La Caisse nationale de l'Assurance Maladie (Cnam) et l'École polytechnique renouvellent pour 3 ans
leur convention de partenariat
de recherche et développement
dans le domaine des sciences
des données
.
Débutée fin 2014, c
ette collaboration a
abouti à de premiers résultats très prometteurs et a permis de dégager
de nouvelles
approches
d'exploitation des données du Système
n
ational
d'
i
nformation
i
nter
r
égimes de l'Assurance Maladie (Sniiram)
, la
base de données santé
qui comp
i
le tout
es les données de remboursements de soins
en France
.

L'objectif du partenariat est de développer différents types d'algorithmes de traitement des données, permettant notamment d'automatiser la détection
de médic
a

ments aux effets se
condaires
nocifs
,
voire
graves
,
permettant
aux autorités sanitaires d'asseoir leurs décisions sur des faits scientifiquement prouvés
.
Au cours des trois premières années de travaux, Emmanuel Bacry, Professeur de
Mathématiques appliquées
,
responsable de
l'initiative
Data science
à l'É
cole polytechnique
et
Directeur de r
e
cherche au CNRS à l'Université Paris-Dauphine (CEREMADE),
et son équipe
du
Centre de
m
athématiques
a
ppliquées (CMAP -
École polytechnique/CNRS)
ont réussi à prouver
l'efficacité
de leur
approche
visant à détecter les signaux faibles dans l'immense volume de
donnée
s
du Sniiram
(voir encadré).
Les algorithmes d'apprentissage automatique
développés par les chercheurs
ont
en effet
retrouvé
l'ass
o

ciation connue
et prouvée par les méthodes classiques
entre prise de pioglitazone et
le déclenchement d'un
cancer de la vessie chez les
patients souffrant de
diabè
te.

Détecter les effets secondaires potentiels de médicaments

L'objectif du renouvellement du partenariat est d'approfondir les recherches dans le domaine d
e la surveillance des produits de santé
et de
développer
des
algorithmes sur
différents
types de données
.
Le projet de pharm
aco-épidémiologie mené dans le cadre du renouvellement de ce partenariat
sera ainsi
centré sur la recherche et la validation de nouveaux signaux de sécuri
té pour détecter les
effets second
aires potentiels de m
é
dicaments
.
L'étude de l'association entre la prise de rosiglitazone et
l'apparition
d'une insuffisance cardiaque ou d'un infarctus du myocarde est
ainsi en cours.

Un second projet en cours de développement porte sur la prise médicamenteuse et la survenue de fractures chez les personnes âgées de plus de 65 ans. Ce projet cherchera à m
ettre en évidence à partir des
d
onné
es
les
sur-risque
s
de fracture lié
s
à la prise de médicaments
identifiés
c
omme facteur de risque de chute
(anxiolytiques, sédatifs et hypnotiques, antidépresseurs, antipsychotiques,
etc
...
)
.

Analyser les parcours de soins de patients

Au-delà des effets secondaires des médicaments, plusieurs approches orienteront les recherches
afin d'analyser, d'identifier et de qualifier dans la masse de données diff
é
rentes typologies de parcours de soins. Les chercheurs travailleront par exemple à estimer la fréquence des recours au système de santé
ou
décrire des pr
a
tiques cli
niques et

leur adéquation aux recommanda
tions médicales. Pour ce faire, les équipes dévelo
p
peront des
outils de visualisa
tion
et de
clustering
adaptés aux données volumineuses.

Identifier les mésusages et les fraudes

L'identification de situations de mésusage, voire de fraude, est indispensable à l'Assurance Maladie
pour éviter les gaspillages et assurer l'utilisation légitime de ses re
s
sources
. Ce projet vise à mettre en place
de
s
procédures automatisées permettant de balayer en continu la base de données
du Sniiram
pour
identifier
les situations su
s
pectes.
Avec
cette
nouvelle méthode employant le
Big Data
, l'innovation réside dans le fait de croiser les données en recherchant des cas de fraudes
sans a priori
initial sur les
types de
médicaments, la base étant analysée dans son intégralité. Cet outil constitue
un complément précieux, car économe en moyens, à la réalisation

par la CNAM
de travaux spécifiques
destinés à documenter ces éventuels mésusages et ces fraudes.

Pour Emmanuel Bacry, responsable de l'initiative Data Science de l'École polytechnique, « l'int
érêt de ce projet de recherche est de
déployer
, grâce à l'analyse de do
n
nées massives,
de nouvelles pistes
permettant de répondre
aux
enjeu
x
majeur
s
de santé publique.
À terme, un de nos objectif
s
est par exemple de passer
d'analyses ciblées et de tests individuels de médicaments à une détection automatique, simult
a
née et globale d'interactions
nocives
pour les patients
».

Le Sniiram : une des plus grandes bases de données médico-administratives au monde

Créé en 1999 par la loi de financement de la sécurité sociale, le Sniiram regroupe des données issues de 1,2 milliard de feuilles de soins chaque année sur la population couverte par le régime général en France, soit près de 65 millions d'assurés sociaux. Depuis qu'il a été relié aux données du Programme de Médicalisation des Systèmes d'Information (PMSI) en 2009 qui centralise la consommation de soins dans les établissements de santé, il rassemble désormais à la fois la consommation de soins de ville et celles liées aux hospitalisations.

Cette base de données, constituée de 200 millions de mégaoctets d'informations liées les unes aux autres (âge, sexe, médicaments, médecins traitants, durée des traitements, etc.), est l'une des rares à couvrir la quasi-totalité de la consommation de soins d'un pays.
En charge de la gestion du Sniiram pour l'Etat, l'Assurance Maladie s'appuie depuis plusieurs années sur cette base de données pour mener à bien ses missions d'assureur solidaire en santé tant au niveau de la qualité des soins que de la gestion des politiques de santé. Elle l'exploite également à des fins de santé publique et de pharmaco épidemiologie, en lien avec les autorités de santé.