Predicting H1N1 Vaccine Uptake

Machine Learning To Address Vaccine Hesitancy

Overview

This project utilizes data from the National Flu Survey (NHFS 2009) to predict the likelihood of individuals receiving the H1N1 flu vaccine. By analyzing historical vaccination patterns, the study aims to provide insights into contemporary vaccination behaviors, particularly relevant in the context of emerging health crises such as the COVID-19 pandemic.

Problem Statement: Predict vaccine uptake and uncover factors influencing vaccination behaviors.

Limitations; Vaccine hesitancy challenges public health, leading to lower immunization rates and outbreaks.

Goal: Predict H1N1 vaccine uptake and identify key influencing factors to guide public health strategies.

Business Understanding ; Vaccination remains critical for disease prevention

Data Understanding;

The dataset utilized in this project comprises 38 columns and 26,707 rows, sourced from Kaggle. This dataset originates from the National Flu Survey (NHFS) of 2009, a comprehensive survey conducted to understand public attitudes and behaviors regarding influenza vaccination, specifically targeting the H1N1 flu vaccine.

The value counts indicate that the distribution of respondents across the "seasonal_vaccine" classes is more balanced compared to the distribution across the "h1n1_vaccine" classes.

Choosing H1N1 as the target variable is justified by its greater public health relevance, potential to generate actionable insights for future pandemics, and the unique challenges it presents for modeling vaccine uptake. It aligns with the goal of understanding vaccination behavior in crisis contexts, which has broader applications beyond routine immunization.

Data Cleaning & Preprocessing

Dropping Irrelevant Columns: Columns such as respondent_id, employment_industry, employment_occupation, hhs_geo_region, census_msa, and seasonal_vaccine were dropped based on relevance to H1N1.

Handling Missing Values: Missing values were filled using median for continuous variables, and mode for categorical variables.

One-Hot Encoding: Categorical features were transformed into a numerical representation using one-hot encoding.

Feature Scaling: Numerical variables were scaled using StandardScaler to normalize the data and ensure equal weighting for features across the models.

Modeling Approach & Evaluation

Handling Class Imbalance: SMOTE (Synthetic Minority Over-sampling Technique) was applied to address class imbalance in the dataset.

Baseline Model: Logistic Regression was used as a baseline model to establish a minimum standard of performance.

Advanced Models: Decision Tree and Random Forest models were implemented with hyperparameter tuning to improve prediction accuracy

Performance Metrics

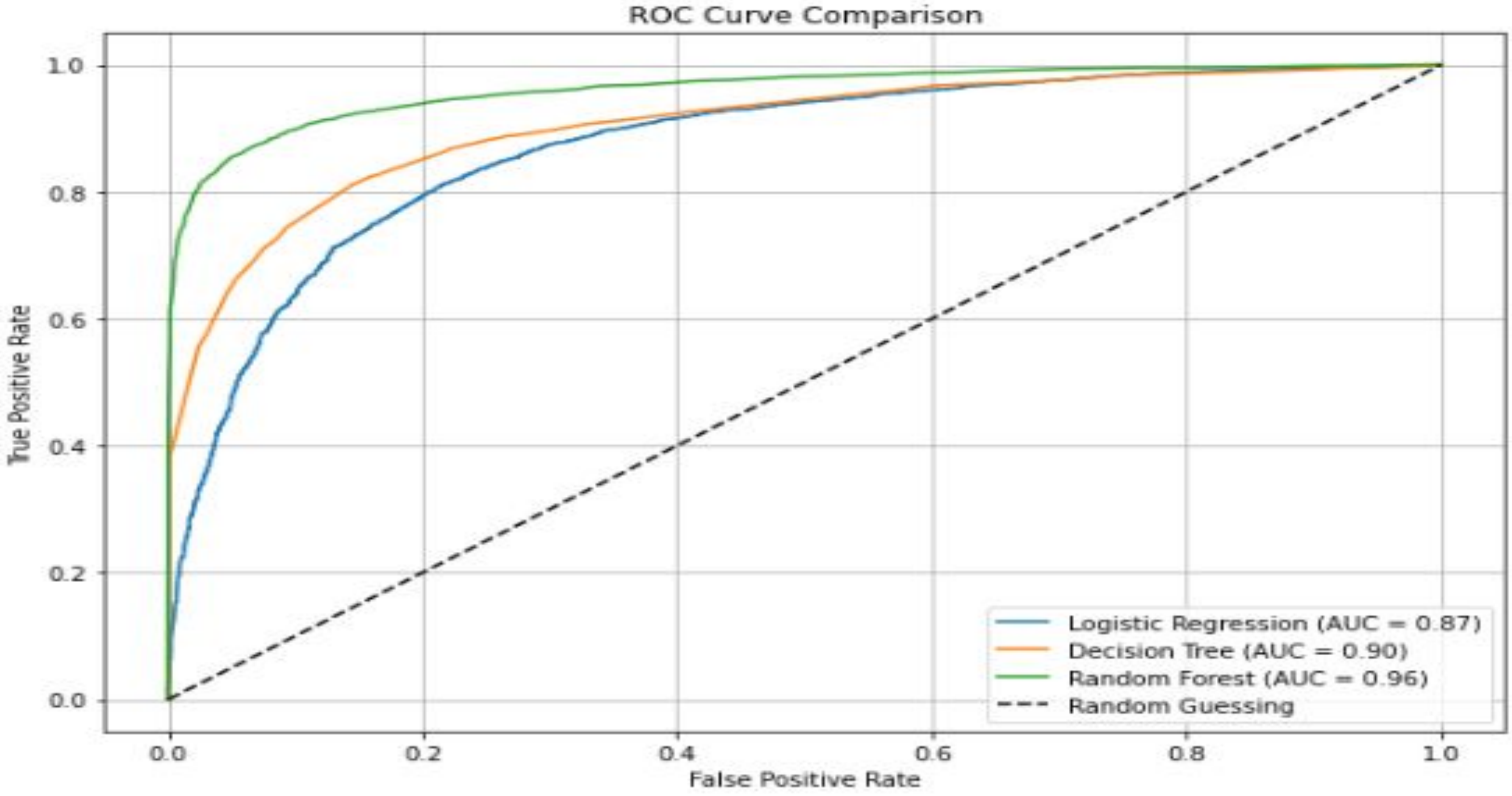
The models were evaluated using metrics such as `**accuracy, precision, recall, F1-score, and ROC AUC score**` to assess their performance in predicting H1N1 vaccine uptake.

`**Logistic Regression (AUC = 0.87):** A reasonably strong classifier. It performs well but is not the best compared to the other two models. It may slightly underperform in scenarios requiring higher precision or recall.

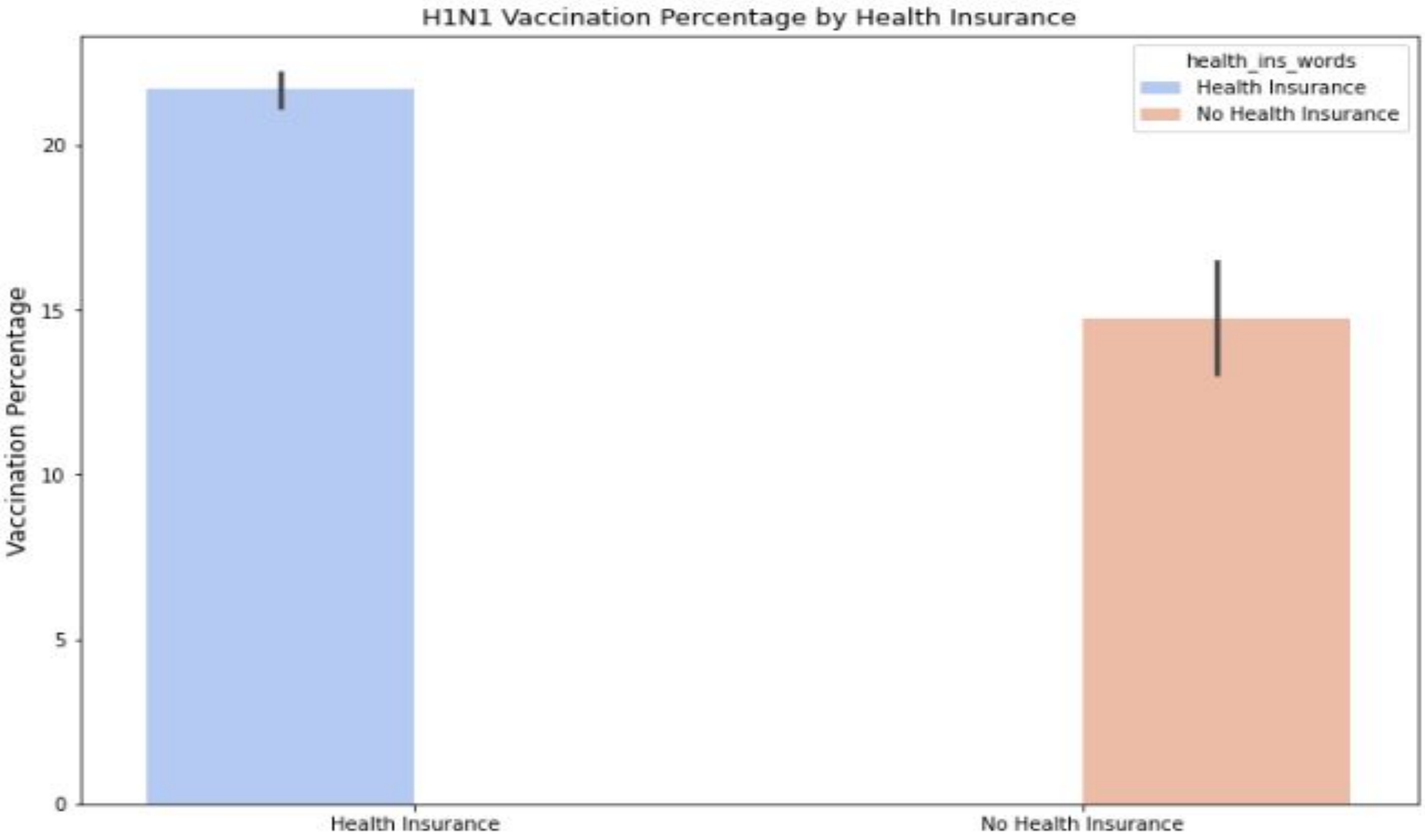
Decision Tree (AUC = 0.90): Better than Logistic Regression. Likely captures non-linear relationships in the data. However, Decision Trees can overfit on training data, so this result might need validation on unseen data.

Random Forest (AUC = 0.96): The best-performing model among the three. High AUC indicates that it can distinguish between classes with high accuracy. Random Forest's ensemble approach minimizes overfitting and improves generalization.`

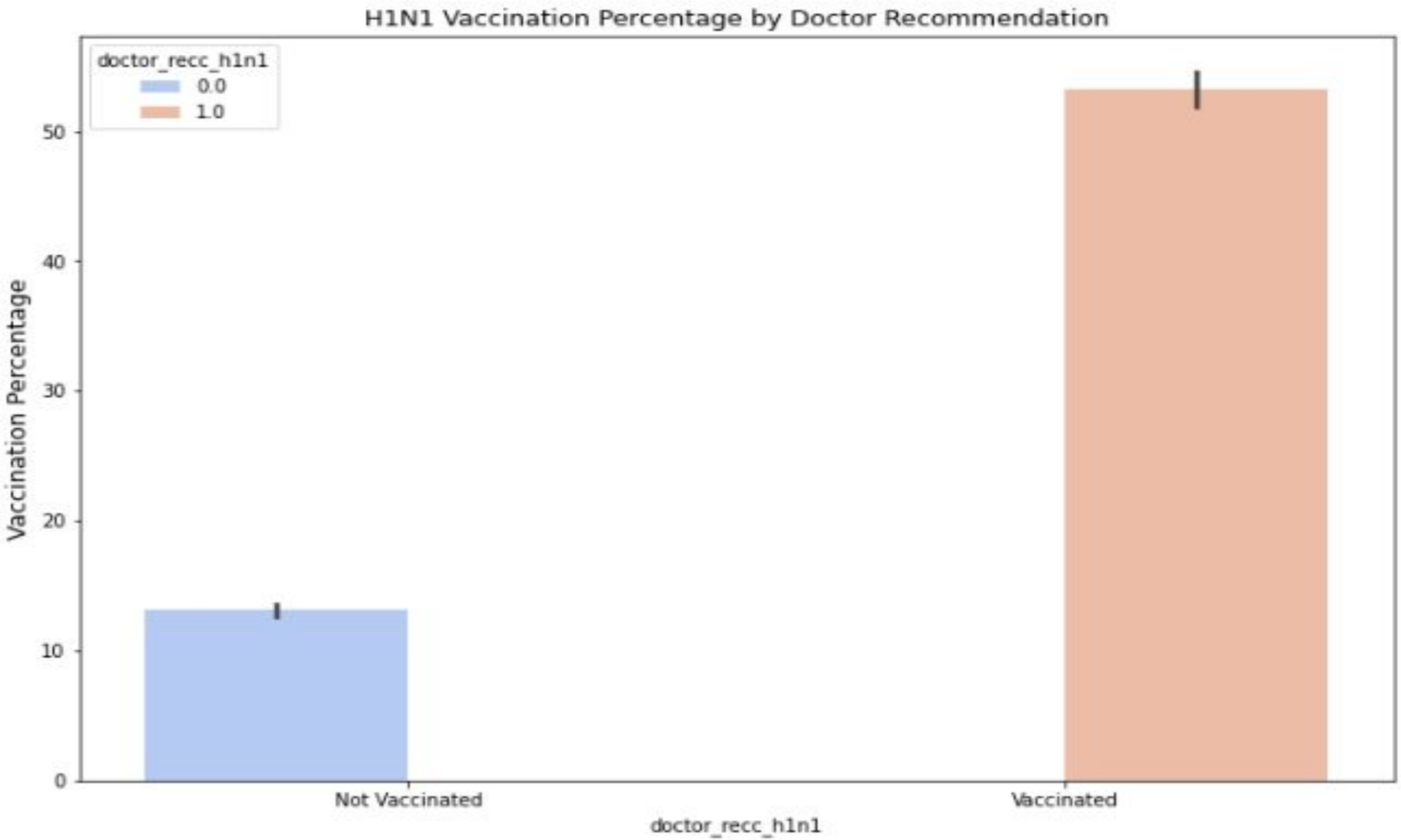
Performance metric visual



HEALTH INSURANCE WAS A DETERMINING FACTOR IN VACCINE INTAKE



DOCTOR RECOMMENDATIONS WAS INFLUENTIAL



RECOMMENDATIONS

Public Health: Increase awareness campaigns and address vaccine misinformation.

Healthcare Collaboration: Engage providers to promote vaccinations actively.

Policy Initiatives: Enhance health insurance access and affordability.

Innovative Access: Mobile clinics and community-driven outreach to underserved areas.

The implemented preprocessing and Random Forest model deliver a reliable and robust solution for predicting vaccine uptake.This high score indicates that the model is highly effective at distinguishing between individuals who received the H1N1 vaccine and those who did not. Its ensemble approach minimizes the risk of overfitting, making it a reliable choice for predicting vaccine uptake in real-world scenarios. Stakeholders can confidently use these insights to make data-driven decisions in public health strategies, resource allocation, and vaccine campaigns.

CONCLUSIONS

This project provides a comprehensive understanding of vaccination behaviors and informs strategies to combat vaccine hesitancy effectively. The insights derived from the predictive models can assist public health officials and policymakers in enhancing vaccination rates and addressing public health challenges.

Long-Term Implications: The ability to accurately predict vaccination uptake can significantly enhance public health responses, especially in the context of emerging health crises. By understanding the factors that drive vaccine acceptance, stakeholders can implement more effective public health campaigns that foster community trust and participation.

Strategic Recommendations: Given the superior performance of the Random Forest model, stakeholders should consider leveraging this model for future public health initiatives aimed at increasing H1N1 vaccination rates. The insights derived from this model can guide targeted interventions and resource allocation.

Thank You;