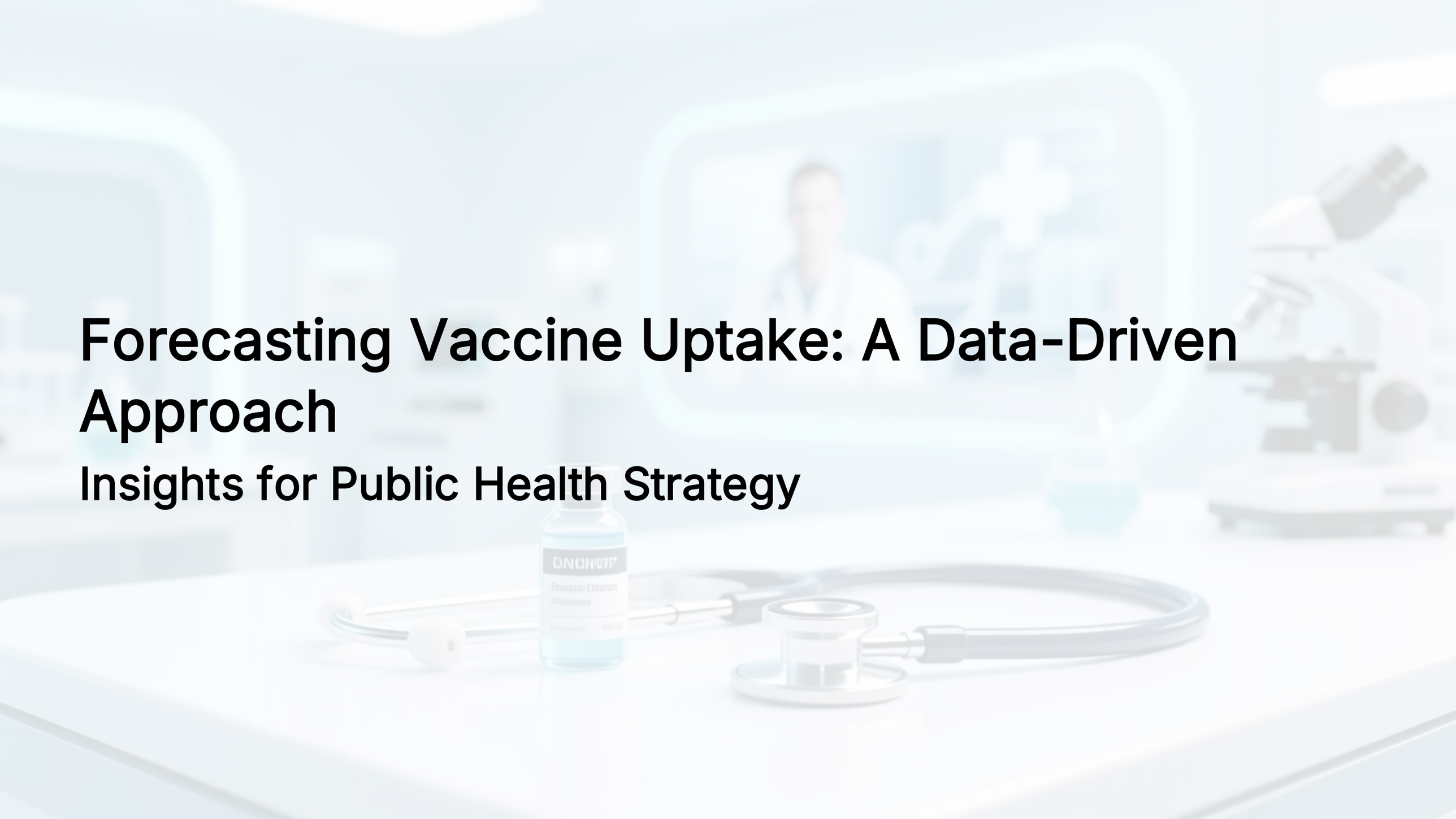
# Forecasting Vaccine Uptake: A Data-Driven Approach

## Insights for Public Health Strategy



# Executive Summary

This report presents a data-driven approach to predicting H1N1 vaccine uptake using machine learning. By analyzing survey data from the National 2009 H1N1 Flu Survey, we developed an optimized XGBoost model that identifies individuals most likely to receive the vaccine. The model achieves 62% recall while maintaining balanced precision, successfully addressing the challenge of imbalanced data (80% unvaccinated, 20% vaccinated). Key findings reveal that physician recommendations are the strongest predictor of vaccine uptake, followed by risk perception and access factors. These insights enable public health officials to allocate limited resources more effectively, prioritize provider outreach, and tailor communication strategies to maximize vaccination rates and protect public health.


# Project Objectives

## Predict H1N1 vaccine uptake

Build a model to identify individuals most likely to receive the H1N1 flu vaccine

## Identify key influencing factors

Understand which social, economic, demographic, and behavioral factors drive vaccination decisions

## Optimize for high recall

Catch the majority of vaccinated individuals to guide targeted public health campaigns

## Enable resource-efficient targeting

Help public health efforts focus limited resources on high-impact outreach

## Support equitable vaccination strategy

Identify underserved populations for targeted intervention

# Data Overview

Our analysis leverages data from the National 2009 H1N1 Flu Survey, providing a detailed foundation for understanding vaccine uptake dynamics during that critical period.

1

### Data Sources

National 2009 H1N1 Flu Survey - a comprehensive phone survey conducted in late 2009 and early 2010 by the United States.

2

### Dataset Scope

Respondents from the National 2009 H1N1 Flu Survey who answered questions about H1N1 and seasonal flu vaccine receipt.

3

### Key Variables

Social, economic, and demographic background; opinions on risks of illness and vaccine effectiveness; behaviors towards mitigating transmission; vaccination status for both H1N1 and seasonal flu vaccines.

4

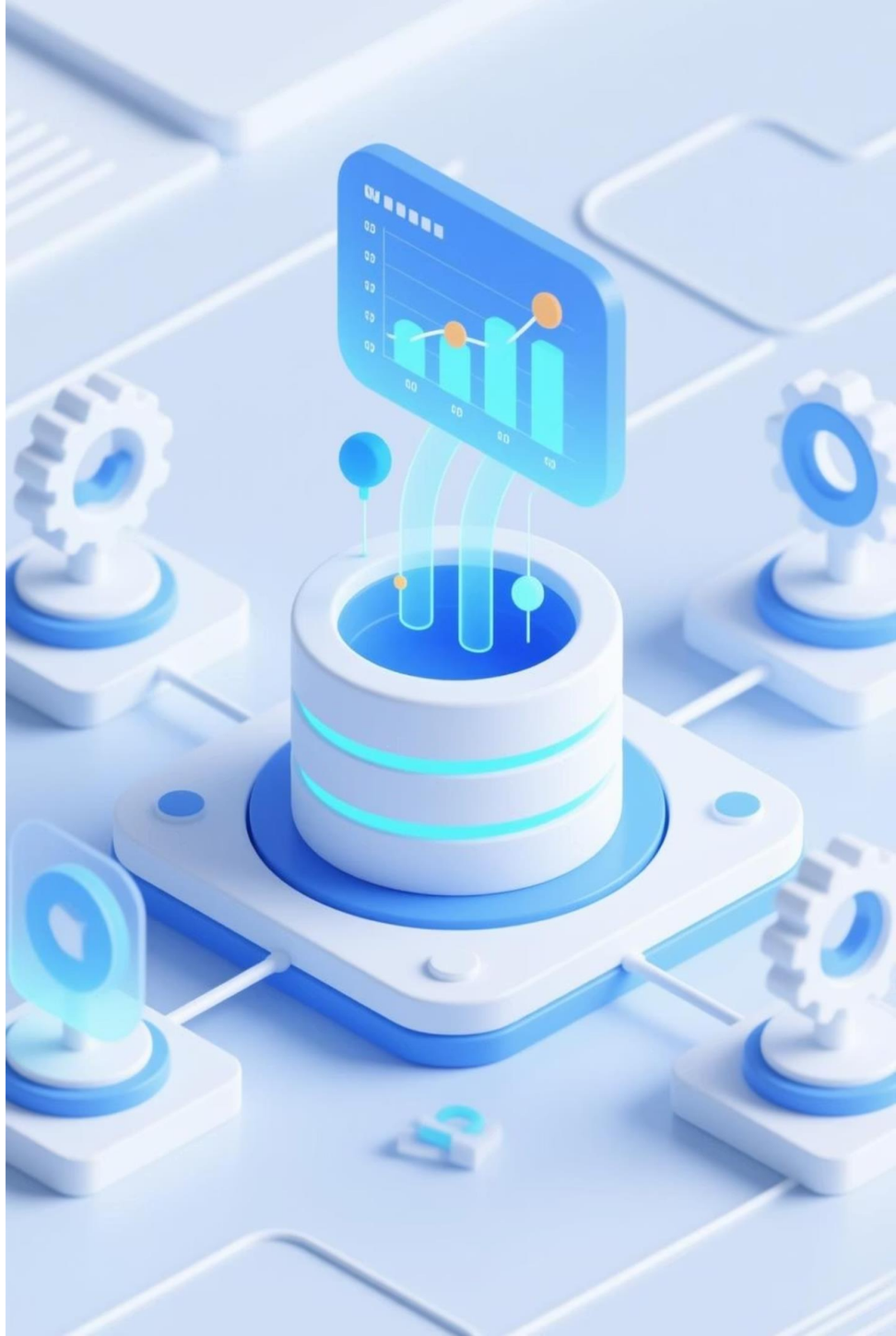
### Data Quality

Clean, validated survey data with comprehensive coverage of respondent characteristics and vaccination outcomes.

5

### Analysis Period

Late 2009 to early 2010, capturing the critical period following the H1N1 vaccine rollout in October 2009.



# Key Findings

### High Recall Achievement

Optimized model achieves 62% recall, catching the majority of vaccinated individuals while maintaining balanced precision

### Doctor Recommendation Dominates

Physician recommendation is nearly 3x more influential than any other factor in predicting vaccine uptake

### Risk Perception Drives Behavior

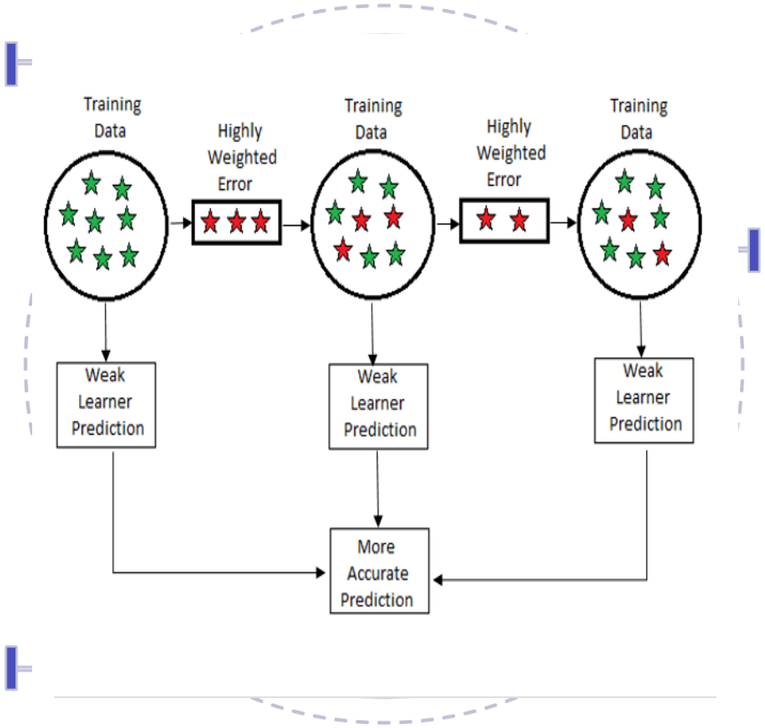
Perception of H1N1 risk outweighs belief in vaccine effectiveness as a motivator

### Imbalanced Data Challenge Solved

Successfully addressed the 80/20 imbalance (80% unvaccinated, 20% vaccinated) through cost-sensitive learning

# Our Winning Model: XGBoost Classifier (Extreme Gradient Boosting)

Feature EngineeringCraft informative predictors




Ensemble LearningCombine weak learners

PredictionProduce probability estimates

This model was selected for its superior performance in handling imbalanced data and achieving high recall rates.

# How Our Model Works

XGBoost works like a team of expert advisors learning from each other's mistakes. Each advisor starts by making predictions, then learns where they went wrong. The next advisor focuses specifically on fixing those mistakes. This process repeats, with each new advisor building on the insights of the previous ones. By combining all their refined predictions, we get an exceptionally accurate forecast. This approach is especially powerful when dealing with imbalanced data—like our vaccine uptake challenge where only 20% of people get vaccinated. The model learns to pay special attention to the minority group, ensuring we don't miss important patterns.

# Model Performance Results

## ROC-AUC

Improved from 0.73 to 0.86

**Drastic improvement in discriminative power**

## Recall

Improved from 52% to 62%

**Caught 10% more positive cases**

## Precision

From 72% to 61%

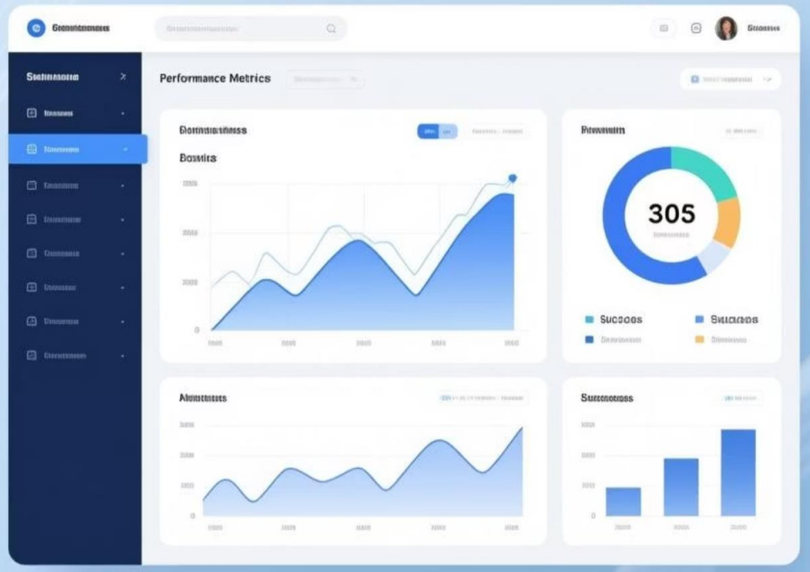
**Accepted trade-off to reduce missed opportunities**

## Accuracy

From 85% to 84%

**Negligible drop**

The optimized model successfully balances the need to catch cases (Recall) with accuracy (Precision), solving the imbalanced data challenge.



# Key Drivers of Vaccine Uptake

## Doctor's Recommendation is King

Nearly 3x more influential than any other factor. If a doctor recommends the vaccine, patients are highly likely to comply.

## Fear Over Logic

Perception of risk outweighs belief in vaccine effectiveness. Patients are more motivated by fear of getting sick than by the logic of the cure.

## Access Matters

Health insurance and access to vaccination sites are important secondary factors.

# Recommendations

### Prioritize Provider Outreach

Shift marketing budget from general awareness to equipping doctors with reminder scripts and talking points, as physician recommendation is the strongest lever for conversion.

### Target Risk-Based Messaging

For populations not regularly visiting doctors, focus campaigns on the susceptibility and severity of H1N1, rather than just vaccine safety.

### Deploy at Threshold 0.60

Implement the model with a strict 0.60 decision threshold to filter out weak false positives while retaining high detection rates.

### Focus on Underserved Populations

Use model predictions to identify and target demographic groups and regions with lower vaccination rates.

### Establish Community Partnerships

Collaborate with community leaders and trusted local organizations to build vaccine confidence in hesitant populations.

# Next Steps

- Implement targeted outreach campaigns in identified regions
- Establish partnerships with community organizations
- Deploy forecasting model for real-time monitoring
- Schedule quarterly reviews to assess progress