Predicting H1N1 Vaccine Uptake

TABLE OF CONTENTS

Business Understanding

The goal of the "Flu Shot Learning: Predict H1N1" competition is to predict whether individuals received H1N1 vaccines based on their demographic information, opinions, and health behaviors. This predictive model has significant implications for public health initiatives and vaccine campaigns.

Problem Statement

Faced with the reality of how quickly viral diseases like H1N1 can spread, along with their associated health risks and a notable mortality rate, there is a pressing need to develop a model to accurately identify individuals who are more likely to receive their H1N1 and seasonal flu vaccines.

1

OBJECTIVES

To predict the probability for each individual receiving H1N1

Data Understanding

The dataset is derived from the National 2009 H1N1 Flu Survey and contains information on 26,707 respondents, with each row representing one individual's responses.

Data Preprocessing

The following steps were taken in the preparation of the data ready for modelling.

1. Data Cleaning: Selected relevant columns, handled missing values by dropping rows with NaNs.
2. Train-Test Split: Split the dataset into training (80%) and testing (20%) sets.
3. Standardization: Standardized the feature set using StandardScaler.
4. Handling Class Imbalance: Applied SMOTE to balance the class distribution in the training data.

Modeling

The following models were used during the developing of a machine learning model that can predict whether individuals received the H1N1 vaccine based on demographic information, health behaviors, opinions, and medical data.

1. Logistic Regression
2. Decision Tree Classifier
3. Random Forest Classifier
4. K-Nearest Neighbors (KNN) Classifier

Evaluation

In this project, multiple evaluation metrics were used to assess the performance of different models in predicting H1N1 vaccine uptake.

These metrics provided insights into the accuracy, precision, recall, and overall predictive power of the models.

CONCLUSIONS

Based on the evaluation metrics (accuracy, ROC-AUC, classification reports), the Random Forest Classifier outperformed other models with the highest test accuracy and robust ROC-AUC scores.

Random Forest Classifier results:

- Train Accuracy: 0.79
- Test Accuracy: 0.80
- AUC: 0.78 (train), 0.77 (test)

This model can help public health campaigns identify and target populations that are less likely to receive the vaccine, thereby improving vaccination coverage.

THANKS!

Do you have any questions?

lynn90952@gmail.com
+2547 08908681
https://github.com/Lynn-rose/phase-3-project