

# PREDICTING H1N1 VACCINE UPTAKE


By Tracy Otieno





# INTRODUCTION

**Goal:** The project aims to develop a predictive model that can identify the key factors influencing H1N1 vaccine uptake, enabling organizations to design targeted interventions that increase vaccine adoption




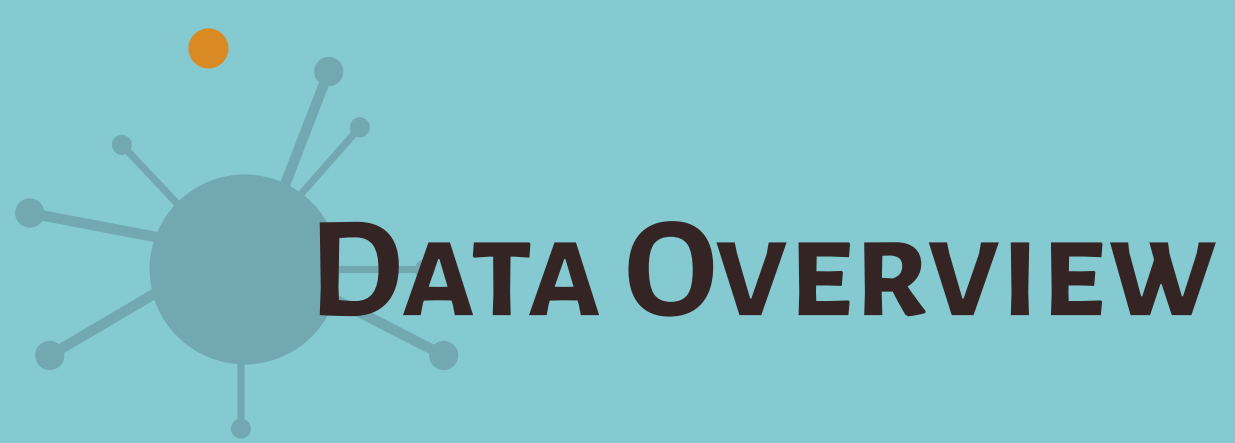


## BUSINESS VALUE

**Why This Matters:** Increasing vaccine uptake is essential for public health, especially in preventing the spread of contagious diseases like H1N1.

**Business Impact:** Understanding the factors that influence vaccine uptake can help health organizations develop effective campaigns, allocate resources more efficiently, and ultimately improve public health outcomes.





## DATA OVERVIEW

**Data Set Size:** 26,707 observations with 38 variables

**Key Variables:** Perceived risk of H1N1, doctor recommendations, behavioral factors, employment status, and health indicators

**Missing Data:** 6% of the data was missing, which was handled appropriately during preprocessing






## KEY INSIGHTS FROM DATA EXPLORATION



**Class Imbalance:** The dataset has a significant class imbalance, with the majority of people not receiving the vaccine.

**High Correlation:** Behavioral factors like large gatherings and time spent outside the home correlate highly with vaccine uptake.

**Perceived Risk & Doctor Recommendations:** People who perceive a higher risk of H1N1 or receive a doctor's recommendation are more likely to get vaccinated.






# METHODOLOGY

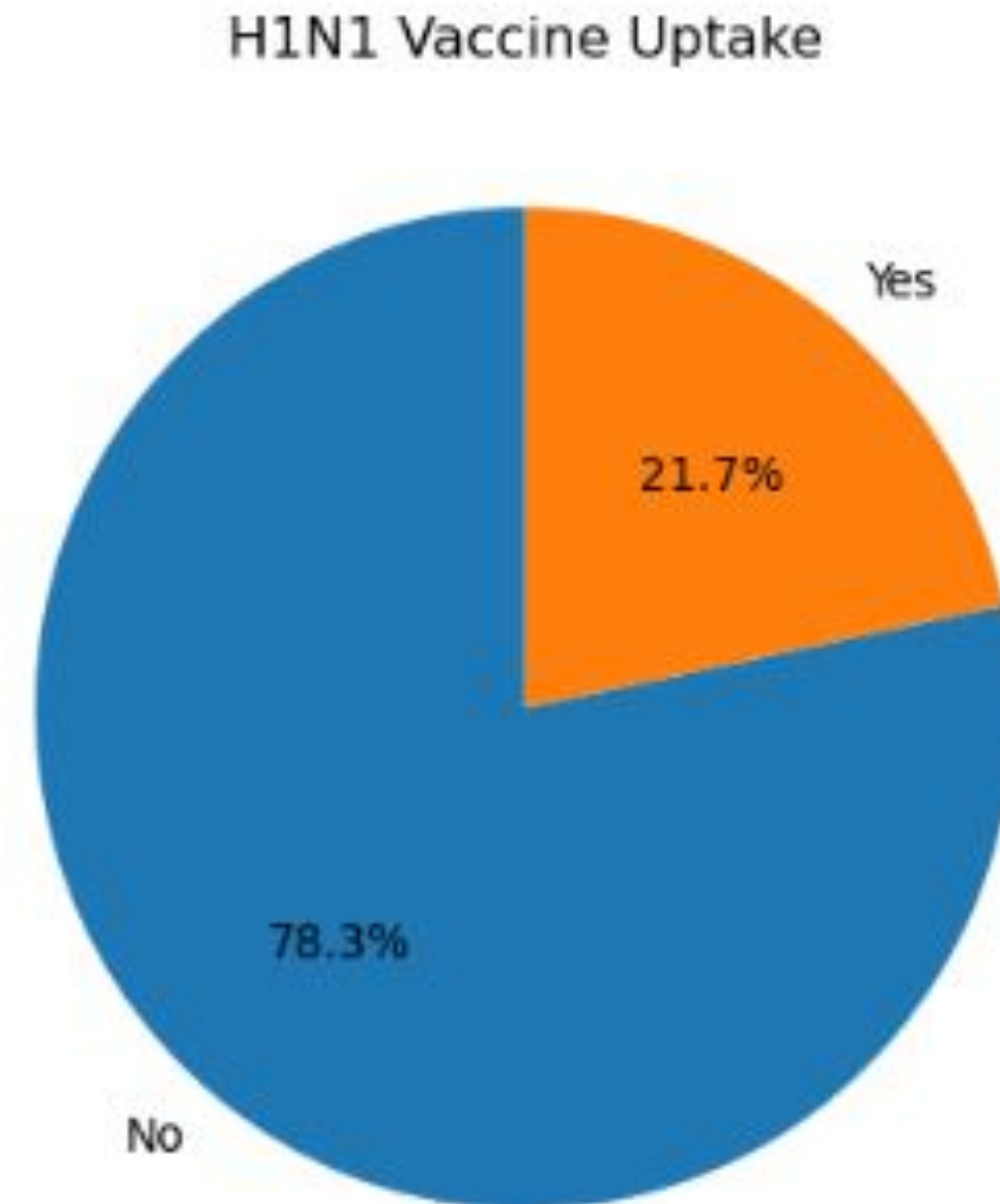
**Data Preprocessing:** SMOTE and NearMiss techniques were used to address class imbalance, followed by scaling of continuous variables using Standard Scaler.

**Model Selection:** Logistic regression, decision tree, and random forest models were tested, with logistic regression emerging as the best performer in terms of accuracy, precision, recall, and F1-score.



# DATA VISUALIZATIONS

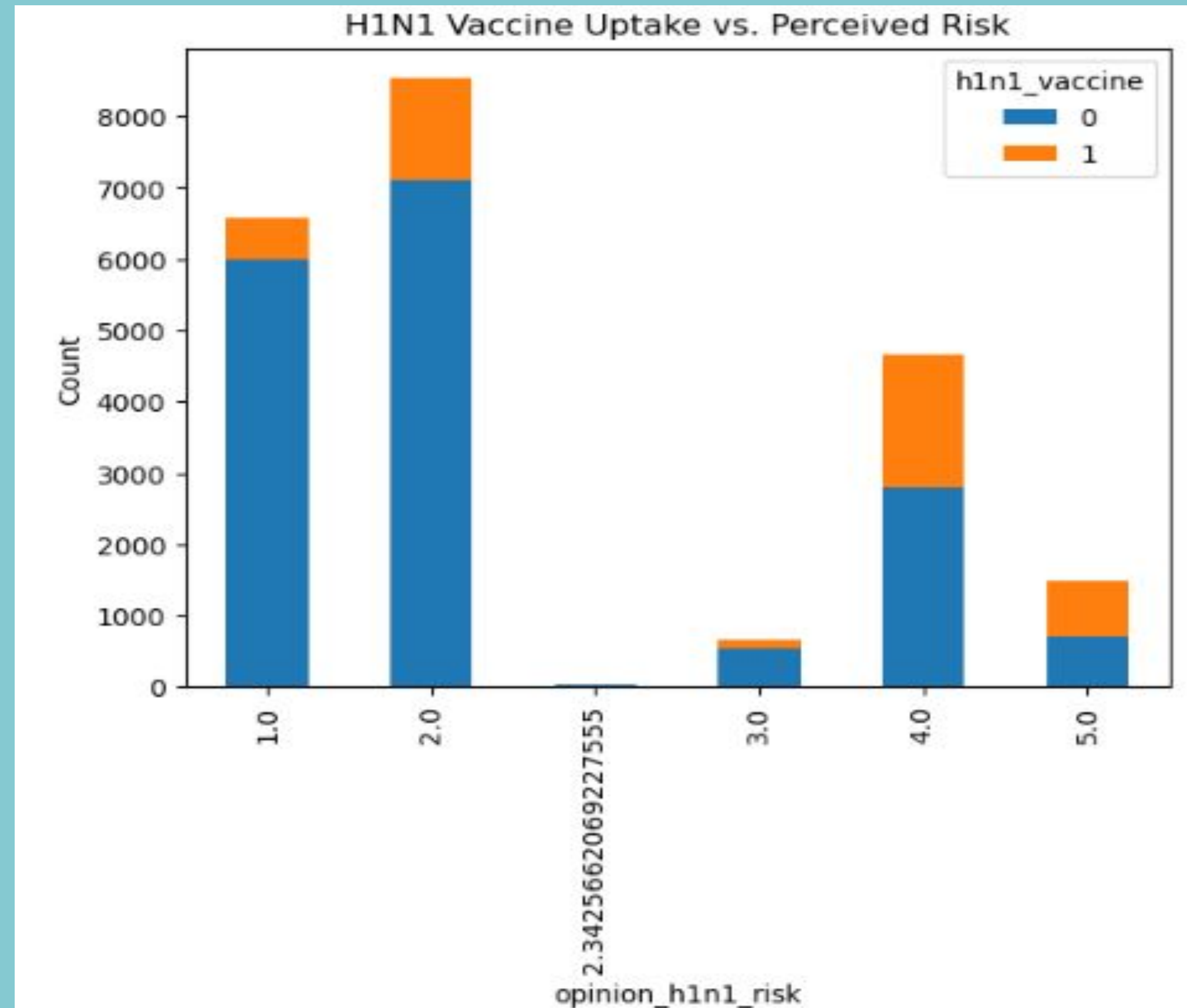
The target variable H1N1 Vaccine has a class imbalance with with class No being the highest at 78.3%





# VACCINE UPTAKE VS. PERCEIVED RISK

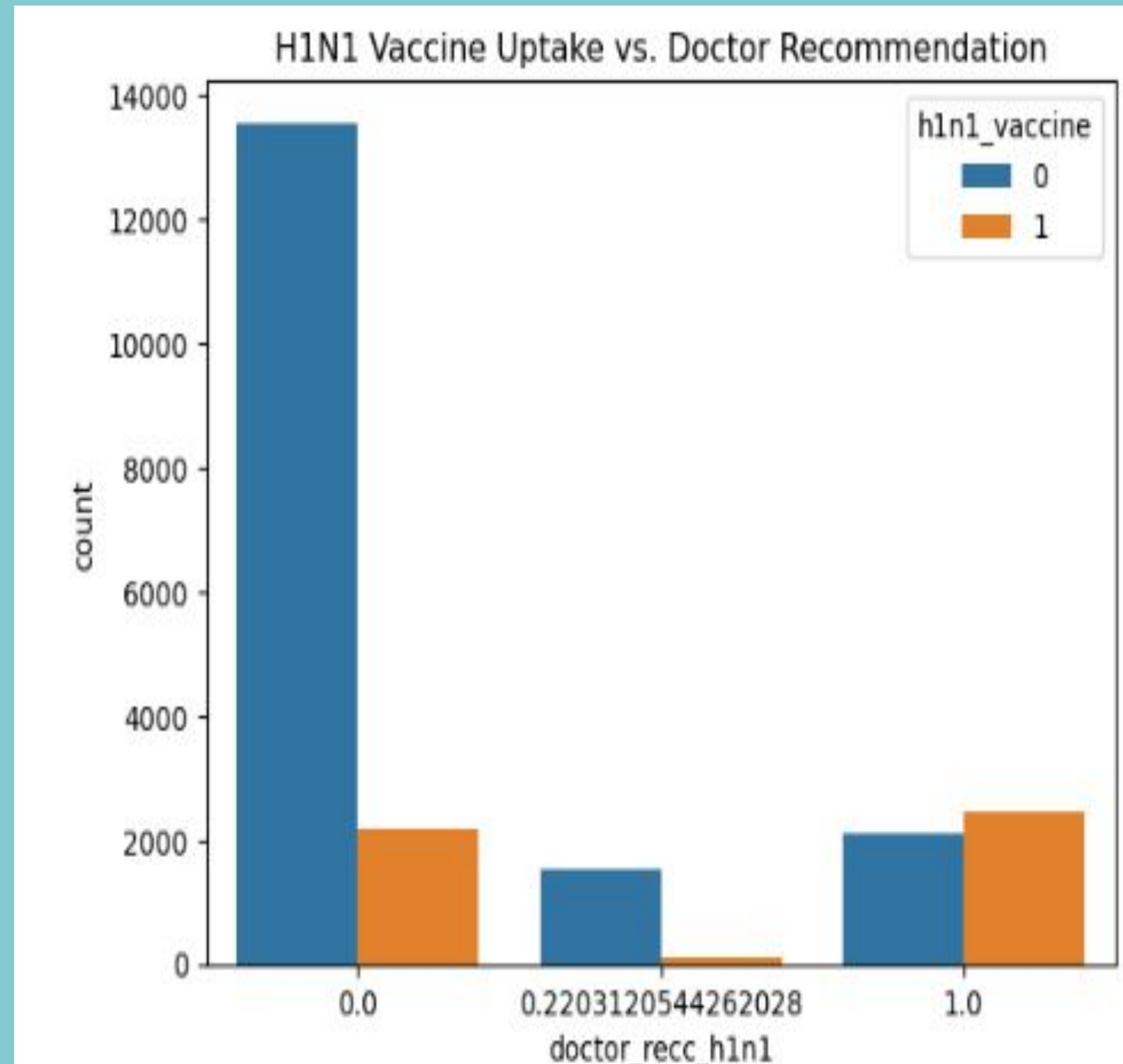
The bar chart illustrates the relationship between the perceived risk of H1N1 (on a scale of 1 to 5) and vaccine uptake. It shows that higher perceived risk (levels 4 and 5) correlates with increased vaccine uptake, whereas lower perceived risk (levels 1 and 2) is associated with significantly fewer people choosing to vaccinate.





# H1N1 VACCINE UPTAKE VS. DOCTOR RECOMMENDATION

The bar chart highlights the influence of doctor recommendations on H1N1 vaccine uptake. Individuals who received a doctor's recommendation (value 1) showed a significantly higher proportion of vaccine uptake than those who did not (value 0), where most chose not to vaccinate.






## MODEL RESULTS

**Best Model:** Logistic Regression with a test accuracy of 87.22%, and balanced performance across both classes (diabetes and non-diabetes).

**Other Models:** Random Forest and Decision Tree showed good performance but had some issues like overfitting or lower accuracy.

**Confusion Matrix:** Logistic regression achieved 3,055 true negatives and 2,942 true positives, indicating good classification performance.






# CONCLUSION

## Key Takeaways:

- Factors like perceived risk, doctor recommendations, and behavioral characteristics are strong indicators of vaccine uptake.
- Logistic regression performed the best, offering a reliable model for predicting vaccine uptake.

**Business Value:** This model helps health organizations target interventions more effectively, especially for at-risk groups, improving vaccination rates.





# RECOMMENDATIONS

- **Increase Awareness of Perceived Risk:** Public health campaigns should focus on raising awareness about the risk of H1N1 to motivate individuals to get vaccinated.
  - **Leverage Doctor Recommendations:** Encourage doctors to recommend the vaccine to patients, as this significantly increases uptake.
  - **Optimize Behavioral Factors:** Incorporate behavioral data (e.g., large gatherings, time spent outside) into targeted interventions.
- 