**PREDICTING THE PROBABILITY THAT A PERSON WILL RECEIVE THE H1N1 AND SEASONAL FLU VACCINES USING MACHINE LEARNING MODELS**

**BUSINESS UNDERSTANDING**

In 2009, the H1N1 influenza pandemic, commonly referred to as "swine flu," spread across the world, resulting in an estimated 151,000 to 575,000 deaths globally in its first year. To combat the virus, a vaccine was made publicly available in October 2009. To gain insight into vaccination patterns and behavior, the United States conducted the National 2009 H1N1 Flu Survey, which was a phone survey that asked participants about their H1N1 and seasonal flu vaccine status, as well as their social, economic, and demographic information, opinions on illness and vaccine risks, and behaviors for reducing transmission. By gaining a better understanding of the relationships between these factors and vaccination patterns, this information can be used to guide future public health initiatives.

**Main Objective**

The goal is to predict how likely individuals are to receive their H1N1 and seasonal flu vaccines.

# **The metric of success**

The performance will be evaluated according to the area under the receiver operating characteristic curve (ROC AUC) for each of the two target variables.

**Data Understanding**

**Data Source**

Data that was used was downloaded from [Driven Data](https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/)

**Data Description**

The information in the data is as follows:

* h1n1\_concern - Level of concern about the H1N1 flu.
* h1n1\_knowledge - Level of knowledge about H1N1 flu.
* behavioral\_antiviral\_meds - Has taken antiviral medications.
* behavioral\_avoidance - Has avoided close contact with others with flu-like symptoms.
* behavioral\_face\_mask - Has bought a face mask.
* behavioral\_wash\_hands - Has frequently washed hands or used hand sanitizer.
* behavioral\_large\_gatherings - Has reduced time at large gatherings.
* behavioral\_outside\_home - Has reduced contact with people outside of own household.
* behavioral\_touch\_face - Has avoided touching eyes, nose, or mouth.
* doctor\_recc\_h1n1 - H1N1 flu vaccine was recommended by doctor.
* doctor\_recc\_seasonal - Seasonal flu vaccine was recommended by doctor.
* chronic\_med\_condition - Has any of the following chronic medical conditions: asthma or an other lung condition, diabetes, a heart condition, a kidney condition, sickle cell anemia or other anemia, a neurological or neuromuscular condition, a liver condition, or a weakened immune system caused by a chronic illness or by medicines taken for a chronic illness.
* child\_under\_6\_months - Has regular close contact with a child under the age of six months.
* health\_worker - Is a healthcare worker.
* health\_insurance - Has health insurance.
* opinion\_h1n1\_vacc\_effective - Respondent's opinion about H1N1 vaccine effectiveness.
* opinion\_h1n1\_risk - Respondent's opinion about risk of getting sick with H1N1 flu without vaccine.
* opinion\_h1n1\_sick\_from\_vacc - Respondent's worry of getting sick from taking H1N1 vaccine.
* opinion\_seas\_vacc\_effective - Respondent's opinion about seasonal flu vaccine effectiveness.
* opinion\_seas\_risk - Respondent's opinion about risk of getting sick with seasonal flu without vaccine.
* opinion\_seas\_sick\_from\_vacc - Respondent's worry of getting sick from taking seasonal flu vaccine.
* age\_group - Age group of respondent.
* education - Self-reported education level.
* race - Race of respondent.
* sex - Sex of respondent.
* income\_poverty - Household annual income of respondent with respect to 2008 Census poverty thresholds.
* marital\_status - Marital status of respondent.
* rent\_or\_own - Housing situation of respondent.
* employment\_status - Employment status of respondent.
* hhs\_geo\_region - Respondent's residence using a 10-region geographic classification defined by the U.S. Dept. of Health and Human Services. Values are represented as short random character strings.
* census\_msa - Respondent's residence within metropolitan statistical areas (MSA) as defined by the U.S. Census.
* household\_adults - Number of other adults in household, top-coded to 3.
* household\_children - Number of children in household, top-coded to 3.
* employment\_industry - Type of industry respondent is employed in. Values are represented as short random character strings.
* employment\_occupation - Type of occupation of respondent. Values are represented as short random character strings.

# **Data Preparation**

## **Loading Libraries**

Load the libraries necessary for cleaning and analysis

## **Loading the Data**

Load the dataset from the CSV file. The names of the CSV files are: “test\_set\_features.csv, “training\_set\_features.csv”, “training\_set\_labels.csv” and “submission\_format.csv”

The shape of the dataset is: training set labels(26707rows,3 columns), training set features (26707 rows, 36 columns), and test set features (26707 rows, 36 columns).

## **Cleaning the Data**

Data cleaning is the process of preparing data for analysis by removing irrelevant or incorrect information. This type of information usually reinforces a false belief, which can have a negative impact on the model or algorithm into which it is fed.

The steps for the data cleaning process are:

1. Ensure that there are no duplicates in the data.
2. Ensure that the dataset data types are correct.
3. Ensure that there are no missing values in the dataset.

# **Modelling**

To predict the likelihood of respondents in the survey taking the h1n1 vaccine and seasonal vaccine, machine learning algorithms were used.

The ultimate accuracy was determined using logistic regression, decision trees, and XGBoost.

# 

# **Conclusion and Recommendation**

## **Conclusion**

1. People who received the seasonal vaccine outperformed those who received the H1N1 vaccine.
2. Women received both immunizations at a higher rate than males.
3. White folks received both immunizations at a higher rate than other races.
4. People with health insurance chose seasonal vaccines over H1N1 vaccines.
5. More People chose the seasonal vaccine over the H1N1 vaccine based on doctor’s recommendations.
6. There were fewer persons seeking both immunizations than predicted.
7. In addition, children were not as well immunized as planned.

## Recommendations

Based on the conclusion drawn from the data, here are a few recommendations that could be made:

1. Promote the importance of the seasonal flu vaccine over the H1N1 vaccine, highlighting its higher efficacy and safety record.
2. Encourage greater immunization rates among males and people of other races, perhaps through targeted outreach and education initiatives.
3. Address the issue of access to health insurance, as it appears to impact vaccine choice.
4. Encourage more people to seek recommendations from their healthcare providers, as this seems to be a significant factor in vaccine choice.
5. Address the low immunization rate among children by increasing awareness of the importance of protecting this vulnerable population and providing easier access to vaccines for families with young children.
6. Consider additional strategies to increase overall immunization rates, such as offering incentives or increasing accessibility to vaccines.
7. Further analysis and research may be needed to understand and address the underlying issues more effectively.