

|   |   |
|---|---|
| <b>Project Title</b>                      | <b>Vaccine Usage analysis and prediction</b>                            |
| <b>Skills take away From This Project</b> | <b>Python scripting, Pandas,Data Visualisation and Machine Learning</b> |
| <b>Domain</b>                             | <b>Data analysis and ML</b>   |

### **Problem Statement:**

This project aims to predict the likelihood of people taking an H1N1 flu vaccine using Logistic Regression. It involves analyzing a dataset containing various features related to individuals' behaviors, perceptions, and demographics, and building a predictive model to determine vaccine acceptance.

Predict the probability of individuals taking an H1N1 flu vaccine based on their characteristics and attitudes. This can help healthcare professionals and policymakers target vaccination campaigns more effectively.

### **Approach:**

You'll be able to access the **Vaccine dataset** from the below URL.

<https://raw.githubusercontent.com/nethajinirmal13/Training-datasets/main/Vaccine.csv>

### Dataset Information:

| 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)  |

|                           |  |
|---------------------------|--|
| 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, 5Respondent 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   |

|                     |  |
|---------------------|--|
|                     | is effective or not, 4=Thinks it is somewhat effective, 5=Thinks it is highly effective)   |
| is_seas_flu_risky   | What respondents 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)   |

|                |  |
|----------------|--|
| 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 receive the h1n1 vaccine or not(1,0) - (Yes, No) |

### **TASK 1: DATA ENGINEERING**

- 1.Understand the complete dataset and Engineer it as per the needs
- 2.Do Exploratory data analysis to get deeper understanding about the attributes.

### **TASK 2 : DASHBOARD DEVELOPMENT**

The Plotly or streamlit analytical Dashboard is expected to be developed that has at least 5 types of dynamic charts and at least 3 data filtering options.

### **TASK 3 : MODEL DEVELOPMENT**

1. Training a Logistic Regression model (or any suitable model)for prediction.
2. Model tuning and evaluation to improve predictive performance.
3. Developing a predictive model that can help in targeting vaccination campaigns effectively.
4. Create a model serving api once the model is ready.
5. Develop an application using flask or plotly or streamlit that can accept features as input and provide prediction to the end user.
6. Host the application using any free hosting service provide (example: render.com, pythonanywhere,etc).Refer to the following article for hosting.
  - a. <https://medium.com/@ahossack07/create-and-deploy-plotly-dash-apps-to-the-internet-for-free-49ebca9633da>
  - b. [https://www.youtube.com/watch?v=H16dZMYmvqo&ab\\_channel=Plotly](https://www.youtube.com/watch?v=H16dZMYmvqo&ab_channel=Plotly)

## Project Evaluation metrics:

- You are supposed to write code in a modular fashion (**in functional blocks**)
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on **GitHub**. (Mandatory)
- You have to keep your **GitHub** repo public so that anyone can check your code. (Mandatory)
- Proper readme file you have to maintain for any project development (Mandatory)
- You should include basic workflow and execution of the entire project in the readme file on **GitHub**
- Follow the coding standards: <https://www.python.org/dev/peps/pep-0008/>
- You need to Create a Demo video of your working model and post in **LinkedIn** (Mandatory)

