Who's getting vaccinated?

Hyewon Jeong
@ Metis

Motivation

- **Vaccination** is the **most effective way** to prevent infectious diseases
- Despite evidence of the effectiveness and safety of vaccinations, some people in the U.S. are reluctant to receive vaccinations.

Vaccine hesitancy Problem

27 % **Of adults**

in the U.S. would NOT get a COVID-19 vaccination


Question
Who is getting vaccinated?

About Data

The National 2009 H1N1 Flu Survey

A phone survey asked respondents whether they had received the H1N1 vaccine (swine flu), including followings:

- social, economic, and demographic backgrounds
- opinions on risks of illness and vaccine effectiveness
- behaviors towards mitigating transmission

Details

- Target: whether vaccinated for H1N1
  - 22 % of target variables - positive
- 35 Categorical Features

Methodology

STEP 1

EDA & Pre-Processing

- Missing data imputation (i.e., mode)
- Categorical variable treatment (one hot encoding vs. ordinal encoder)

STEP 2

Baseline & Feature Engineering

- Feature importance
- Class imbalance (SMOTE, Weighting, Threshold)
- Tuning hyperparameter (Grid Search)

STEP 3

Model Comparison

- Logistic Regression
- Naïve Bayes
- Random Forest
- XGBoost

STEP 4

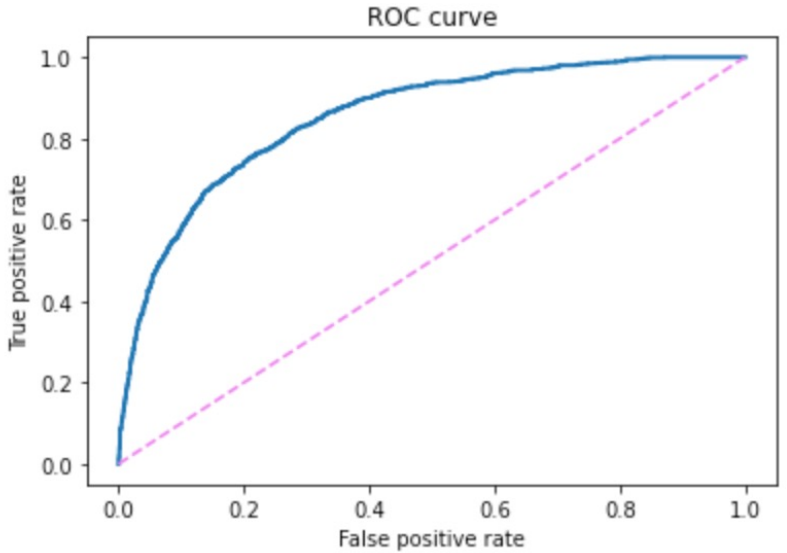
Best Performance Model Selection

Baseline

Logistic Regression

Recall Score: 0.442

ROC AUC Score: 0.854



Notable Feature Engineering:

Class imbalance + Grid Search CV

Logistic Regression

Recall Score

0.46 ⟶ **0.73**

Naïve Bayes

Recall Score

0.57 ⟶ **0.69**

Random Forest

Recall Score

0.58 ⟶ **0.74**

Model Comparison

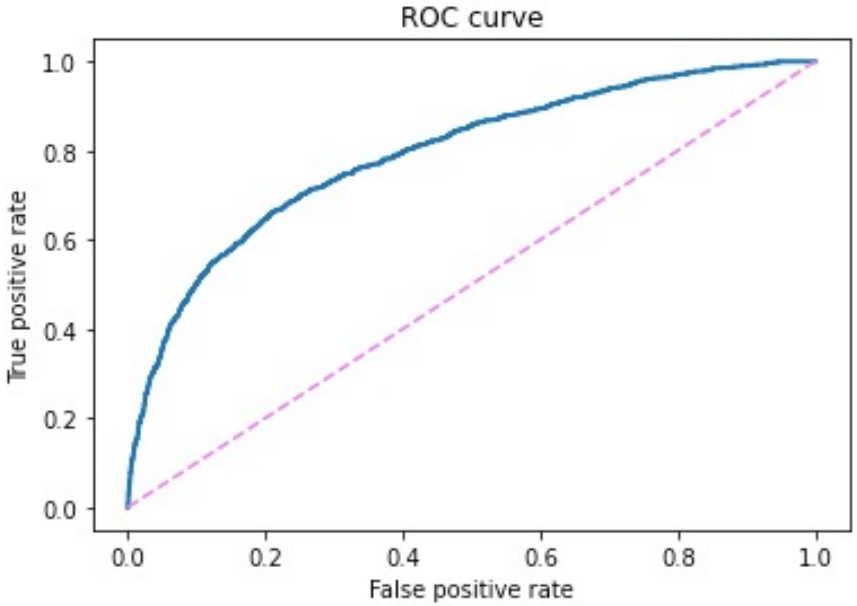
(hyperparameters in all models except for Baseline were tuned with Grid Search CV)

Scores	Recall
Baseline	0.442
Logistic Regression (with selected features)	0.696
Logistic Regression (with all features)	0.730
Naïve Bayes	0.687
Random Forest	0.742
XGBoost	0.470

BEST Performance Model

Random Forest

The Score on Test Data: 0.75
The Score on Recall : 0.70

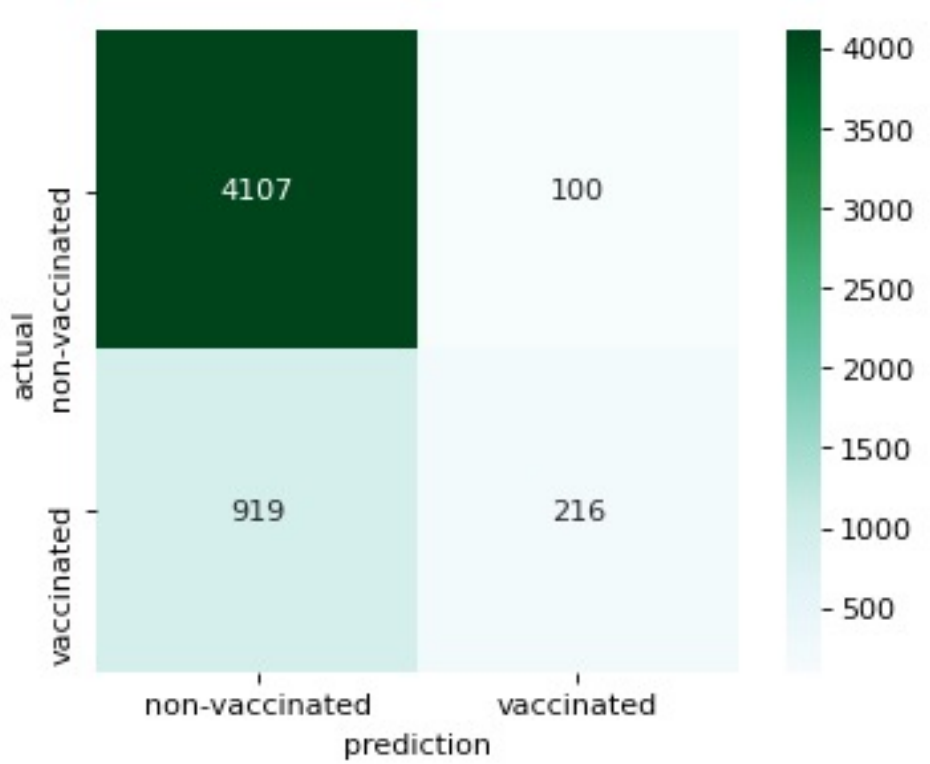


ROC AUC score: 0.80

BEST Performance Model

Random Forest

The Score on Test Data: 0.75
The Score on Recall : 0.70



Future Studies

Advanced Feature Engineering 01

- Feature selection
- Further aggregating or grouping categories
  - Employment occupation and industry
  - geolocation

Other Ensembling Methods 02

Thank you!

Contact Information:

email: heeyewonj@gmail.com
github: hyewonjng