Predicting Heart Disease

Flatiron School, Data Science, Flex Program

by

Milad Shirani

Contents

Project Overview

1. One of the main causes of death in the world is heart disease.
2. In this work, we want to classify cases that will have heart disease
3. We used several categorical models such as Decision Tree, Random Forese, XGBoost, LightGBM, CatBoost

Data

1. We use 319795 data points for our analysis
2. Data is coming from kaggle.com and it contains information about
   - BMI, Smoking, Alcohol Drinking
3. Some of categorical variables are:
   - Diabetic, Asthma, Race, Sex,
4. Some of numerical variables are:
   - BMI, Sleep Time
5. Data preparation contains
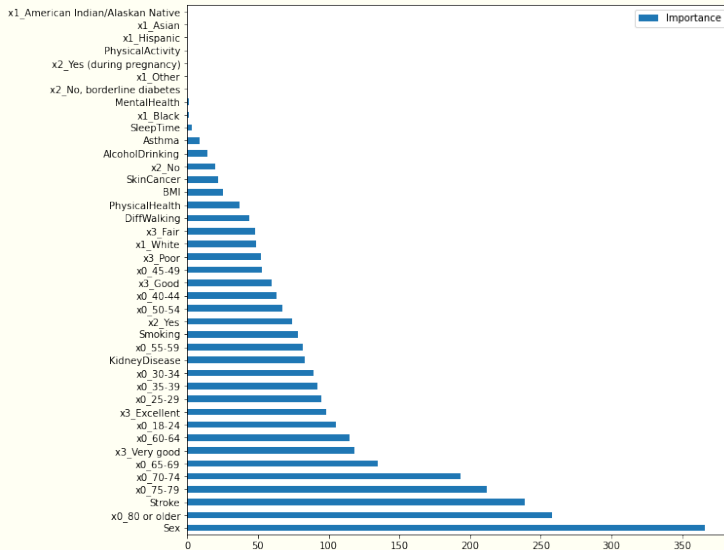   - Converting Categorical to Numerical, Resampling the data

1. We used LightGBM as our final model
2. We considered recall-score as our metric
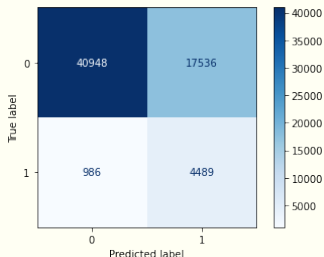
	Precision	Recall	F1-score
0	0.98	0.70	0.82
1	0.20	0.82	0.33

Modeling And Results

Conclusion

1. Heart disease is one of the main causes of death in the world
2. We used categorical regression models to fit and train the data
3. We concluded that
   - Men have higher risk of having heart disease than women.
   - Age category 80 and above are in higher risk.
   - To reduce the chance of heart disease
     - Exercise more
     - Do not smoke

Next Steps

1. Adding other features from other data sets might improve the model and predictions
2. Trying other different values for tuning hyperparameters

Thank You!