H1N1 Vaccination Uptake Prediction

BY JACINTA CHEPKEMOI

Overview

The primary goal of this project is to utilize data from the National 2009 H1N1 Flu Survey to predict individuals' likelihood of receiving the H1N1 vaccine. Given the significance of vaccination in combating infectious diseases, understanding the factors influencing vaccination uptake is paramount for public health efforts to improve immunization rates. By analyzing demographic, socio-economic, and attitudinal factors alongside health behaviors, this project seeks to provide actionable insights for future vaccination promotion strategies.

Outline

- Business Understanding
- Exploratory Data Analysis
- Modelling
- Evaluation
- Conclusion
- Recommendations
- Future Work

Business Understanding

Our project centers on examining the H1N1 vaccination uptake among respondents of the National 2009 H1N1 Flu Survey. Vaccination represents a crucial measure for achieving herd immunity and reducing the spread of infectious diseases. The survey data offers a comprehensive view of respondents' backgrounds, opinions, and health behaviors, enabling us to identify key factors influencing vaccination decisions. However, challenges such as handling missing data and selecting appropriate machine learning models necessitate a systematic approach to address them effectively.

Problem Statement

Vaccine hesitancy poses a significant obstacle to public health efforts, leading to decreased immunization rates and heightened vulnerability to infectious diseases. Understanding the determinants of individuals' decisions regarding vaccine uptake is essential for designing targeted interventions and promoting community immunity. In this context, our project aims to predict the likelihood of individuals receiving the H1N1 flu vaccine using machine learning techniques and data from the National Flu Survey (NHFS 2009).

Objectives

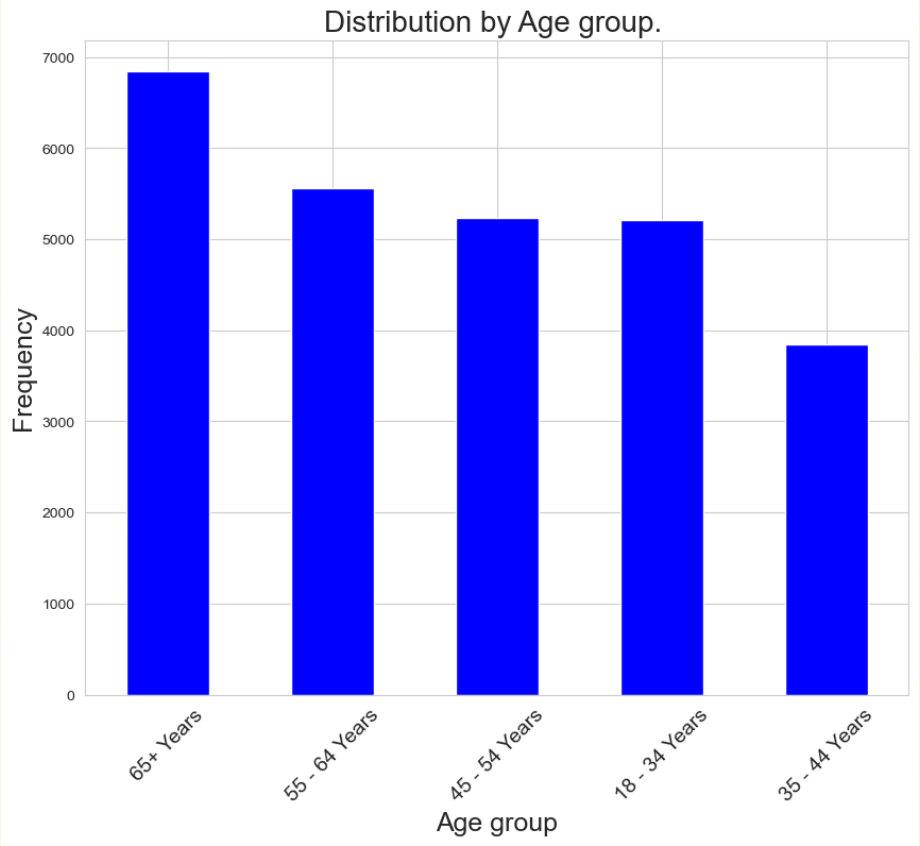
The objectives encompass several key areas:

1. Assessing feature importance within the dataset to identify critical factors for vaccination promotion.
2. Determining influential factors affecting H1N1 vaccine acceptance, including doctor recommendations and perceptions of vaccine effectiveness.
3. Evaluating the performance of different machine learning algorithms in predicting H1N1 vaccine uptake.
4. Analyzing the implications of predictive models to derive actionable insights for public health professionals and policymakers to enhance vaccination rates.

EXPLORATORY DATA ANALYSIS

Univariate Analysis

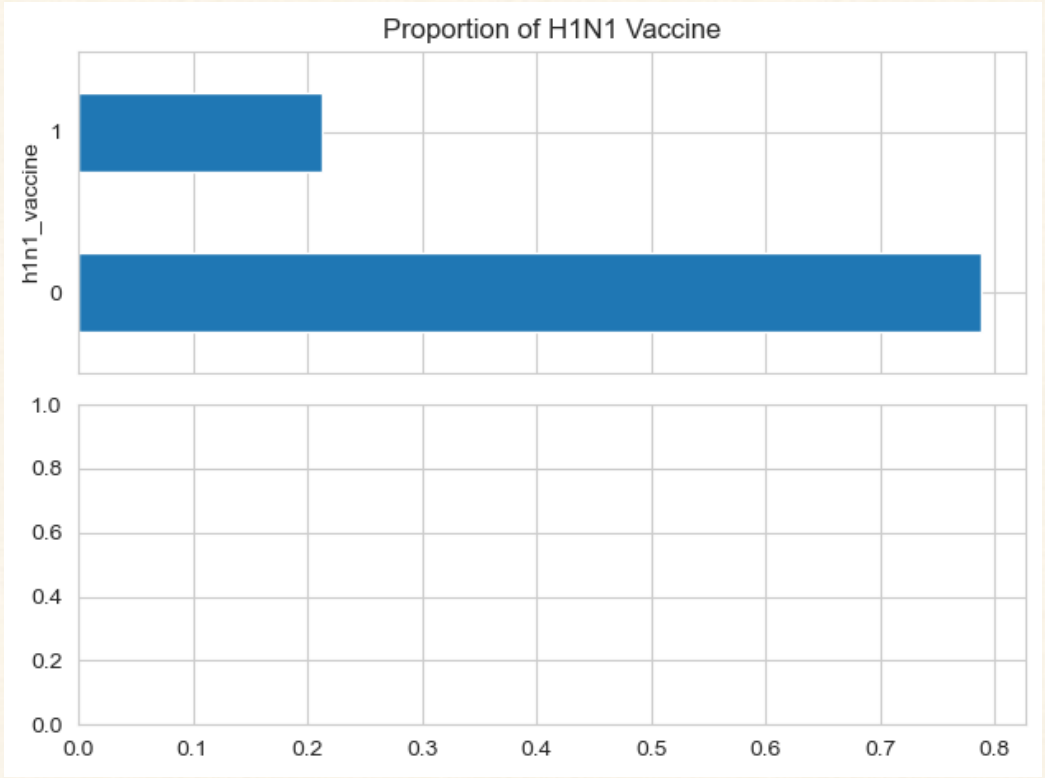
The plot displays the distribution of respondents' age groups, showing that the majority of respondents fall into the '65+ Years' category, with 6843 respondents. Following that, '55 - 64 Years' has the second-largest representation with 5563 respondents. '45 - 54 Years' and '18 - 34 Years' are close in count, with 5238 and 5215 respondents, respectively. '35 - 44 Years' has the fewest respondents at 3848, making it the smallest age group category.



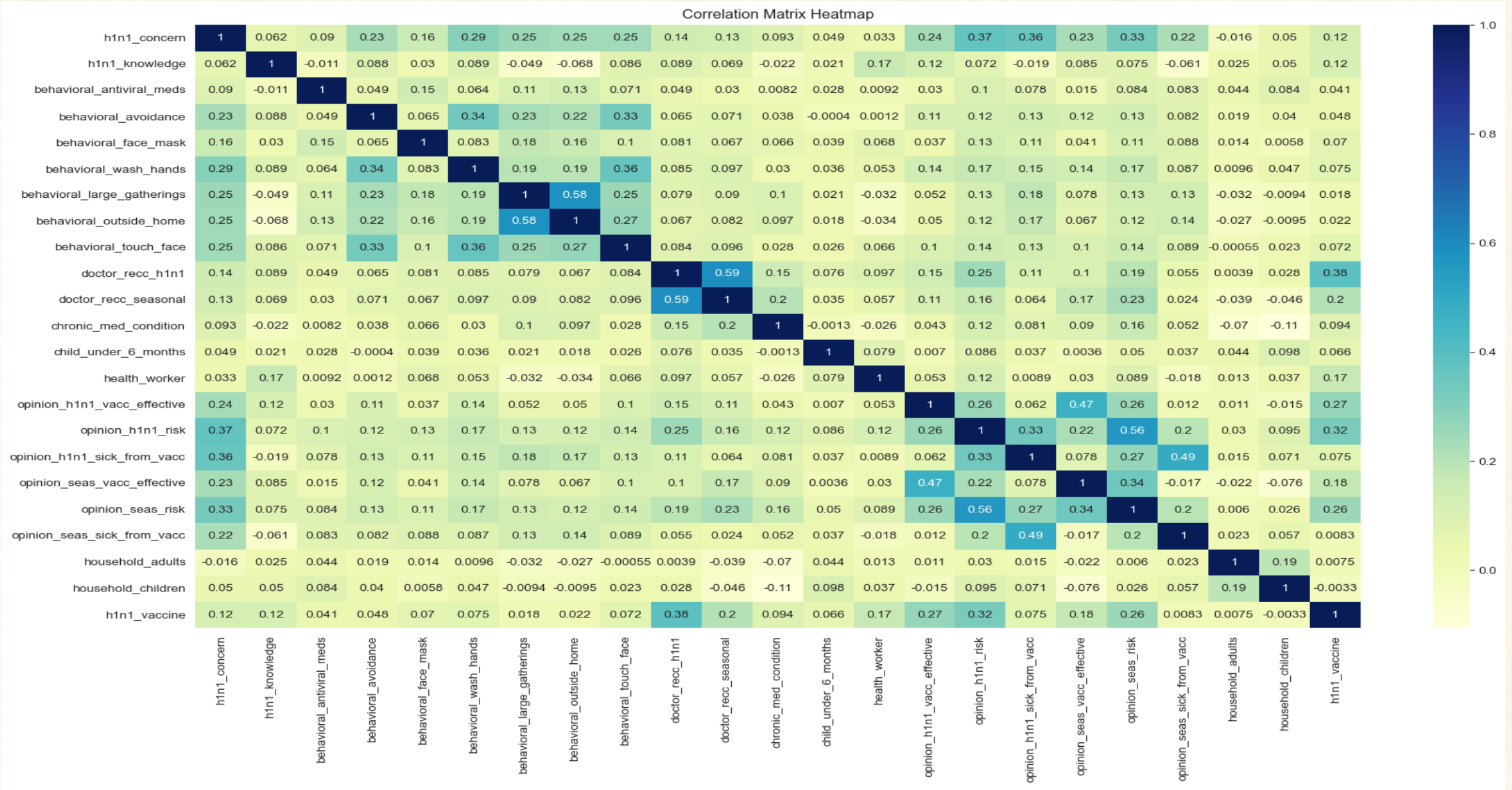
EXPLORATORY DATA ANALYSIS

Bivariate Analysis

The plots illustrate the distribution of respondents' vaccination statuses for H1N1 vaccine.'0' represents those who have not been vaccinated, accounting for 21,033 respondents, while '1' represents those who have been vaccinated, totaling 5,674 respondents. This distribution highlights that a majority of respondents in the dataset have not been vaccinated against H1N1.



Multivariate Analysis



EXPLORATORY DATA ANALYSIS

The heatmap visualizes the relationships between various factors in a public health survey about H1N1 (swine flu). The color intensity indicates the strength and direction of the correlation. Positive correlations (dark blue) suggest that factors tend to increase together or decrease together. Conversely, negative correlations (beige) show opposing trends. There's a weak negative correlation between concern about H1N1 and the number of adults or children in the household, which might be due to a perceived lower risk or a focus on protecting others. These findings offer valuable insights into public health behaviors and opinions, but it's important to remember that correlation doesn't imply causation.

The most significant predictors include recommendations from doctors (num_doctor_recc_h1n1), opinions on the vaccine's effectiveness (num_opinion_h1n1_vacc_effective), and perceived risk of H1N1 (num_opinion_h1n1_risk). These factors likely highlight the importance of trusted medical advice and personal beliefs about the vaccine's benefits and risks in driving vaccination decisions. Conversely, the least significant predictors include behavioral factors like going outside the home (num_behavioral_outside_home), household adults (num_household_adults), and employment status (cat_employment_status_Unemployed).

MODELLING

Evaluated models:

- **Decision Tree**: 75% accuracy, struggled with minority class.
- **KNN**: 81% accuracy, better balance.
- **Logistic Regression**: 83% accuracy, handled class imbalance well.
- **Random Forest**: 83% accuracy, strong but faced class challenges.
- **Gradient Boosting**: 84% accuracy, best overall performance.

Model Tuning

Hyperparameters of the Gradient Boosting Classifier were optimized using grid search, improving predictive performance through cross-validation.

The tuned model and preprocessing steps were integrated into a deployment pipeline for seamless use by public health professionals and policymakers.

EVALUATION and CONCLUSION

Gradient Boosting Classifier achieved the highest accuracy (84%) and balanced performance. Logistic Regression and Random Forest were also strong performers. Decision Tree and KNN were less effective.

Key predictors of vaccination include doctor recommendations, vaccine effectiveness opinions, and perceived H1N1 risk. Gradient Boosting Classifier is the preferred model for this dataset.

Public health campaigns should focus on education, tailored strategies for different demographics, and emphasizing medical advice and vaccine effectiveness. Gradient Boosting Classifier is recommended for prediction, with Logistic Regression and Random Forest as alternatives.

Further tuning of the Gradient Boosting Classifier, exploring ensemble methods, integrating additional data sources, and collaborating with public health experts to enhance model effectiveness.

Thank You

This comprehensive presentation has been prepared by Jacinta Chepkemoi, utilizing data-driven insights to enhance public health strategies for increasing H1N1 vaccine uptake. For further inquiries or collaborations, please connect with me on LinkedIn. I am eager to collaborate and contribute to advancements in public health and vaccination efforts.
www.linkedin.com/in/jacinta-chepkemoi-9936b0220