

SEASONAL FLU- VACCINE UPTAKE ANALYSIS

Sep 1, 2024



Business Problem

The primary business problem is to predict the likelihood that individuals will receive the seasonal flu vaccines in the **United States** based on their backgrounds, opinions, and health-related behaviors. By understanding the factors that influence vaccine uptake, we can help the public health organizations and healthcare providers to design more effective vaccination campaigns and help policy makers allocate resources more efficiently. This will ultimately improve vaccination rates thus reducing the spread of these viruses. |

Project Objectives



Which factors most influence the decision to get vaccinated?



How can public health interventions be targeted more effectively?



Can we accurately predict who is likely to get vaccinated?



How can we help improve the seasonal vaccine uptake by leveraging the analytical insights and features selected through this project.

Data Understanding

This dataset contains information about individual backgrounds, opinions, and health behaviors in the USA

. It includes data collected during spread of the National 2009 H1N1 and seasonal flu virus outbreak survey.

- The dataset contains 36 columns and 26707 rows. This means there are 36 different variables each with 26707 records.
- There are three main data types in the data; float, integer and object.

Data Cleaning



Imputed missing values.



Created a subset of the main dataset consisting features to use during the analysis.

- Imputed the categorical columns with mode
- Imputed numerical columns with mean
- Handled Outliers using log transformation.
- Encoded the data to deal with categorical columns

MODELLING

I started off with logistic regression as my first model and then gradually improved the **Decision Trees Model**

- ✓ Scaling
- ✓ Encoding
- ✓ Feature Selection
- ✓ Using random classifier to re-train the model

Decision Trees Model

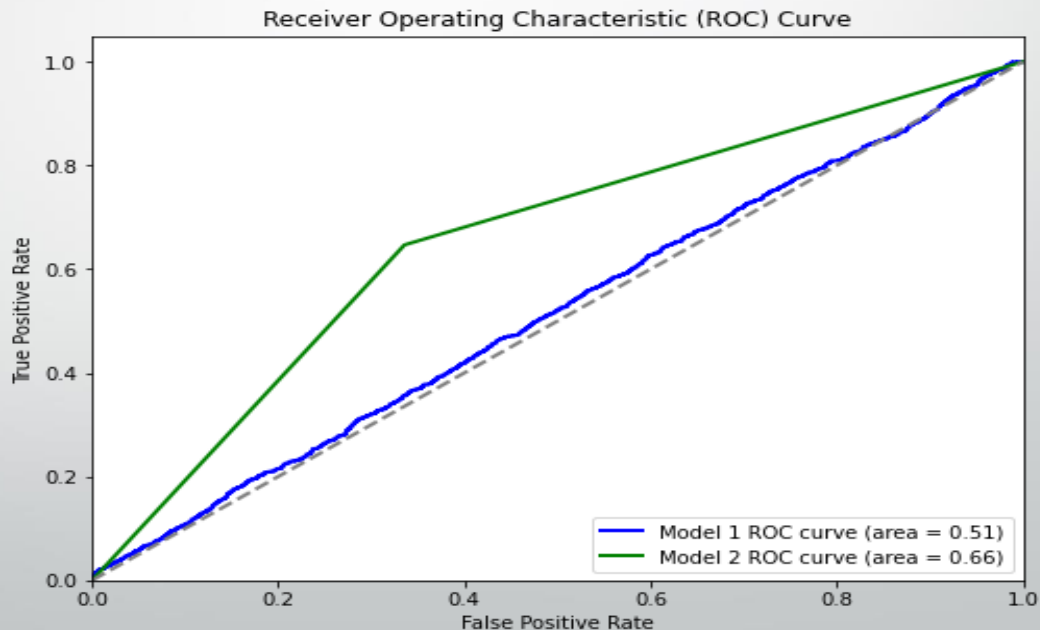
Become our best performing model

This is seen in the performance of the model through metrics such as:

Model Summary:

ROC AUC curve– 0.66

Confusion matrix

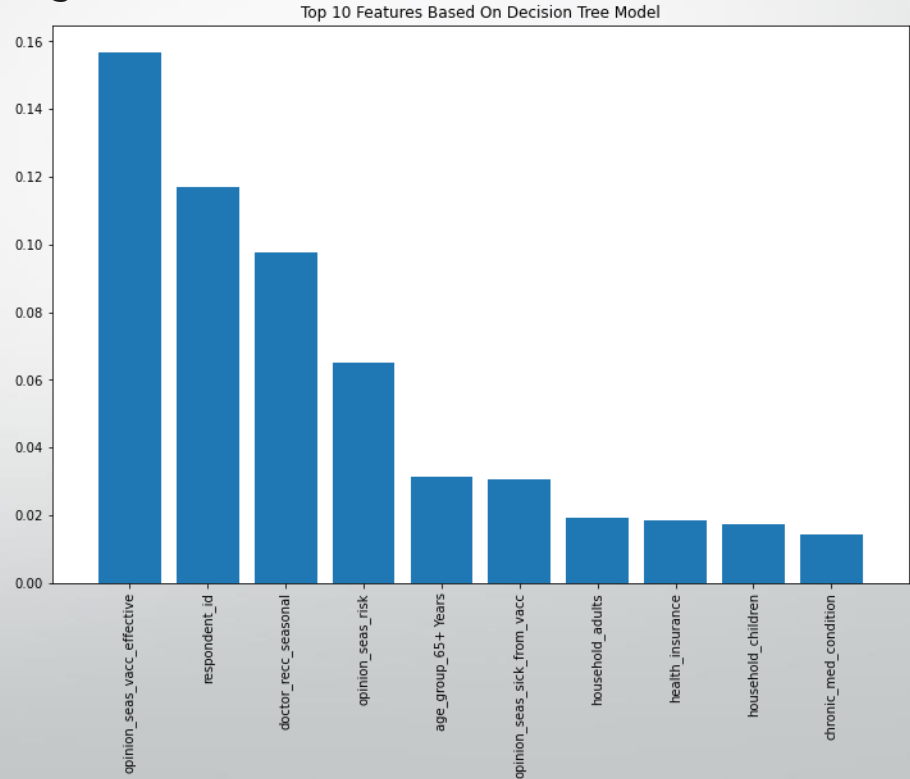


Final Model

Model Improvement:

- Hyper parameter tuning.
- Dropping statistically insignificant variables.

Model Summary
*Top-10 Important
Features used*



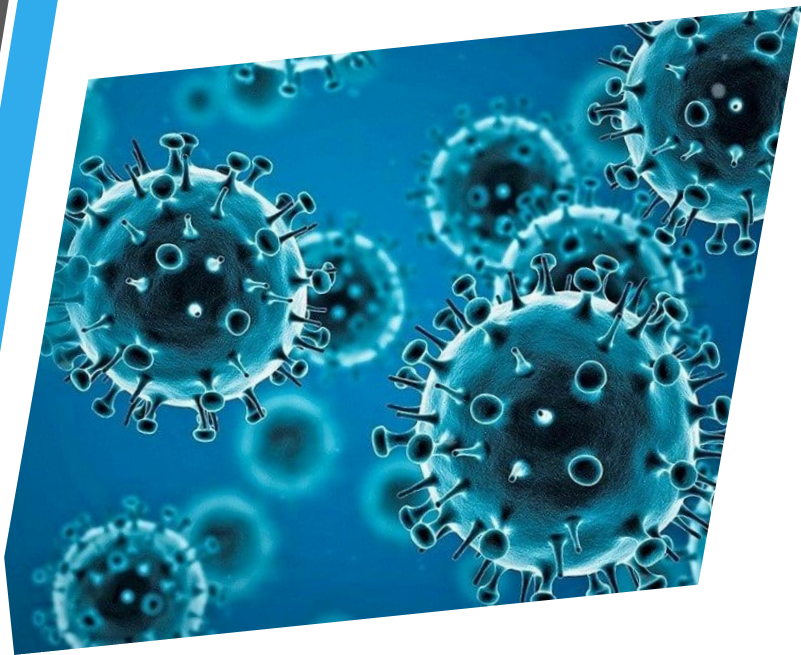
Conclusion



The model identifies key factors influencing house prices and explains 66% of the vaccine uptake variation with a high degree of statistical significance based on our ROC curve



Important factors affecting vaccine uptake include; opinion, doctors recommendations, and age-group above 65 years



Recommendations

- ▶ Focus on key features analyzed when setting health policies.
- ▶ Utilize the decision tree model to generate an optimal advertising/incentive strategy.
- ▶ Compare predicted and actual vaccine uptake for proper planning
- ▶ Leverage analytical insights derived to enhance decision-making.



Next Steps

The background features a dark gray bar chart with bars of varying heights, labeled Q1, Q2, and Q3. Overlaid on this is a large, semi-transparent circular gauge with a blue and white border. The gauge has labels Q2, Q3, and Q4 along its bottom arc, and a numerical label 1,000 near the top right. The overall aesthetic is professional and data-oriented.

- Analyze temporal dynamics to capture seasonal trends and long-term fluctuations.
- Collaborate with policy makers, health organizations, and healthcare providers to validate the analysis and gain deeper understanding of population behavior towards vaccination.
- Refine the model: Explore more regression techniques to improve our model accuracy.
- Stay updated and iterate continuously to track new data and industry trends to ensure the model remains efficient.