Vaccination Drivers:

PREDICTING WHO GOT THE 2009 H1N1 PANDEMIC FLU VACCINE

Vaccination Drivers
Outline

- Business Problem
- Data Understanding
- Models
- Results
- Recommendations

Business Problem

- As the current COVID-19 pandemic demonstrates, getting people to take recommended vaccinations can be an absurdly difficult prospect.
- We're becoming increasingly aware of person and social factors influencing their decisions, but can personal medical and demographic factors help us predict who might or might not vaccinate?
- By accurately predicting those individuals less likely to adhere to medical advice during pandemics, we can focus our outreach strategies to combat influences that run counter to prevailing medical practice.

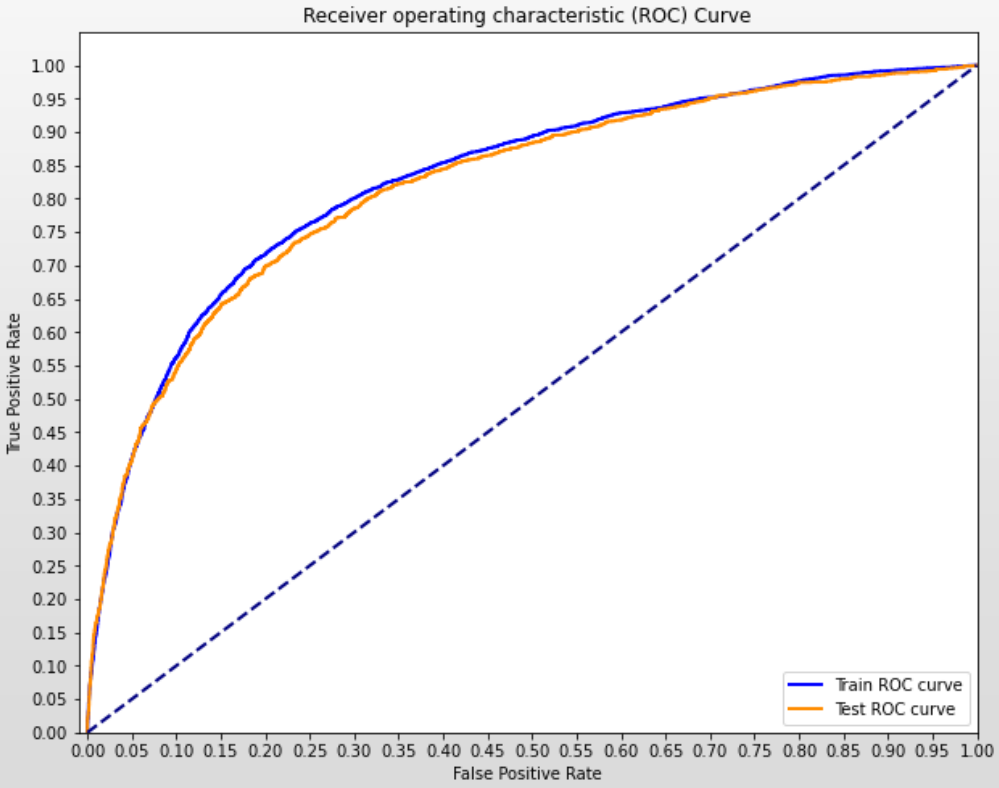
Vaccination Drivers: Data Understanding

- Data from three files: a training features, training labels, and test labels
- Training features dataset contained both categorical and numerical features, and required relatively extensive cleaning. Some features pertained to a class that was not pertinent to our problem, and were removed

Class	Training Data
Did not take vaccine (0)	21033
Took vaccine (1)	5764

Vaccination Drivers
Logistic Regression Model

- Basic classification algorithm
- Logistic regression is best used for binary classification problems—i.e. there are only two classes we're attempting to predict
- Acceptable accuracy and AUC score, but has issues predicting the minority class. However, given our business problem, this isn't the issue it first appears to be.



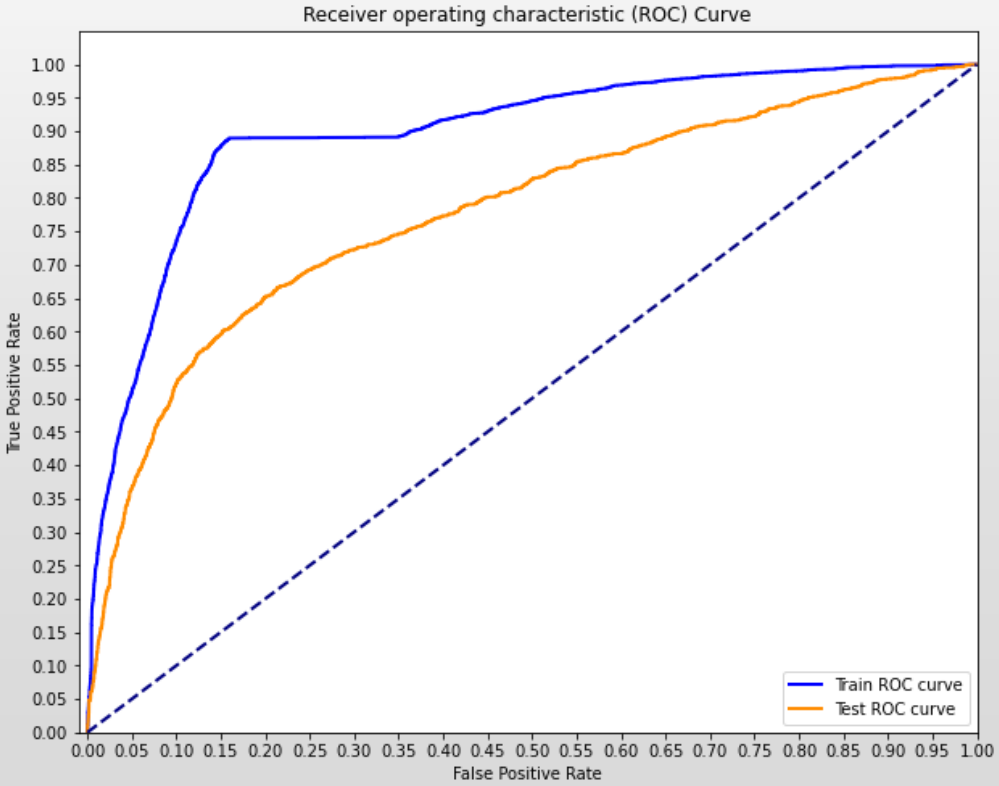
ROC curve for Logistic Regression Model

Vaccination Drivers SVM model

Notes

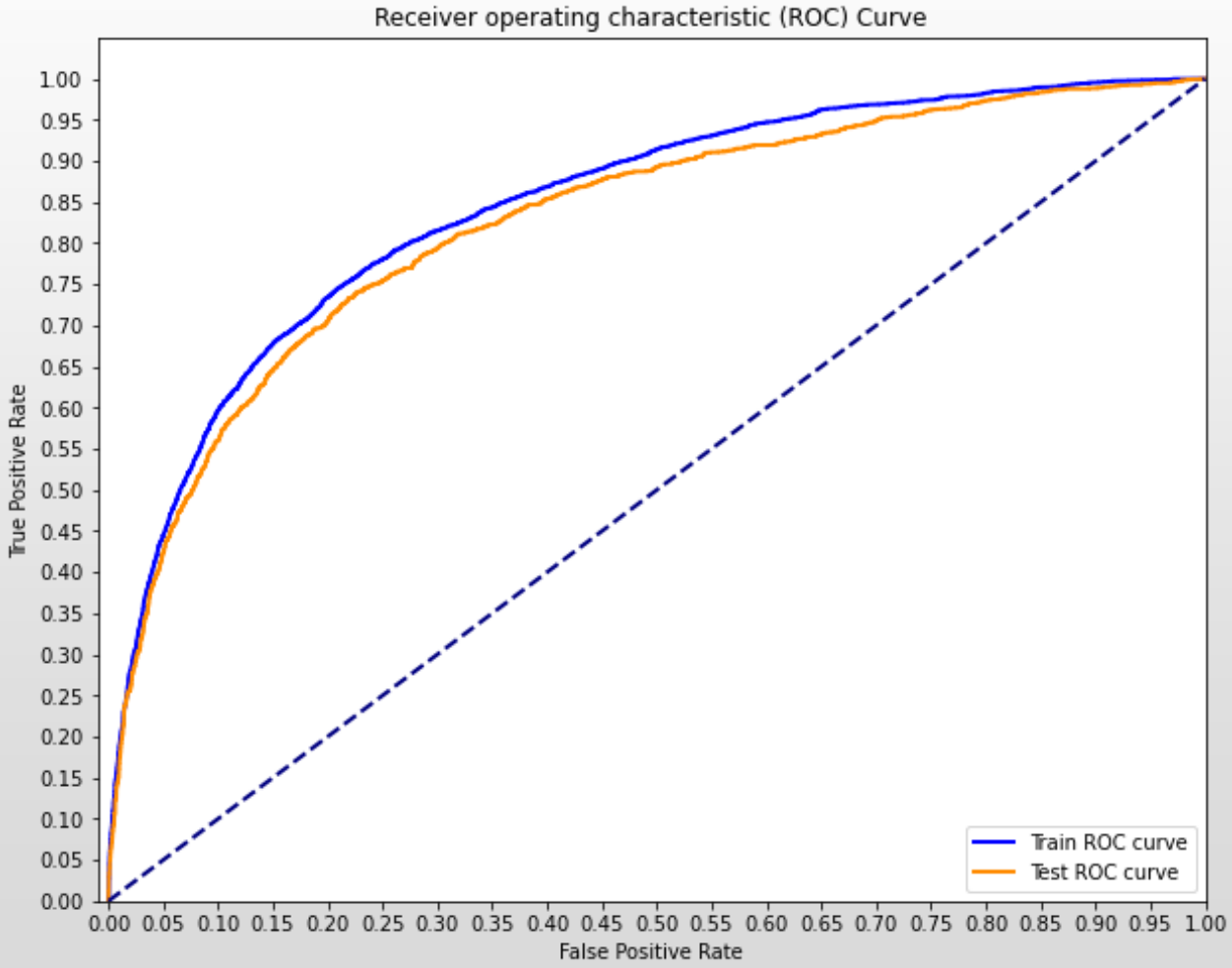
- SVM are more complicated models that are often quite accurate
- Not great for large datasets
- Results slightly worse than logistic regression in vanilla model. Changing kernel to ‘poly’ helped bump AUC up 0.01-0.02 points.

ROC Curve for Support Vector Machine Model



Vaccination Drivers
Gradient Boosting Classifier

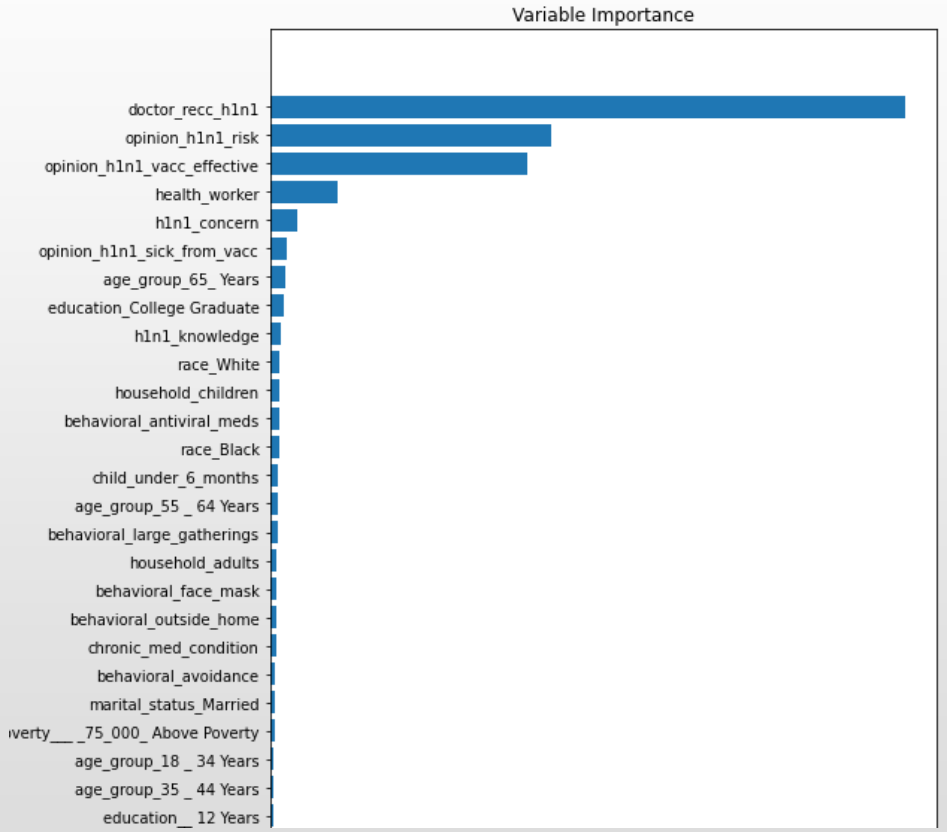
- GradientBoostingClassifier is an ensemble of decision trees…
- Results similar to the other models, with only a slightly higher auc score of 0.83
- As with other models, it handled the minority class poorly with a very narrow range of specificity-sensitivity balance



Results

Some findings:

- Doctor recommendation was by far most influential in the decision to vaccinate
- Socio-economic features also at play, as lower-income and POC tending not to vaccinate
- Few differences in rates across age and genders
- Final Class Predictions
  0 (unvaccinated): 23,200
  1 (vaccinated): 3,508



Cropped plot of feature importances

- As noted previously, we were unable to achieve an auc score higher than 0.83.
- The class data was moderately skewed, with 1's representing a true minority of responses.
- It was not included in the notebook, but we had at one point used SMOTE, and it had no effect whatsoever on test auc.
- The issue may lie in lack of features: there simply isn't enough that makes classes distinct enough to score higher.

Vaccination Drivers

Limitations and for-next-times

Thank you!

Questions?
Comments?

Matthew Noonan

mcn9284@gmail.com

925-917-0396