Predicting Heart Disease

By: Eric Wolf

Problem Identification

Cardiovascular diseases (CVDs) stands as the number one cause of death worldwide

A dataset merging five large scale studies has been created identifying 11 common identifying features

Primary Goal: The creation of a machine learning model capable of identifying patients with a high probability of heart disease

The Data

A compilation of five large scale health studies from across the globe focusing on 12 key attributes related to heart failure

- ○ Cleveland: 303 observations
- ○ Hungarian: 294 observations
- ○ Switzerland: 123 observations
- ○ Long Beach VA: 200 observations
- ○ Stalog (Heart) Data Set: 270 observations
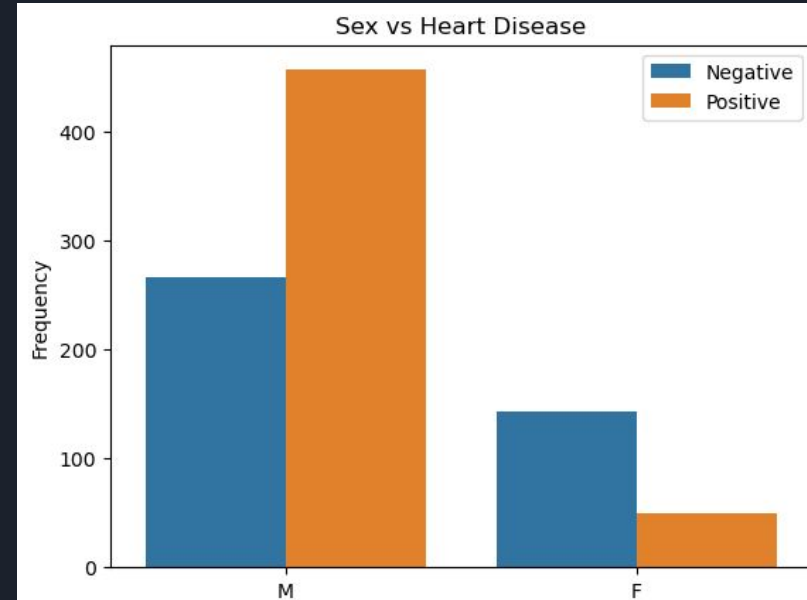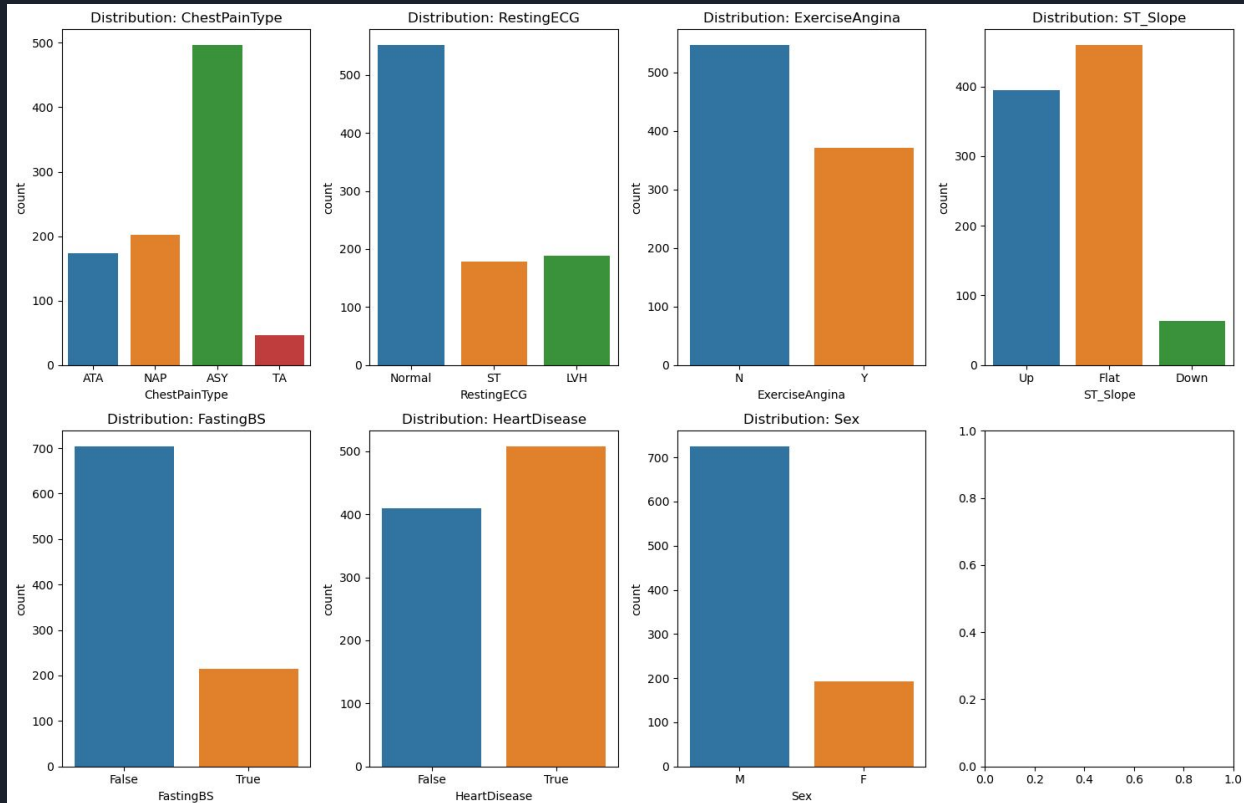
Data Wrangling & EDA

Categorical features:
- Sex
- Chest Pain Type
- Resting ECG
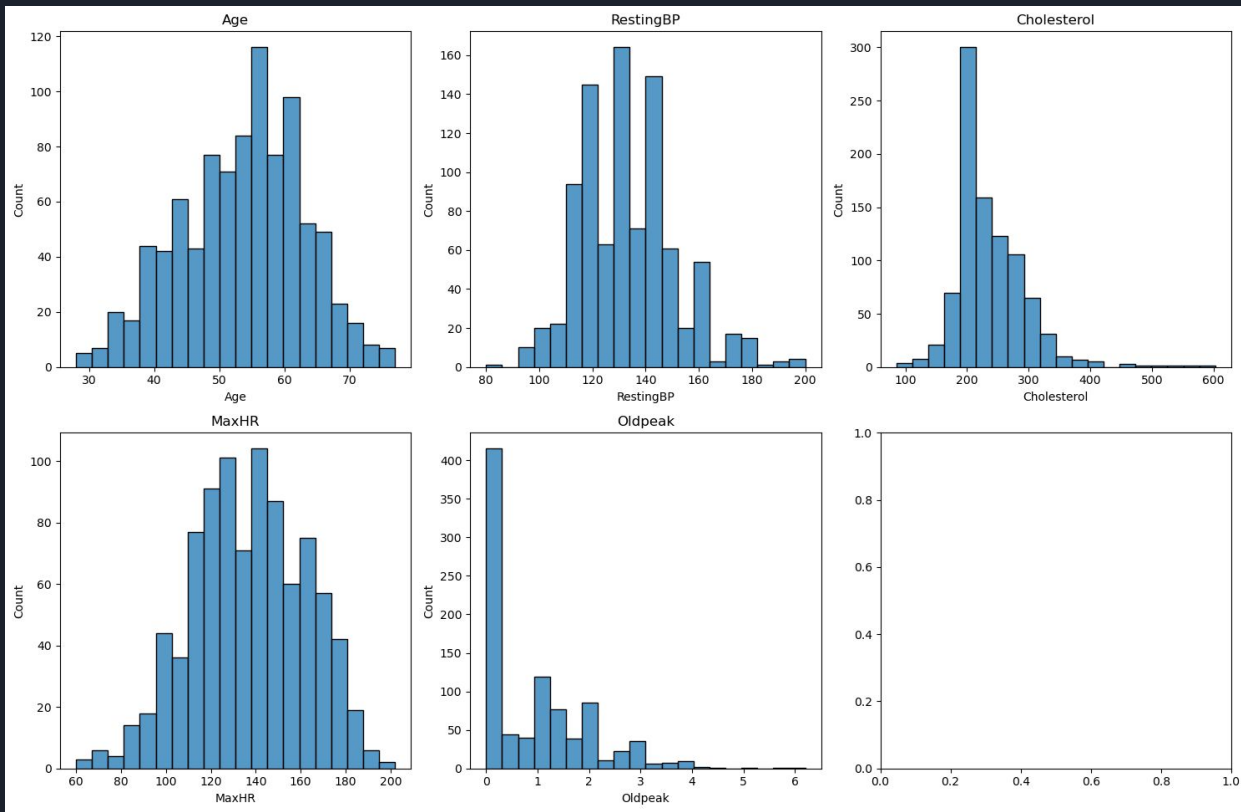- Exercise Angina
- ST Slope

Numerical features:
- Age
- Resting Blood Pressure
- Cholesterol
- Fasting BS
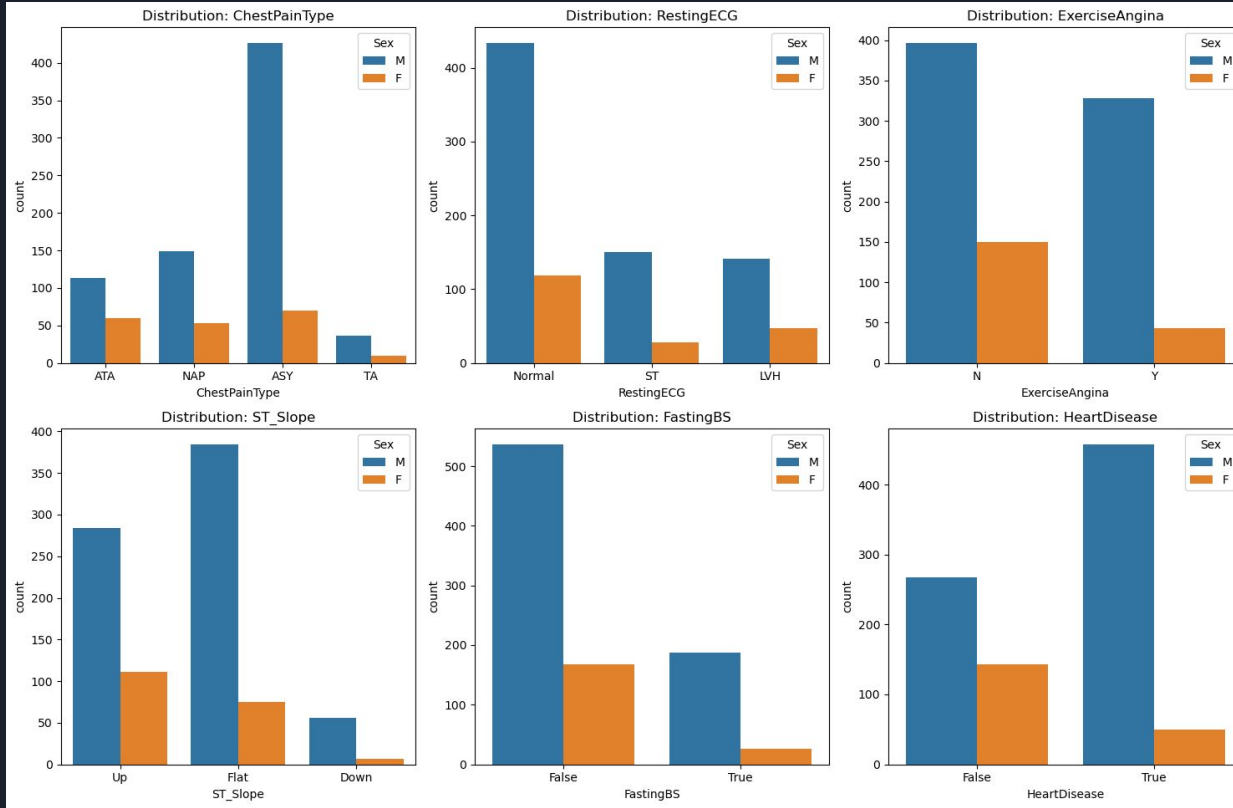- Maximum Heart Rate
- Oldpeak
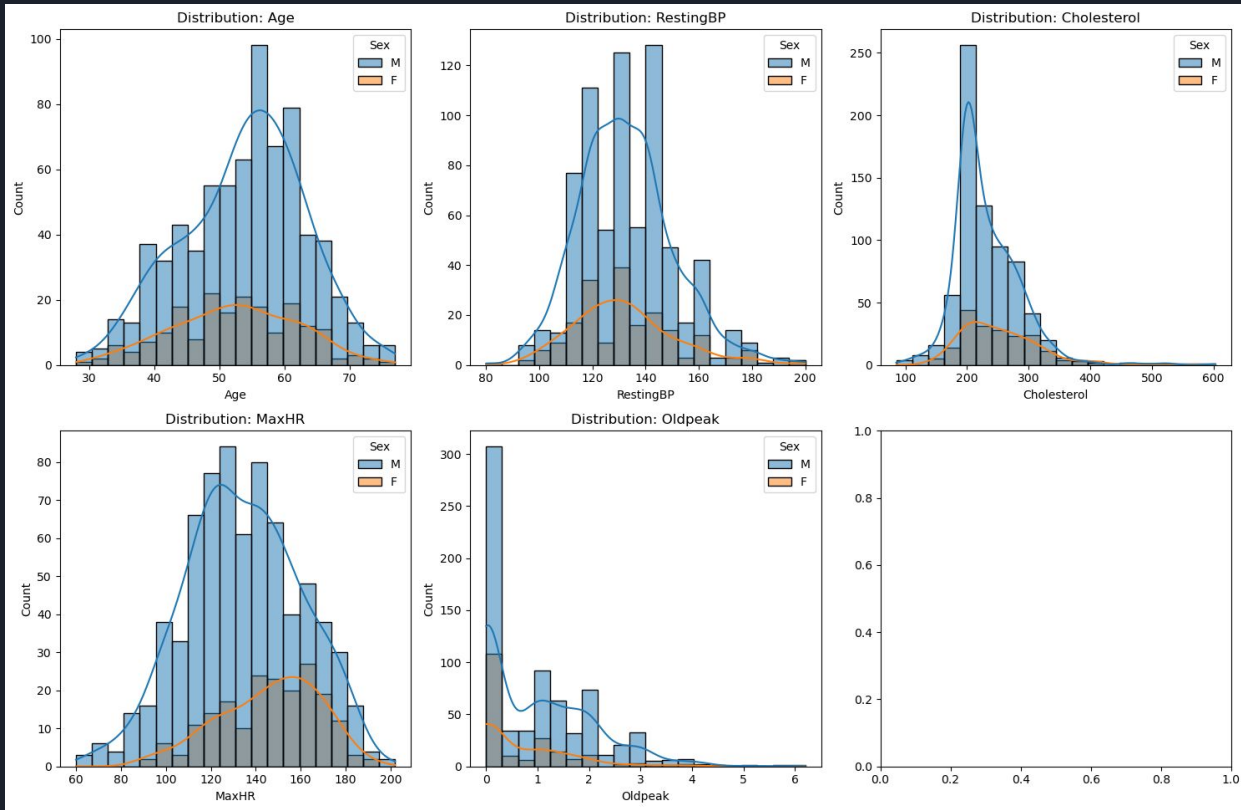- Presence of Heart Disease

Categorical Features

Numerical Features

Categorical Features by Sex

Numerical Features by Sex

EDA (cont.)

Categorical features ranked by Cramer's V score:

- ○ ST Slope
- ○ Chest Pain Type
- ○ Exercise Angina
- ○ Sex
- ○ Fasting BS
- ○ Resting ECG (Moderate association)

Numerical features ranked by t-test:

- ○ Oldpeak
- ○ Max HR
- ○ Age
- ○ Resting BP
- ○ Cholesterol showed no significant relationship.

Note: Cholesterol had many missing values.

Modeling

Dummy Classifier

Logistic Regression

KNN

Random Forest

Gradient Boosting

Modeling Results
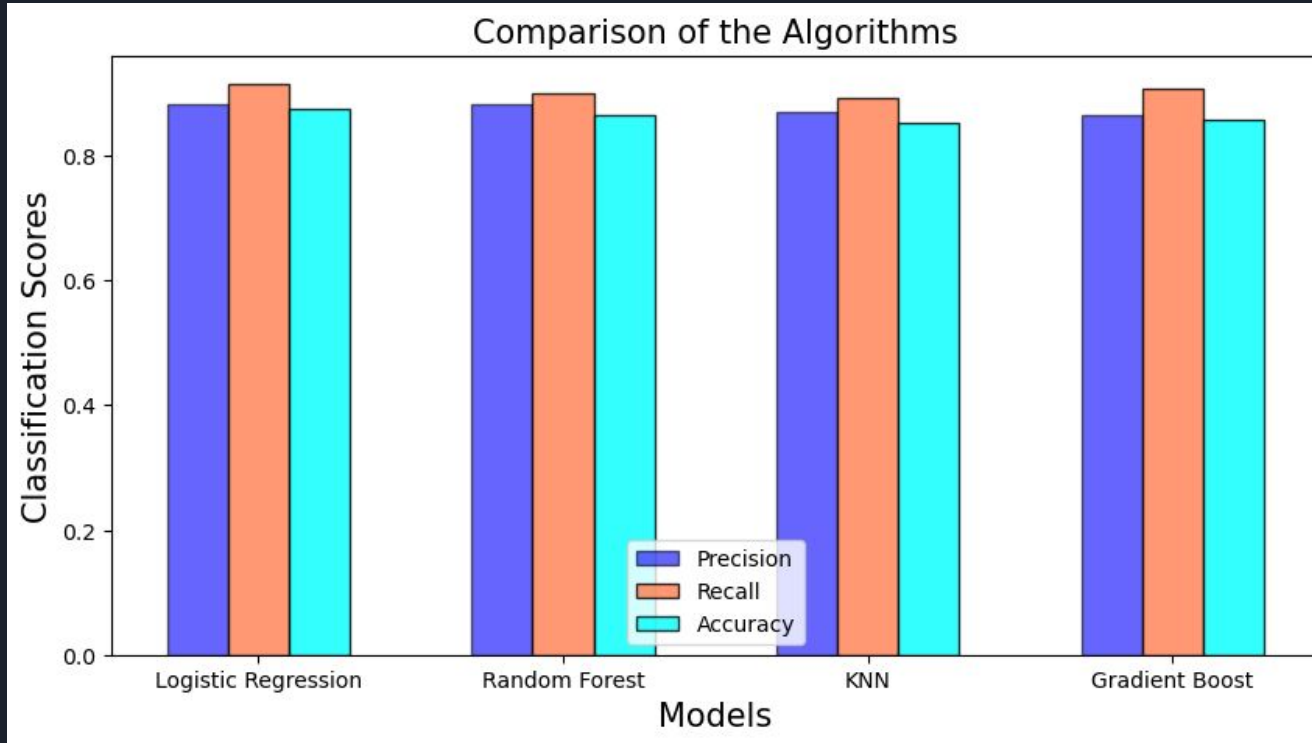
Logistic Regression

- ○ Recall: 91.43%

Gradient Boosting

- ○ Recall: 90.71%

Random Forest

- ○ Recall: 90.00%

K-Nearest Neighbors

- ○ Recall: 89.29%

Comparison of the Algorithms

Future Improvements

Incorporate Supplementary Patient Data:
- ○ Consider adding height, weight, pre-existing heart conditions
- ○ Enhance predictive capabilities
- ○ Offer a more comprehensive risk assessment

Explore Advanced ML Techniques:
- ○ Explore neural networks
- ○ Ideal for complex pattern recognition
- ○ Potential for higher predictive performance

Develop User-Friendly Interface:
- ○ Create an intuitive healthcare interface
- ○ Facilitate real-world model deployment
- ○ Benefit both patients and healthcare providers

Questions?