

Synthèse : Création de scores de santé tenant compte des risques émergents aux Etats-Unis

En raison du réchauffement climatique, les températures moyennes à la surface du globe vont continuer d'augmenter avec des périodes de chaleur intense, et, dans certaines régions, des épisodes de froid extrême. Parallèlement, la fréquence et l'intensité d'événements naturels tels que les inondations, les ouragans, les tempêtes ou encore les sécheresses prolongées seront amenées à croître dans les prochaines années, comme le soulignent de nombreux rapports du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). Ce constat s'applique également aux pics de pollution de l'air, notamment en milieu urbain, où la concentration de particules fines et d'ozone atteint régulièrement des niveaux excédant les seuils recommandés par les organismes de santé publique.

Or, il est désormais établi par la communauté scientifique que le dérèglement climatique et la pollution atmosphérique ont un impact significatif sur la santé des populations exposées : ils favorisent le développement de maladies chroniques, cardiovasculaires et pulmonaires, touchant surtout les populations les plus vulnérables. Ces risques environnementaux, inégalement répartis au sein de la population, accentuent les disparités d'accès aux soins, en particulier dans des systèmes non universels comme celui des États-Unis. Cette situation met en exergue une double problématique de souveraineté et de dépendance : les systèmes de santé, et notamment les acteurs privés de l'assurance, doivent choisir entre subir l'évolution des risques environnementaux, au prix d'inégalités accrues dans l'accès aux soins et leurs remboursements, ou s'outiller pour mieux anticiper et gérer ces risques.

Si la littérature est abondante sur les effets sanitaires du climat et de la pollution sur la santé, elle est en revanche très pauvre en ce qui concerne leur intégration dans les modèles de scores de santé ou de tarification en assurance santé. Ce mémoire vise à développer des scores de santé, à partir de données d'assureurs santé américains, intégrant le climat et la pollution, dans le but de quantifier leurs impacts sur la santé. Plusieurs modèles statistiques et de *machine learning* (ML) sont mobilisés pour créer des scores pertinents et y intégrer ces risques émergents. L'analyse porte sur :

- la pertinence des scores : le score doit être fidèle à la réalité, c'est-à-dire qu'il doit refléter la santé réelle des individus ;
- la capacité des scores à mesurer l'influence du climat et de la pollution : les modèles de scores doivent être facilement interprétables pour pouvoir mesurer directement l'impact de certains indicateurs climatiques ou de pollution sur la santé ; et
- l'équité des scores, notamment vis-à-vis des différentes origines ethniques, afin d'identifier d'éventuels biais et de garantir une utilisation responsable de ces outils en assurance santé.

Ce mémoire s'articule en trois parties. La **première partie** fournit des éléments de contexte nécessaires à la compréhension du sujet. Elle débute par une présentation détaillée du système d'assurance santé américain, qui, malgré les réformes récentes, peine à couvrir équitablement l'ensemble des citoyens. Elle décrit également l'utilisation et l'encadrement des scores de santé aux États-Unis, en insistant sur leur fonctionnement, leur encadrement réglementaire et les débats éthiques qu'ils suscitent. Cette partie s'achève

par une revue de littérature approfondie des effets du climat et de la pollution sur la santé des populations exposées. La **deuxième partie** de ce mémoire est consacrée à la description des modèles permettant de créer des scores de santé intégrant explicitement des variables climatiques et de pollution. Elle détaille leur principe, leurs avantages et leurs limites. Les modèles retenus, tels que la régression linéaire, les modèles linéaires généralisés et XGBoost, satisfont le critère d'interprétabilité des impacts sous-jacents. Un volet spécifique est dédié dans cette partie aux méthodes qui permettront d'analyser l'équité des modèles développés. La **troisième partie** complète le mémoire avec la mise en œuvre concrète des modèles présentés dans la partie précédente. A partir d'une base de données de santé d'assureurs américains, enrichie de données climatiques et de pollution, cette partie s'attache à implémenter les différents scores de santé et à comparer leurs performances. Elle analyse la capacité de ces modèles à satisfaire à la fois les critères de performance et d'interprétabilité. Le cas échéant, une analyse fine des effets des indicateurs climatiques et de pollution sur les scores de santé développés est réalisée. Enfin, une attention particulière est portée à l'évaluation de l'équité des scores produits, permettant d'identifier d'éventuels biais dans la modélisation.

Résultats majeurs

Les scores de santé développés dans ce mémoire intègrent climat et pollution pour mesurer leur impact sur la santé des assurés américains. Les principaux facteurs ayant des effets sur les scores de santé créés sont l'état de santé antérieur, l'âge et le genre de l'assuré et, dans une moindre mesure, certains facteurs environnementaux comme l'exposition aux vagues de chaleur, aux particules fines ($PM_{2.5}$) ou au NO_2 . Les résultats obtenus avec ces premières applications sont encourageants, bien que les approches présentées restent perfectibles : la maille temporelle ainsi que les zones géographiques pourraient être affinées pour mieux capter l'impact des risques émergents sur la santé. Enfin, la pertinence des scores implémentés est à mettre au regard de l'équité vis-à-vis de l'origine ethnique, de telle sorte qu'il n'y ait pas de différenciation entre les Afro-Américains et les Caucasiens dans la tarification par exemple.

Les résultats encouragent à approfondir ces travaux, en mobilisant des données plus exhaustives et des méthodes plus robustes, tout en gardant à l'esprit les coûts associés. Au-delà de l'intérêt scientifique, cette démarche vise à renforcer la souveraineté des assureurs et des systèmes de santé, en leur permettant d'anticiper et de gérer les risques liés au climat et à la pollution, plutôt que d'en subir passivement les conséquences. Il convient aujourd'hui de choisir entre dépendance aux aléas environnementaux et construction d'outils efficaces, interprétables et équitables pour l'avenir.