Predicting H1N1 and Seasonal Flu Vaccine Uptake

A DATA-DRIVEN APPROACH TO SUPPORT PUBLIC HEALTH DECISIONS

Introduction

- • Vaccination uptake remains a challenge in public health.
- • This project aims to predict the likelihood of individuals getting vaccinated for H1N1 and seasonal flu.
- • Understanding these patterns helps allocate resources and design effective outreach campaigns.

Business Understanding

- Problem: Low uptake of H1N1 and seasonal vaccines during pandemic
- - Objective: Predict vaccine uptake and identify key influencing factors
- - Importance: Inform targeted interventions for better vaccination campaigns

Project Goals

- • Identify key drivers behind vaccine uptake.
- • Build predictive models for H1N1 and seasonal flu vaccines.
- • Provide actionable insights to improve vaccination strategies.
- • Support public health stakeholders in decision-making.

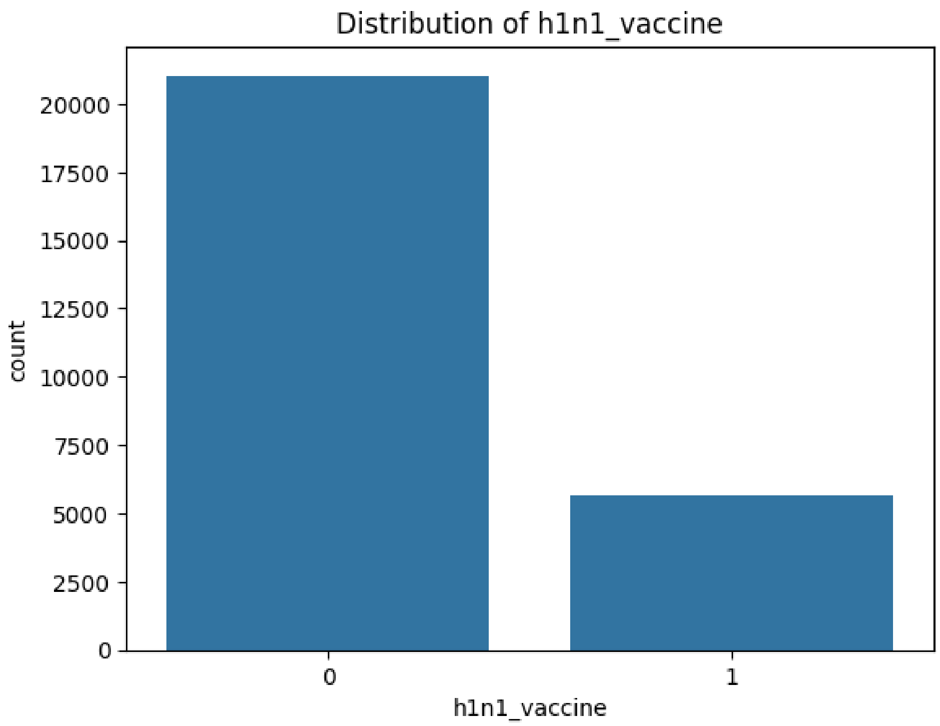
Data Overview

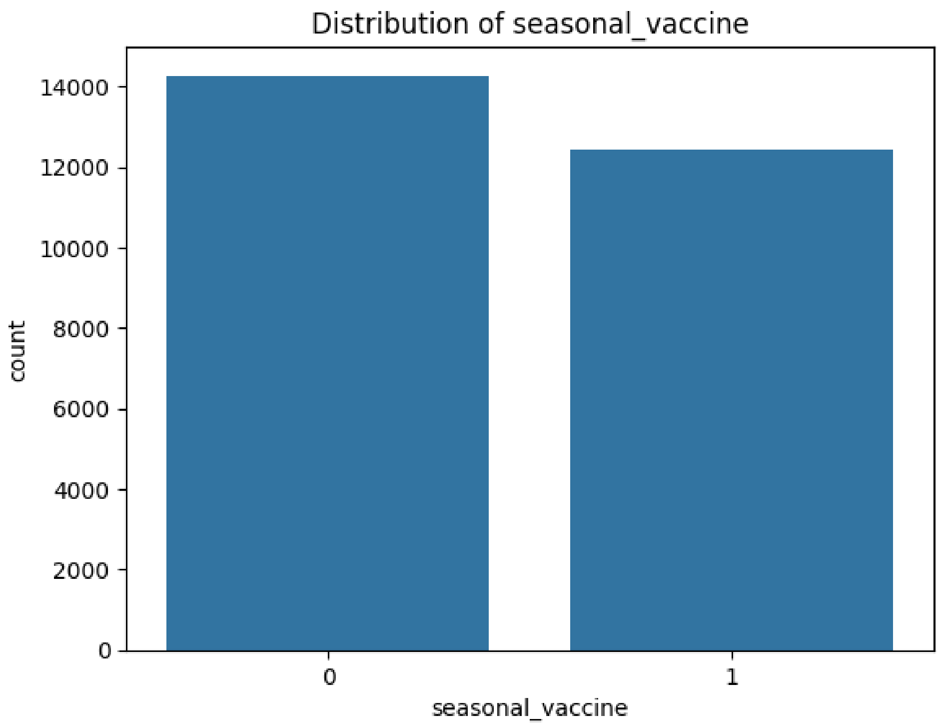
- • Dataset: 26,707 survey respondents from the National H1N1 Flu Survey.
- • Features include demographics, behaviors, opinions, and access to healthcare.
- • Target Variables:
- – H1N1 Vaccine Uptake (Yes/No)
- – Seasonal Flu Vaccine Uptake (Yes/No)

EDA - Univariate

- - Vaccine uptake distributions:
- • H1N1 vaccine: 21% uptake
- • Seasonal vaccine: 47% uptake
- - Preventive behaviors (mask use, handwashing) show strong variation

H1n1 distribution

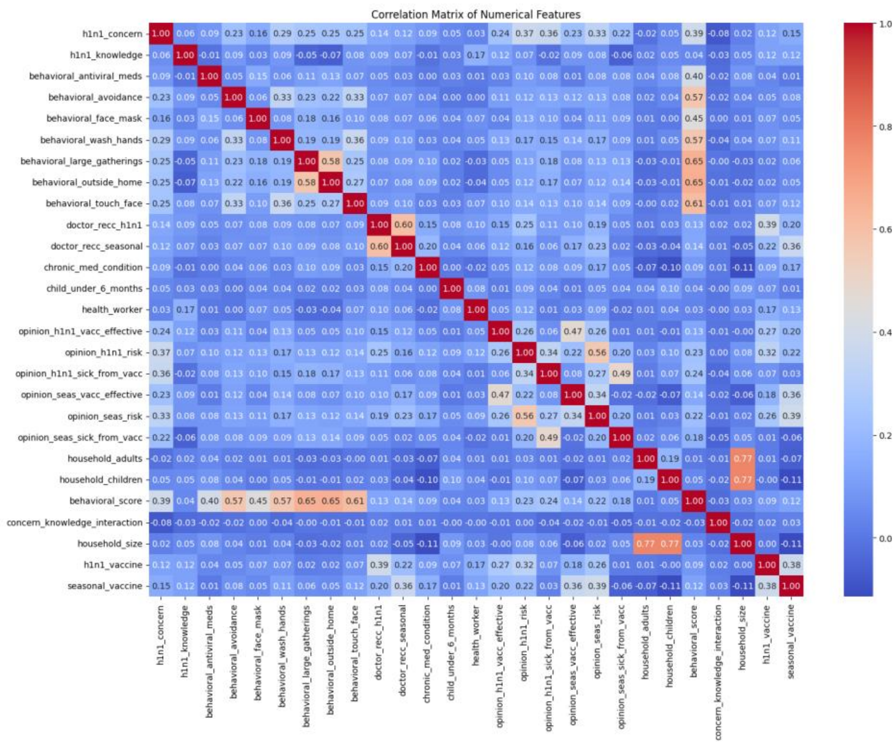



Distribution of seasonal_vaccine
count
14000
12000
10000
8000
6000
4000
2000
0
0
1
seasonal_vaccine

EDA - Multivariate

- - Correlation heatmap insights:
- • Doctor recommendation strongly linked to both vaccines
- - Interaction effects: higher knowledge + concern → more likely vaccinated

Corelation of features



Key Insights from EDA

- - Doctor’s recommendation = strongest driver of uptake
- - Preventive behaviors influence uptake
- - Socioeconomic factors (income, education, region) matter
- - New engineered features improved predictive signal

Methods

- • Data Preprocessing:
- – Missing value handling
- – Feature encoding and scaling
- – New feature creation (Behavioral Score, Concern-Knowledge Interaction, Household Size)
- • Modeling Approach:
- – Random Forest Classifier for both targets
- – Train-test split and cross-validation
- – Evaluation with F1-score and AUC-ROC

Modeling Approach

- \- Baseline: Logistic Regression → moderate performance
- \- Final: Random Forest Classifier
- \- Reasons: handles mixed features, robust, interpretable
- \- Validation: Stratified Cross-validation
- \- Metrics: F1-score & ROC-AUC emphasized

Results & Feature Importance

- - Random Forest balanced performance on both targets
- - Top features:
-     • Doctor's recommendation
-     • Concern × Knowledge interaction
-     • Behavioral score
-     • Age group & income
- - (Feature importance plot placeholder)

Results

- • Seasonal Flu Vaccine: ~47% uptake
- • H1N1 Vaccine: ~21% uptake (high imbalance)
- • Key Influential Features:
- – Doctor recommendations
- – Perceived risk and concern levels
- – Preventive behaviors (handwashing, masks)
- – Household size and income

- • Random Forest performed best, capturing complex feature interactions.

Insights & Recommendations

- • Concern and knowledge strongly influence vaccination decisions.
- • Doctor recommendations remain a critical driver.
- • Lower uptake in certain income and household groups suggests targeted outreach.

- Recommendations:
- • Use model insights to guide vaccination campaigns.
- • Focus on education and communication strategies.
- • Prioritize vulnerable households for outreach.

Conclusion

- • Predictive modeling helps identify factors influencing vaccine uptake.
- • Random Forest models provide actionable insights.
- • Public health stakeholders can use these findings to improve vaccination coverage.

- Next Steps:
- • Address class imbalance with SMOTE or ADASYN.
- • Incorporate external data for improved predictions.
- • Deploy models into decision-support dashboards.