Cardiovascular Disease Prediction

Aktan Abdygaziev

1

# Why Cardiovascular Disease(CVD)?

- ✤ CVD is #1 cause of death globally

- ✤ Tobacco use, unhealthy diet and obesity

- ✤ Physical inactivity and harmful use of alcohol

- ✤ **GOAL**: Try several models and see which one performs the best
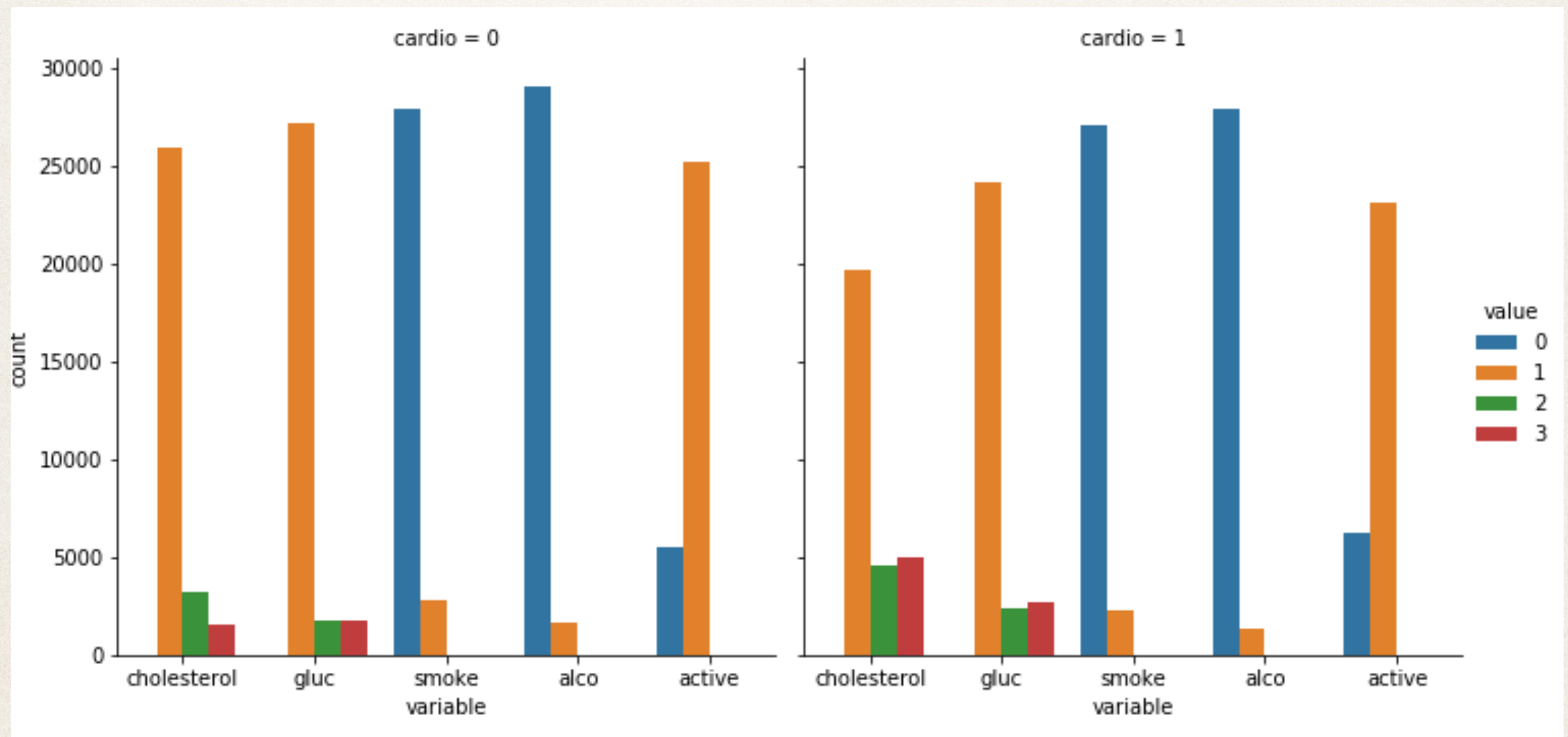
About the Data

✤ Source: Kaggle

✤ Size: 70k instances, 13 features

✤ Continuous: Age, gender, blood pressure,BMI

✤ Binary: cholesterol lvl, glucose lvl —> 1-3
           alcohol ;physical activity —> 0-1

Initial Exploration
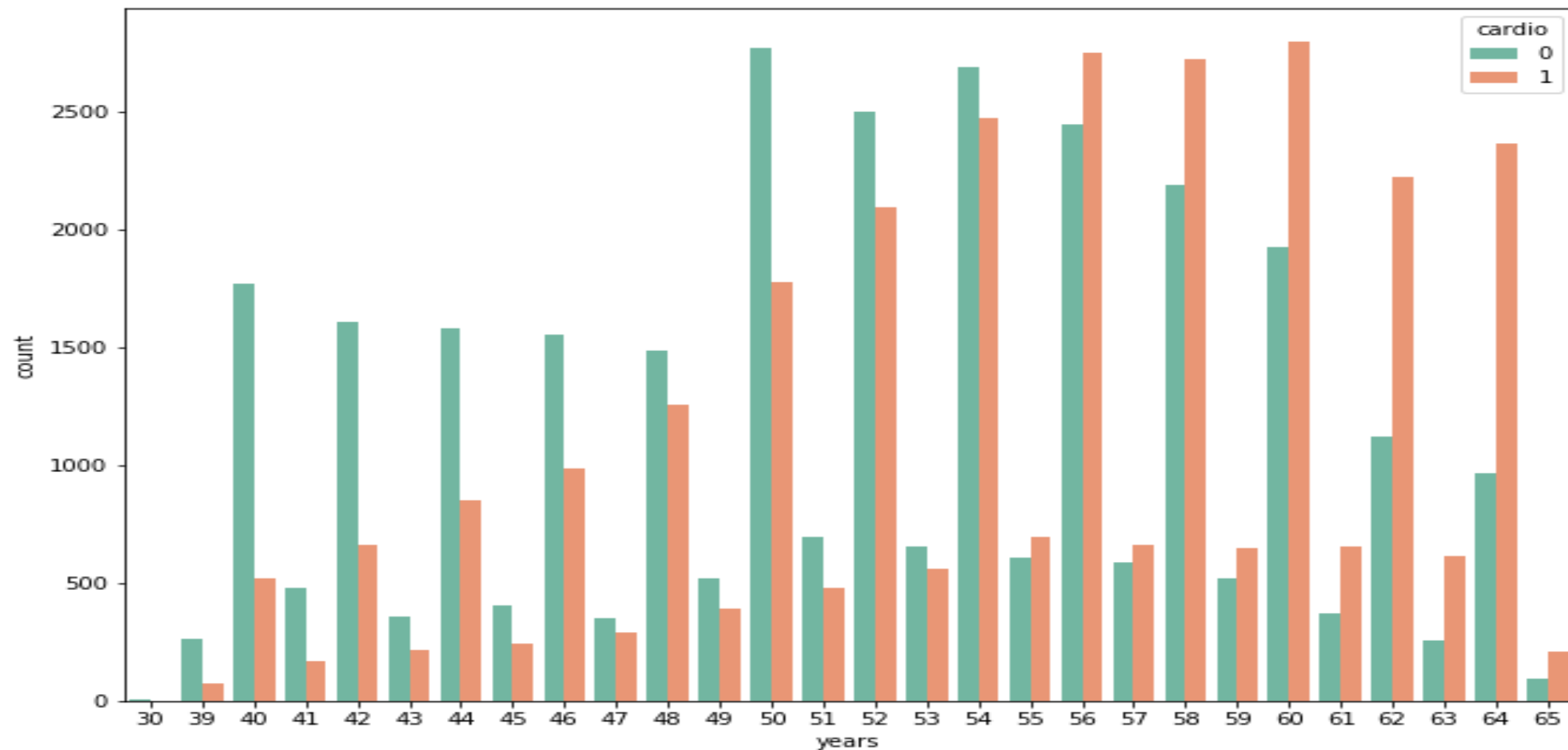
As you age the probability of having CVD is increases

Initial Exploration

Peoples with CVD have more cholesterol, and tend to be less active

Modelling

	Precision	Recall	Accuracy	AUC_ROC
Ada Boost	0.63	0.76	0.72	0.73
Extra Trees	0.64	0.76	0.72	0.73
Random Forest	0.67	0.74	0.73	0.73
KNN	0.67	0.71	0.70	0.71
SVM	0.64	0.75	0.72	0.72
Naive Bayes	0.62	0.74	0.71	0.71
Decision Tree	0.65	0.75	0.72	0.73
Logistic Regression	0.65	0.74	0.72	0.72

Decision Tree Classifier
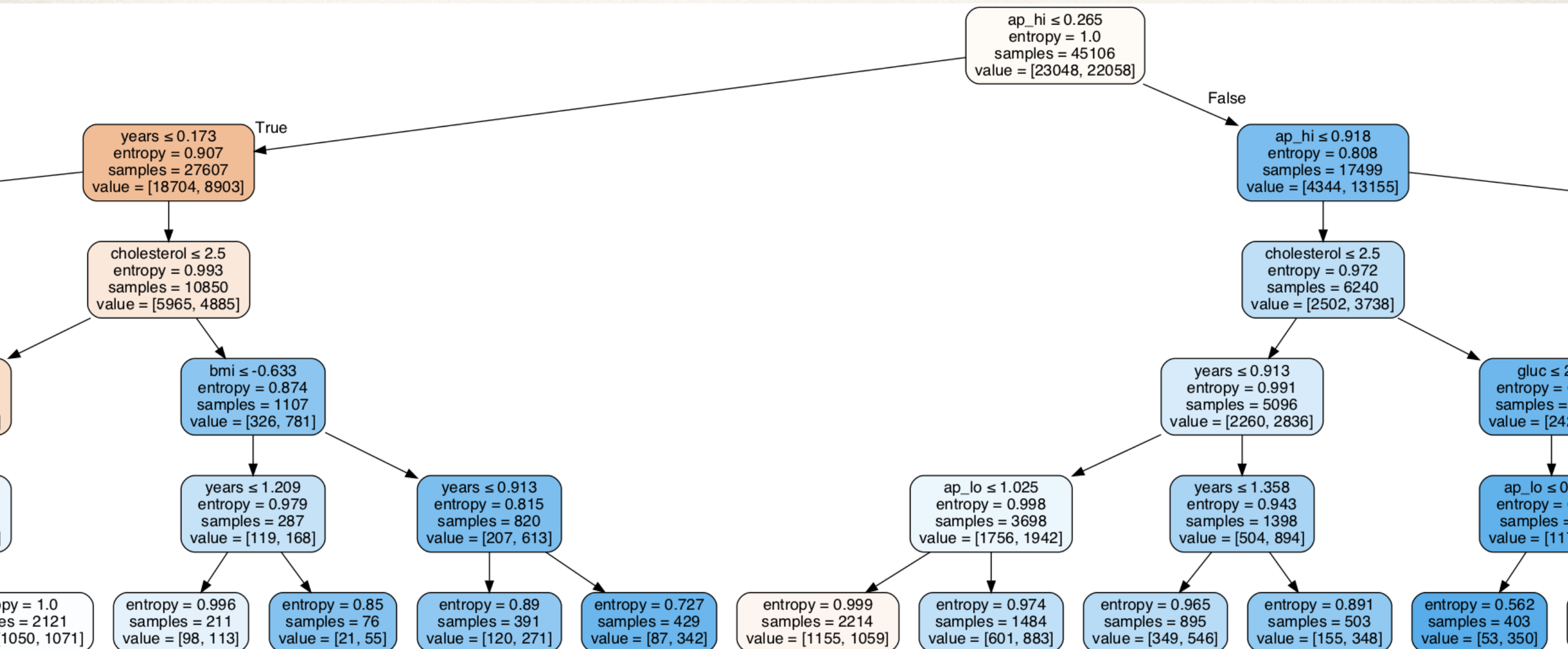
Recommendations

* Reduce alcohol and smoking

* Eat healthy food: low fat diet

* Exercise