UNVEILING PATTERNS IN H1N1 AND SEASONAL FLU VACCINE UPTAKE

By Winnie Mbuve Nzuve

Overview

This project focuses on analyzing H1N1 and seasonal flu vaccine uptake patterns during the 2009 pandemic, considering respondents' backgrounds, opinions, and health behaviors. The U.S. survey from late 2009 to early 2010 captured data on social, economic, and demographic factors, alongside perceptions of illness risks and vaccine effectiveness.

By examining this dataset, the goal is to uncover insights into the factors shaping individual vaccination choices, providing valuable knowledge for future public health efforts. The competition offers an accessible opportunity for participants of all levels to engage in data science and contribute to understanding vaccination patterns during a crucial public health event.

Content

Business Understanding

The project aims to harness the power of machine learning to predict individuals' likelihood of receiving H1N1 and seasonal flu vaccines. By leveraging comprehensive datasets containing information on respondents' backgrounds, opinions, and health behaviors, the project endeavors to unravel the intricate determinants influencing vaccination decisions.

Problem Statement

The primary problem is predicting individuals' likelihood to receive H1N1 and seasonal flu vaccines. By doing so, public health entities can tailor their campaigns, communication strategies, and resource distribution to increase vaccination rates. The goal is to mitigate the spread of infectious diseases and reduce the severity of flu outbreaks.

Objectives

- Develop predictive models for H1N1 and seasonal flu vaccine likelihood.
- Identify key features influencing vaccine decisions.
- Provide actionable insights for public health officials and healthcare providers.
- Contribute to increasing overall vaccination rates and improving preventive healthcare.
- Achieve these objectives within a 6-day timeline through efficient data analysis and model development.

Data Understanding

01. Two datasets available from Data-Driven: one on H1N1 vaccine features and the other indicating patients who took either H1N1 or seasonal flu vaccine.

02. Target: Predict likelihood of individuals getting H1N1 and seasonal flu vaccines based on backgrounds, opinions, and health behaviors.

03. Predictors include demographic, behavioral, and opinion-related features (e.g., h1n1_concern, h1n1_knowledge, behavioral_antiviral_meds).

04. Data-Driven controls the data sources for both datasets.

05. Data distribution covers 26,707 respondents with 38 features. Over 26,000 observations provide sufficient data for building predictive models without requiring resampling methods.

Modeling

Models Used

logistic regression

This statistical method is well-suited for binary classification tasks, such as predicting the probability of a categorical outcome based on one or more predictor variables.

Decision Tree

Decision tree classification was utilized to predict whether patients would opt for the H1N1 vaccine or the seasonal vaccine. This method involves segmenting the dataset into subsets based on various attributes, ultimately providing interpretable results to understand the factors influencing vaccine acceptance.

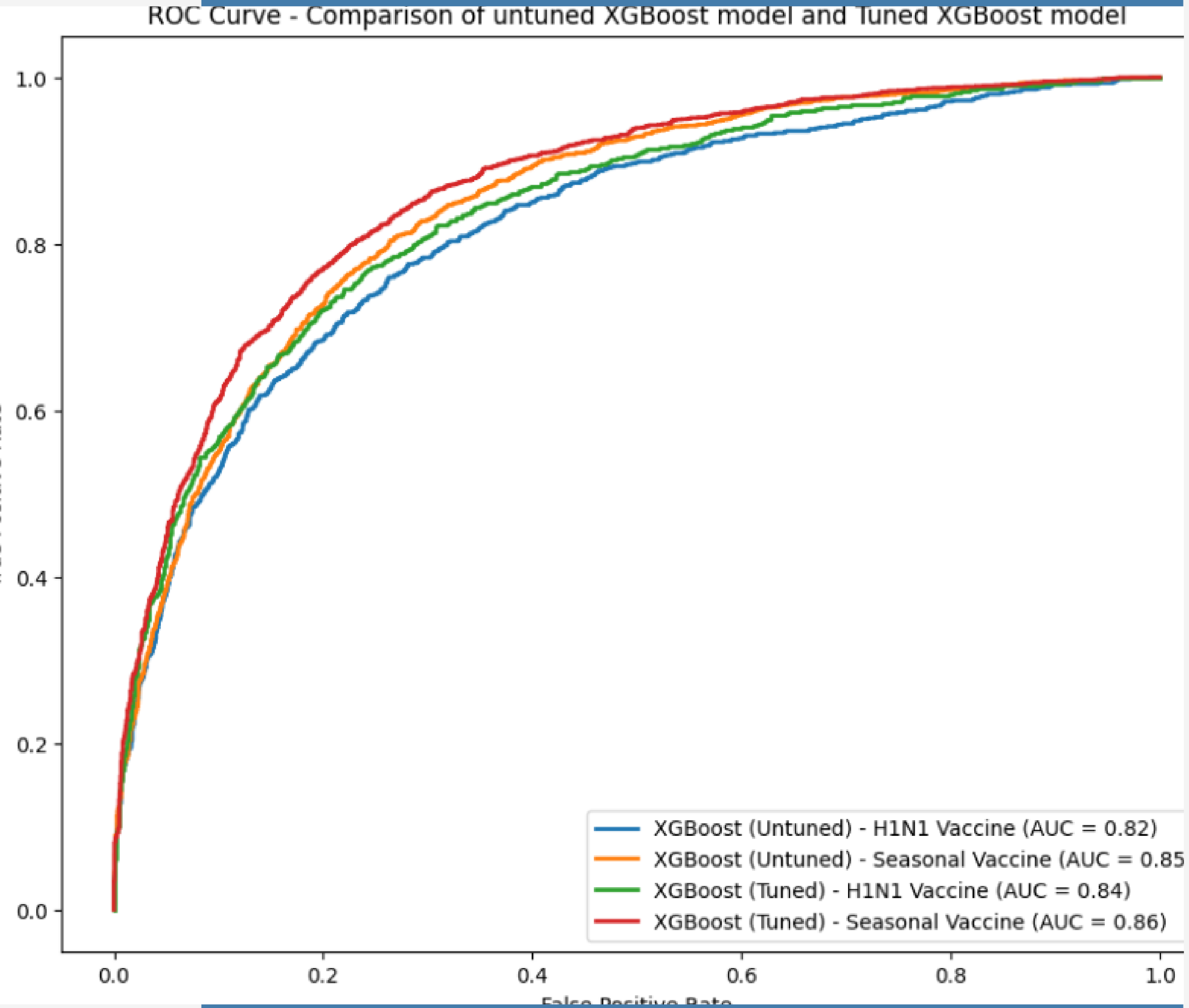
Random Tree

Random forest classification is a machine learning algorithm that operates by constructing multiple decision trees during training and outputting the mode of the classes as the prediction. This approach allows for robust predictions by aggregating the results from multiple decision trees, making it suitable for classification tasks like vaccine uptake prediction.

XGboost

This powerful algorithm was selected due to its ability to handle complex data relationships, its robustness against overfitting, and its effectiveness in classification tasks.

Model Comparison

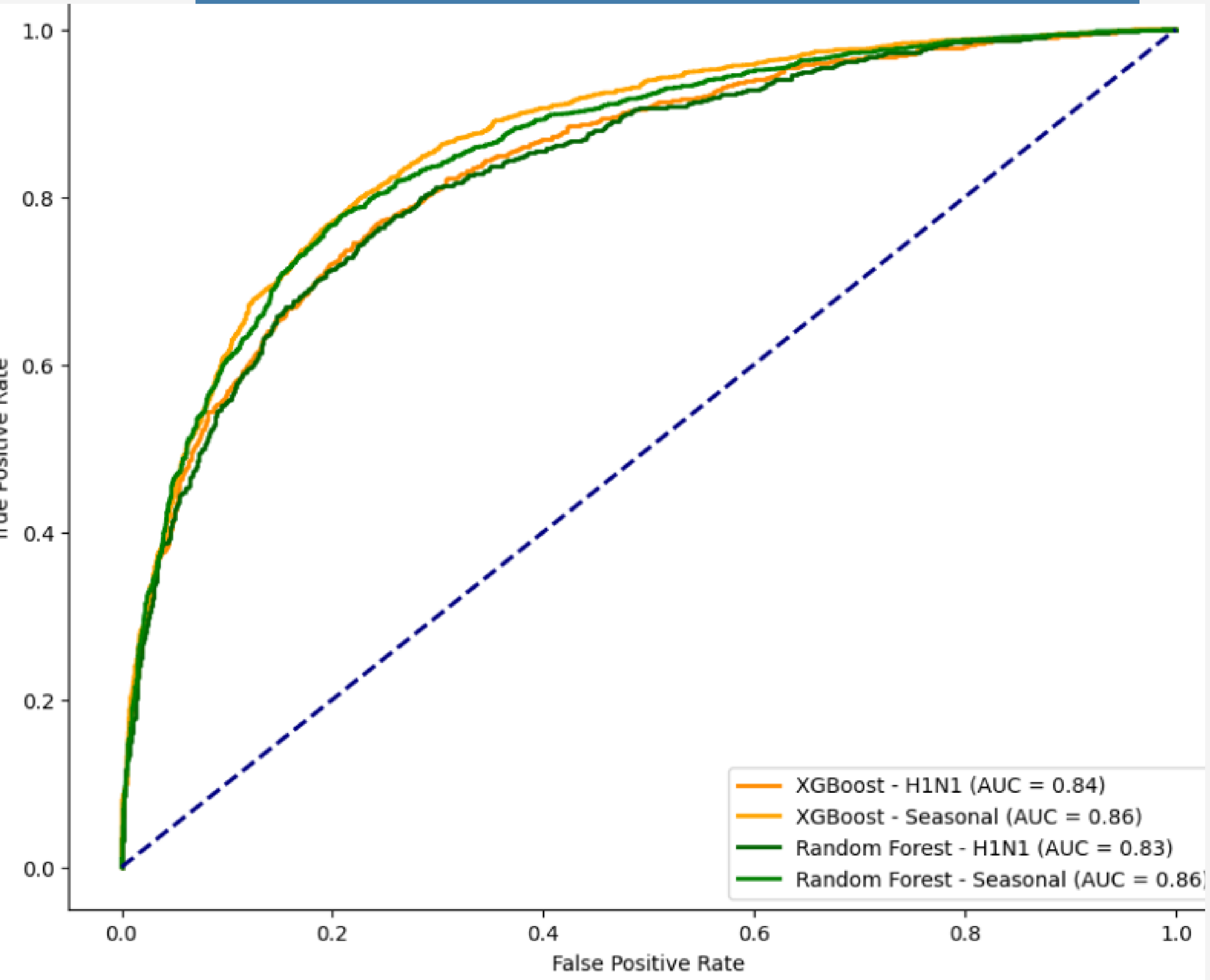


The model performance comparison between the XGBoost before and after tuning by evaluating the Area Under the Curve (AUC) metrics for predicting H1N1 and seasonal flu vaccine uptake.

Before tuning, the XGBoost model exhibited an AUC of 0.82 for H1N1 vaccine and 0.85 for the seasonal vaccine, indicating good discriminative ability. After tuning, the XGBoost model's AUC improved to 0.84 for H1N1 vaccine and 0.86 for the seasonal vaccine.

These enhancements in AUC values signify the effectiveness of model tuning in optimizing predictive performance. The improved AUC values for both vaccines suggest that the tuned XGBoost model provides better discrimination between positive and negative cases, enhancing its ability to accurately identify individuals likely to receive the vaccines.

Model Performance

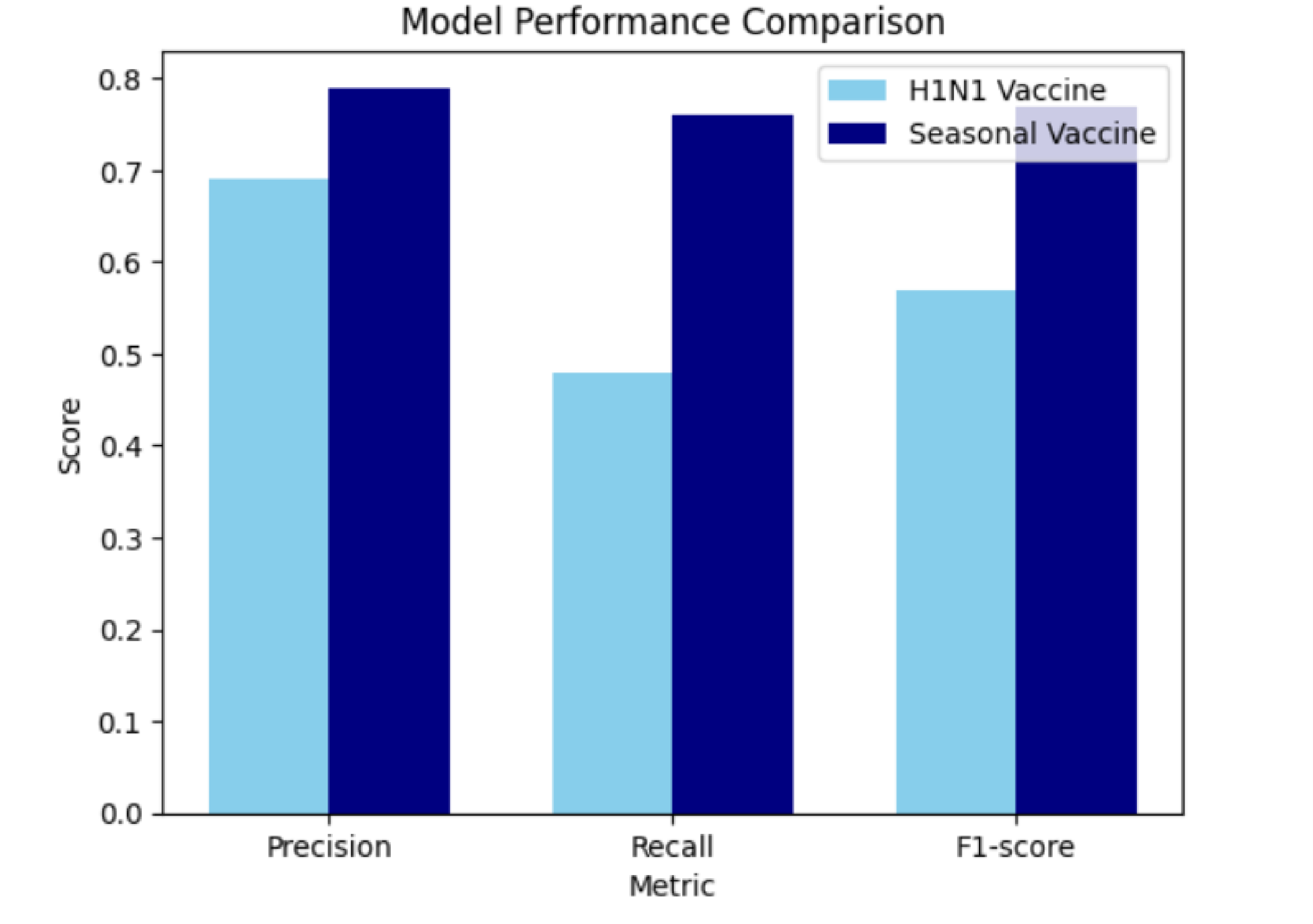


After thorough evaluation and tuning, the XGBoost model demonstrated superior performance compared to the Random Forest model in predicting vaccine uptake for both H1N1 and the seasonal flu.

For the H1N1 vaccine, the tuned XGBoost model achieved an AUC of 0.84, outperforming the Random Forest model, which attained an AUC of 0.83. Similarly, in predicting seasonal flu vaccine uptake, the XGBoost model excelled with an AUC of 0.86, surpassing the Random Forest model's AUC of 0.86.

These results highlight the efficacy of XGBoost in optimizing model parameters and capturing complex patterns within the data, leading to improved predictive accuracy. The tuned XGBoost model is thus identified as the preferred choice for predicting vaccination uptake in the given dataset, offering higher discriminative ability and enhanced performance over the Random Forest model.

Model Performance Comparison



From the graph, we can observe that the precision, recall, and F1-score are generally higher for the H1N1 vaccine model compared to the seasonal vaccine model.

This suggests that the H1N1 vaccine model performs better in accurately predicting whether patients will receive the H1N1 vaccine, while the seasonal vaccine model may have slightly lower accuracy in its predictions. Despite this difference, both models demonstrate relatively strong performance across all metrics.

ANALYSIS

- Higher Uptake: The HIN1 vaccine vaccine appears to have a higher uptake compared to the H1N1 vaccine. This is evident from the higher support (number of instances) for the seasonal vaccine in the dataset.
- Prediction Confidence: Despite the higher uptake of the seasonal flu vaccine, both models exhibit similar predictive performance for both vaccines. This indicates that the models are equally effective at predicting individuals who took each vaccine.
- Balanced Metrics: The F1-scores, precision, and recall values for both vaccines are relatively balanced, indicating that the models have a reasonable trade-off between correctly identifying individuals who took the vaccine and minimizing false positives.
- Similar Accuracy: The accuracy of both models is also comparable, suggesting that they perform similarly overall in correctly predicting vaccine uptake.

Conclusions

- H1N1 vaccine uptake is higher compared to the seasonal flu vaccine, evident from the higher support in the dataset.
- Both vaccines are predicted with similar confidence levels, indicating the effectiveness of the models in predicting vaccine uptake.
- The higher accuracy for the H1N1 vaccine model may be attributed to specific factors strongly correlated with its uptake, such as doctor's recommendations and perceived risk.
- Tailoring vaccination promotion strategies based on vaccine-specific factors is crucial for improving uptake rates.
- Targeted interventions may be needed to address challenges associated with seasonal flu vaccine uptake, as reflected in the slightly lower accuracy of the model.

Recommendations

- Conduct more in-depth analysis to understand why certain demographic groups might be more hesitant to receive vaccines.
- Explore additional data sources that might provide more insights into vaccine hesitancy and uptake, such as social media data or electronic health records.
- Continuously monitor vaccine uptake trends and update the models as new data becomes available to ensure their predictive accuracy remains high.
- Collaborate with healthcare providers and public health agencies to implement targeted vaccination campaigns based on the findings of the models.
- Further analysis could be done to identify the specific demographic or behavioral factors that significantly influence vaccine uptake. This could involve feature importance analysis or exploring additional data sources.

THANK YOU