Flu shot acceptance: learning from a previous pandemic

Improving public health by identifying targets for vaccine uptake promotion

Christina Rudolf, Raymond Boateng, Juliane Berek

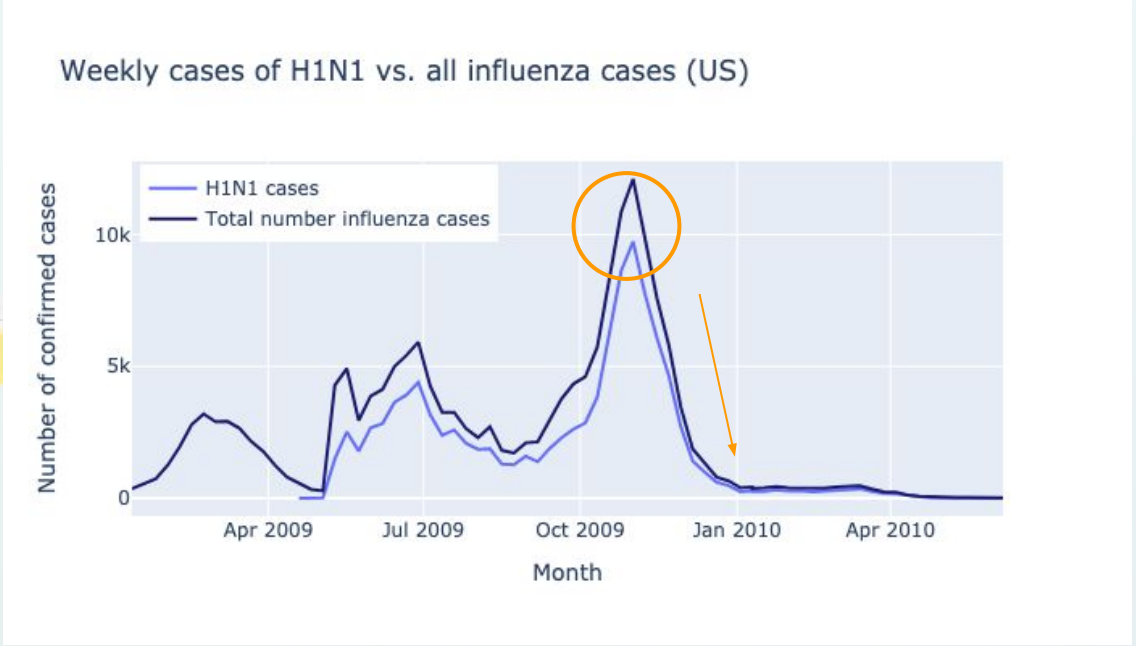
Setting the Scene: The H1N1 Pandemic

- Pandemic of swine influenza, caused by **H1N1 influenza** virus
- **June 2009 to August 2010** (according to WHO declaration)
    - Most recent pandemic prior to COVID19

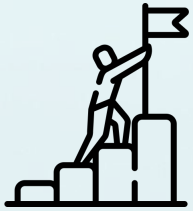
President Barack Obama being vaccinated against H1N1 flu on 20 December 2009 ® Public Domain

Setting the Scene: The H1N1 Pandemic

- Real number of cases estimated
  11 - 21% of the global population
- Total number of deaths in the U.S. was 12,469



Objectives



- **Increasing vaccination rates** for seasonal and pandemic flu in the overall population
    - **Reducing the burden of influenza** by decreasing hospitalisations/ deaths

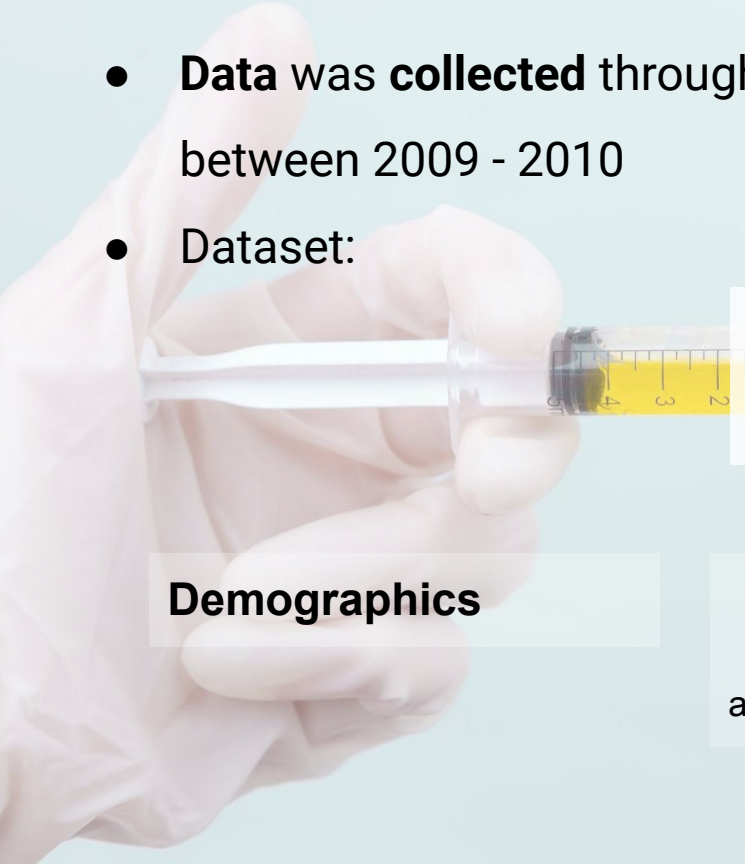
- Identify **factors** that determine the chance of **getting vaccinated**
- Identify **groups with lower likelihood for getting vaccinated**, in order to target them with promotions

- Determining differences of seasonal vs. H1N1 (pandemic) vaccinations

Data Overview

Our Sample: 27.000 Respondents

- **Data** was **collected** through the **National 2009 H1N1 Flu Survey** in the U.S. between 2009 - 2010
- Dataset:



Vaccination status
- H1N1
- Seasonal flu

Demographics

Attitudes and knowledge
about H1N1 and seasonal flu

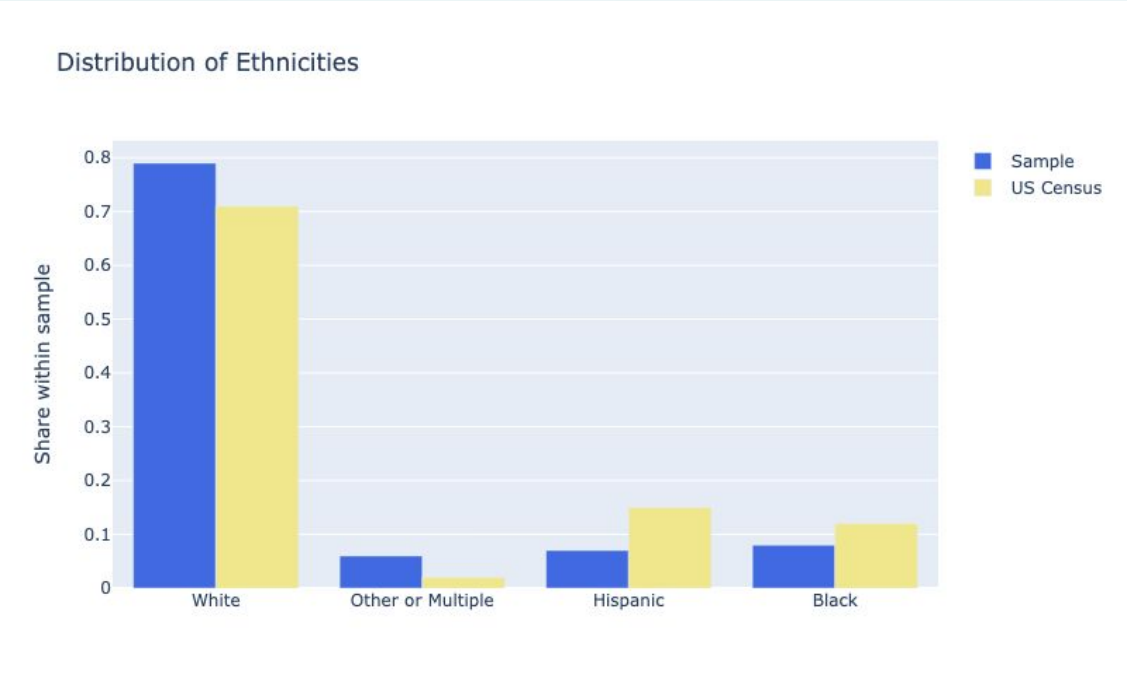
Healthcare Information

The Sample - 27,000 of the US population

The sample is biased towards older population.

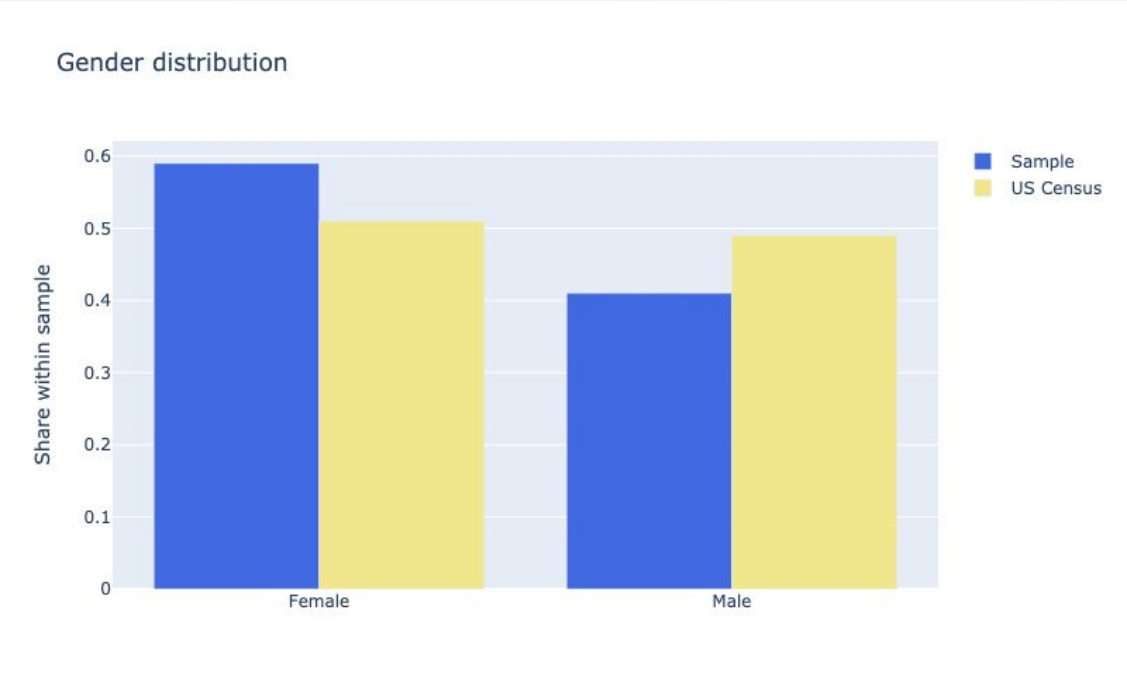
Ethnicities

Non-white population is underrepresented.



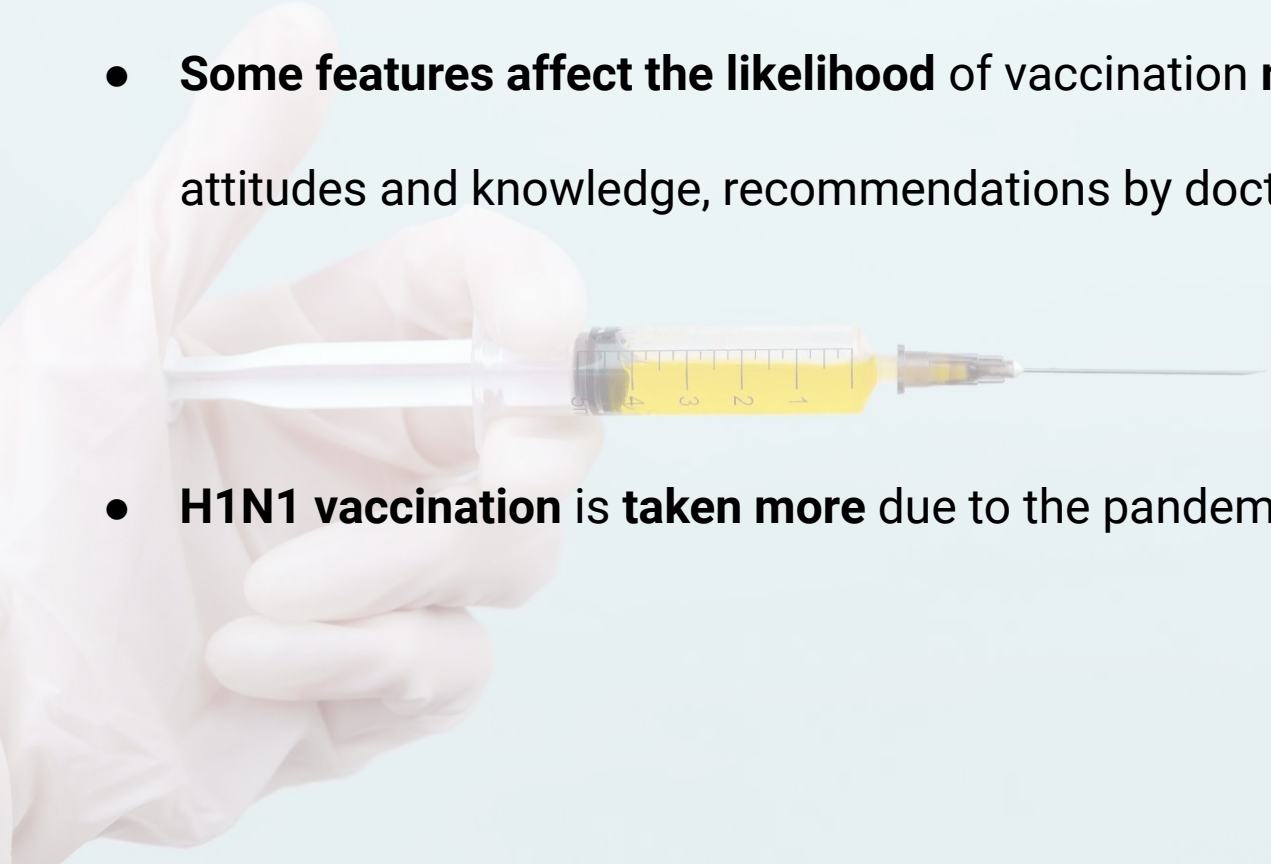
Gender

Men are underrepresented.



Hypotheses

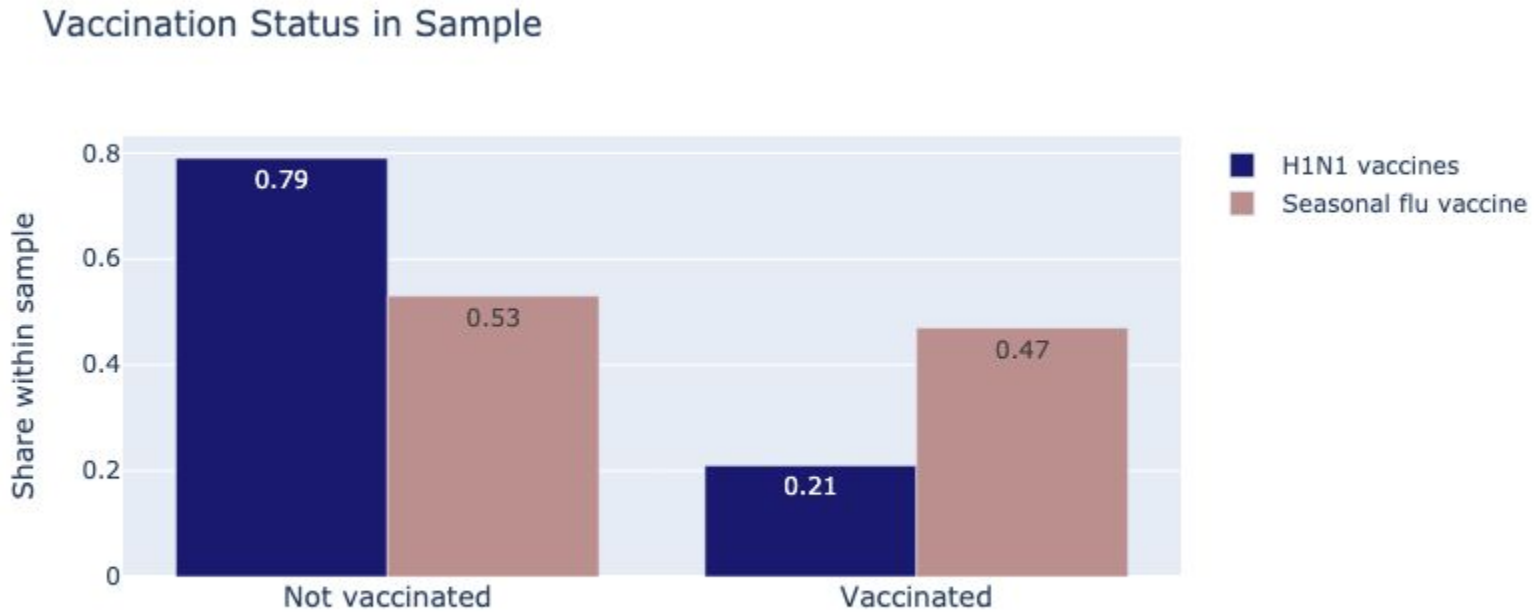
- **Some features affect the likelihood** of vaccination **more than others**, e.g. attitudes and knowledge, recommendations by doctors



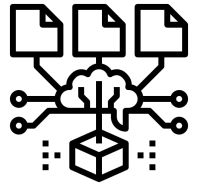
- **H1N1 vaccination** is **taken more** due to the pandemic context

Distribution of H1N1 and Seasonal Flu Vaccines

The seasonal flu vaccine is taken more than the H1N1 flu vaccine



Overview of Prediction Model Development



Imputing missing data
Drop all missing values
Most frequent value per column
KNN imputation
Bayesian Ridge Regression imputation
Decision Tree Regressor imputation
No imputation

Hyperparameter tuning
For H1N1 vaccine prediction

Feature selection
Not including opposite vaccine as a feature
Opposite vaccine included as a feature

Data balancing (H1N1 vaccine)
No balancing
Upsampling minority class
Downsampling majority class

BEST PREDICTION MODEL

Modelling approach
Predict H1N1 and seasonal vacc together (multilabel modelling)
Predict each vaccine individually

Feature engineering, multicollinear features
Combination or dropping

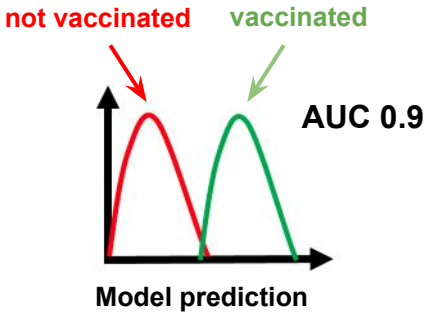
Predictive algorithm comparison
Logistic regression
K nearest neighbours
SVM
Random forest classifier

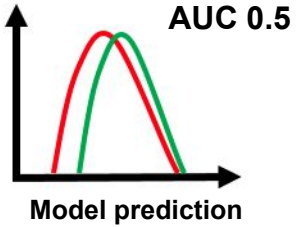
Outcomes of Our Model

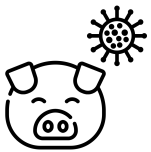
Performance of Best Prediction Model

How to decide on best model?

→ AUC score

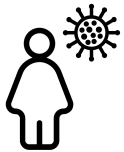






H1N1 flu vaccination prediction

AUC score 0.81



Seasonal flu vaccination prediction

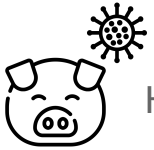
AUC score 0.8

Why do we want a prediction model?

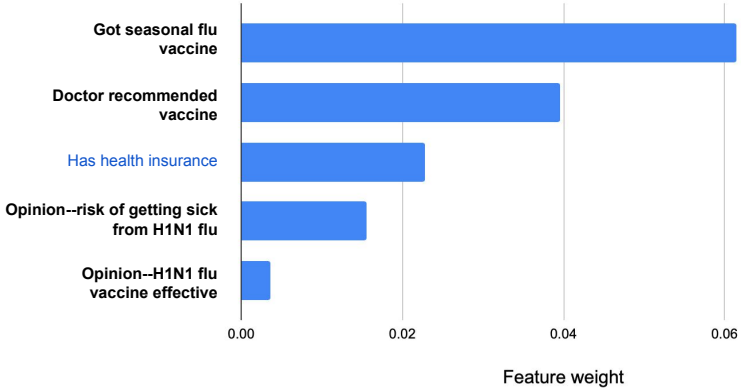
To identify important features for vaccine uptake!

https://medium.com/greyatom/lets-learn-about-auc-roc-curve-4a94b4d88152

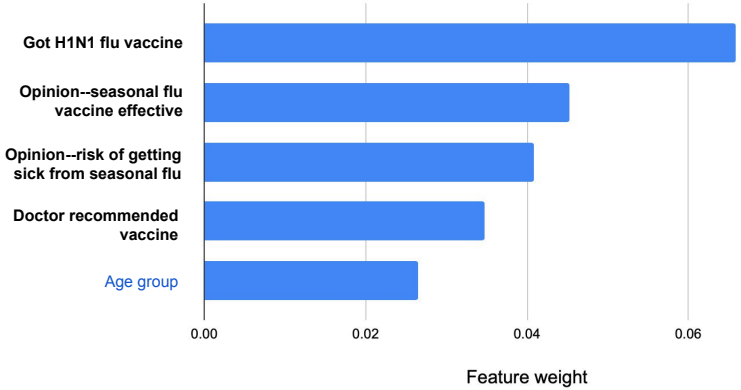
Important Features for Vaccine Uptake



H1N1 flu vaccination, top 5 drivers

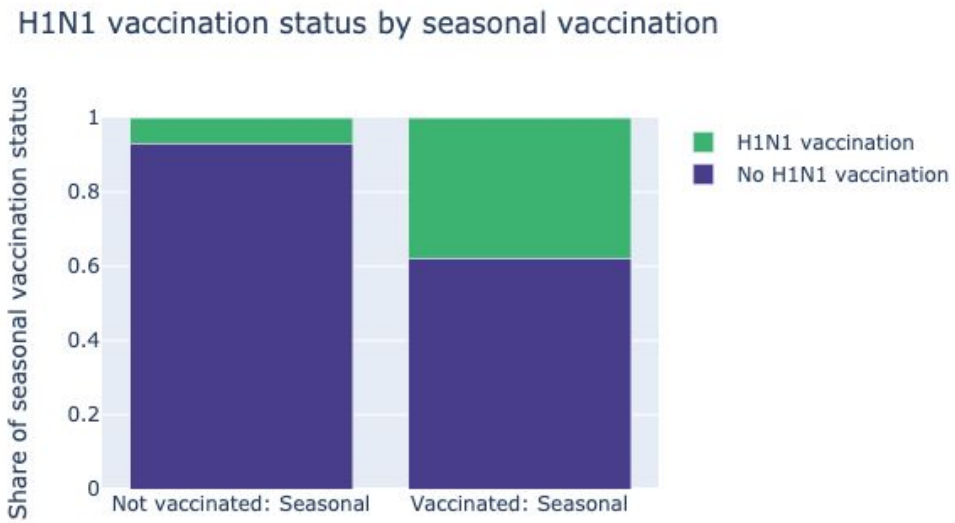


Seasonal flu vaccination, top 5 drivers

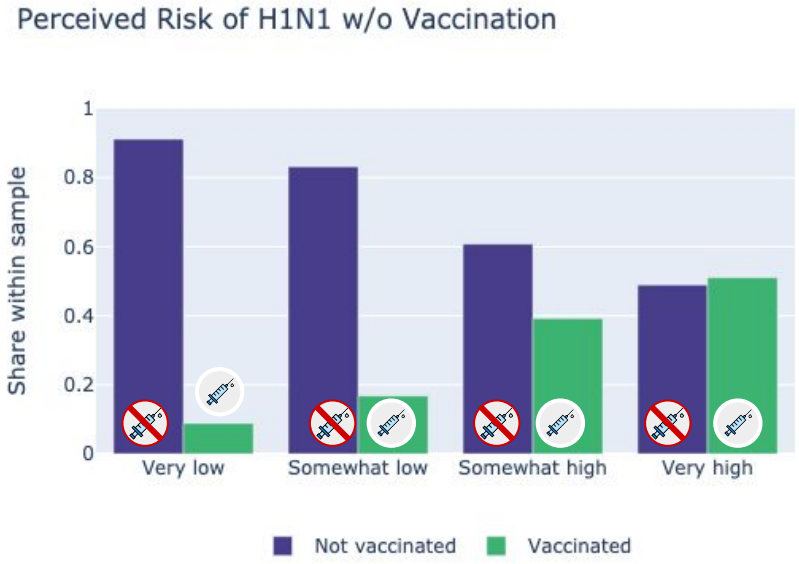


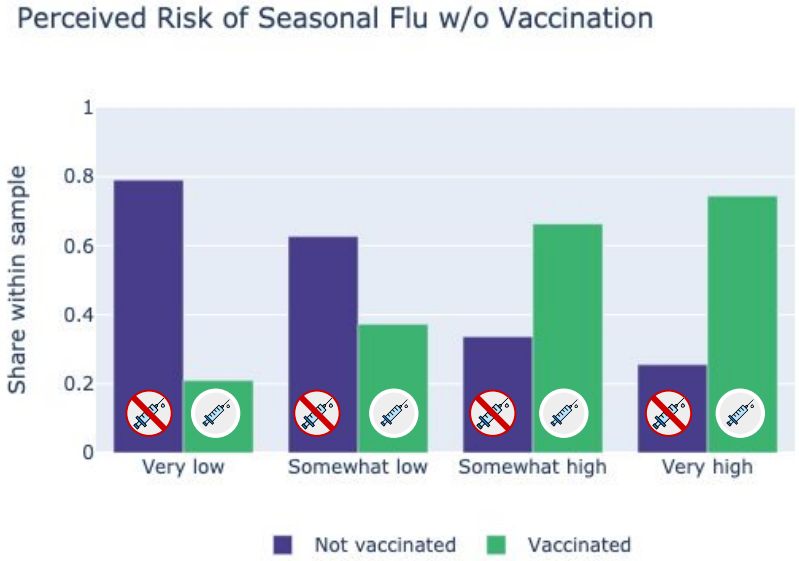
Effect of Taking the Other Vaccine

Taking the other vaccine corresponds with **greater uptake** of our vaccine of interest



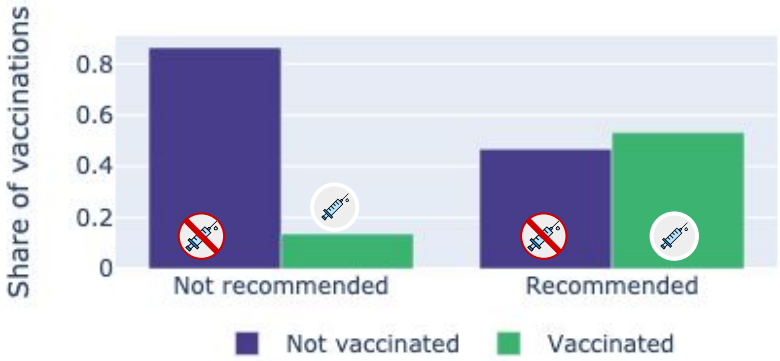
Risk Perception Increases the Uptake of Vaccinations





Positive Effect of Doctor's Recommendation

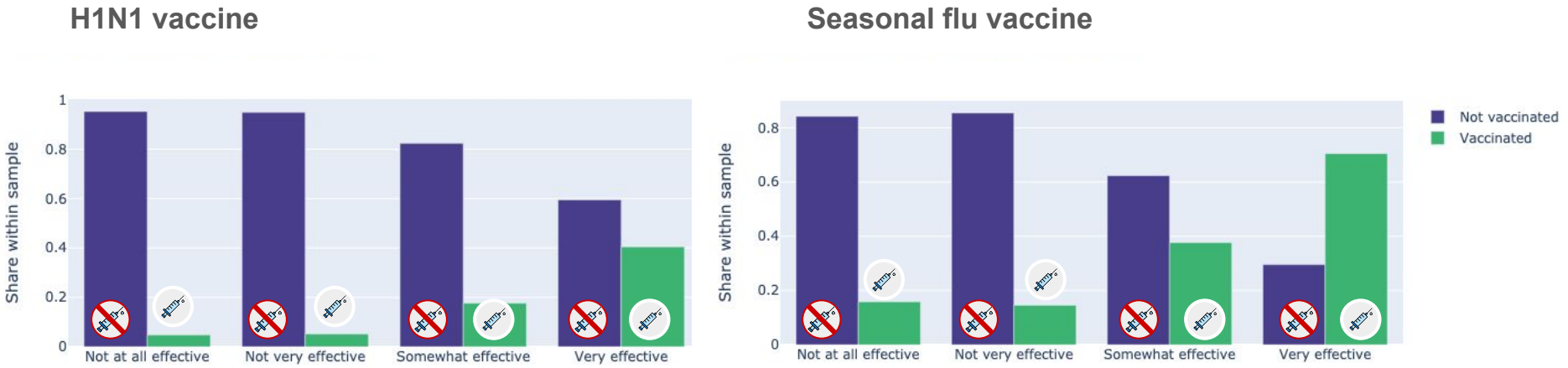
Doctor’s recommendation, H1N1 vaccine



Doctor’s recommendation, seasonal flu vaccine

Opinion of Vaccine Effectiveness

A higher opinion of whether a vaccine is effective leads to higher vaccination rates.



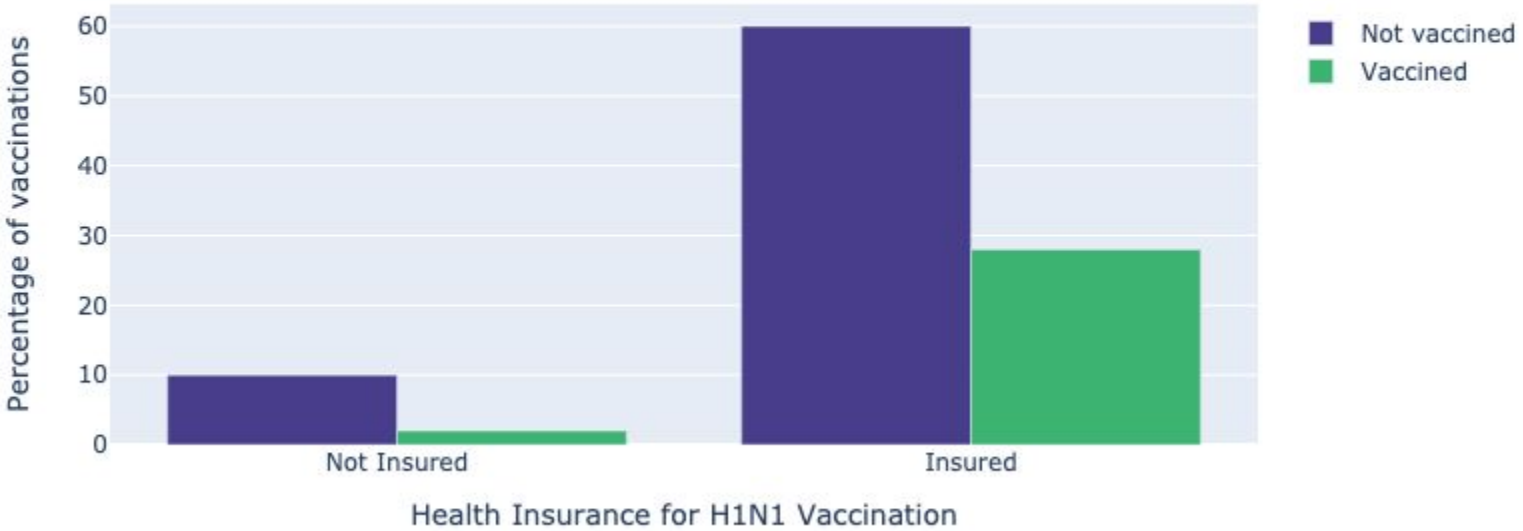
Lessons for a Pandemic

Concerns and knowledge

Increased concern and knowledge about H1N1 flu leads to greater vaccination uptake

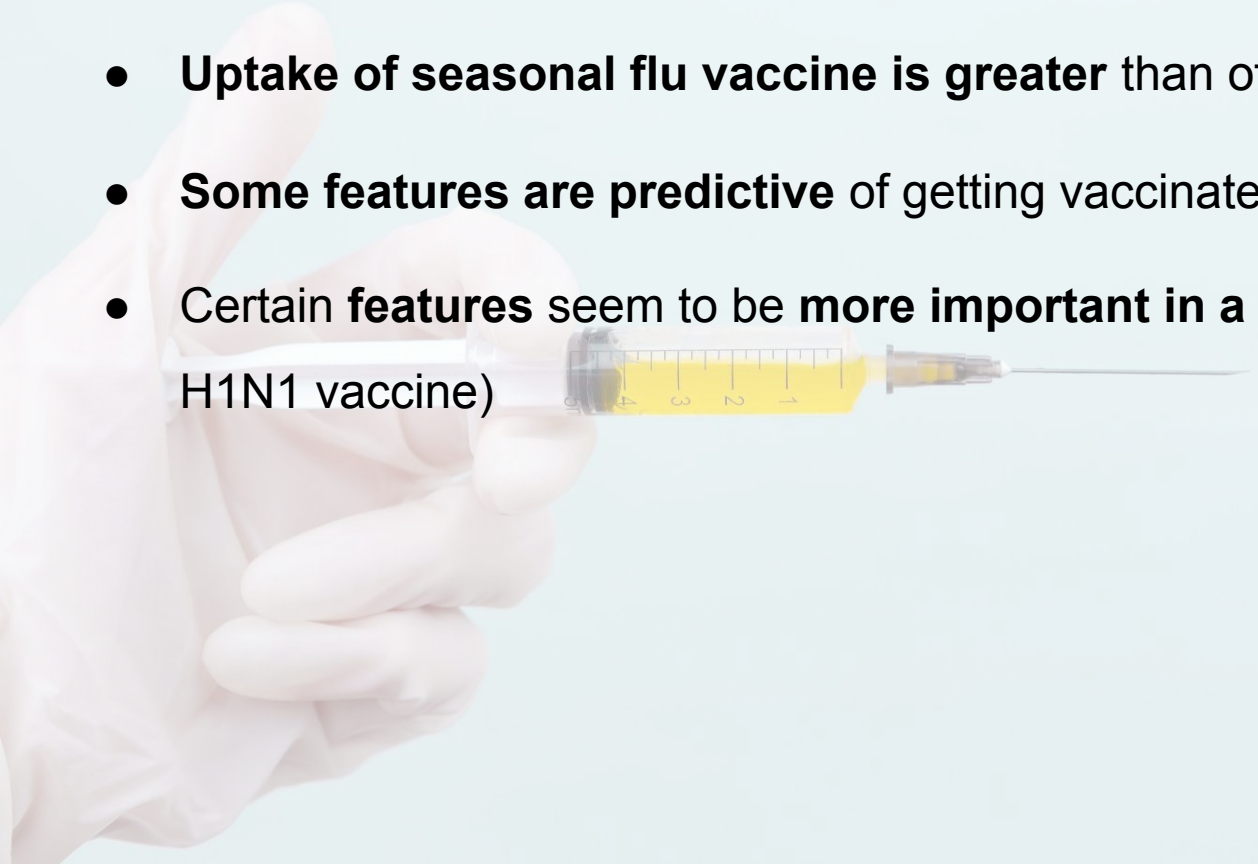
Impact of Health Insurance

Having health insurance coverage has a big positive impact on H1N1 vaccination rates



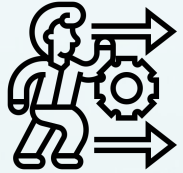
Summary

- **Uptake of seasonal flu vaccine is greater** than of the H1N1 vaccine
- **Some features are predictive** of getting vaccinated
- Certain **features** seem to be **more important in a pandemic context** (for H1N1 vaccine)



Recommendations

- **Raising awareness** of just how contagious each flu is can encourage people to get vaccinated
- **Encourage doctors** to actively recommend the vaccines
- **Educational campaign** to promote the effectiveness of the vaccine
- Additionally, in a pandemic:
    - Education about the **risks of pandemic disease**
    - Ensuring good **health insurance coverage**



Future work and perspectives

- Explore potential **endogenous variables**
- **Cluster analysis** to identify groups with low vaccination rates
- **SHAP analysis**--what features are important on an individual level?
- How do our predictions **translate to disease burden** from influenza?

Dashboard

http://flushot-dashboard.herokuapp.com/

Link

Thank you!

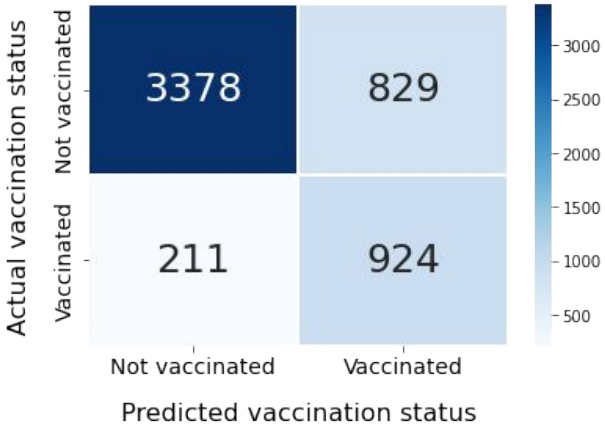
Questions?

Back-up slides

Performance of Best Prediction Model

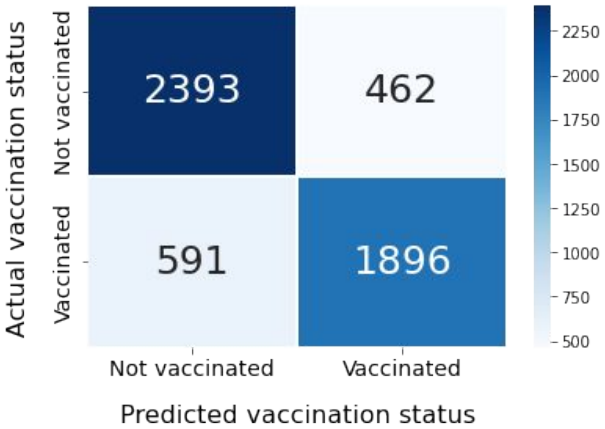
H1N1 flu vaccination prediction

AUC score 0.81



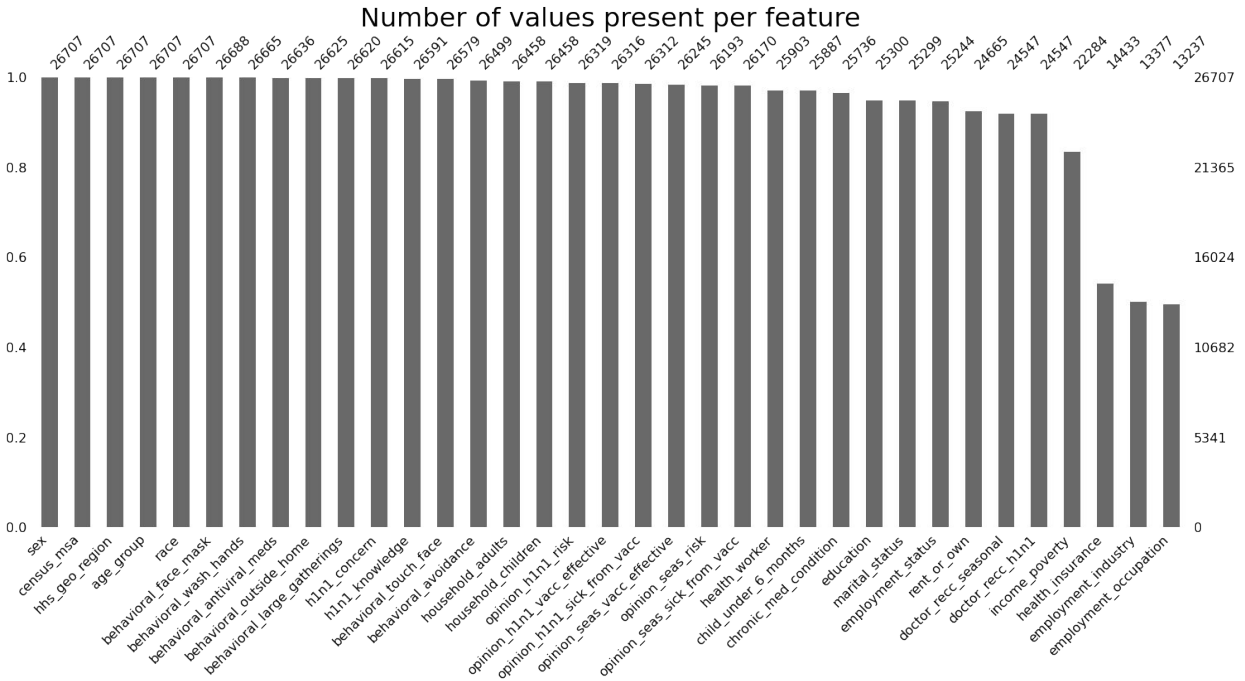
Seasonal flu vaccination prediction

AUC score 0.8



Data completeness and imputation strategy

- Five features have complete data
- Four features have >10% of data missing
- Imputation strategies were compared
- Best outcome: using raw data



Imputation/data cleaning techniques

1. Drop all missing data
    - Dropped columns with too many missing values (‘health_insurance’, ‘employment_industry’, ‘employment_occupation’, ‘income_poverty’, ‘marital_status’, ‘employment_status’)
    - Next, dropped all rows with null values
    - 26707 values → 21853 values
2. No cleaning/imputation
3. Imputation with most frequent values per column (SimpleImputer)
4. Multivariate imputation--KNN (KNNImputer)
5. Multivariate imputation with IterativeImputer (default--ridge regression)
6. Multivariate imputation with IterativeImputer (decision tree classifier)

Imputation/data cleaning techniques--outcomes

Outcome = logistic regression, multilabel prediction

	AUC, H1N1 vaccination	AUC, seasonal flu vaccination
Drop all missing data	0.69	0.78
No cleaning/imputation	0.72	0.78
Imputation with most frequent values per column	0.69	0.78
Multivariate imputation--KNN	0.68	0.78
Multivariate imputation, ridge regression	0.69	0.77
Multivariate imputation, decision tree classifier	0.68	0.78