

# Vaccine Usage Prediction

## Abstract:

Subjects receiving the same vaccine often show different levels of immune responses and some may even present adverse side effects to the vaccine. Systems vaccinology can combine omics data and machine learning techniques to obtain highly predictive signatures of vaccine immunogenicity and reactogenicity. Currently, several machine learning methods are already available to researchers with no background in bioinformatics.

## Problem Statement:

Predict how likely it is that the people will take an H1N1 flu vaccine using Logistic Regression.

## Dataset Information:

The data is about the technical specifications of cars.

Column	Description
unique_id	Unique identifier for each respondent
h1n1_worry	Worry about the h1n1 flu(0,1,2,3) 0=Not worried at all, 1=Not very worried, 2=Somewhat worried, 3=Very worried
h1n1_awareness	Signifies the amount of knowledge or understanding the respondent has about h1n1 flu - (0,1,2) - 0=No knowledge, 1=little knowledge, 2=good knowledge
antiviral_medication	Has the respondent taken antiviral vaccination - (0,1)
contact_avoidance	Has avoided any close contact with people who have flu-like symptoms - (0,1)
bought_face_mask	Has the respondent bought mask or not - (0,1)
wash_hands_frequently	Washes hands frequently or uses hand sanitizer - (0,1)

## PG Program in Analytics

### Problem Statement – Logistic Regression

avoid_large_gatherings	Has the respondent reduced time spent at large gatherings - (0,1)
reduced_outside_home_cont	Has the respondent reduced contact with people outside their own house - (0,1)
avoid_touch_face	Avoids touching nose, eyes, mouth - (0,1)
dr_recc_h1n1_vacc	Doctor has recommended h1n1 vaccine - (0,1)
dr_recc_seasonal_vacc	The doctor has recommended seasonal flu vaccine - (0,1)
chronic_medical_condition	Has any chronic medical condition - (0,1)
cont_child_undr_6_mnth	Has regular contact with child the age of 6 months - (0,1)
is_health_worker	Is respondent a health worker - (0,1)
has_health_insur	Does respondent have health insurance - (0,1)
is_h1n1_vacc_effective	Does respondent think that the h1n1 vaccine is effective - (1,2,3,4,5)- (1=Thinks not effective at all, 2=Thinks it is not very effective, 3=Doesn't know if it is effective or not, 4=Thinks it is somewhat effective, 5=Thinks it is highly effective)
is_h1n1_risky	What respondents think about the risk of getting ill with h1n1 in the absence of the vaccine- (1,2,3,4,5)- (1=Thinks it is not very low risk, 2=Thinks it is somewhat low risk, 3=don't know if it is risky or not, 4=Thinks it is a somewhat high risk, 5=Thinks it is very highly risky)
sick_from_h1n1_vacc	Does respondent worry about getting sick by taking the h1n1 vaccine - (1,2,3,4,5)- (1=Respondent not worried at all, 2=Respondent is not very worried, 3=Doesn't know, 4=Respondent is somewhat worried, 5=Respondent is very worried) -
is_seas_vacc_effective	Does respondent think that the seasonal vaccine is effective- (1,2,3,4,5)- (1=Thinks not effective at all, 2=Thinks it is not very effective, 3=Doesn't know if it

## PG Program in Analytics

### Problem Statement – Logistic Regression

	is effective or not, 4=Thinks it is somewhat effective, 5=Thinks it is highly effective)
is_seas_flu_risky	What respondenst think about the risk of getting ill with seasonal flu in the absence of the vaccine- (1,2,3,4,5)- (1=Thinks it is not very low risk, 2=Thinks it is somewhat low risk, 3=Doesn't know if it is risky or not, 4=Thinks it is somewhat high risk, 5=Thinks it is very highly risky)
sick_from_seas_vacc	Does respondent worry about getting sick by taking the seasonal flu vaccine - (1,2,3,4,5)- (1=Respondent not worried at all, 2=Respondent is not very worried, 3=Doesn't know, 4=Respondent is somewhat worried, 5Respondent is very worried)
age_bracket	Age bracket of the respondent - (18 - 34 Years, 35 - 44 Years, 45 - 54 Years, 55 - 64 Years, 64+ Years)
qualification	Qualification/education level of the respondent as per their response -(<12 Years, 12 Years, College Graduate, Some College)
race	Respondent's race - (White, Black, Other or Multiple ,Hispanic)
sex	Respondent's sex - (Female, Male)
income_level	Annual income of the respondent as per the 2008 poverty Census - (<= 75000–AbovePoverty,> 75000–AbovePoverty,>75000, Below Poverty)
marital_status	Respondent's marital status - (Not Married, Married)
housing_status	Respondent's housing status - (Own, Rent)
employment	Respondent's employment status - (Not in Labor Force, Employed, Unemployed)
census_msa	Residence of the respondent with the MSA(metropolitan statistical area)(Non-MSA, MSA-Not Principle, CityMSA-Principle city) - (Yes, no)

# PG Program in Analytics

## Problem Statement – Logistic Regression

no_of_adults	Number of adults in the respondent's house (0,1,2,3) - (Yes, no)
no_of_children	Number of children in the respondent's house(0,1,2,3) - (Yes, No)
h1n1_vaccine	Dependent variable)Did the respondent received the h1n1 vaccine or not(1,0) - (Yes, No)

### Scope:

- Exploratory data analysis
- Data Pre-processing
- Training logistic regression model with MLE and SGD for prediction

### Learning Outcome:

The students will get a better understanding of how the variables are linked to each other and how the EDA approach will help them gain more insights and knowledge about the data that we have and train Logistic Regression using with MLE and SGD.