Predicting Vaccination Willingness

Heeyewon Jeong
Metis

Motivation

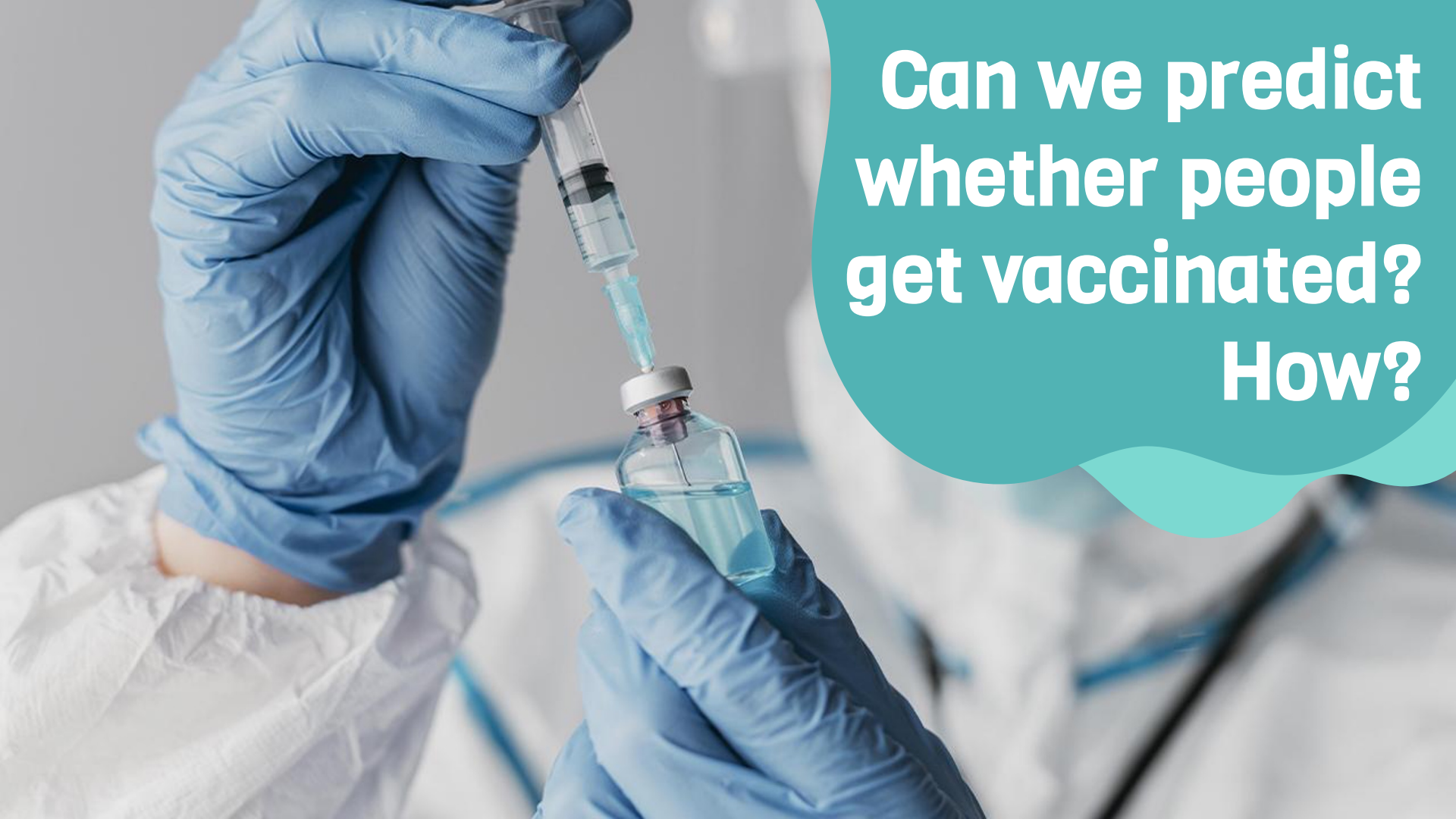
- Vaccination is the most effective way to prevent infectious diseases
- Despite evidence of the effectiveness and safety of vaccinations, a number of people in the U.S. are reluctant to receive vaccinations.

Vaccine hesitancy

- Vaccine hesitancy gained worldwide attention as the COVID-19 pandemic sept the globe
- Many citizen in the United States were unwilling to take vaccinations against COVID-19 due to the following reasons:
  - Concerns about safety
  - Side effects
  - General mistrust of the government

27 %

Of adults in the U.S. would definitely not get a COVID-19 vaccination



Can we predict whether people get vaccinated? How?

About Data

The National 2009 H1N1 Flu Survey

A phone survey asked respondents whether they had received the H1N1 vaccine, 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

- Missing data imputation (i.e., mode)
- Categorical variable treatment (dummy vs ordinal variables)

STEP 2

Baseline & Feature Engineering

- Feature importance
- Class imbalance
- Tuning hyperparameter

STEP 3

Model Comparison

- Logistic Regression
- Naïve Bayes
- Random Forest
- XGBoost

STEP 4

Best Performance Model Selection

Baseline

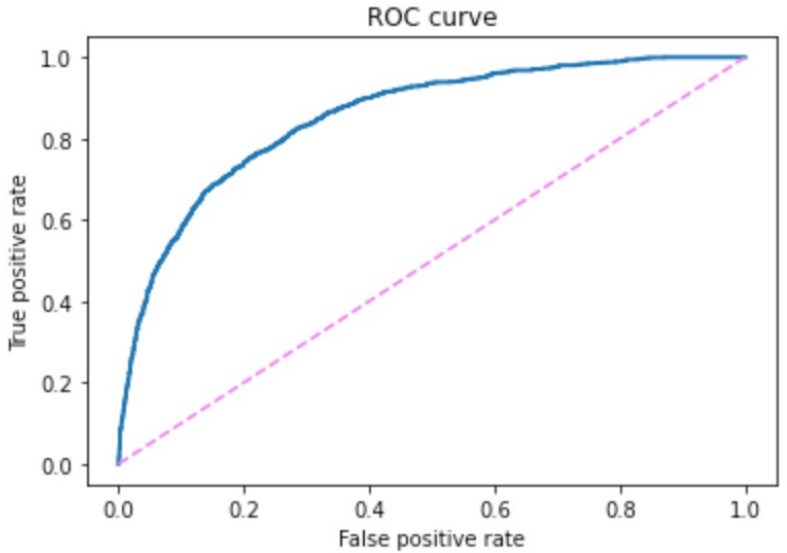
Logistic Regression

Score on Training : 0.835

Score on Validation: 0.848

Recall Score: 0.442

ROC AUC Score: 0.854



Model Comparison

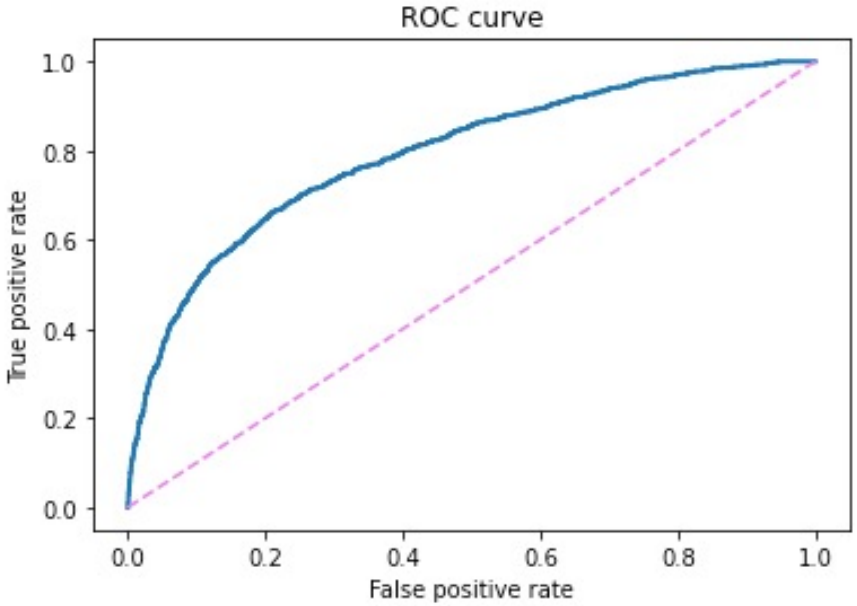
(hyperparameters in all models except for Baseline were tuned with Grid Search CV)

Scores	Train	Validation	Recall
Baseline	0.835	0.848	0.442
Logistic Regression (with selected features)	0.829	0.843	0.403
Logistic Regression (with all features)	0.835	0.843	0.461
Naïve Bayes	0.791	0.801	0.573
Random Forest	0.791	0.811	0.058
XGBoost	0.850	0.844	0.470

Best Performance Model

Naïve Bayes

The Score on Test Data: 0.802
The Score on Recall : 0.558



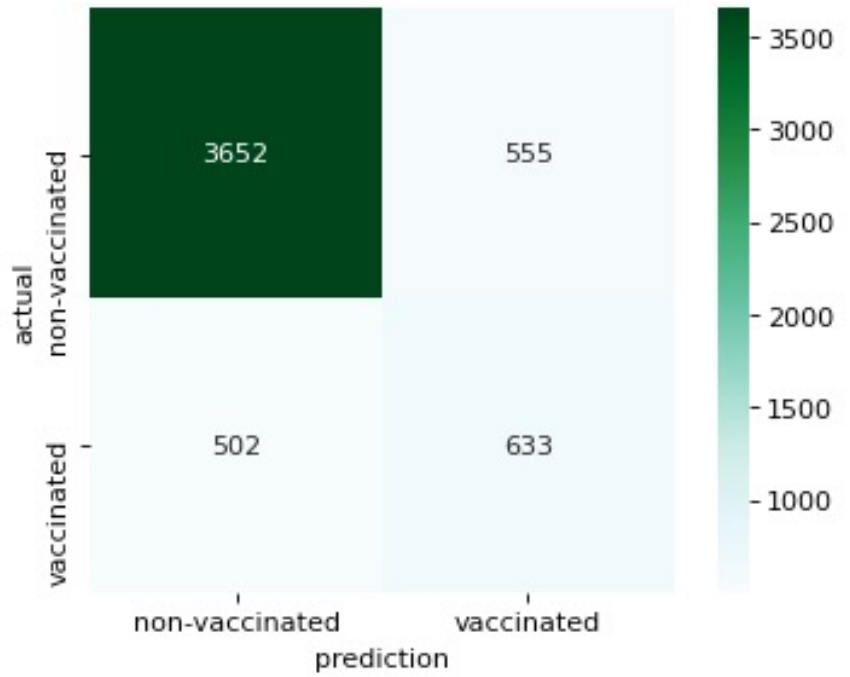
ROC AUC score: 0.792

Best Performance Model (con't)

Naïve Bayes

The Score on Test Data: 0.802
The Score on Recall : 0.558

ROC AUC score: 0.792



Thank you!