Predicting Vaccine Uptake Using Machine Learning

- Phase 3 Classification Project

- • Predicting H1N1 and Seasonal Flu Vaccination
- • Data-Driven Public Health Insights
- • End-to-End Predictive Modeling

Business Problem

- The Challenge:
- • Identify groups less likely to get vaccinated
- • Improve outreach strategies
- • Allocate resources efficiently

- Objective:
- Build a predictive model to estimate vaccination probability.

Why This Is a Classification Problem

- • Target = Yes (1) or No (0)
- • We predict categories, not numeric values
- • Output is probability between 0 and 1

- Binary Classification Task

Data Overview

- • ~26,700 respondents
- • 36 features per respondent
- • 2 target variables

- Feature types:
- • Demographics
- • Health behavior
- • Risk perception
- • Geographic info

Modeling Approach

- 1. Data understanding
- 2. Cleaning & preprocessing
- 3. Train/validation split
- 4. Baseline model
- 5. Hyperparameter tuning
- 6. Ensemble comparison
- 7. Final model selection

Models Built

- 1. Logistic Regression (Baseline)
- • Interpretable
- • Fast

- 2. Tuned Logistic Regression
- • Optimized performance

- 3. Random Forest (Ensemble)
- • Captures nonlinear relationships

Evaluation Metrics

- Due to class imbalance we used:

- • ROC-AUC (ranking quality)
- • F1 Score (precision & recall balance)
- • Accuracy (secondary metric)

Key Results

- • Tuned model improved baseline performance
- • Random Forest captured complex patterns
- • Vaccination behavior is predictable

- Strong signal found in behavioral & demographic features.

Key Insights

- Strong predictors included:
- • Age group
- • Health worker status
- • Risk perception
- • Income & education
- • Geographic region

Business Recommendations

- 1. Target younger & lower-income groups
- 2. Increase risk-awareness campaigns
- 3. Tailor messaging by region
- 4. Use probability scores for outreach prioritization

Limitations

- • Survey-based self-reported data
- • No time-series information
- • Not causal, only predictive
- • Some missing data challenges

Next Steps

- 1. Deploy as decision-support tool
- 2. Integrate into outreach platforms
- 3. Track outcomes & retrain periodically
- 4. A/B test intervention strategies