Predicting H1N1 Vaccination

Predict how likely individuals are to receive their H1N1 vaccination.

Prepared by : Ang Su Yiin
Date : 15 November 2020

Content

Overview and objective

Data preparation and exploration

Type of results

Basis of model assessment

Selected model

Model Improvement

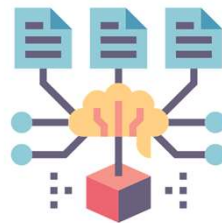Recommendations

Objective and approach

- To build a **predictive model** to predict how **likely individuals are to receive their H1N1 vaccination** based on individuals' demographics and opinions on vaccine.

Clean and analyse

Models

Results

Survey data by NHFS 2009

Build multiple predictive models

Select the best model based on objective

Data preparation and exploration

Data	Data quality	Best predictor

Data

27,707 observations
38 variables

79% Not vaccinated
21% vaccinated

Data quality

Missing values
3 variables > 40%
Others 0.08% - 16.5%

GREAT MODEL
GARBAGE DATA
GARBAGE RESULTS

2 variables with random
characters - no explanation

Best predictor

TRUST ME....
I'M A DOCTOR

Doctors'
recommendation

4 type of results
False positive is most detrimental

| Target (Actual)		Outcome (Predicted)	
		0 – Not vaccinated	**1 - Vaccinated**
	0 – Not vaccinated	**True Negative** - Actual: Not Vaccinated - Predicted: Not Vaccinated	**False Positive** - Actual: Not Vaccinated - Predicted: Vaccinated
	1 – Vaccinated	**False Negative** - Actual: Vaccinated - Predicted: Not Vaccinated	**True Positive** - Actual: Vaccinated - Predicted: Vaccinated

Basis of our model assessment

Misclassification rate

High Negative predicted value

Strong model

Predicted

Actual

NO YES

NO

True Negatives

False Positives

YES
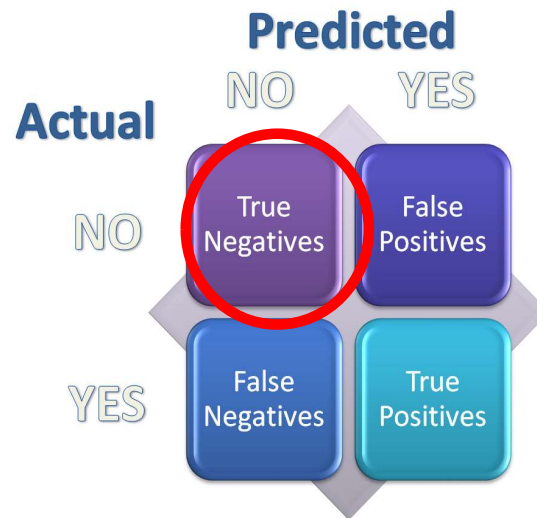
False Negatives
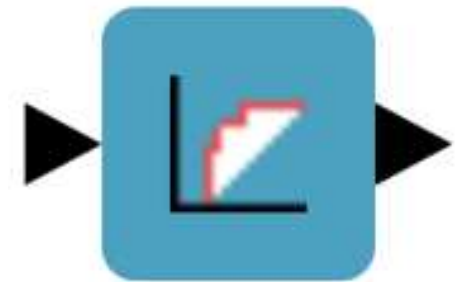
True Positives

Selected model – Recursive Partitioning Model

Attributes	Model results
True Positive	1,048
False Positive	482
True Negative	5828
False Negative	655
Misclassification	14.2%
Negative Predicted Value	89.9%
True Negative Rate	92.4%

Selected model: Recursive Partitioning with 5 branches

Results of assessment:

- 3th best misclassification rate

- Highest True Negative Rate

- ROC Curve of 0.837 (strong model)

Model improvement – True Negative Rate can be improved using different cut off values
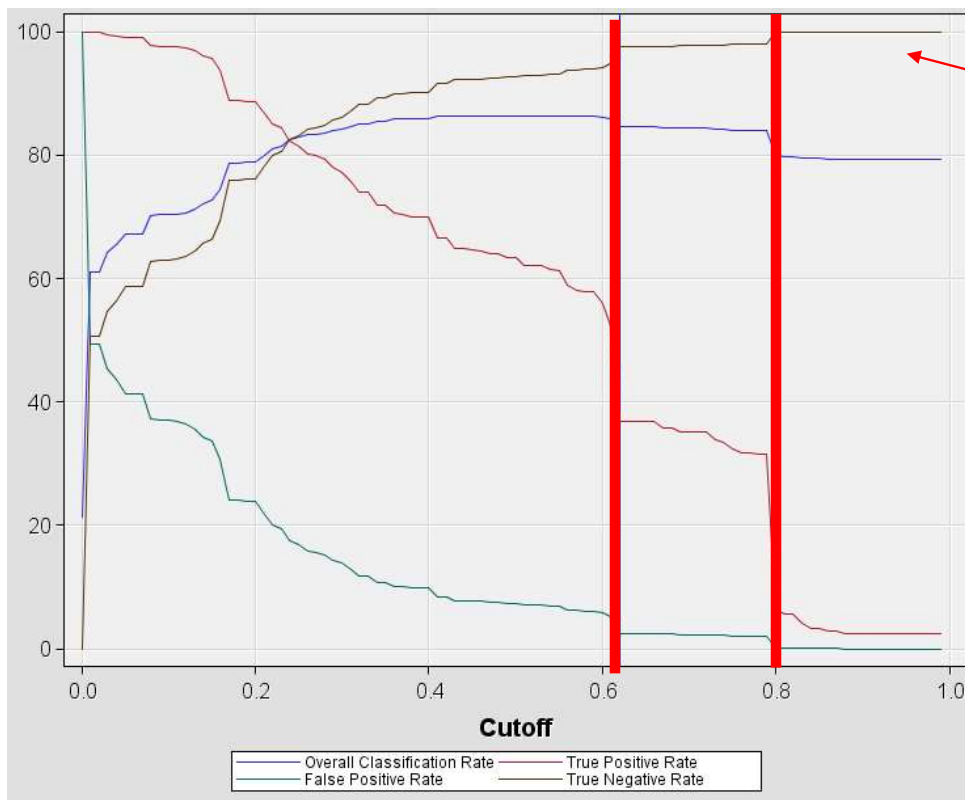
True Negative line

The higher the cut off point, the high the True Negative rate.

3 options to select from :

- Base model

- Cut off 0.62

- Cut off 0.40

Cut off value depends on benchmarks and public health policies

Model Improvement: Cutoff Comparison

Attributes	No Cutoff	Cutoff 0.62	Cutoff 0.79
True Positive	1047	558	495
False Positive	482	219	166
True Negative	5828	6092	6145
False Negative	655	1144	1207
Misclassification	14.2%	17%	17.1%
Negative Predicted Value	89.9%	84.2%	83.6%
True Negative Rate	92.4%	96.5%	97.4%

The most detrimental cases are the False Positive: Wrongly predicted to be vaccinated but actually not vaccinated

Maximise True Negative would reduce False Positive, but increase False Negative

Recommendations

- **Improve data quality**
  - ➢ Reduce missing data
  - ➢ Collect more relevant data that would be good predictors for the prediction model

- Perform **cost-benefit analysis** of different cut off values – selecting a specific cut off value would largely depends on resource available and cost associated with investigation.

- Perform **further analysis** (i.e clustering analysis) on predicted unvaccinated individuals to assess the best and most efficient way of incentivising these individuals to get vaccinated.