



INSPIRE
REGION EST
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Utilisation du machine learning : expérience au CHRU de Nancy

Dr. Arnaud FLORENTIN

Médecin de santé publique - MCU-PH





INSPIRE
REGION EST
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Surveillance des ISO : limites

- Sous sa forme « classique »
 - Consommatrice de temps, donc coût +++
 - Repose sur un système déclaratif (disponibilité, sensibilisation, investissement des ≠ acteurs)



- Intérêt pour **(semi-) automatiser** la surveillance par le biais du **système d'information hospitalier**
 - Présence de nombreuses données collectées dans les bases de données



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES

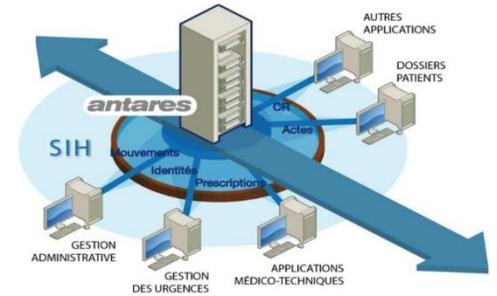


Inserm



Systeme d'information hospitalier

- Une donnée saisie, visible par tous
- Données administratives
- Données produites par les soins
 - Dossier Patient (PMSI, séjours, ...)
 - Logiciels métiers : laboratoire, radiologie, réanimation, pharmacie, blocs opératoires...
- Informatisation tardive mais croissante des données de soins => **opportunité +++**





INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Surveillance et SIH : 1^{ère} étape => système expert

- Première intégration testée : **laboratoires**
 - Notamment, prélèvement **bactériologique**
- Déployé en routine au CHRU de Nancy en 2008
 - Recueil des interventions grâce à la liste de code acte CCAM
 - Recueil des prélèvements bactériologiques
 - Croisement des données => Suspicion d'infection
 - Intégration en 2010 d'autres données biologiques (CRP, PNN) avec création d'un algorithme décisionnel



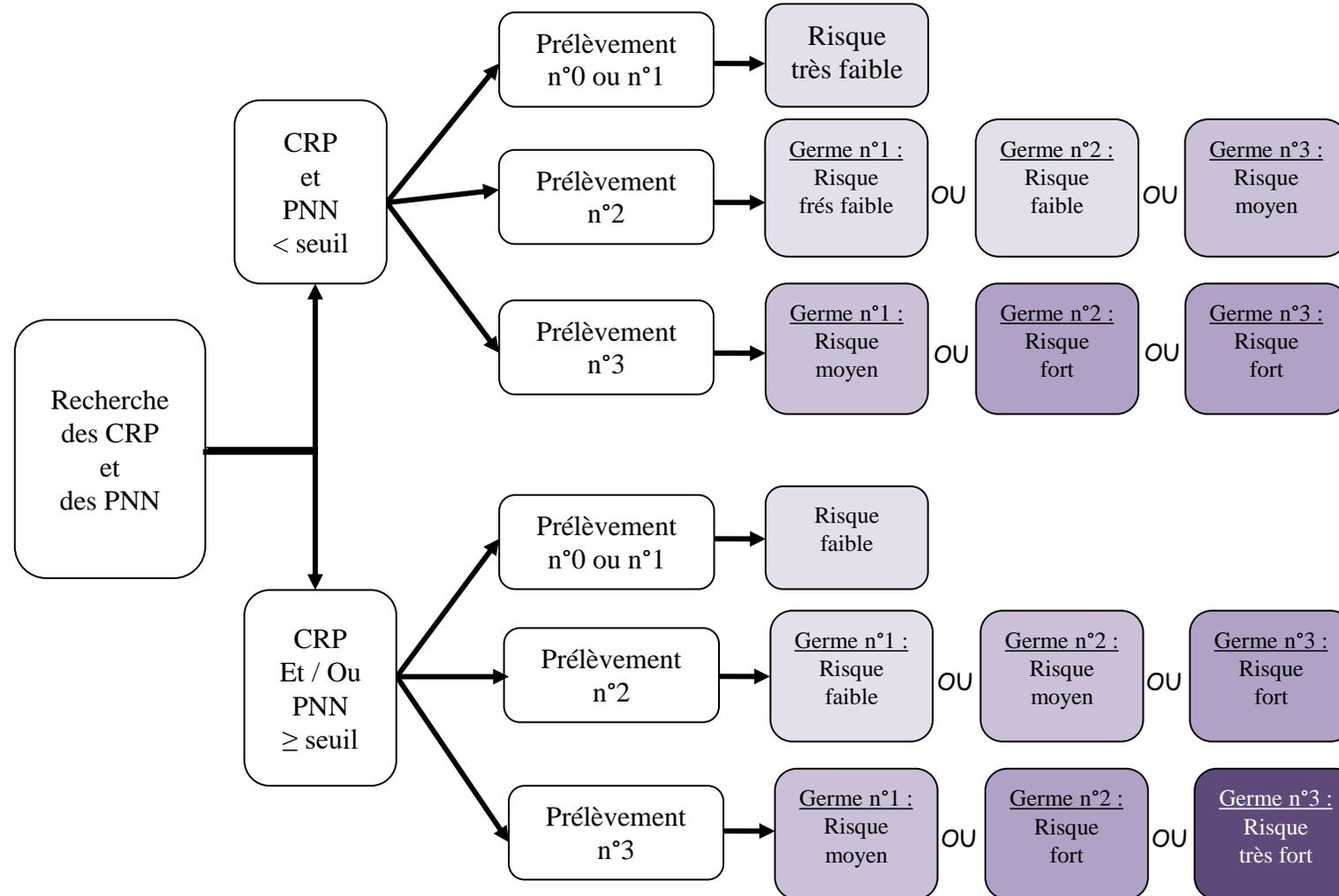
INSPIRE
REGION EST
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Surveillance et SIH : 1^{ère} étape => système expert



(1) Hautemanière A, Florentin A, Hunter PR, Bresler L, Hartemann P. Screening for surgical nosocomial infections by crossing databases. J Infect Public Health. 2013 Apr;6(2):89–97.



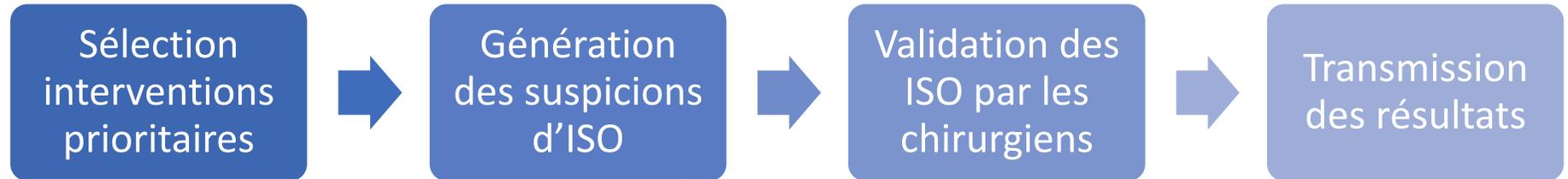
INSPIRE
 REGION EST
 INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
 INTERVENTIONS
 & INSTRUMENTS DE MESURE COMPLEXES



Inserm



Organisation de la surveillance des ISO au CHRU de Nancy



Basée sur les codes CCAM

Données issues du PMSI

Données recueillies à M+2

CRP > 100mg/L
 PNN > 12000 par mm³

Prélèvement bactériologique positif
 Nature du micro-organisme isolé

(1)	4	3	2	1
Se (%)	40,54	51,35	54,05	54,05
Sp (%)	95,42	94,66	94,66	93,13

Performances moyennes

Nécessité d'une nouvelle approche

(1) Hautemanière A, Florentin A, Hunter PR, Bresler L, Hartemann P. Screening for surgical nosocomial infections by crossing databases. J Infect Public Health. 2013 Apr;6(2):89–97.



INSPIRE
 INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
 INTERVENTIONS
 & INSTRUMENTS DE MESURE COMPLEXES



Inserm



Surveillance et SIH : d'autres approches

Infect Control Hosp Epidemiol. 2014 Nov;35(11):1330-5. doi: 10.1086/678422. Epub 2014 Sep 30.

Matching bacteriological and medico-administrative databases is efficient for a computer-enhanced surveillance of surgical site infections: retrospective analysis of 4,400 surgical procedures in a French university hospital.

Leclère B¹, Lasserre C, Bourigault C, Juvin ME, Chaillet MP, Mauduit N, Caillon J, Hanf M, Lepelletier D; SSI Study Group.

RESULTS: Sensitivity results varied significantly between the three algorithms, from 25% (95% confidence interval, 17-33) when using only administrative codes to 87% (80%-93%) with the bacteriological data and 90% (85%-96%) with the combined algorithm. Fewer variations were observed for specificity (91%-98%), PPV (21%-25%), and NPV (98% to nearly 100%). Overall, performance statistics were higher for deep SSIs than for superficial infections.

○ Comparaison de multiples sources

- Déclaration du chirurgien, PMSI, prescription d'antibiotique et prélèvements bactériologiques

Evaluation Study of Different Strategies for Detecting Surgical Site Infections Using the Hospital Information System at Lyon University Hospital, France

Solweig Gerbier-Colomban, MD,*†, Monique Bourjault, Nurse,* Jean-Charles Cêtre, MD, PhD,*†, Jacques Baulieux, MD, PhD,‡ and Marie-Hélène Metzger, MD, PhD*†

Méthode de collecte	Sensibilité	Spécificité
Notification chirurgien	<u>18.4%</u> (IC: 7.9 - 31.6)	100%
Prélèvements bactériologiques +	63.2% (IC: 47.3 - 78.9)	95.1% (IC: 92.9 - 97.1)
Prescription ATB	68.4% (IC: 52.6 - 81.6)	87.5% (IC: 84.3 - 90.7)
Code diag (CIM-10)	26.3% (IC: 13.2 - 42.1)	99.5% (IC: 98.8 - 100)
Combinaison des 3	<u>86.8%</u> (IC: 76.3 - 97.4)	85.5% (IC: 82.1 - 89.0)



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES

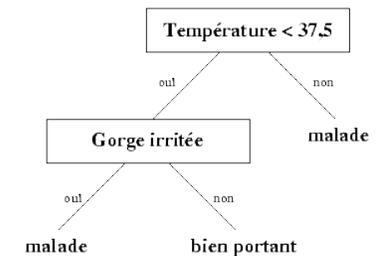


Inserm



Comment s'affranchir des modèles a priori / systèmes experts ?

- Intérêt du datamining / du machine learning
- Plusieurs approches disponibles
 - **Apprentissages non supervisés**
 - Règles d'association, clustering, réduction des dimensions
 - **Apprentissages supervisés**
 - Arbres de décision, forêts aléatoires, régressions linéaires ou logistiques
 - **Apprentissage par renforcement**
 - **Deep-learning**
 - Réseaux de neurones multicouches





INSPIRE
 INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
 INTERVENTIONS
 & INSTRUMENTS DE MESURE COMPLEXES



Inserm



Intérêt et faisabilité des techniques de datamining

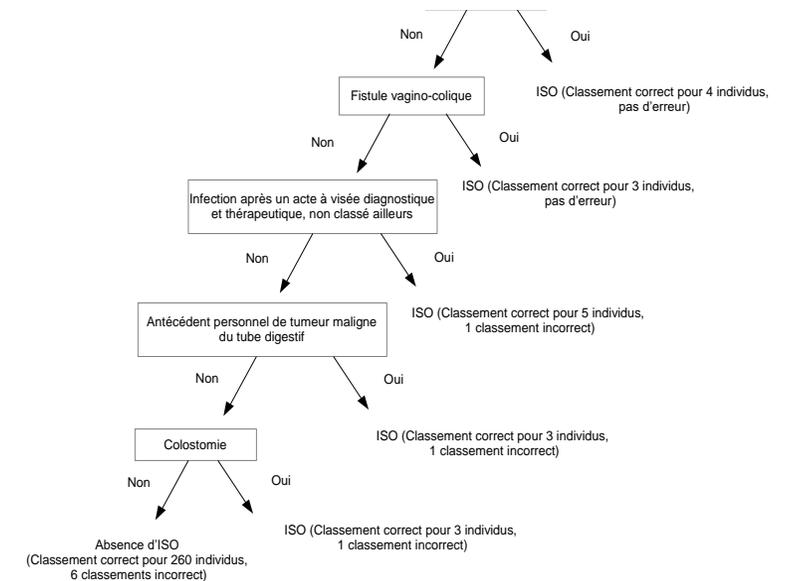
Travaux de thèse d'exercice Dr. D. PIVOT

- Etude rétrospective de 278 dossiers chirurgicaux (166 de neurochirurgie / 112 de chirurgie digestive) => 21 ISO
- Collecte des données de biologie et du PMSI
- 807 variables au total (code CCAM / diag. principal / diag. Associés ...)

Règles d'association

Règles d'association	Support	Confiance	Lift
- Prélèvement per-opératoire au niveau du site opératoire → ISO	6	1	13.24
- Iléus paralytique → ISO	4	1	13.24
- Fistule vagino-colique → ISO	3	1	13.24
- Intervention chirurgicale avec abouchement externe à l'origine de réactions anormales du patient ou de complications ultérieures, sans mention d'accident au cours de l'intervention → ISO	3	1	13.24
- Germe isolé ET Prélèvement per-opératoire au niveau du site opératoire → ISO	6	1	13.24
- CRP > 100 mg/ ET Autres interventions chirurgicales à l'origine de réactions anormales du patient ou de complications ultérieures, sans mention d'accident au cours de l'intervention → ISO	6	1	13.24
- Intervention chirurgicale avec abouchement externe à l'origine de réactions anormales du patient ou de complications ultérieures, sans mention d'accident au cours de l'intervention ET Germe isolé → ISO	5	1	13.24
- Chirurgie carcinologique ET Germe isolé → ISO	4	1	13.24
- Chirurgie carcinologique ET Prélèvement per-opératoire au niveau du site opératoire → ISO	4	1	13.24

Arbre de décision



Performance

Tableau 19 : Synthèse des résultats des différentes techniques d'apprentissage supervisé

	Arbres de décision		Règles de classement	
	SimpleCART	J48	RIPPER	PART
Précision (%)	95,2	92	90,1	92
Taux d'erreur (%)	5,03	6,8	7,5	6,8
Sensibilité (%)	95	93,2	92,4	93,2
Aire sous la courbe ROC (%)	64	70,8	57,9	69

Apport du machine learning pour la détection des ISO

○ Travaux de master 2 Dr. J. LIZON

Machine learning

2497 variables dont 680 codes CCAM et 1760 codes CIM-10

Méthodes d'apprentissages supervisées

Découpage échantillon : apprentissage (70%) et validation (30%)

Modèle aléatoire

- Forêt aléatoire⁽¹⁾
 - Ajout d'une randomisation
 - Hasard dans le choix des variables qui interviennent dans la construction des modèles
- 500 arbres de décision

Modèle adaptatif

- AdaBoost⁽²⁾
 - Adaptation des modèles construits en donnant plus de poids aux observations mal ajustées ou mal prédites
 - 50 itérations



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm





INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Apport du machine learning pour la détection des ISO

○ Population

- 3900 patients inclus issus de 10 services de chirurgie
 - Moyenne âge : 56 ans +/- 19,8 [18-105]
 - Sexe ratio : 0,78
- Ayant eu une chirurgie entre 1^{er} janvier et 30 juin 2015

○ Taux d'infection du site opératoire (spécialité min/max)

	Nombre d'interventions	ISO validées	Taux d'ISO (%)
Chirurgie thoracique	118	16	13,54
Chirurgie gynécologique et obstétrique	639	3	0,47
Totalité	3900	192	4,92



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Apport du machine learning pour la détection des ISO

- Performance des modèles
- Gain en sensibilité et en spécificité

	Système expert	Forêt aléatoire	AdaBoost
Sensibilité (%)	67,8	81,4	79,7
Spécificité (%)	92,2	97,8	97,7
VPP (%)	31,5	66,7	67,1
VPN (%)	98,1	99,0	98,9
AUC	0,827	0,988	0,979
→ NRI	0	0,212	0,176
Erreur classification (%)	9	3	3

$p < 10^{-5}$

$p < 10^{-4}$ $p = 0,43$



INSPIRE
 REGION EST
 INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
 INTERVENTIONS
 & INSTRUMENTS DE MESURE COMPLEXES



Inserm

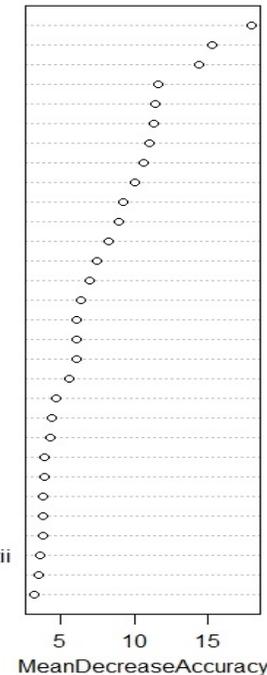


Apport du machine learning pour la détection des ISO

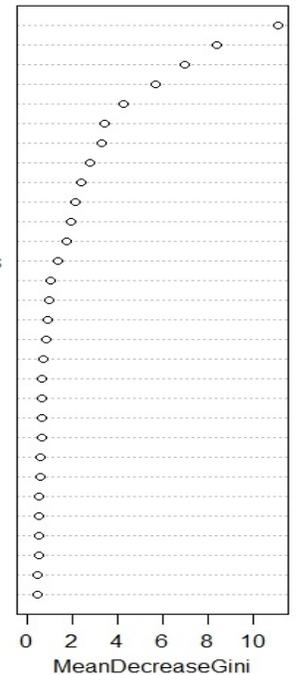
○ Pertinences variables : modèle Forêt aléatoire

- T814 (Complication infectieuse d'actes de diagnostiques ou thérapeutiques)
- L022 (Présence d'un abcès cutané)
- T8138 (Désunion de la plaie opératoire)
- QZJA011 (Evacuation collection profonde)
- L088 (infection de la peau ou sous-cutanée)
- T81 (Complications d'actes à visée diagnostique et thérapeutique, non classées ailleurs)
- ZCJA004 (Évacuation de plusieurs collections intraabdominales, par laparotomie)

T814
 L022
 score_global
 score_germes
 Staphylococcus.aureus
 score_pvt
 Escherichia.coli
 score_crp
 Enterococcus.faecalis
 T8138
 Staphylococcus.epidermidis
 Klebsiella.pneumoniae
 Pseudomonas.aeruginosa
 duree_hospit.5j
 QZJA011
 Morganella.morganii
 L088
 T81
 N823
 J188
 score_pnn
 HHQE005
 ZCJA004
 Clostridium.perfringens
 J985
 HZHE002
 X0QK023
 Corynebacterium.kroppenstedtii
 Y834
 var_coelio



T814
 score_global
 score_germes
 score_pvt
 score_crp
 L022
 T81
 score_pnn
 T8138
 Escherichia.coli
 Staphylococcus.aureus
 Enterococcus.faecalis
 Staphylococcus.epidermidis
 K650
 Y832
 Klebsiella.pneumoniae
 R65
 B952
 duree_hospit.5j
 Pseudomonas.aeruginosa
 U82
 QZJA011
 Y836
 ZBQK002
 T818
 K913
 ZZQP001
 J18
 L088
 B957





INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm

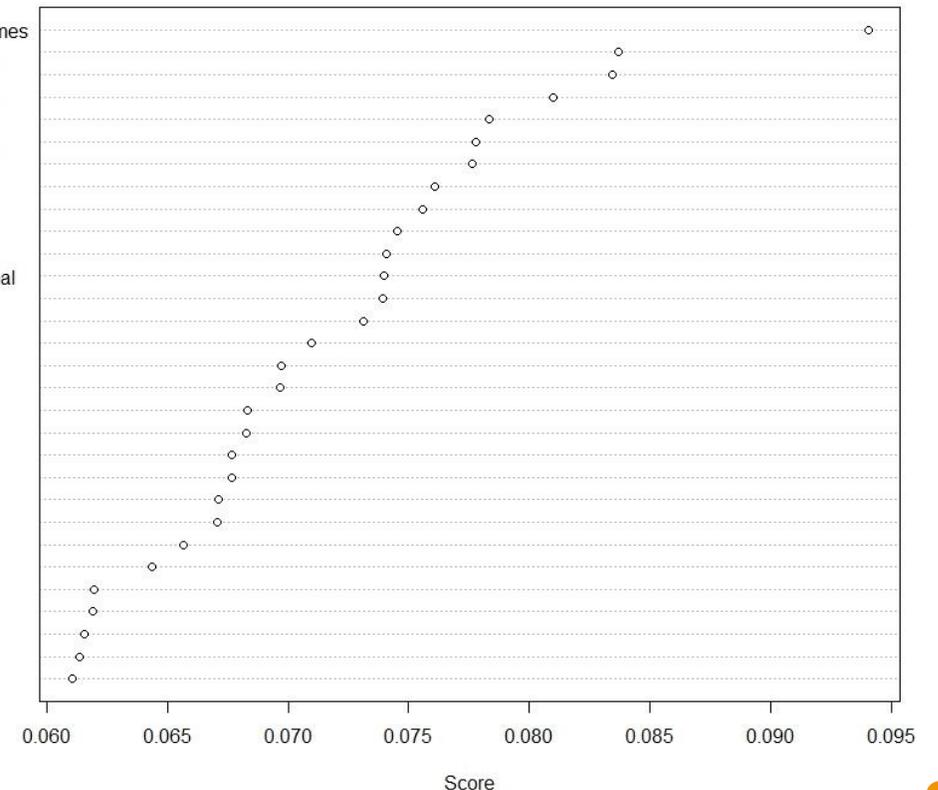


Apport du machine learning pour la détection des ISO

○ Pertinences variables : modèle Adaboost

- L022 (Présence d'un abcès cutané)
- T8138 (Désunion de la plaie opératoire)
- T818 (Autre complication d'actes de diagnostics ou thérapeutiques)
- JJQX004, HGQX008, AZQX005 (examen ana-path à visé carcinologique)
- N61 (affections inflammatoires du sein)

score_germes
score_plvt
score_crp
score_pnn
Z590
JJQX004
HGQX008
N61
L022
Y834
T818
score_global
G25
K413
T8138
H55
G119
AZQX005
ZZQN002
Z951
T810
N40
C17
Z902
N823
K409
Z488
Z8670
EAQJ001
Y836





INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Avantages du machine learning

- Construction possible de modèles de détection
 - Intégration de nombreuses données médicales / administratives...
 - 2497 variables explicatives
 - Pertinences des variables retenues par les modèles
 - Meilleure performance de ces modèles par rapport au système expert antérieur
 - Modèle applicable au quotidien avec de légères manipulations
- En mode fouille de données
 - Découverte de codes intéressants pour un système expert
 - Evènement infectieux potentiel après la chirurgie
 - T814, LO22, QZJA011 => diagnostic direct ?! Probabilité forte dans le modèle
 - Amélioration codage
 - Sensibilisation des TIM au codage des complications infectieuses
 - Améliorer l'exhaustivité et la qualité des observations des chirurgiens (CRH, courrier de consultation, etc.)



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Limites du machine learning

- Validation nécessaire des modèles :
 - Au niveau local : forte dépendance aux pratiques cliniques et de codage
 - Dans le temps : évolution des pratiques/codage
- A intégrer dans une surveillance
 - Semi-automatisée : « alertes » à confirmer cliniquement
 - Automatisée : en ayant conscience des limites (faux positifs ou négatifs)
 - Patients non détecté/déTECTABLE (non hospitalisé)
- Problème technique
 - Surajustement / dérive
- Délai de retour aux équipes médico-soignantes
 - Dépendant de la production des données pour le PMSI



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Perspective du machine learning

- Nombreuses études basées sur l'apport des données du PMSI, notamment les Infections associées aux soins
 - Règles d'associations ⁽¹⁾ voir Text-mining ⁽²⁾
- Intégration des modèles avec les données disponibles en temps réel (pseudo-réel)
 - Entrepôt de données
 - Développement des logiciels métier en surveillance

(1) Brossette SE, Sprague AP, Hardin JM, Waites KB, Jones WT, Moser SA. Association rules and data mining in hospital infection control and public health surveillance. J Am Med Inform Assoc JAMIA. 1998 Aug;5(4):373-81.

(2) Campillo-Gimenez B, Garcelon N, Jarno P, Chaplain JM, Cuggia M. Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France. Stud Health Technol Inform. 2013;192:572-5.



INSPIRE
INTERDISCIPLINARITÉ EN SANTÉ PUBLIQUE,
INTERVENTIONS
& INSTRUMENTS DE MESURE COMPLEXES



Inserm



Merci pour votre attention

