Diabetes Classification for Local Health Department

Benjamin Toler
March 31, 2021

Summary

- Classification modeling of CDC survey predicts diabetes with 76% accuracy and identifies the following as key indicators:
  - General Health
  - Blood Pressure
  - BMI

Overview

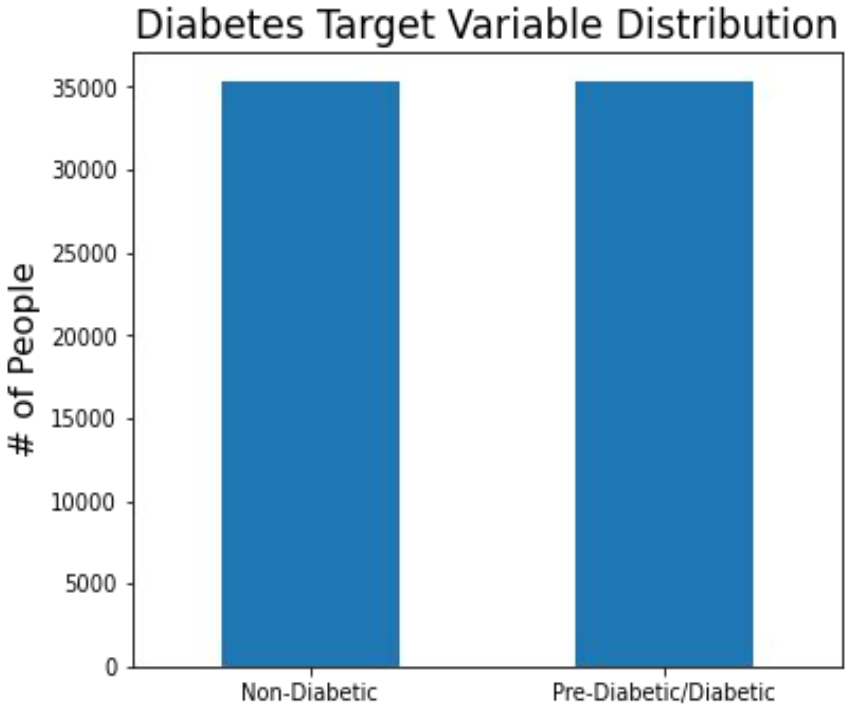
- Business Problem
- Data
- Modeling
- Results
- Conclusions
- Next Steps

Business Problem

- Local Health Department wants to improve the health of their community by addressing the prevalence of diabetes
- A classification model based on surveys can help with early detection of disease and identify key risk factors
- Constructing health initiatives based on this analysis can help prevent diabetes which improves community health and lowers healthcare costs

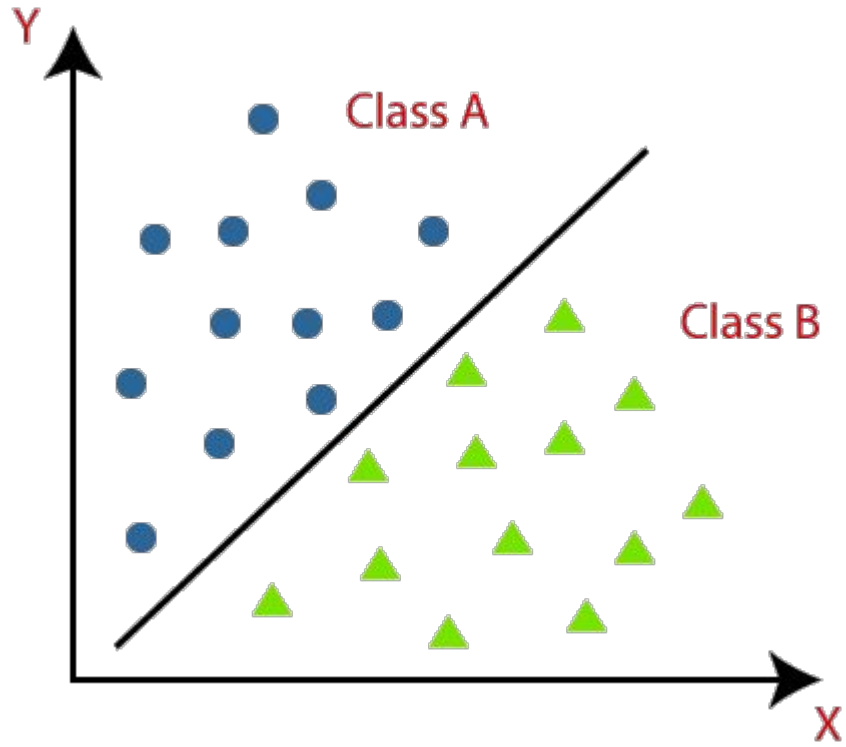
Data

- CDC's Behavioral Risk Factor Surveillance System
  - Surveys of 70,692 people
- Target Variable: Diabetes
  - Balanced Dataset
- 27 Features:
  - Health
  - Lifestyle



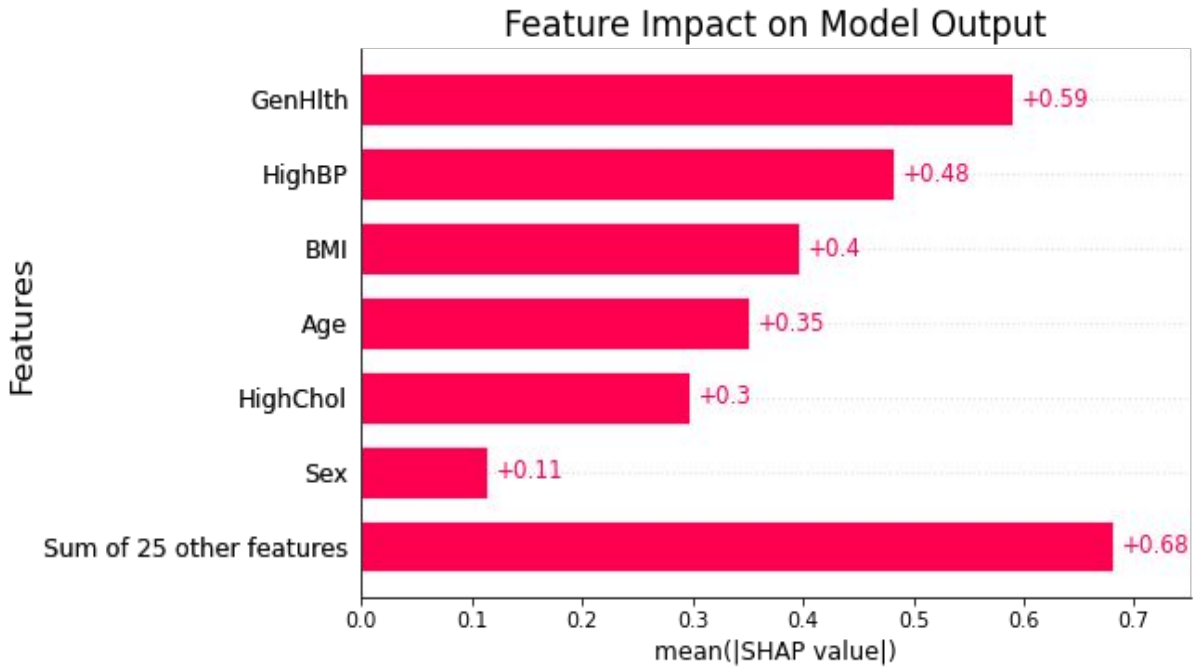
Model

- Classification Modeling
- Evaluation metrics:
  - Accuracy
  - Recall
- Iterative Approach
  - Hyperparameter tuning
- Final Model: XGBoost



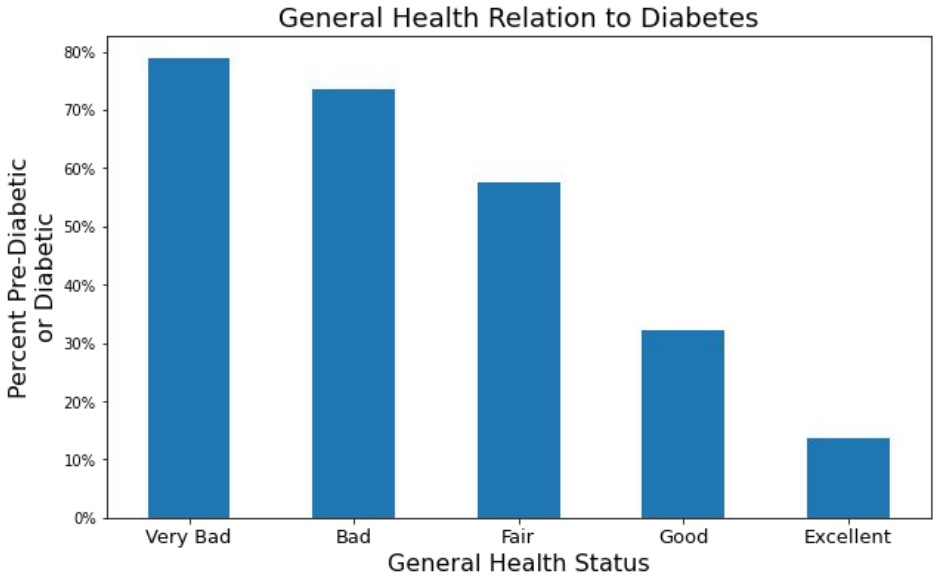
Results

- Model Accuracy: 76%
- Model Recall: 80%
- Important Indicators of diabetes:
  - General Health
  - High Blood Pressure
  - Body Mass Index



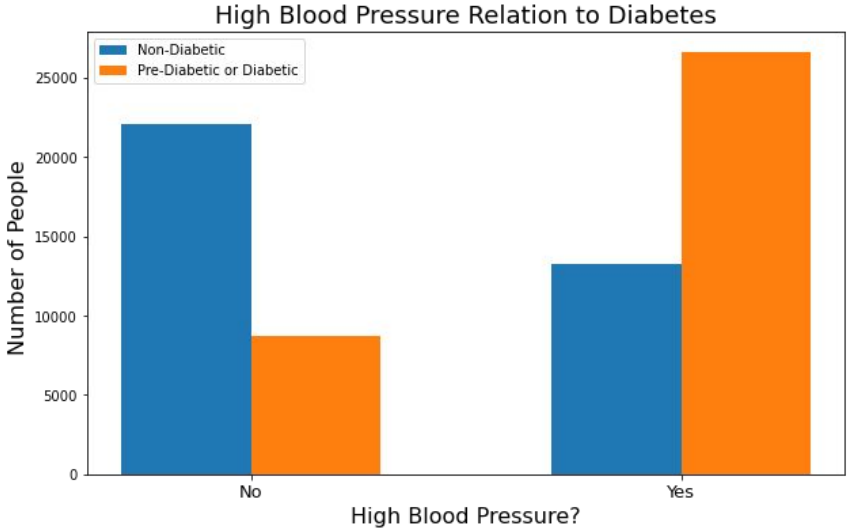
Recommendations

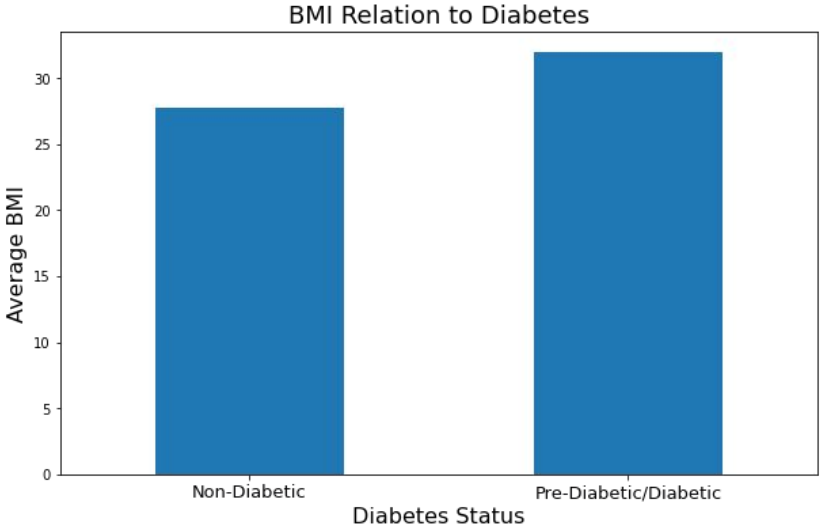
- Provide survey online for community to take
  - Recommend seeing a doctor for those predicted to have diabetes by model
- Educate community on link between general health and diabetes



Recommendations

- Start an education campaign to inform people about dietary and lifestyle choices specific to reducing blood pressure and losing weight
- Invest in programs that increase access to healthy food options





Next Steps

- Search for additional survey data to improve model accuracy and recall
- Perform more specific analysis of how different diets impact diabetes
- Analyze the relationship between environmental factors and diabetes