



The United States National 2009 H1N1 Flu Survey, conducted in late 2009 and early 2010, collected data on vaccination uptake for both H1N1 and seasonal flu. The survey explored respondents' vaccination status alongside information about their socioeconomic and demographic backgrounds, health behaviors, and opinions on vaccine efficacy and illness risk. This project therefore seeks to guide future public health strategies by providing valuable insights into the factors influencing the vaccination patterns through a thorough analysis of the dataset provided.

## Problem Statement

---

Understanding the factors that influence vaccination decisions is essential for designing effective public health campaigns. Despite the availability of vaccines for the H1N1 and seasonal flu, uptake rates varied significantly among different population groups. The challenge lies in identifying and analyzing the individual and social factors that determine vaccine adoption, which is critical for improving vaccination rates and achieving herd immunity during pandemics.

## General Objective

---

To understand which factors influence the uptake of the seasonal flu vaccine.

## Specific Objectives

---

1. To analyze the relationship between demographic factors, opinions on vaccine effectiveness, and health behaviors in influencing vaccination status.
2. To identify key socioeconomic and behavioral barriers to accessing vaccines.
3. To develop predictive models for determining the likelihood of individuals receiving H1N1 and seasonal flu vaccines.

## Research Questions

---

1. What are the key demographic, behavioral, and opinion-based factors influencing vaccination decisions?
2. What socioeconomic barriers hinder vaccine accessibility and uptake?
3. How accurately can predictive models identify individuals likely to receive or forgo vaccination?

## Data Understanding

---

The data is composed of approximately 26,000 instances of individual data and vaccine decision information and the files obtained from DRIVEN DATA (Source: CDC, NCRID and NCHS (2012), National 2009 H1N1 Flu Survey). This data was collected over the phone between late 2009 and June 2010. Furthermore, the Cross Industry Standard Process for Data Mining (CRISP-DM) was used for the analysis of data.

<https://www.datascience-pm.com/crisp-dm-2/>.

More data exploration revealed the following:

The train feature dataset contains 26707 rows and 36 columns; The test dataset contains 26708 rows and 36 columns; The target dataset contains 26707 rows and 3 columns. The data was then cleaned and some columns were removed as needed. To investigate the factors influencing individuals' decisions to receive the H1N1 and seasonal flu vaccines using data on their backgrounds, health behaviors, and opinions, with the goal of informing public health strategies to enhance vaccine uptake.

## Modelling

There are four main models that were used in this project:

Logistic Regression; Decision trees; Random forests; XGBoost Algorithm.

## Model Evaluation Summary

The Random Forest Model (best\_model3) has demonstrated commendable performance in predicting the uptake of the seasonal flu vaccine. It has strong evaluation metrics and an ROC curve with strong discriminatory power. A better balance on the data may give more promising results. ![all models roc curve] (images copy/ROC\_curve.png)

## Recommendations

Embrace personalized outreach as a campaign tool so as to target individuals and mould their perception towards immunization. Public campaigns should be geared towards bringing onboard more younger people as it seems that they are less likely to get the seasonal flu vaccines. The public health sector should continue encouraging doctors to recommend suitable vaccines to their clients. This modelling and analysis process has shown that people are highly likely to listen to their doctor's advice. For further improvements:



### Releases

No releases published

[Create a new release](#)

### Packages

No packages published

[Publish your first package](#)

### Languages

● Jupyter Notebook 100.0%