MORINGA

Discover · Grow · Transform

10/03/2024

H1N1 VACCINATION PREDICTION

A MACHINE LEARNING APPROACH

Table of Contents

INTRODUCTION

BUSINESS UNDERSTANDING

- The project aims to predict H1N1 flu vaccine reception based on data from the National 2009 H1N1 Flu Survey. By identifying key factors influencing vaccination decisions, we seek insights to inform public health strategies. The analysis explores connections between individuals' backgrounds, opinions, and health behaviors to flu vaccination choices. This initiative contributes to global vaccination efforts against infectious diseases, drawing lessons from the 2009 H1N1 flu outbreak to address current challenges, including the COVID-19 pandemic.

STAKEHOLDER UNDERSTANDING

- Assist public health officials seeking insights into vaccination patterns and aid various stakeholders interested in optimizing vaccine distribution strategies.

BUSINESS CONTEXT

BUSINESS UNDERSTANDING	The project aims to predict H1N1 flu vaccine reception based on data from the National 2009 H1N1 Flu Survey. By identifying key factors influencing vaccination decisions, we seek insights to inform public health strategies. The analysis explores connections between individuals' backgrounds, opinions, and health behaviors to flu vaccination choices. This initiative contributes to global vaccination efforts against infectious diseases, drawing lessons from the 2009 H1N1 flu outbreak to address current challenges, including the COVID-19 pandemic.
STAKEHOLDER UNDERSTANDING	Assist public health officials seeking insights into vaccination patterns and aid various stakeholders interested in optimizing vaccine distribution strategies.

PROJECT GOALS

OBJECTIVE # 1

Determine how likely individuals are to receive the H1N1 vaccines

OBJECTIVE # 2

Determine what features are most important to my model

OBJECTIVE # 3

Determine how well my classification models are able to predict vaccine intake

DATA UNDERSTANDING

DATA SOURCE

Obtained from The National Flu Survey (NHFS, 2009)

DATA PREPERATION

EDA, visualization, correlation analysis.

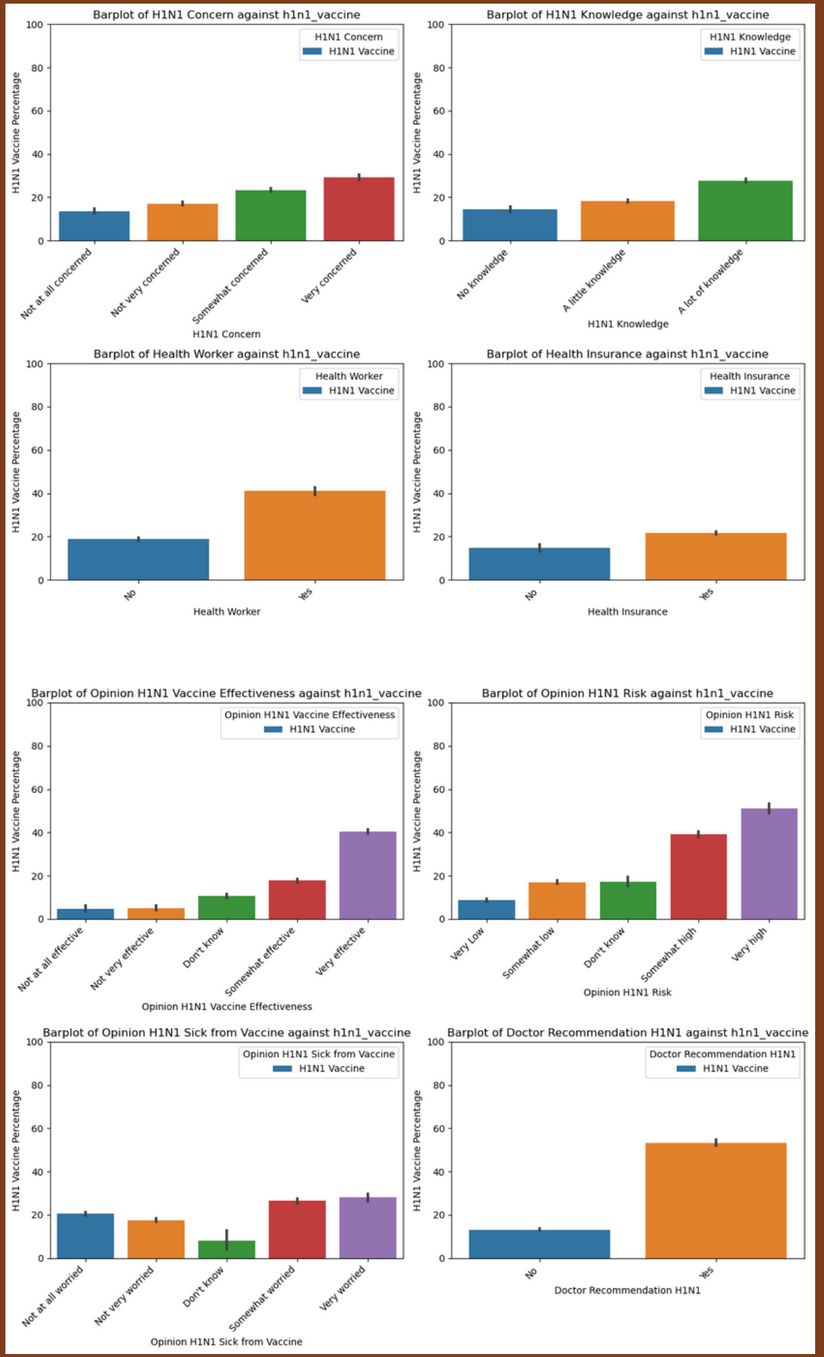
ATTRIBUTES

26,707 Individuals, 35 Unique Features, 79% Unvaccinated

EXPLORATORY DATA ANALYSIS

DATA VISUALIZATIONS

- The graphs represent the distribution of the target variable H1N1 Vaccine against features with the highest correlation
- Our distribution has a class imbalance with the vaccinated only compromising of 21%



DATA MODELLING

MODELLING

01

BASELINE MODEL

02

LOGISTIC REGRESSION MODEL

03

RANDOM FOREST MODEL

04

HYPERPARAMETER TUNING RANDOM FOREST

LOGISTIC REGRESSION MODEL

- The Logistic Regression model achieved an accuracy of approximately 83.75% on the test set, meaning that the model correctly predicted the vaccination status for around 83.75% of the individuals in the test set.
- For the positive class (1 - indicating vaccination), the precision is 0.70, implying that when the model predicts an individual to be vaccinated, it is correct about 70% of the time.
- The precision for the negative class (0 - not vaccinated) is 0.86, suggesting that when the model predicts an individual not to be vaccinated, it is correct about 86% of the time.

RANDOM FOREST MODEL

- The Random Forest model achieved an accuracy of approximately 83.90% on the test set, which implies that the model correctly predicted the vaccination status for around 83.90% of the individuals in the test set.
- For the positive class (1 - indicating vaccination), the precision is 0.70, implying that when the model predicts an individual to be vaccinated, it is correct about 70% of the time.
- The precision for the negative class (0 - not vaccinated) is 0.86, suggesting that when the model predicts an individual not to be vaccinated, it is correct about 86% of the time.

Comparison Between Logistic Regression and Random Forest

- The Logistic Regression Model got an accuracy score of 83.75% whereas the Random Forest Model got a 83.90% score. Both models exhibit similar accuracy on the test set, with the Random Forest model slightly outperforming the Logistic Regression model.
- In the Precision and Recall for Class 1 (Vaccinated) results, both models show similar performance in identifying individuals who received the vaccination, with the Random Forest model having a slightly higher recall.
- In the Precision and Recall for Class 0 (Not Vaccinated) results, both models demonstrate high precision and recall for individuals who did not receive the vaccination, with no significant differences between them.
- The Random Forest model performed slightly better in terms of false negatives and false positives, with fewer instances of misclassifications.

HYPERPARAMETER TUNING RANDOM FOREST

- The Tuned Random Forest Model achieves an accuracy of 84.05%, which is slightly higher than the untuned Random Forest model (83.90%). The hyperparameter tuning has led to a modest improvement in overall accuracy.
- In the Precision and Recall for Class 1 (Vaccinated) results, the Tuned Random Forest Model maintains similar precision for identifying individuals who received the vaccination compared to the untuned model, but the recall has slightly increased. It is successfully identifying a higher proportion of actual positives.
- In the Precision and Recall for Class 0 (Unvaccinated) results, the Tuned Random Forest Model maintains high precision and recall for individuals who did not receive the vaccination, with no significant changes compared to the untuned model.

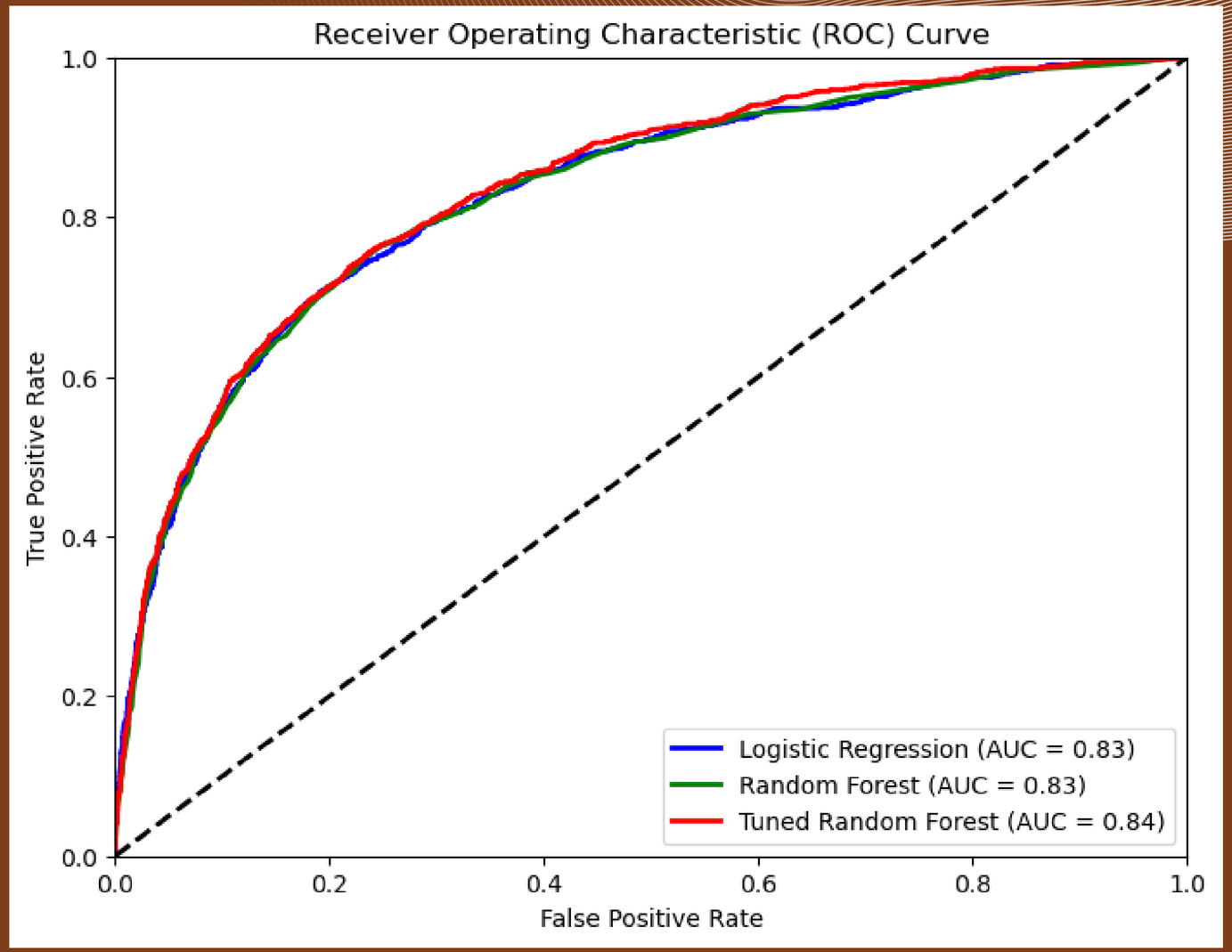
MODEL EVALUATION

- The Tuned Random Forest Model achieves an accuracy of 84.05%, which is slightly higher than the untuned Random Forest model (83.90%). The hyperparameter tuning has led to a modest improvement in overall accuracy.
- In the Precision and Recall for Class 1 (Vaccinated) results, the Tuned Random Forest Model maintains similar precision for identifying individuals who received the vaccination compared to the untuned model, but the recall has slightly increased. It is successfully identifying a higher proportion of actual positives.
- In the Precision and Recall for Class 0 (Unvaccinated) results, the Tuned Random Forest Model maintains high precision and recall for individuals who did not receive the vaccination, with no significant changes compared to the untuned model.
- TIn summary, the Tuned Random Forest Model demonstrates a subtle enhancement in key metrics, providing a more refined and optimized solution for predicting H1N1 vaccination status. The adjustments made during hyperparameter tuning contribute to improved model performance, particularly in capturing positive instances.

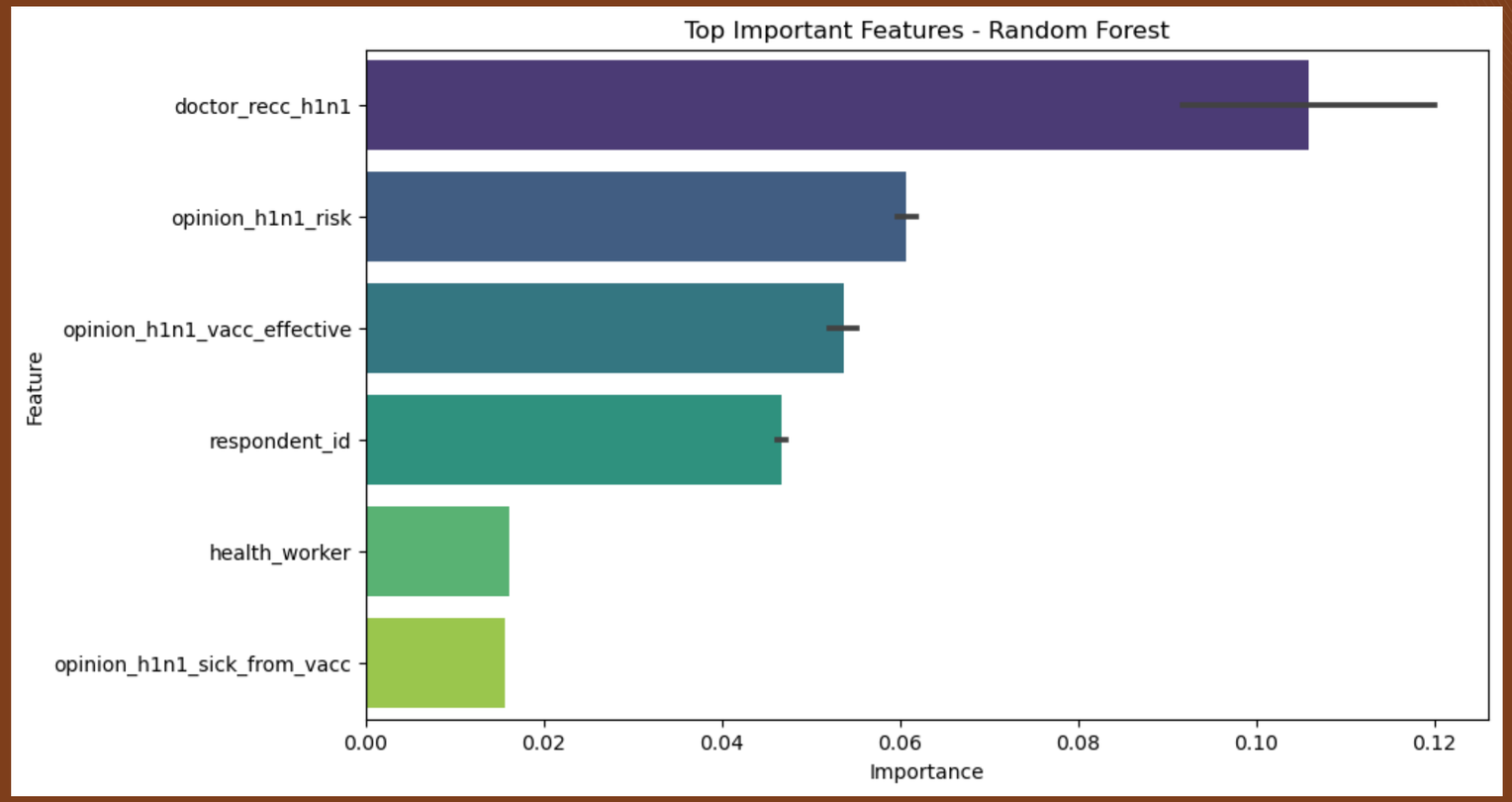
THE ROC AUC (RECEIVER OPERATING CHARACTERISTIC - AREA UNDER THE CURVE) METRIC

The ROC AUC (Receiver Operating Characteristic - Area Under the Curve) metric is measure of how well a model distinguishes between vaccinated and non-vaccinated individuals.
Performance Rank:

- Tuned Random Forest Model: ROC AUC Score of 0.84
- The Random Forest Models: 0.83
- Logistic Regression: 0.83
- The Tuned Random Forest Model performed the best among all the models in the ROC AUC measure.



THE TOP IMPORTANT FEATURES



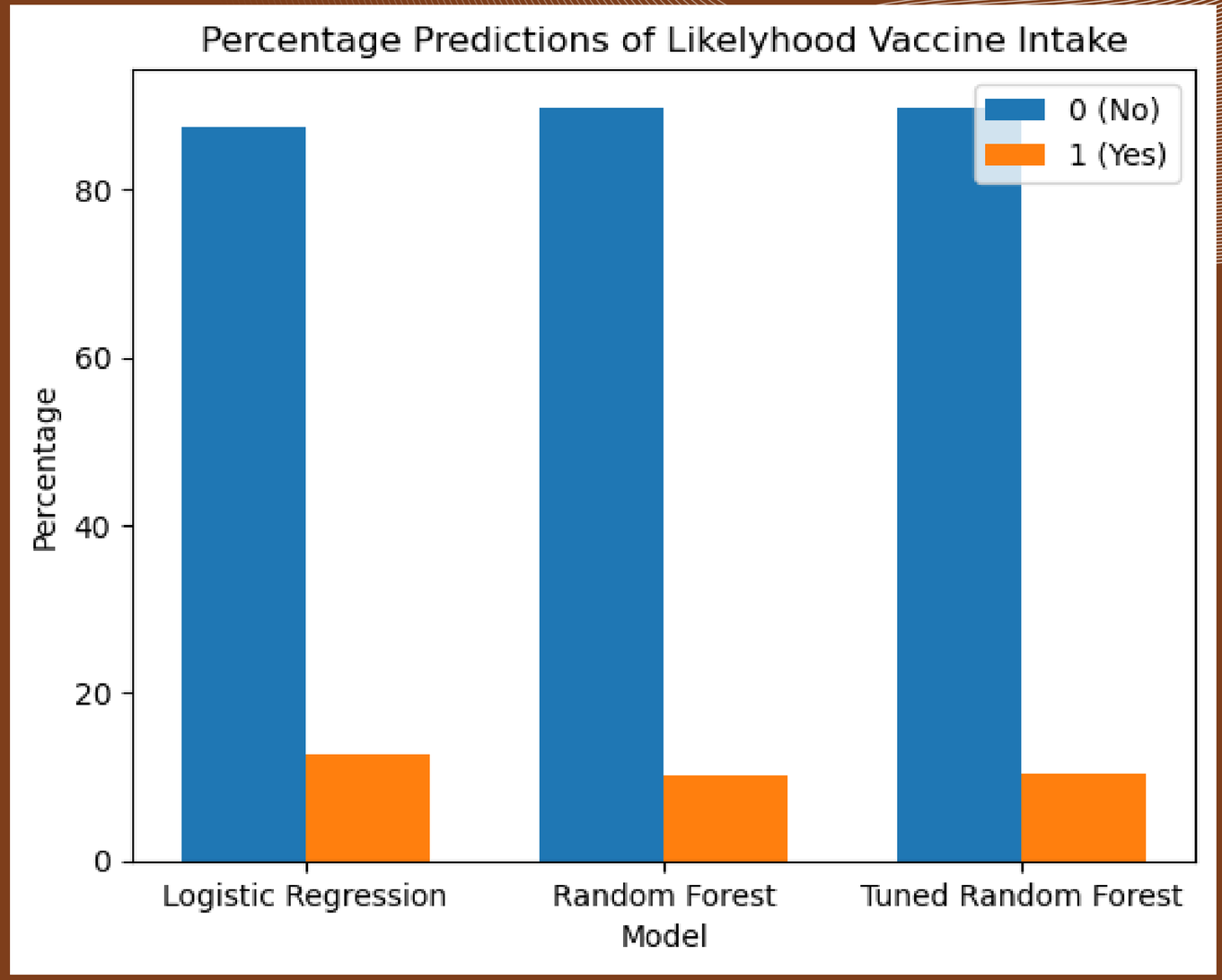
The top five factors that determine whether a person has or will get vaccinated:

- Getting a Doctor's Reccommendation
- Respondent's opinion of risk of getting h1n1 flu without vaccine
- Respondent's opinion about seasonal flu vaccine effectiveness
- Is a healthcare worker
- Respondent's worry of getting sick from taking H1N1 vaccine

TESTING OUR MODELS ON A NEW DATASET

The models were tested on a new dataset to see how likely individuals are to receive their H1N1 vaccine

- Percentage likely to receive vaccine (Logistic Regression): 12.66%
- Percentage likely to receive vaccine (Random Forest): 10.91%
- Percentage likely to receive vaccine (Tuned Random Forest): 10.37%



PREDICTIVE RECOMMENDATIONS

- The models might not be capturing certain factors influencing vaccine acceptance adequately. It's crucial to consider additional external factors or variables not present in the current dataset that could impact the predictions.
- The predictions might not be useful in situations where external factors, such as evolving public health policies, change rapidly and are not reflected in the dataset.

BUSINESS MODIFICATION SUGGESTIONS

- Regularly update the model with the latest data to ensure it reflects the current vaccination landscape.
- Explore the inclusion of additional features or external data sources that could enhance the model's predictive power.
- Consider re-evaluating and updating the model periodically to adapt to changing circumstances and improve overall performance.
- In summary, the models provide insights into vaccine acceptance likelihood, but their conservative nature suggests caution in relying solely on these predictions. Regular updates and consideration of external factors are essential for maintaining model relevance and accuracy.

Resource Page

Project By: DANIEL MURUTHI

https://github.com/Daniel-Muruthi/flu_vaccination.git

MORINGA

Discover · Grow · Transform