

Predicting Seasonal Flu Vaccination

Max Steele

Purpose

Guiding Questions for Predictive Modeling

1. Which factors have the strongest relationships with vaccination status?
 2. How can I make accurate predictions for both classes, while maximizing recall for non-vaccinators?
 3. Which population groups should pro-vaccine campaigns focus on?
-

Obtain, Scrub, & Explore

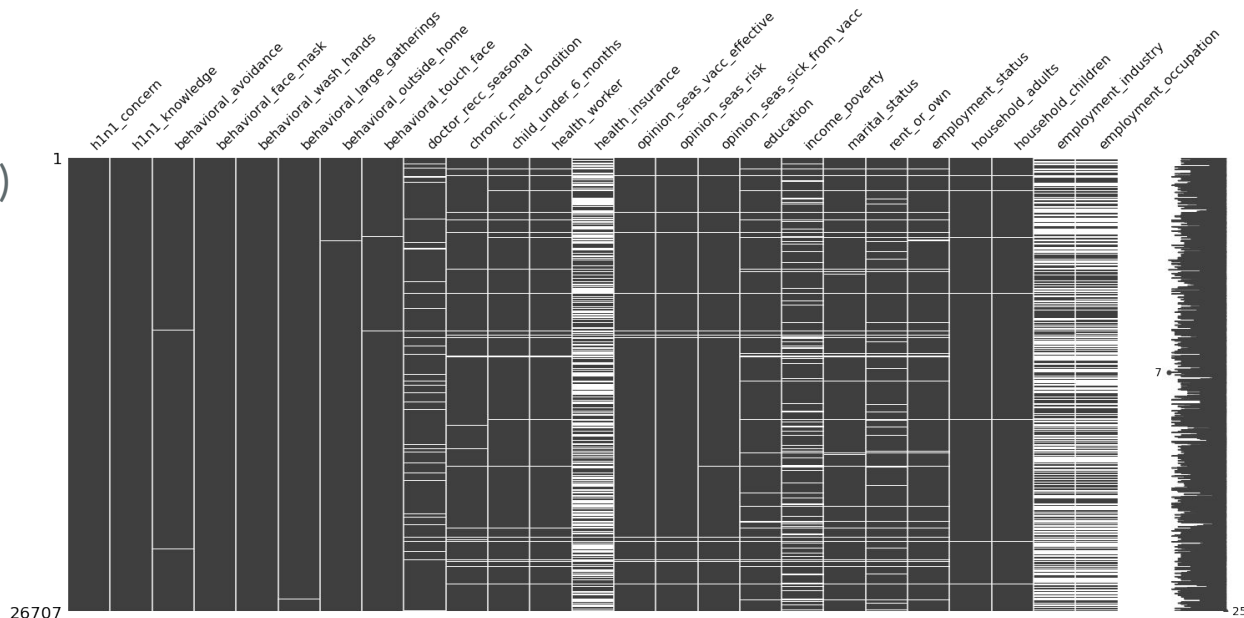
- National 2009 H1N1 Flu Survey

- Trends in missing values

- < 5% of column, records dropped (6.6%)
- > 5% of column, filled with 'missing'

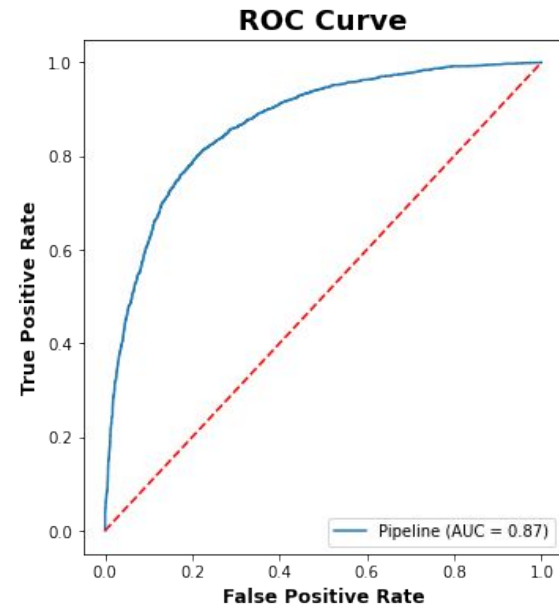
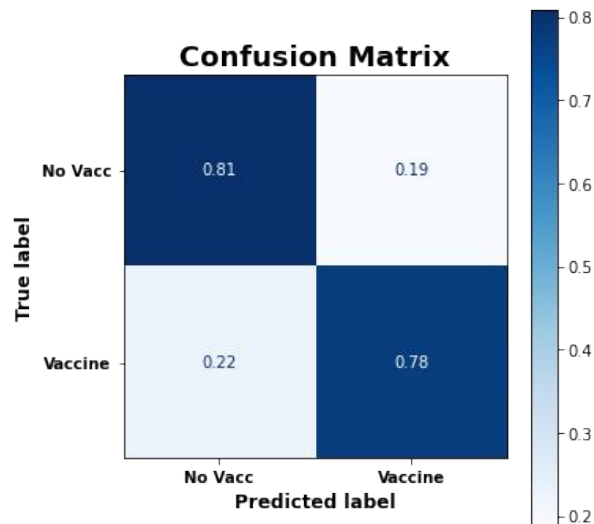
- Feature Engineering

- behav_score
- behav_to_risk
- high_risk_cat



Model

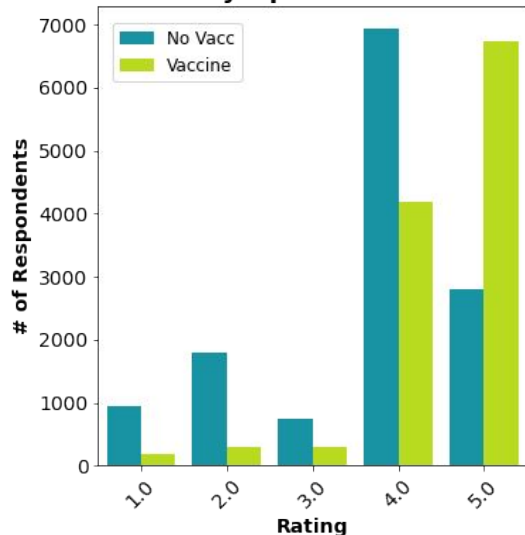
- Compared multiple classifiers:
 - Logistic regression
 - Decision trees
 - Random forests
 - XGradient Boost
 - Stacking
- Evaluated based on
 - Accuracy
 - Recall of No Vacc
 - ROC AUC
- Top models
 - XGradient Boost
 - Stacking classifier



Stacking Classifier Final Estimator Coeffs	
Best Log Reg	1.621
Best Random Forest	1.289
Best XGB	2.621

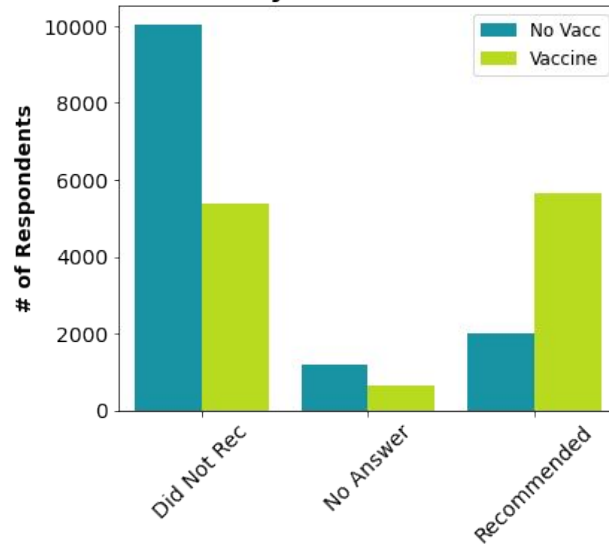
Recommendations - Top 2 Predictors

Vaccination By Opinion of Effectiveness



- Increase public awareness of vaccine effectiveness
 - Evidence-based

Vaccination By Doctor Recommendation

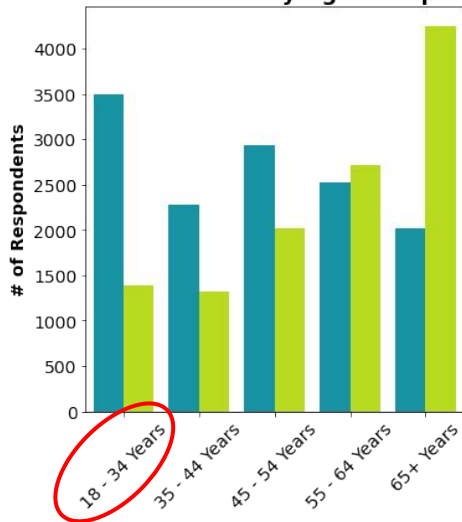


- Doctors need to recommend every year

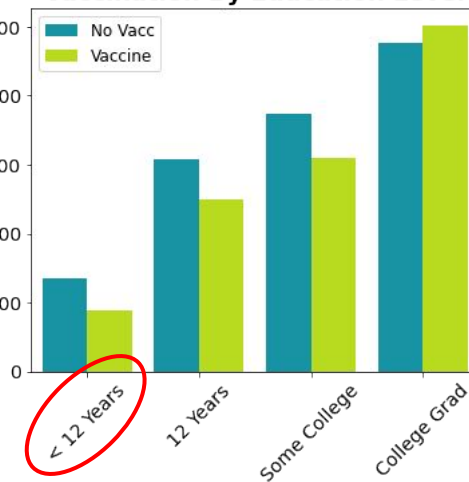
Recommendations - Key Demographics

- Encourage and provide easy, affordable access to vaccines for individuals
 - 18 -34 years old
 - Did not graduate high school
 - No health insurance
 - Below poverty line

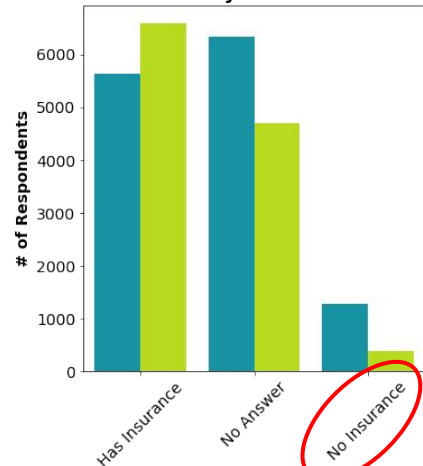
Vaccination By Age Group



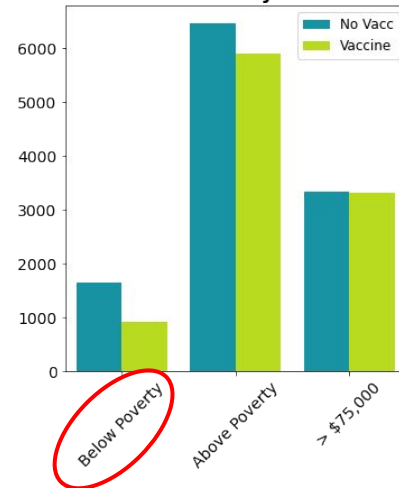
Vaccination By Education Level



Vaccination By Health Insurance



Vaccination By Income



Summary

- Data collected on seasonal flu vaccination in 2009 during H1N1 pandemic
 - Opinion of effectiveness and doctor recommendation strongly influence choice to vaccinate
 - Focus on vaccinating:
 - 18 -34 years old
 - Did not graduate high school
 - No health insurance
 - Below poverty line
-

Future Work

- Update model by adding more recent data
 - Attitudes have changed towards vaccination
- Find/ collect more data to address underrepresentation of minorities

Thank you!

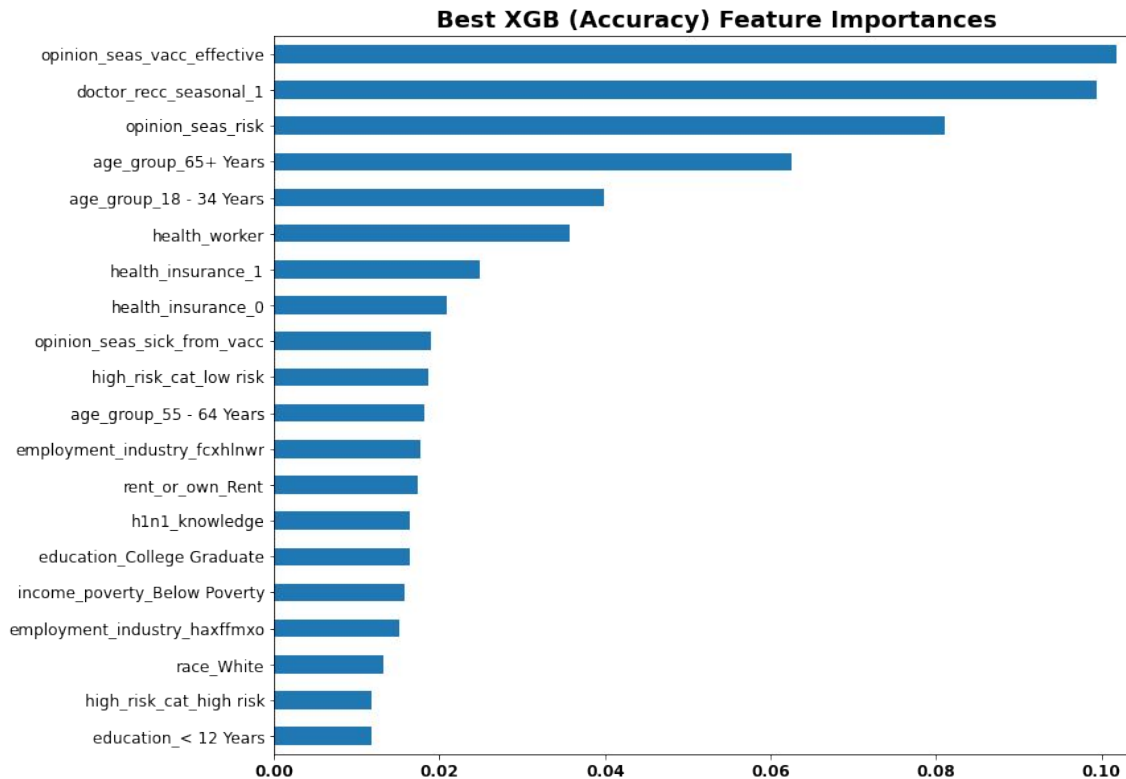
Questions?

Max Steele

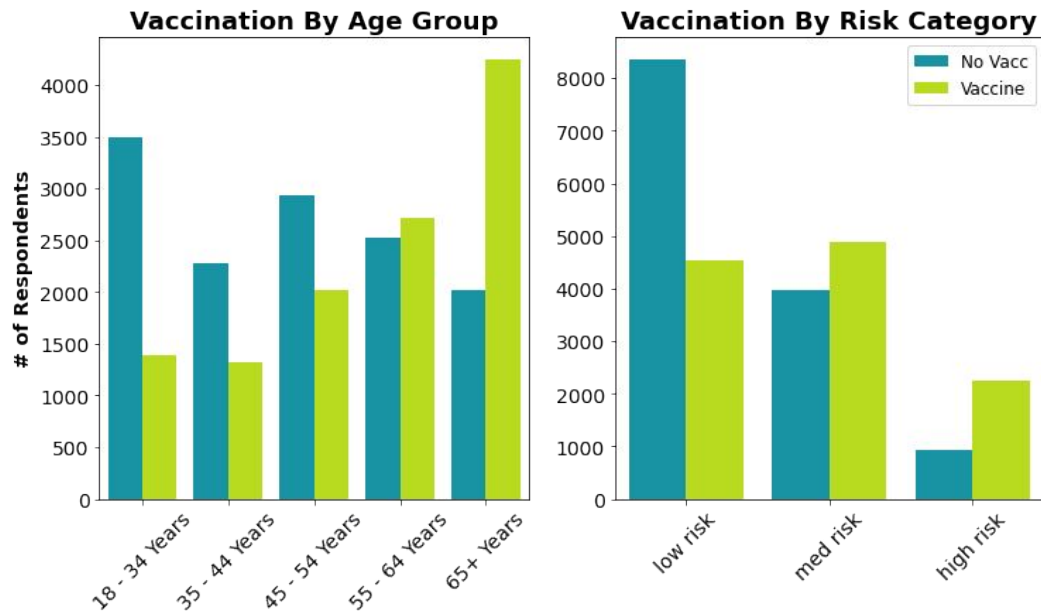
<https://github.com/zero731>

Appendix I

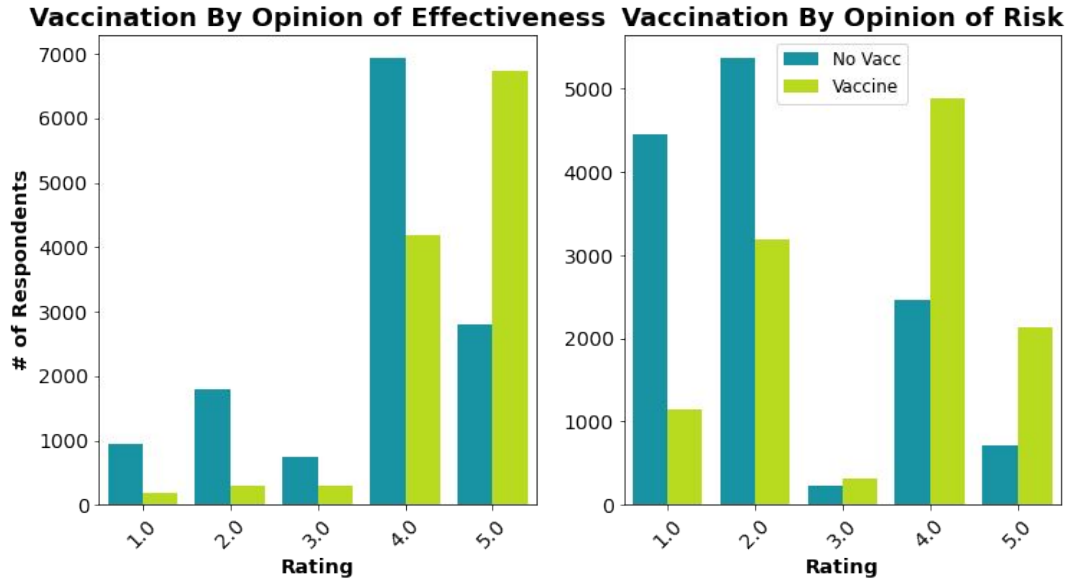
- Top predictors of vaccination status in best model:



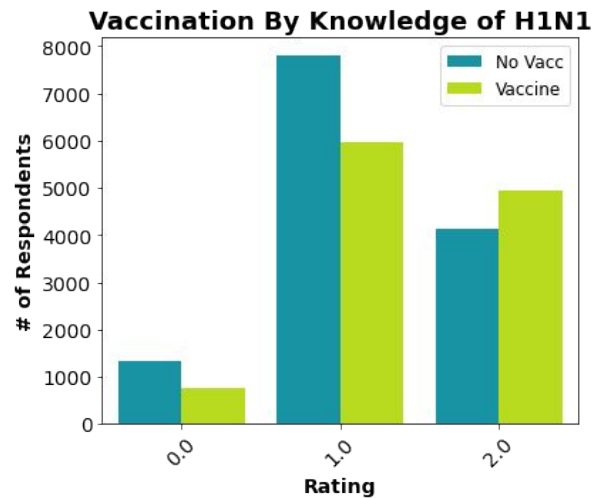
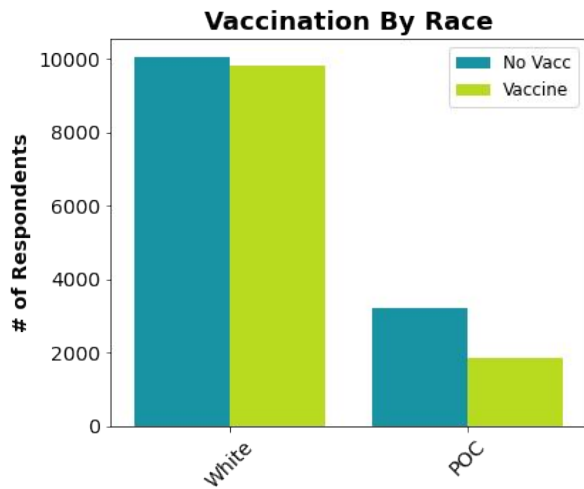
Appendix II



Appendix III

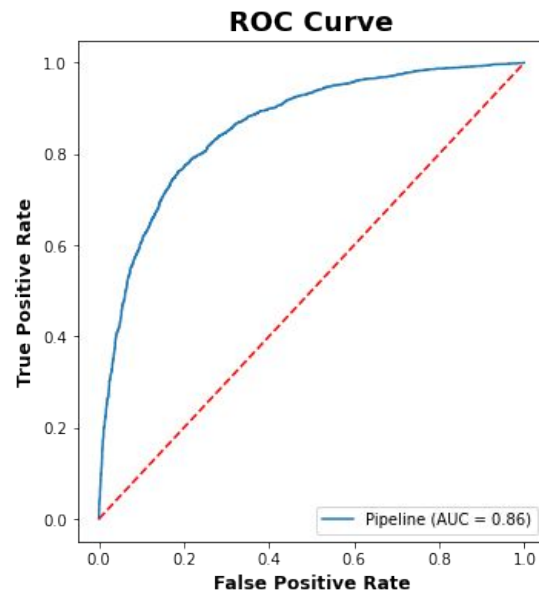
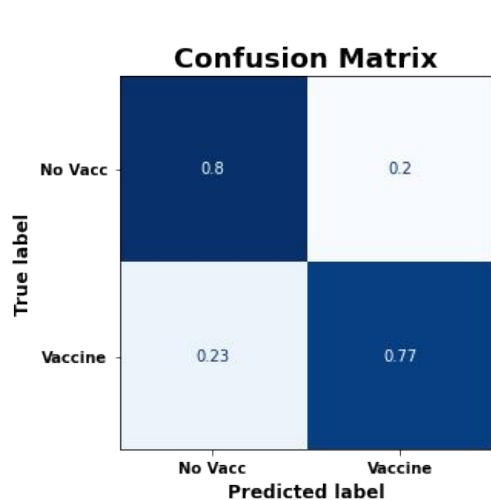


Appendix IV



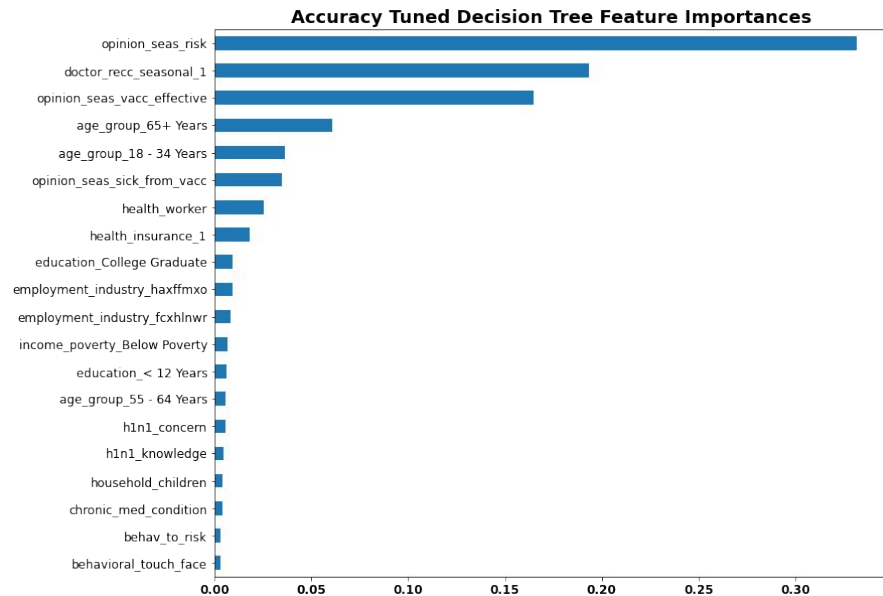
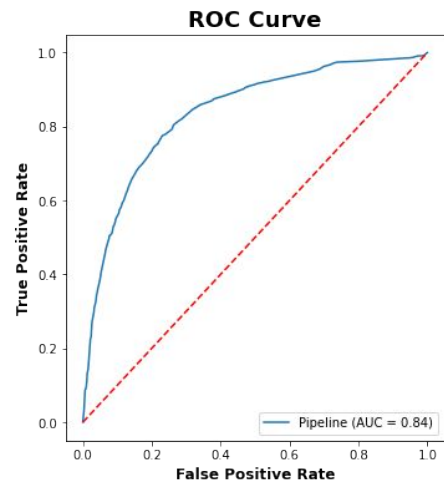
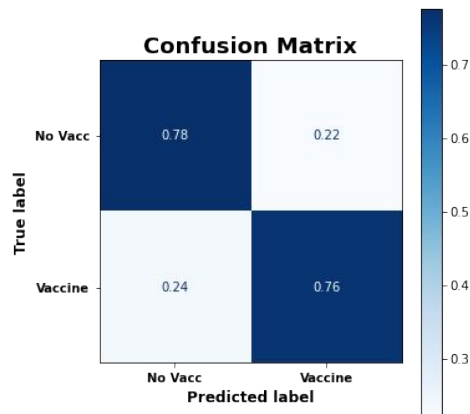
Appendix V

Best Logistic Regression



Appendix VI

Best Decision Tree



Appendix VII

Best Random Forest

