## Apprentissage sur le vaccin contre la grippe : prédire les vaccins contre le H1N1 et la grippe saisonnière

Merci pour ce retour d'expérience. Vous avez vécu là un drôle d'entretien, certainement pas facile à vivre!

Il faut bien comprendre que les data ce n'est pas : j'ai un fichier donc il est analysable, j'ai des données donc elles sont correctes, j'ai des données il y a un problème ou bien j'ai des données et je vais trouver où est le problème…

On analyse en général des données quand on est sûr qu'elles peuvent donner quelque chose sinon c'est le démarrage de toutes les bêtises liés à la folie de l'entreprise d'analyser ces données alors qu'elle ne peuvent rien apporter (porte ouverte à tous les biais et à toutes les erreurs d'interprétation). Je pense qu'une entrevue où on demande à quelqu'un de regarder où est le problème en 15 secondes est complètement farfelue. On se sert d'ordinateurs à notre époque pour analyser les données. Alors, bien sûr, on peut toujours s'amuser à recevoir un jeune informaticien et à l'intimider mais de toute façon l'analyse des données ce n'est pas de l'informatique et seulement de l'informatique, l'analyse des données c'est une démarche bien particulière qui demande des compétences bien particulières et je dirais pour ajouter une touche finale que chaque fichier peut constituer un cas particulier.

Vous pouvez toujours regarder ma vidéo "les Data sciences qu'est-ce que c'est; à quoi ça sert" où j'explique en 4 minutes quelle est cette démarche.

Encore merci pour ce retour d'expérience très intéressant. J'espère que vous avez pu vous en remettre et que vous avez trouvé une autre opportunité de job.

Bien cordialement.

Pascal

# C’est noté. Pensez à avoir une analyse assez complète :

# Analyse exploratoire, visualisation,

# Réduction de dimension, statistiques descriptives,

# Et toute la partie prédictive bien étudiée

# Chaque section doit contenir une analyse de votre part expliquant et motivant les choix et décrivant les méthodes.

# Ce rapport représente 4 mois d’alternance pour certains de vos collègues. De plus, rappelez-vous qu’il est associé à une note éliminatoire de 8/20. Il est important que vous ne preniez pas ce travail comme un TP, mais comme un véritable projet.

# Description du problème

## Votre objectif est de prédire la probabilité que les individus reçoivent leur vaccin contre le H1N1 et la grippe saisonnière. Plus précisément, vous devez prédire deux probabilités : une pour et une pour .h1n1\_vaccine seasonal\_vaccine

## Chaque ligne de l’ensemble de données représente une personne qui a répondu à l’Enquête nationale sur la grippe H1N1 de 2009.

## Étiquettes

Pour cette compétition, il existe deux variables cibles :

* h1n1\_vaccine - Si le répondant a reçu le vaccin contre la grippe H1N1.
* seasonal\_vaccine - Si le répondant a reçu le vaccin contre la grippe saisonnière.

Les deux sont des variables binaires : 0 = Non ; 1 = Oui. Certains répondants n’ont reçu aucun vaccin, d’autres n’en ont reçu qu’un seul et certains ont reçu les deux. Ceci est formulé comme un problème multi label (et *non* multiclasse).01

## Les fonctionnalités de cet ensemble de données

Vous disposez d’un jeu de données avec 36 colonnes. La première colonne est un identifiant unique et aléatoire. Les 35 fonctionnalités restantes sont décrites ci-dessous.respondent\_id

Pour toutes les variables binaires :0 = Non ;1 = Oui.01

* h1n1\_concern - Niveau d’inquiétude concernant la grippe H1N1.
  + 0 = Pas du tout concerné; 1 = Pas très inquiet; 2 = Quelque peu inquiet; 3= Très inquiet.123
* h1n1\_knowledge - Niveau de connaissance de la grippe H1N1.
  + 0 = Aucune connaissance; 1 = Un peu de connaissance; 2= Beaucoup de connaissances.12
* behavioral\_antiviral\_meds - A pris des médicaments antiviraux. (binaire)
* behavioral\_avoidance - A évité tout contact étroit avec d’autres personnes présentant des symptômes pseudo-grippaux. (binaire)
* behavioral\_face\_mask - A acheté un masque facial. (binaire)
* behavioral\_wash\_hands - S’est fréquemment lavé les mains ou a utilisé du désinfectant pour les mains. (binaire)
* behavioral\_large\_gatherings - A réduit le temps passé lors des grands rassemblements. (binaire)
* behavioral\_outside\_home - A réduit les contacts avec des personnes extérieures à son propre ménage. (binaire)
* behavioral\_touch\_face - A évité de se toucher les yeux, le nez ou la bouche. (binaire)
* doctor\_recc\_h1n1 - Le vaccin contre la grippe H1N1 a été recommandé par le médecin. (binaire)
* doctor\_recc\_seasonal - Le vaccin contre la grippe saisonnière a été recommandé par le médecin. (binaire)
* chronic\_med\_condition - A l’une des conditions médicales chroniques suivantes: asthme ou une autre affection pulmonaire, diabète, une maladie cardiaque, une maladie rénale, une anémie falciforme ou une autre anémie, une affection neurologique ou neuromusculaire, une affection hépatique ou un système immunitaire affaibli causée par une maladie chronique ou par des médicaments pris pour une maladie chronique. (binaire)
* child\_under\_6\_months - Avoir des contacts étroits réguliers avec un enfant de moins de six mois. (binaire)
* health\_worker - Est un travailleur de la santé. (binaire)
* health\_insurance - A une assurance maladie. (binaire)
* opinion\_h1n1\_vacc\_effective - Opinion du répondant sur l’efficacité du vaccin H1N1.
  + 1 = Pas du tout efficace; = Pas très efficace; = Je ne sais pas; = Quelque peu efficace; = Très efficace.2345
* opinion\_h1n1\_risk - Opinion du répondant sur le risque de contracter la grippe H1N1 sans vaccin.
  + 1 = Très faible; = Un peu faible; = Je ne sais pas; = Un peu élevé; = Très élevé.2345
* opinion\_h1n1\_sick\_from\_vacc - L’inquiétude du répondant de tomber malade en prenant le vaccin H1N1.
  + 1 = Pas du tout inquiet; = Pas très inquiet; = Je ne sais pas; = Quelque peu inquiet; = Très inquiet.2345
* opinion\_seas\_vacc\_effective - Opinion du répondant sur l’efficacité du vaccin contre la grippe saisonnière.
  + 1 = Pas du tout efficace; = Pas très efficace; = Je ne sais pas; = Quelque peu efficace; = Très efficace.2345
* opinion\_seas\_risk - Opinion du répondant sur le risque de contracter la grippe saisonnière sans vaccin.
  + 1 = Très faible; = Un peu faible; = Je ne sais pas; = Un peu élevé; = Très élevé.2345
* opinion\_seas\_sick\_from\_vacc - L’inquiétude du répondant de tomber malade en prenant le vaccin contre la grippe saisonnière.
  + 1 = Pas du tout inquiet; = Pas très inquiet; = Je ne sais pas; = Quelque peu inquiet; = Très inquiet.2345
* age\_group - Groupe d’âge du répondant.
* education - Niveau d’éducation autodéclaré.
* race - Race du répondant.
* sex - Sexe du répondant.
* income\_poverty - Revenu annuel du ménage du répondant par rapport aux seuils de pauvreté du Recensement de 2008.
* marital\_status - État matrimonial du répondant.
* rent\_or\_own - Situation du logement du répondant.
* employment\_status - Statut d’emploi du répondant.
* hhs\_geo\_region - Résidence du répondant selon une classification géographique de 10 régions définie par le département américain de la Santé et des Services sociaux. Les valeurs sont représentées sous forme de courtes chaînes de caractères aléatoires.
* census\_msa - Résidence du répondant dans les régions statistiques métropolitaines (MSA) telles que définies par le recensement des États-Unis.
* household\_adults - Nombre *d’autres* adultes dans le ménage, codé en haut à 3.
* household\_children - Nombre d’enfants dans le ménage, codé en haut à 3.
* employment\_industry - Type d’industrie dans laquelle le répondant est employé. Les valeurs sont représentées sous forme de courtes chaînes de caractères aléatoires.
* employment\_occupation - Type de profession du répondant. Les valeurs sont représentées sous forme de courtes chaînes de caractères aléatoires.

1.h1n1\_concern

2.h1n1\_knowledge

3.behavioral\_antiviral\_meds

4.behavioral\_avoidance

5.behavioral\_face\_mask

6.behavioral\_wash\_hands

7.behavioral\_touch\_face

8.health\_worker

9.health\_insurance

10. opinion\_h1n1\_vacc\_effective

11. opinion\_h1n1\_risk

12. opinion\_h1n1\_sick\_from\_vacc

13. opinion\_seas\_vacc\_effective

14. opinion\_seas\_risk

15. opinion\_seas\_sick\_from\_vacc

16. marital\_status

1- [NLP : une classification multilabels simple, efficace et interprétable | OCTO Talks !](https://blog.octo.com/nlp-une-classification-multilabels-simple-efficace-et-interpretable/)