Predicting H1N1 and Seasonal Flu Vaccine Uptake

A Data-Driven Approach Using Survey-Based Modeling

ESTERINA KANANU

Project Overview

- **Objective:** Predict likelihood of H1N1 and Seasonal Flu vaccine uptake.
- **Motivation:** Improve public health interventions through targeted messaging.
- **Dataset:** National 2009 H1N1 Flu Survey (NHFS).

Business Problem & Stakeholders

- **Problem:** Vaccine hesitancy weakens pandemic response.
- **Goal:** Identify drivers of vaccine uptake for better outreach
- **Stakeholders:**
  - Public health officials (e.g. CDC, WHO)
  - Policy makers
  - Health communication teams

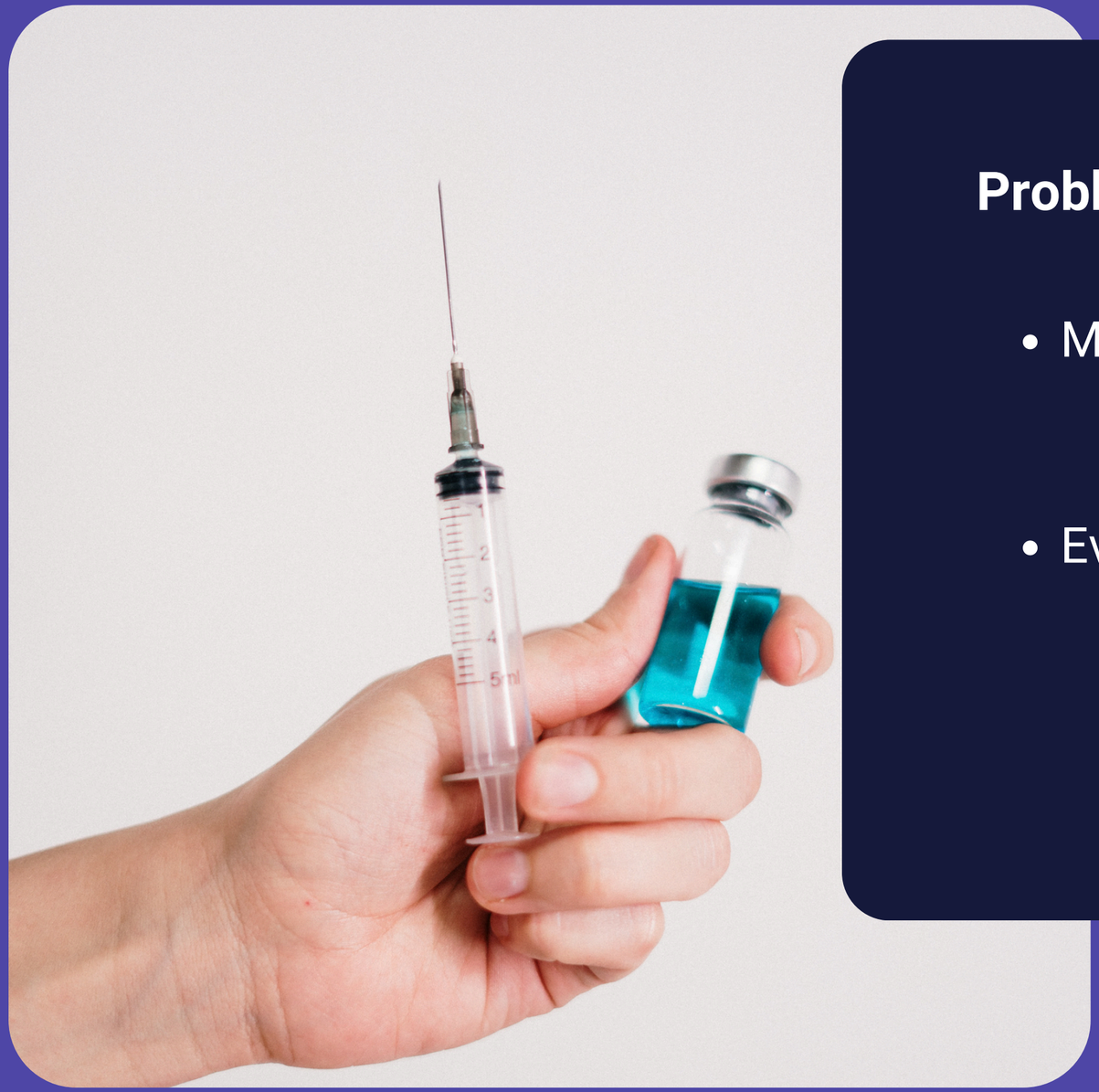
Data and Features

- Source: DrivenData Flu Shot Learning competition.
- Key Features:
  - Demographics: Age, sex, income, education
  - Behavior: doctor visits, mask use
  - Opinions: risk perception, vaccine effectiveness
  - Targets: h1n1_vaccine, seasonal_vaccine (binary)

DATA PREPARATION

- **Handled missing values:**
  - Dropped high-missing columns
  - Categorical → 'Unknown'
  - Numeric → median
- **Encoding:**
  - Label + One-hot encoding
- **Scaled numeric features with StandardScaler**

MODELING APPROACH



Problem Type: Multi-label classification

- Models Used:
  - Logistic Regression (Baseline)
  - Random Forest (Advanced)
- Evaluation Metrics:
  - Accuracy, F1-Score, AUC (ROC Curve)

Model Comparison

Accuracy Comparison

Vaccine Type	Logistic Regression	**Random Forest**
H1N1	84.6%	85.1%
Seasonal	78.2%	78.4%

AUC Score

Vaccine Type	**Logistic Regression**	Random Forest
H1N1	0.8451	0.8576
Seasonal	0.8503	0.8528

- Random Forest slightly outperforms Logistic Regression, especially for H1N1.
- AUC scores improved marginally → better discrimination of vaccinated vs. non-vaccinated.
- H1N1 Classification:
- → Logistic Regression recall: 0.45
- → Random Forest recall: 0.44, but better precision and F1-score.

Key Insights from EDA

Perceived Risk

Higher uptake

Strong predictor

Belief in Effectiveness

Demographics:

- Higher income & education → More likely vaccinated
- Older adults & females → Higher seasonal flu uptake

Strategic Recommendations

Targeted Messaging:

- Educate low-risk perception groups
- Reinforce vaccine effectiveness

Demographic Outreach:

- Young adults & men
- Low-income & less educated populations

Model Application:

- **Identify likely non-vaccinated individuals for intervention**

Whats next?

Uptake

Deploy model for real-time campaign planning

Efficacy

Explore XGBoost or boosting methods

Equity

Address class imbalance for minority (vaccinated) class

Questions or comments?

Phone Number

+254726592947

Email Address

soniesterina98@gmail.com

Git hub

https://github.com/esterinasoni/H1N1_vaccine_prediction.git