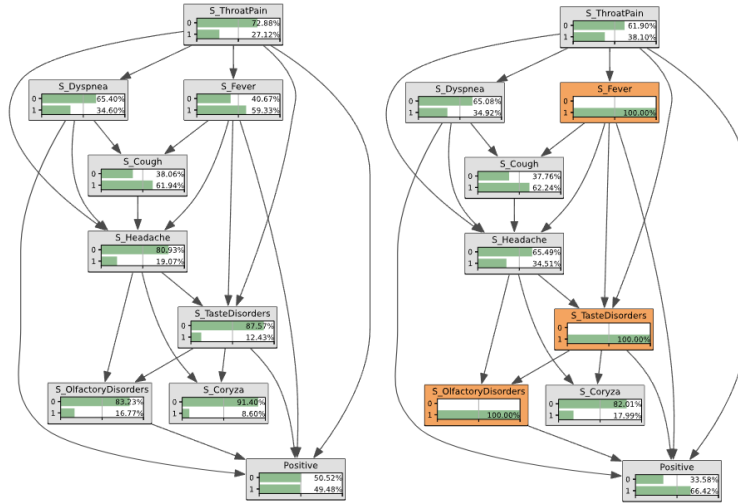


FIGURE 4 – Graphe d'inférence pour l'ensemble des symptômes

### 3.4 Comparaison Greedy Hill et K2

La principale différence entre les deux algorithmes de recherche réside dans l'ordre des symptômes. En utilisant Greedy Hill, le graphe obtenu est l'inverse de celui généré par l'algorithme K2 : les nœuds qui étaient auparavant des causes deviennent des conséquences, bien que les relations entre chaque nœud aient peu changé. Le défi avec le réseau bayésien obtenu par l'algorithme Greedy Hill réside dans le fait qu'il considère un test positif comme l'une des causes principales des autres symptômes.

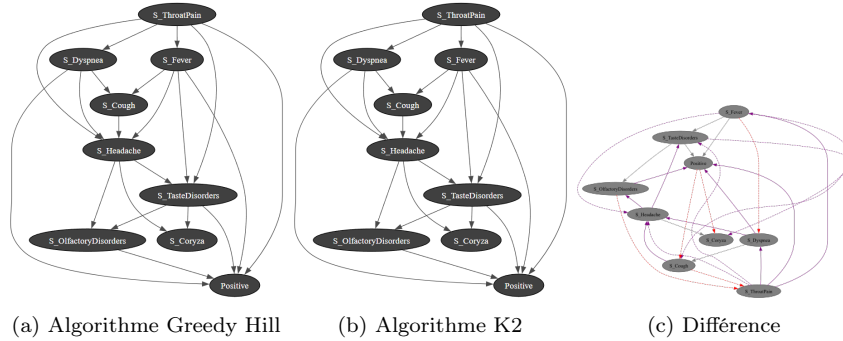


FIGURE 5 – Comparaison algorithme de recherche

## 4 Conclusion

L'analyse des données à l'aide d'un réseau bayésien met en lumière plusieurs aspects, en commençant par les symptômes majeurs tels qu'ils étaient présentés dans les hypothèses. En effet, les troubles du goût, de l'odorat et la toux semblent avoir un impact significatif sur le résultat du test COVID-19.

On observe également l'importance du choix de l'algorithme de recherche. K2 génère des résultats intéressants, en particulier car il permet de choisir un ordre. Cependant, l'ordre sélectionné ici est totalement arbitraire, suscitant donc l'intérêt d'explorer la possibilité de trouver un ordre optimal.

## Références

- [1] Íris Santana, Andressa Silveira, Alvaro Sobrinho, Lenardo Silva, Leandro Silva, Danilo Santos, Edmar Candeia, and Angelo Perkusich. Classification models for covid-19 test prioritization in brazil : Machine learning approach. *Journal of Medical Internet Research*, 23, 04 2021.
- [2] NHS. COVID-19 Symptoms and what to do. <https://www.nhs.uk/conditions/covid-19/covid-19-symptoms-and-what-to-do/>, Mars 2023.
- [3] CDC. Symptoms of COVID-19. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>, Octobre 2022.
- [4] SPF Santé. Que faut-il savoir sur la COVID-19? <https://www.info-coronavirus.be/fr/covid/>, Novembre 2023.
- [5] Sciensano. COVID-19 - Définition de cas et testing. <https://covid-19.sciensano.be/fr/covid-19-definition-de-cas-et-testing>, Novembre 2023.
- [6] Wikipedia, The Free Encyclopedia. Rhinovirus. <https://en.wikipedia.org/w/index.php?title=Rhinovirus&oldid=1180652006>, 2023.
- [7] CDC. Common cold. <https://www.cdc.gov/antibiotic-use/colds.html>, Juin 2023.
- [8] Wikipedia, The Free Encyclopedia. Coronavirus. <https://en.wikipedia.org/w/index.php?title=Coronavirus&oldid=1173402921>, 2023.
- [9] Nurofen. Meningitis. <https://www.nurofen.fr/douleurs-et-ou-fievre-liees-a-l-etat-grippal/>, Février 2023.
- [10] NSH Inform. Meningitis. <https://www.nhsinform.scot/illnesses-and-conditions/infections-and-poisoning/meningitis/>, Juin 2023.
- [11] Python - pandas. <https://pandas.pydata.org/>.
- [12] Python - pygrum. <https://agrum.gitlab.io/>.
- [13] Shahab Behjati and Hamid Beigy. An order-based algorithm for learning structure of bayesian networks. In *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 25–36. PMLR, 11–14 Sep 2018.