A Smart Health Records Application

USING MACHINE LEARNING FOR PERSONALIZED, INTERACTIVE HEALTH

Smart Health Records App

Outline

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

Smart Health Records App

Business Problem

- Multiple heath networks create incomplete, spotty health records. Increased uncertainty, conflicting treatments/procedures.
- Medical Records Database aggregates and combines patient records across healthcare systems
- We will demonstrate the app's value with a simple scenario involving predicting vaccination status

Smart Health Records App

Vaccination Scenario Data Understanding

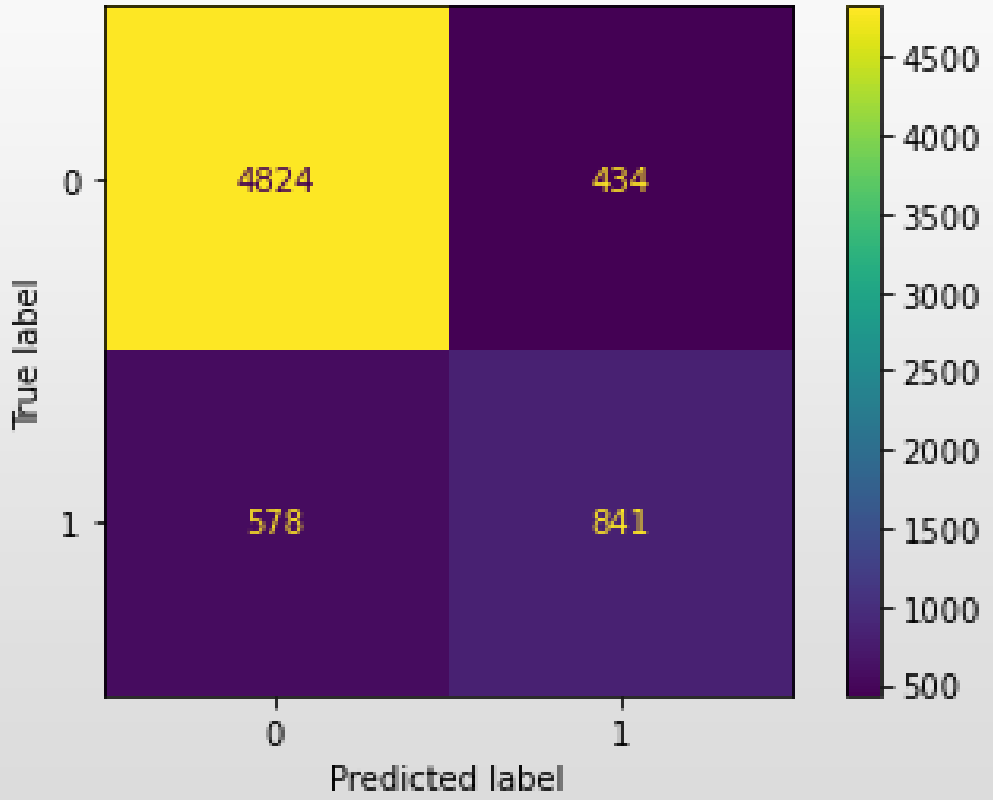
- Data from National Center for Health Statistics
- Survey data taken during 2009 H1N1 flu pandemic
- Shows number of vaccinated and unvaccinated for H1N1 vaccine
- Training Data column = actual classes
- Predicted Data = predicted classes

Class	Training Data	Predicted Data
Did not take vaccine (0)	21033	22970
Took vaccine (1)	5764	3738

Smart Health Records App

Vaccination Scenario: Logistic Regression Model

- Basic binary classification algorithm
- Handles majority class (unvaccinated) very well, minority class reasonable
- Logistic Regression Accuracy Score: 85%



Confusion Matrix for Logistic Regression Model

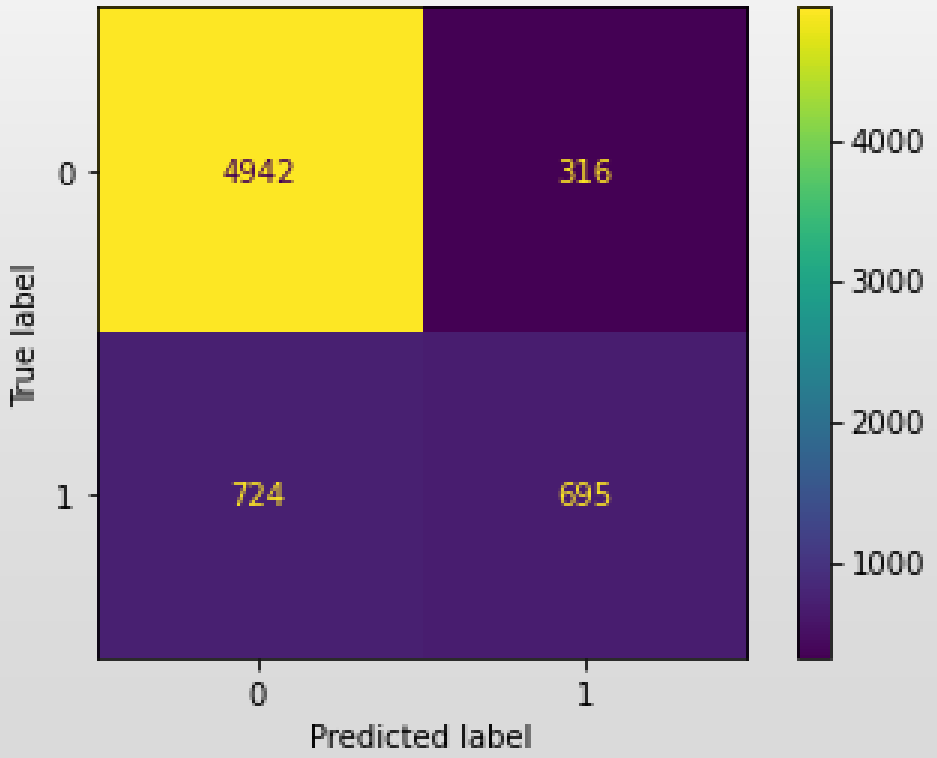
Smart Health Records App

Vaccination Scenario: SVM model

Notes

- SVM are more complicated models that are often quite accurate
- Results similar to previous model
- SVM Accuracy Score: 84%

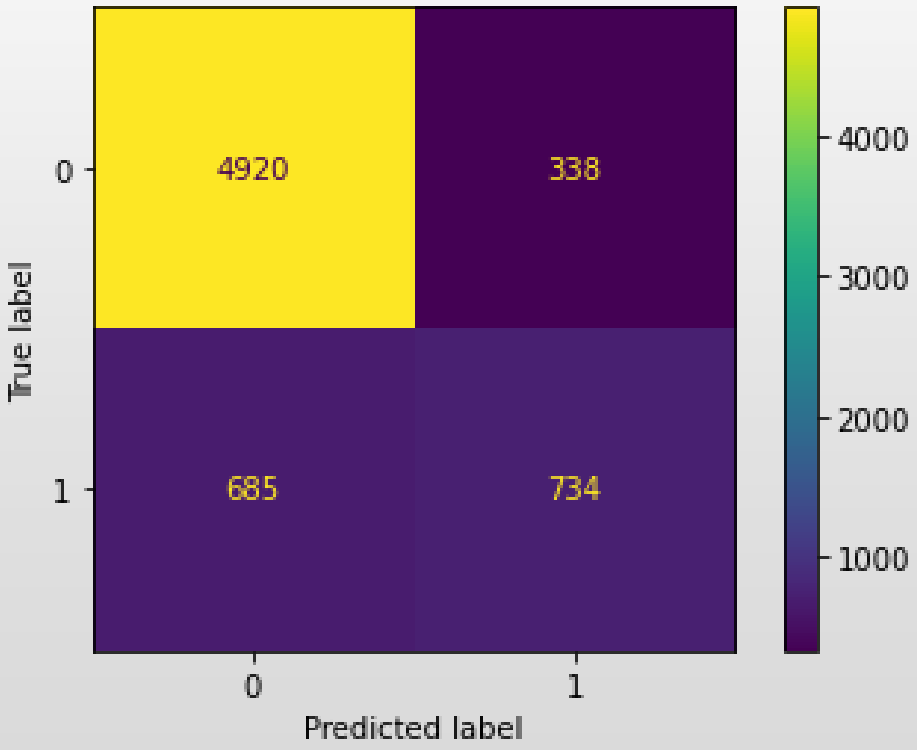
Confidence Matrix for Support Vector Machine Model



Smart Health Records App

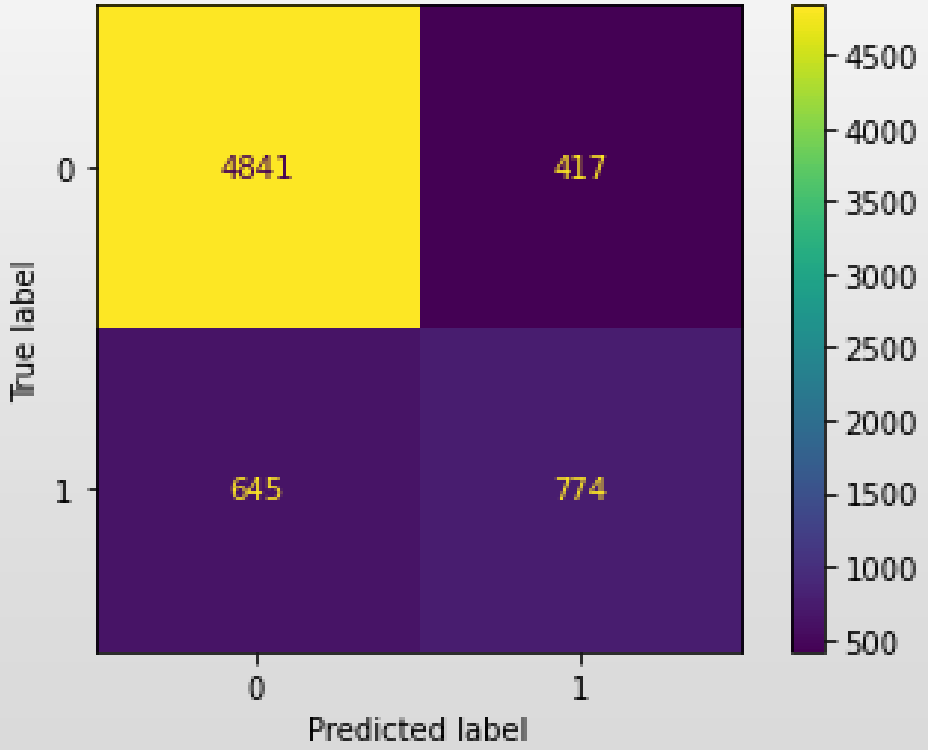
Vaccination Scenario: Gradient Boosting Classifiers

GradientBoostingClassifier



Accuracy: 85%

XGBClassifier



Accuracy: 83%

Smart Health Records App

Vaccination Scenario: Predictive Modeling Results

- All models performed similarly, with accuracy ranging from 77-85%
- Results suggest prediction improvement with increased data and features
- The application would send vaccination reminders and recommendations to the correctly classified patient 85% of the time

Model	Accuracy (%)
Logistic Regression	85%
Support Vector Machine	84%
Decision Tree	77%
Random Forest	84%
XGBClassifier	84%
Gradient Boosting Classifier	84%
Gradient Boosting Classifier (post-tuning)	85%

- Data
  - Does not include social network data or patient history.
  - Include data from many other sources.
  - Develop synthetic data to model effect of social network on medical decisions

- Models
  - Explore other models (Naïve Bayes, K-Nearest Neighbor, for example) or even clustering techniques.
  - Develop ensembles and or voting models

- Future Work
  - Integrate other data streams into predictive models
  - Develop UX/UI and backend

Smart Health Records App

Shortcomings, Limitations, and Future Directions

Thank you!

Smart Health Records App

Questions? Comments?

Feel free to contact me using the information below.

Matthew Noonan

mcn9284@gmail.com

925-917-0396