Seasonal Flu Vaccine: Predictive Model

Caroline Surratt

Phase 3 Project

Flatiron Data Science Flex (June 2023)

Business Understanding

"Immunizations are among the most successful and cost-effective health interventions ever devised."   –World Health Organization

This model can be used to identify features that are the strongest predictors of vaccination status for the seasonal flu.

Healthcare providers can use this knowledge to:

- identify individuals unlikely to get vaccinated
- provide appropriate interventions
- improve overall vaccination rates
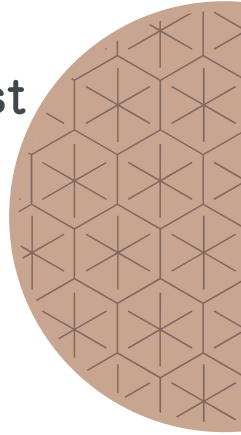
Data Understanding

National 2009 H1N1 Flu Survey

- 26,707 responses

- conducted via telephone

- responses are anonymous

- features include individuals' behavior, demographics, and opinions/knowledge

- target variable indicates vaccination status (binary)

Modeling

Strategy:

- 2 preprocessing strategies were used
- 10 moderately-generic models were created
- Model with the highest overall accuracy was fine-tuned

Evaluation

Primary goal: identify individuals who have not gotten seasonal flu vaccine

False positive: predicting that a person *has not gotten* their seasonal flu vaccine when they *actually have.*

False negative: predicting that a person *has gotten* their seasonal flu vaccine when they *actually have not.*
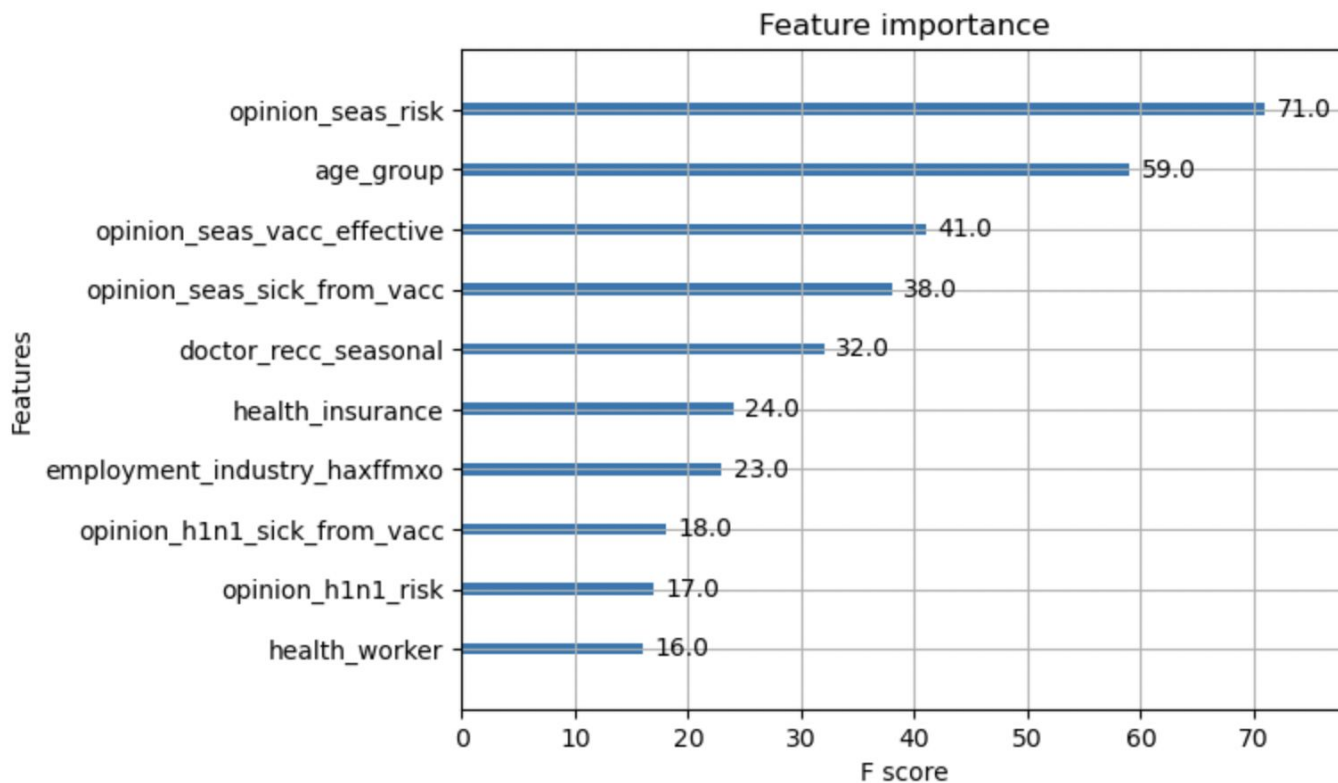
Recall score is more important than accuracy score, as it avoids false negatives.

Evaluation

After tuning, the best model had the following metrics:

- Accuracy: 79.66%
- Recall: 81.84%

Findings

Feature importance

Recommendations

In order to improve vaccination rates, healthcare providers should prioritize impacting the following features:

- Individual's opinion about risk of getting sick with seasonal flu without vaccine
- Individual's opinion about seasonal flu vaccine effectiveness
- Individual's worry of getting sick from taking seasonal flu vaccine
- Recommendation of seasonal flu vaccine by a doctor

Next Steps

- Utilize more recent data (consider COVID implications)

- Consider additional strategies for reducing noise in dataset

Thank you!
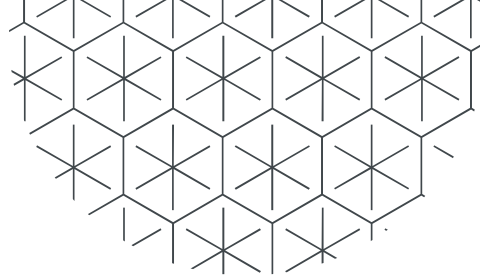
Caroline Surratt

Email: carolinecsurratt@gmail.com

GitHub: @ccsurratt

LinkedIn: linkedin.com/in/carolinesurratt/