Detecting Diabetes with Machine Learning

...

Ian Butler
In Association with the Center for Disease Control

Objective - Increase Accessibility to Diabetes Testing

Complications

- Heart Disease
- Vision Loss
- Lower-Limb Amputation
- Kidney Disease
- Stroke

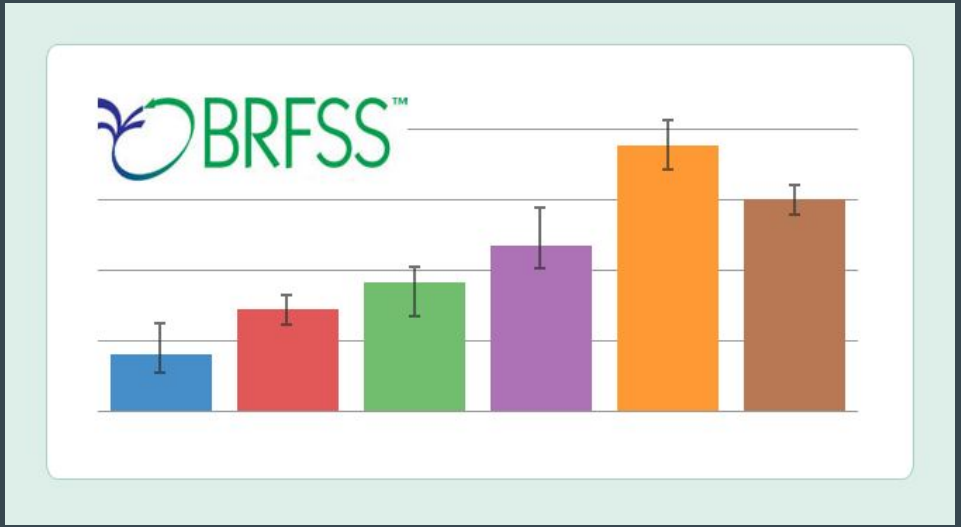
Early Diagnosis

- Lifestyle Changes
- More Effective Treatment

Data - CDC's Behavioral Risk Factor Surveillance System

250,000 Records

- Reported Fields
  - Diet
  - Exercise
  - General Health
- Calculated Fields
  - BMI
  - High Cholesterol
  - High Blood Pressure



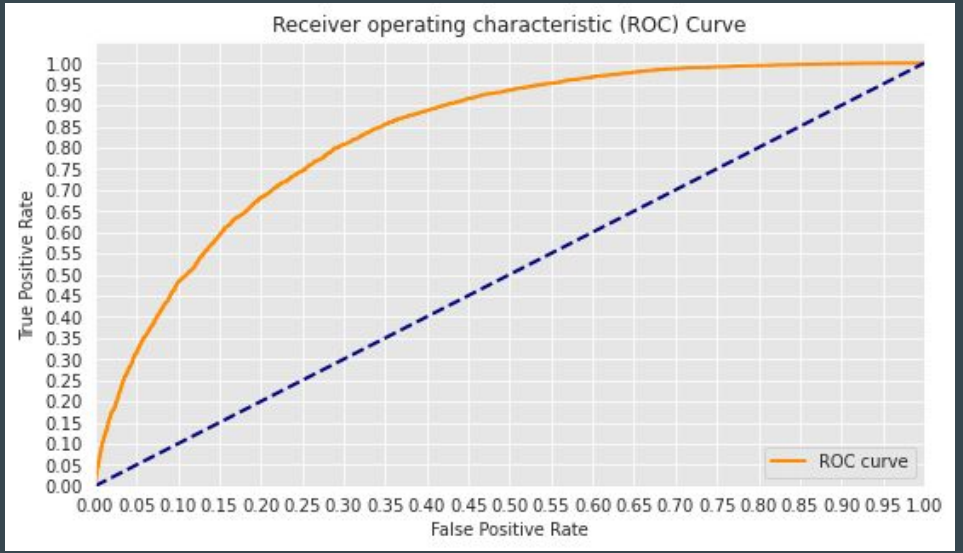
Results - Final Model and Performance

GBC 2

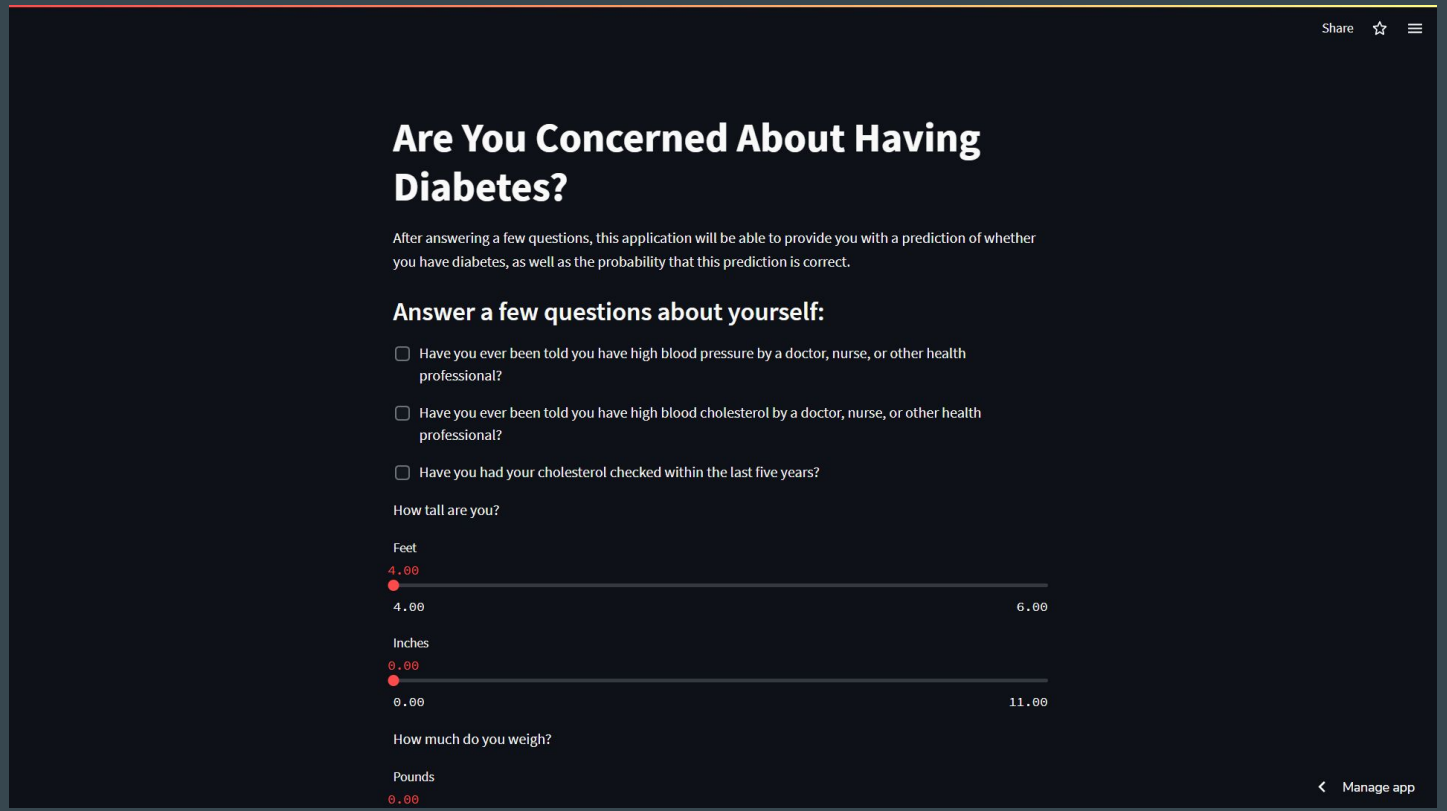
- Gradient Boosting Classifier
- Ensemble Method
- Regression Trees

Performance

- Accuracy: 75%
- Precision: 74%
- Recall: 80%
- F1: 76%
- AUC: 83%



Deployment - Streamlit



Moving Forward - Improving the Application

Next steps:

- Reduce Dimensionality
  - Less Questions
  - Ease of Use

- Multiple Models
  - Variety of Questions
  - Flexibility

Thank you!