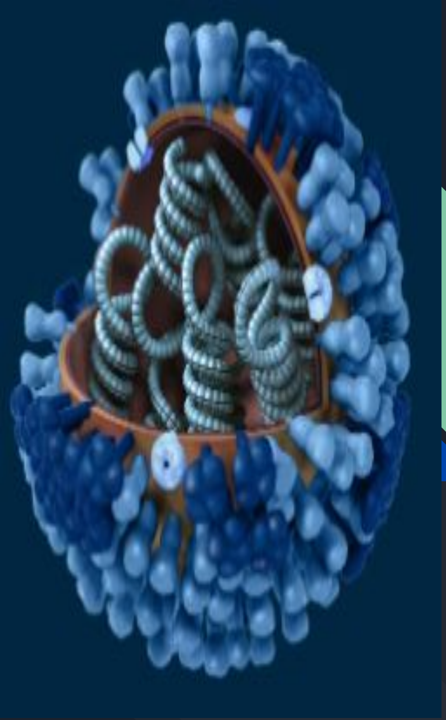
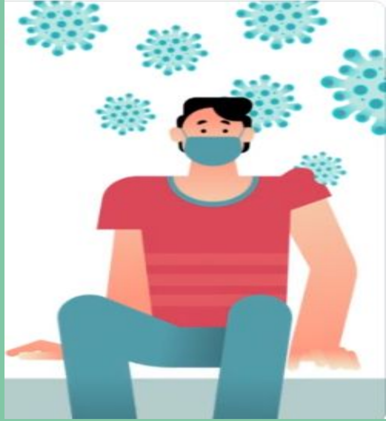


PREDICTION OF FLU SEASONAL VACCINE UPTAKE

By Vitelis Siocha

Introduction



Each year, the recurring flu season compels individual to weigh the decision of whether to receive the flu seasonal flu vaccine.

The goal is to build a predictive model that anticipates the people's choice in flu vaccination each year, utilizing historical data to uncover influential trends and patterns

Business Statement

The primary purpose of the predictive model is to aid in public health planning decisions and resource allocation by government health agencies to combat or reduce flu outbreaks that can overrun health sector.

Objectives

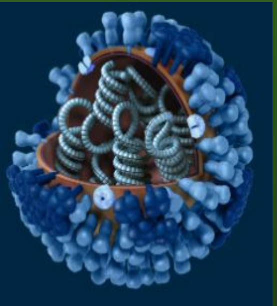
1. To examine the factors influencing individual’s choices to get the annual flu seasonal vaccine.

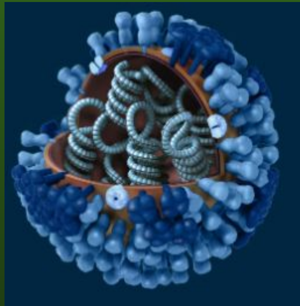
2. To come up with strategies for targeting specific population segments to boost vaccination rates.

3. To explore variables like age categories, gender, and household composition to comprehend statistical significance.

Main Objectives:

1. To construct a predictive model for predicting whether one would opt for seasonal vaccine under specific conditions
2. To offer recommendations on increasing vaccination rates to avoid Health Care overruns by seasonal influenza.





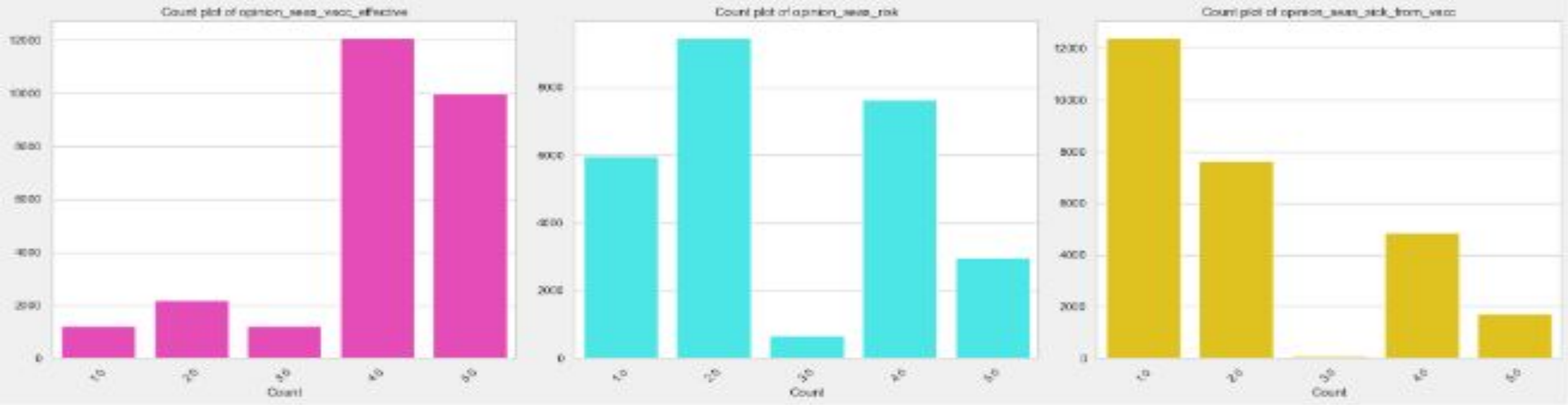
Data

The flu seasonal predictive mode was constructed using H1N1_Flu_Vaccines.csv dataset sourced from Kaggle https://www.kaggle.com/datasets/arashnic/flu-data.

The dataset consisted of 26,707 rows and 38 columns, with 36 columns for predictive features and 2 columns for targets. One target variable (namely, h1n1_vaccine) was dropped and therefore was not used in building predictive model.

Exploratory Data Analysis(EDA)

Distribution of individuals' opinions in terms of effectiveness, risks and falling sick after taking flu seasonal vaccine



1 = Not at all worried, 2 = Not very worried, 3 = Don't know, 4 = Somewhat worried, 5 = Very worried

The data suggests that majority generally consider seasonal vaccines to be somewhat effective, with few believing them to be ineffective. Concerns about vaccination risks are low, with more moderate apprehension for the seasonal vaccine. Many respondents show little fear of getting sick after either vaccine. Overall, vaccine perceptions are positive, and worries about risks and sickness are limited

Modeling

Classifier Models Used Included:

1. Decision Tree Classifier Model (Baseline Model)
2. Logistic Regression Classifier Model
3. Gradient Boosting Classifier Model
4. Extreme Gradient Boosting (XGBoost) Classifier Model
5. Random Forest Classifier Model

Data was divided into train and test sets. Each model was trained on the training set and tested on test set which served as unseen data the model is trying to predict.

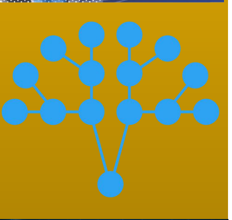
Each classifier model was trained twice without and with optimized hyperparameters to get the best performing model.

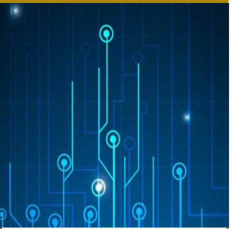
Evaluation of Models

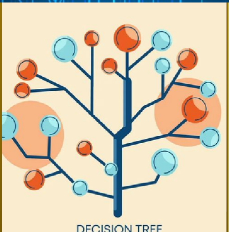
Bases used to evaluate each model:

1. Prediction accuracy on the training and testing/validation sets.
2. Area under the curve using Receiver Operating Characteristic (ROC) curves.
3. Consistency in prediction accuracy of both training and validation sets to avoid underfitting/overfitting.

Modeling Results







THE BEST MODEL:
OPTIMIZED GRADIENT BOOSTING

Closely analyzing the modeling results, optimized Gradient Boosting model emerges as the best choice, offering high accuracy, robust generalization, and strong flu vaccine likelihood predictions, making it the top candidate among the models.

It achieves prediction accuracy of 78.1% on train set, 77.5% on test set, cross-validation of 77.33% and has highest ability of likelihood prediction of 85%. It is the best in most of the bases/metrics of evaluation.

Conclusions

1. Understanding Factors Influencing Flu Vaccine Choices: The analysis has provided valuable insights into the factors that influence individuals' choices to receive the annual flu vaccine.
2. Targeting Specific Population Segments: The analysis offers an opportunity to develop targeted strategies for improving vaccination rates.
3. Exploring Statistical Significance: The analysis explores the statistical significance of variables.
4. Constructing Predictive Models: The project has successfully constructed predictive models to assess whether people are likely to opt for the seasonal flu vaccine under specific conditions or features.
5. The analysis has not only achieved its objectives but also paved the way for evidence-based strategies to government public agencies to enhance flu vaccination rates and tailor interventions to specific population segments.
6. Recommendations have been offered on how to increase flu seasons vaccine uptake.

Limitations

1. The dataset used in the analysis was biased and not representative of a wide demographic. This may lead to poor generalization of the predictive model.
2. The model performance left out the time metric required to train various models and therefore training time of each model was not evaluated.
3. Some population segments have lower vaccination rates due to a lack of awareness about the benefits of the Seasonal Flu Vaccine. This affects model generalization.

Recommendations

1. **Diverse Data Collection:** To mitigate bias, collect data that more accurately represents a wider demographic and geographic range. Ensure data collection from a more diverse set of communities and population groups
2. **Awareness Campaigns:** Develop targeted awareness campaigns to educate people about the benefits of the Seasonal Flu Vaccine, particularly focusing on population segments with lower vaccination rates.
3. **Improve Data Balance**: Work on improving the balance of the dataset to ensure that models generalize well to a broader population. Techniques like oversampling or gathering more data from underrepresented groups may help.
4. **Continual Monitoring**: Maintain continuous monitoring of vaccination rates and effectiveness. This will help adapt strategies in real-time and ensure the most relevant actions are taken and change the opinions of vaccine risks and illness perception among some individuals.
5. **Model Refinement:** To improve predictive models, it is essential to focus on optimizing techniques, particularly for the Gradient Boosting model. Further fine-tuning its parameters has the potential to enhance its performance. Additionally, a thorough evaluation of the time required to train various models is crucial to ensure efficiency and streamline the modeling process.

THANK YOU