Final Project - Data Science Competition

Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines

Zihao Li
University of Wisconsin – Madison
GB 656
12-16-2023

Author Note

ChatGPT and Colab AI were utilized to assist in code correction and composition. Grammar and linguistic precision in this report were also enhanced with the aid of ChatGPT. Beyond these contributions, the entire project was independently conducted, with original writing and analysis performed prior to seeking linguistic refinement to ensure clarity and accuracy of expression.

Report Title

1. Problem Statement
    a. Objective
    b. Use Case for the Vaccine Allocation Model
    c. Value Provided by the Model
    d. Utility to the Government
2. Approach
    a. Data Cleaning
    b. Method Selection
3. Result
4. Reference

Problem Statement:

Objective:

The objective is to utilize training data to develop a model predicting the likelihood of individuals receiving H1N1 and seasonal flu vaccines. This data encompasses personal, familial, and occupational information.

Use Case for the Vaccine Allocation Model:

The government's vaccine allocation model is designed to optimize the distribution efficiency and impact of COVID-19 vaccinations. With a limited vaccine supply, the model strategically prioritizes areas with anticipated higher uptake rates. This strategy aims to deploy vaccines effectively to bolster community immunity and minimize wastage. The analysis includes a thorough review of vaccination data, demographic, and behavioral factors to anticipate vaccine acceptance.

Value Provided by the Model:

This predictive model's significance lies in optimizing COVID-19 vaccine distribution. It assesses individual willingness to vaccinate and calculates regional vaccination rates to inform governmental distribution strategies. This method improves campaign efficacy and promotes fair access to healthcare resources. Furthermore, it enhances cost-efficiency by reducing the likelihood of vaccine expiry.

Utility to the Government:

The model serves multiple government objectives. It ensures optimal use of the vaccine supply chain, supports the commitment to equitable healthcare by guiding resource allocation, and enhances public health education by identifying regions with increased vaccine hesitancy. Aligning vaccine supply with forecasted demand strengthens public confidence, alleviates healthcare burden, and accelerates progress toward herd immunity against COVID-19.

Approach:

1. **Data Cleaning:**
    a. The data cleaning process commenced with the upload of four files, revealing numerous missing values within the dataset. The inclusion of training results facilitated the cleaning process.

    b. The initial strategy entailed removing rows with missing data, which reduced the dataset size from 27,000 to 7,000 entries, negatively impacting the model's performance, as the ROC-AUC score dropped to 0.5029.

    c. To address this issue, the strategy was adjusted to impute missing values by random sampling from the same column. This method, developed with the assistance of ChatGPT and Colab AI, improved the ROC-AUC score to 0.6114, signifying an enhanced predictive model.

2. **Method Selection:**
    a. The analysis commenced with the selection of an appropriate prediction method for the H1N1 and seasonal flu vaccination rates. It was determined to model both rates independently using consistent methods to improve accuracy.

    b. Backward selection was employed to identify the optimal model for each vaccine. The selected features were then used in the prediction models.

    c. The Gradient Boosting Model and Tree classifier were applied to the H1N1 dataset to ascertain which would yield a higher score. The Gradient Boosting Model proved superior and was thus chosen for the H1N1 test data.

    d. This approach was replicated for the seasonal flu vaccine, with the Gradient Boosting Model once again demonstrating better performance, leading to its selection for the seasonal flu test data.

    e. The culmination of this process resulted in the successful prediction of vaccination rates for both vaccines.

Result:

The results were compiled into the concluding document, which was then downloaded and submitted to the competition's website, yielding the scores and outcomes detailed therein.

Partial results:

respondent_id	h1n1_vaccine	seasonal_vaccine
26707	0.356	0.342
26708	0.377	0.295
26709	0.301	0.673

Reference:

1. DrivenData. (n.d.). Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines. Retrieved from https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/
2. https://colab.research.google.com/drive/1dpTbTKD2b5J9lYyIgxoDtpJgLxG0DG3P?usp=sharing