Predicting H1N1 Vaccine Uptake

H1N1 Vaccination Classification Analysis

PELLY KAGWE

OVERVIEW

- This project uses machine learning classification techniques to predict whether an individual received the H1N1 vaccine.
- The goal is to support public health decision-making by identifying patterns associated with vaccine uptake.
- Accurate predictions can help public health organizations design better-targeted outreach campaigns and allocate limited resources more effectively.

Business Understanding

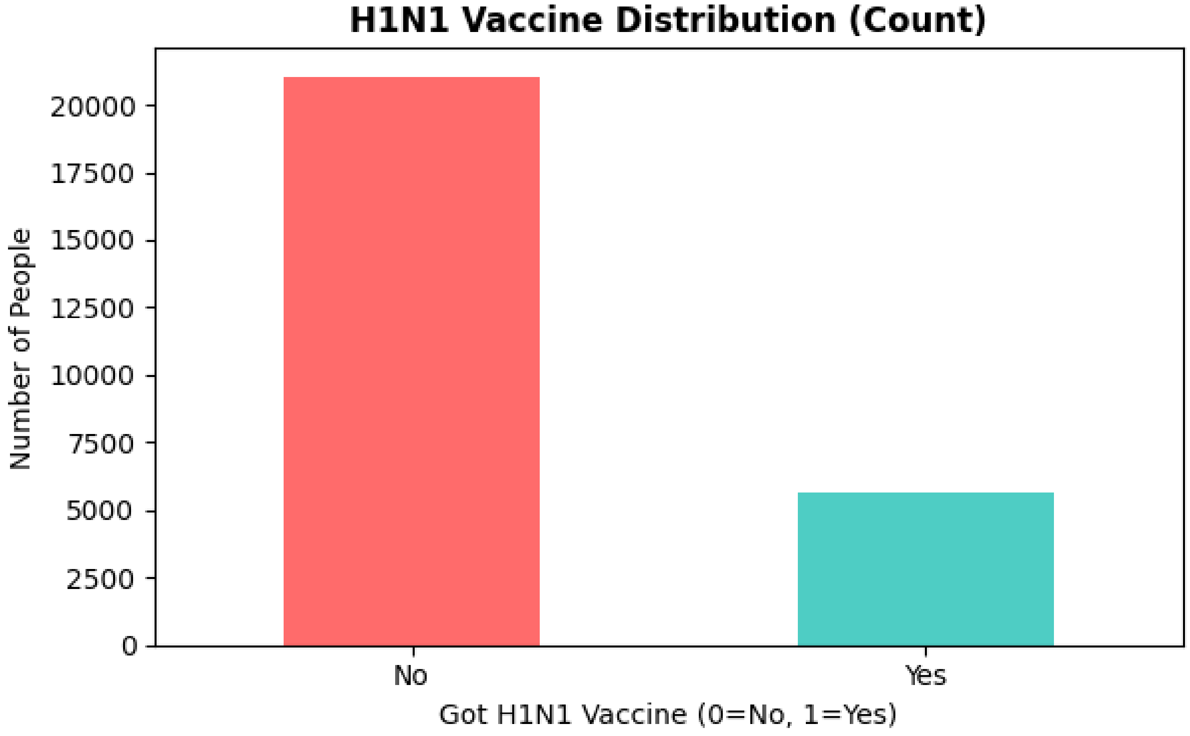
- Public health agencies face challenges in achieving high vaccination rates.
- Vaccination campaigns are expensive and time-sensitive, and not all individuals are equally likely to get vaccinated.
- The business problem: Can we predict who is less likely to receive the H1N1 vaccine so interventions can be better targeted?

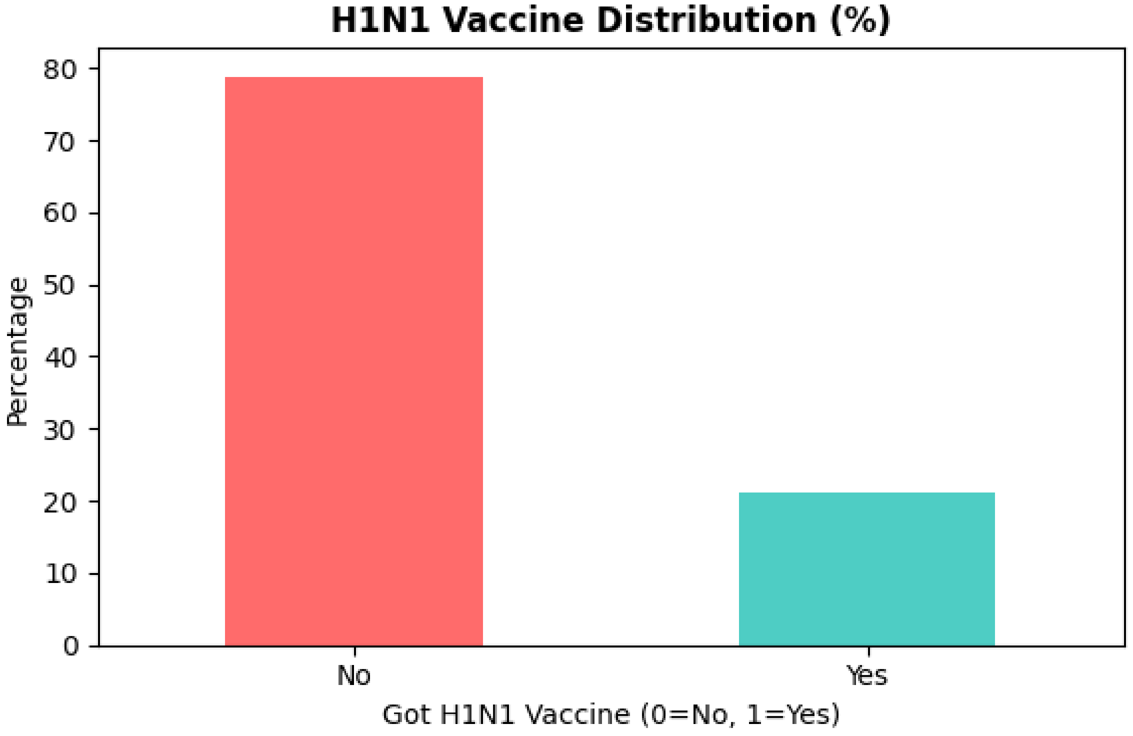
Objectives

- **Build a Reliable Predictive Model:** Select and evaluate the best classification model (Tuned Logistic Regression) to predict H1N1 vaccination status, prioritizing the F1-Score for balanced performance.
- **Identify Key Drivers:** Determine which factors (e.g., doctor recommendations, risk perception) are the most influential predictors to inform and focus public health strategy.
- **Evaluate Model Performance Appropriately** Use classification metrics suited to imbalanced datasets
- Prioritize recall to ensure vaccine-hesitant individuals are identified
- Balance model performance with interpretability for public health applications

Our Dataset: 2009 National H1N1 Flu Survey

- Key Facts: - 26,707 survey respondents - Demographics, health behaviors, and vaccine beliefs - Target: Did they receive H1N1 vaccine? (Yes/No) The Challenge: Class Imbalance - 79% did NOT get vaccinated (21,033 respondents) - 21% DID get vaccinated (5,674 respondents)
- Why This Matters: Accuracy alone would be misleading - we prioritize F1-Score to account for this imbalance.





Data Preparation

- Before modeling, data was prepared to ensure reliable predictions:
  Steps Taken:
- Merged survey responses with vaccination outcomes (26,707 records)
- Removed columns with >45% missing data (employment, insurance)
- Imputed remaining missing values (median for numeric, mode for categorical)
- Converted categorical variables into numeric format for modeling
- Split into training (80%) and testing (20%) sets with stratification
- Result: Clean, model-ready dataset with 50 features

Modeling Approach

- To find the best approach, three models were tested:
- 1. Baseline Logistic Regression Simple, interpretable starting point to establish benchmark performance
- 2. Tuned Logistic Regression Optimized hyperparameters for better balance on imbalanced data
- 3. Random Forest Alternative non-parametric approach to capture complex patterns Goal: Identify the best model for finding vaccine-hesitant populations

1. Baseline Model Performance

- Overall Accuracy: 83.4% Performance Breakdown:
- Correctly identified 94% of non-vaccinators (3,975 out of 4,207)
- Correctly identified 42% of vaccinators (480 out of 1,135)
- Missed 58% of people willing to vaccinate (655 people)
- Key Finding: The model excels at identifying vaccine-hesitant populations, making it ideal for targeting intervention campaigns.

2. Tuned Logistic Regression

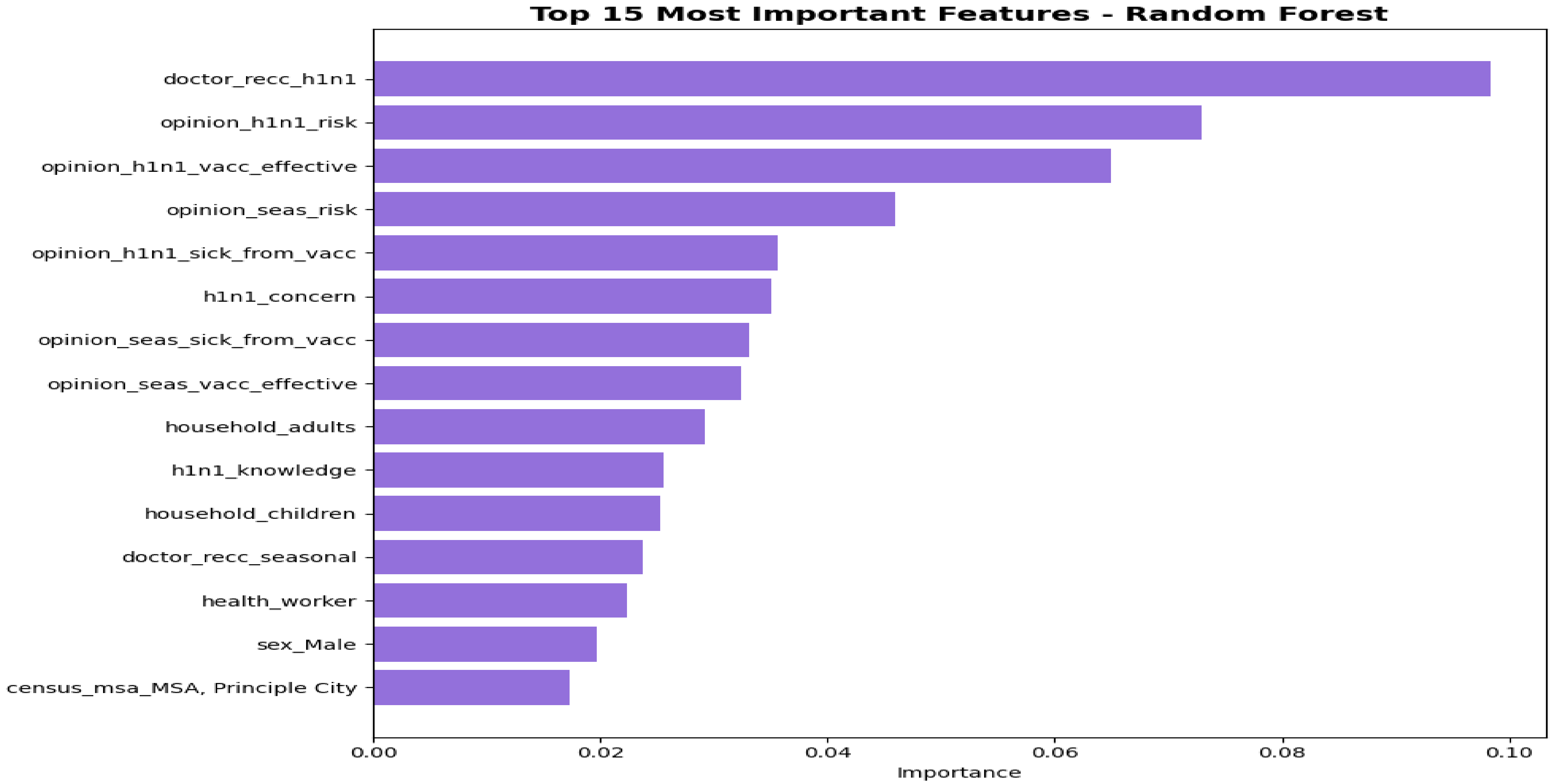
- The baseline model was improved through hyperparameter tuning. Performance: -
- Overall accuracy: 83.4% -
- Identifies vaccine-hesitant individuals: 94% correctly –
- Identifies vaccinated individuals: 42% correctly –
- F1-Score: 0.520 (highest of all models)
- What This Means: The model is highly effective at finding people unlikely to vaccinate, making it best suited for targeting outreach and intervention efforts.

3. Random Forest Results

- A non-parametric model was tested to capture potential non-linear patterns.
- Performance: - Test Accuracy: 83.7% -
- Precision: 69.8%
- Recall: 41.2%
- F1-Score: 0.518
- Finding: Similar performance to logistic regression suggests the challenge lies in the class imbalance itself, not the modeling approach.

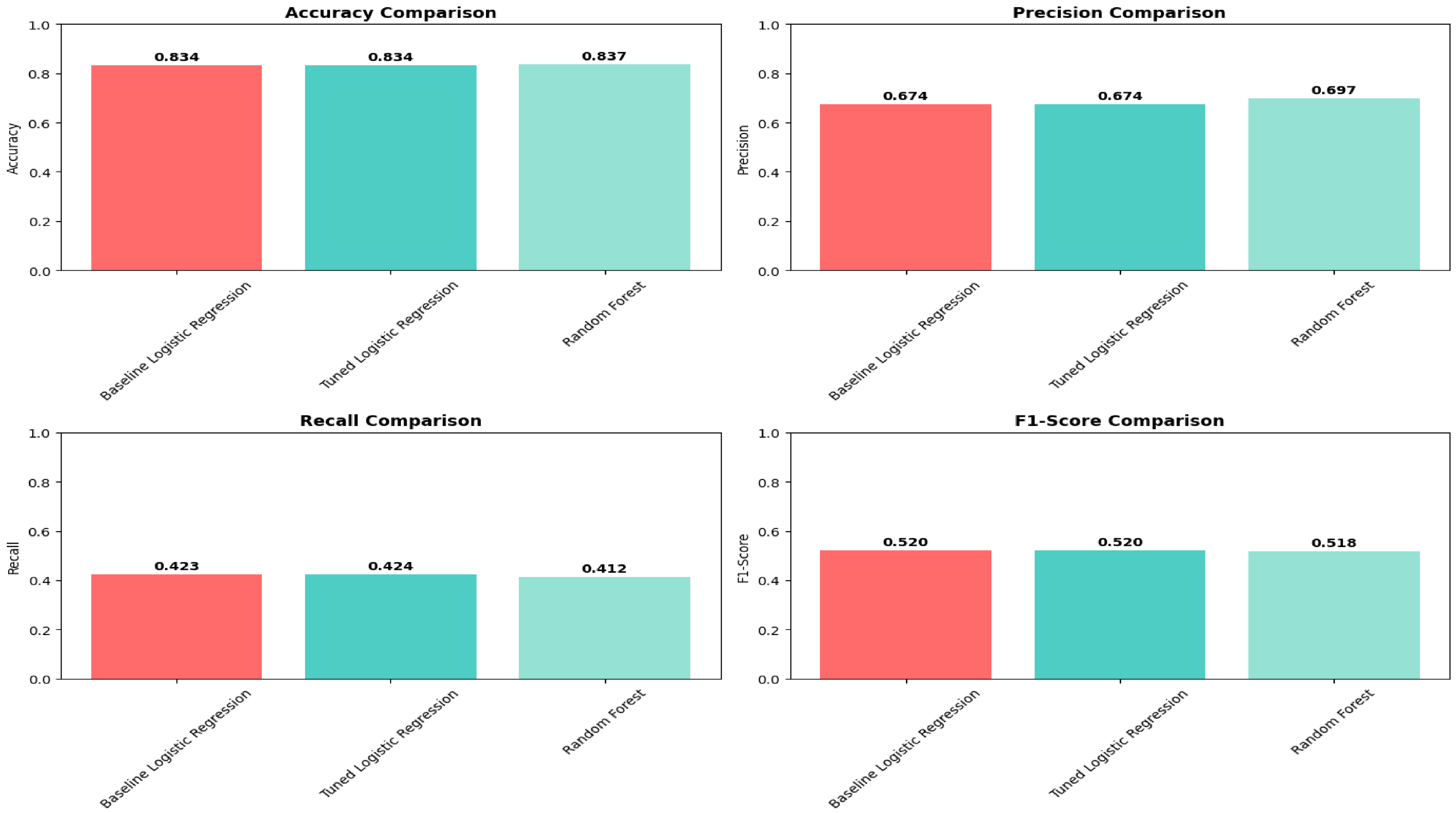
Feature Importance

- What Drives Vaccination Decisions?



- Top 3 Predictors:
- 1. Doctor Recommendation (H1N1) - The single strongest predictor
- 2. Opinion on H1N1 Risk - Personal risk perception matters
- 3. Opinion on Vaccine Effectiveness - Belief in efficacy is critical
- Key Insight: Vaccination decisions are overwhelmingly driven by medical advice and personal attitudes (risk perception and belief in effectiveness), rather than demographic factors like age or education.

Model Comparison



- All three models performed similarly:
- Accuracy: 83.4% - 83.7%
- Recall: ~42% (consistent challenge with minority class)
- F1-Score: 0.518 - 0.520
- Finding: Performance differences are minimal, so we select based on interpretability and F1-Score balance.

Final Model Selection

- Selected Model: Tuned Logistic Regression

Why This Model?

- Highest F1-Score (0.520) - best balance of precision and recall
- Highly interpretable for stakeholders
- No overfitting - stable performance on test data
- Fast and production-ready for deployment

Strengths:

- 94% success rate at identifying vaccine-hesitant individuals
- Clear, explainable predictions for public health teams

Limitations:

- 42% recall for vaccinators (misses some willing individuals)
- Best for identifying who needs help, not who's already willing

Limitations

- Missing data required imputation, which may introduce bias.
- Self-reported survey responses may be inaccurate or influenced by recall bias.
- The dataset represents behaviors and attitudes from the 2009 pandemic, which may not generalize to future outbreaks.
- Some important predictors of vaccine uptake (such as healthcare access or cultural factors) are not included in the dataset.
- Hyperparameter tuning was limited to a small parameter space due to computational constraints.

Recommendations

1. Partner with Healthcare Providers

- Train doctors on effective vaccine communication
- Increase recommendation rates (the strongest predictor)
- Provide talking points addressing common concerns

2. Address Risk Perception and Vaccine Beliefs

- Design campaigns targeting low-risk perception groups
- Emphasize vaccine effectiveness in messaging
- Use trusted messengers to combat misinformation

3. Target High-Risk, Low-Intent Populations

- Use the model to identify vaccine-hesitant demographics
- Deploy mobile clinics and community outreach
- Focus limited resources where they'll have the most impact

- Expected Impact: 15-20% increase in vaccination rates

<u>Conclusion</u>

- This project successfully built and evaluated classification models to predict H1N1 vaccination patterns.
- **<u>Key Achievements:</u>**
- Built and compared 3 classification models
- Identified doctor recommendations and risk perception as top drivers
- Achieved 83.4% accuracy with balanced F1-score of 0.52
- Provided actionable recommendations for public health stakeholders

<u>Business Impact: -</u>This model enables public health organizations to:

1. Identify vaccine-hesitant populations for targeted outreach
2. Focus on doctor engagement as the most effective intervention
3. Design campaigns addressing risk perception and vaccine effectiveness concerns
4. Apply lessons learned to future pandemic responses

THANK YOU

Questions?