Heart Disease Prediction
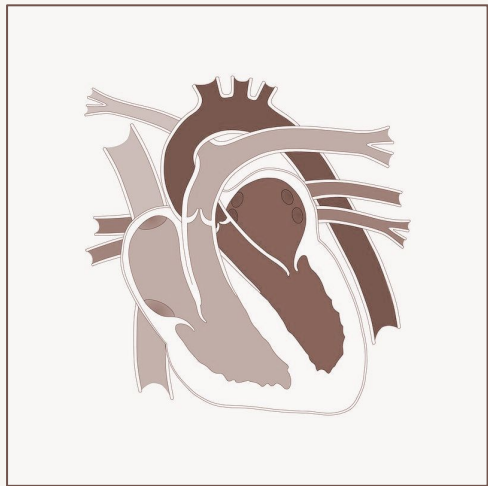
Cristopher Delgado

Agenda

01
Business Understanding

02
Business Problem

03
Data

04
Methods

05
Results

06
Evaluation

Business Understanding

Cardiovascular Diseases

- Leading cause of death globally
- 32% Proportion
- Risk Mitigation

Diagnostic Future

- Machine Learning Algorithms
- Continuous monitoring
- Software Medical Device

Business Problem

Identify Important Aspects

- Age
- Cholesterol
- Electrocardiogram

Develop Medical Devices

- At Home Use
- Clinical Setting

Goal: Predict Heart Disease

Data

Source

- Cleveland clinic (Cleveland, Ohio)
- Hungarian Institute of Cardiology (Budapest, Hungary)
- University Hospitals (Zurich and Basel, Switzerland)
- Veterans Administration Medical Center (Long Beach, California)
- Stalog (Heart) Data Set

Overview

- University of California Irvine Machine Learning Repository
- 5 heart datasets are combined over 11 common features

Distribution of Heart Disease
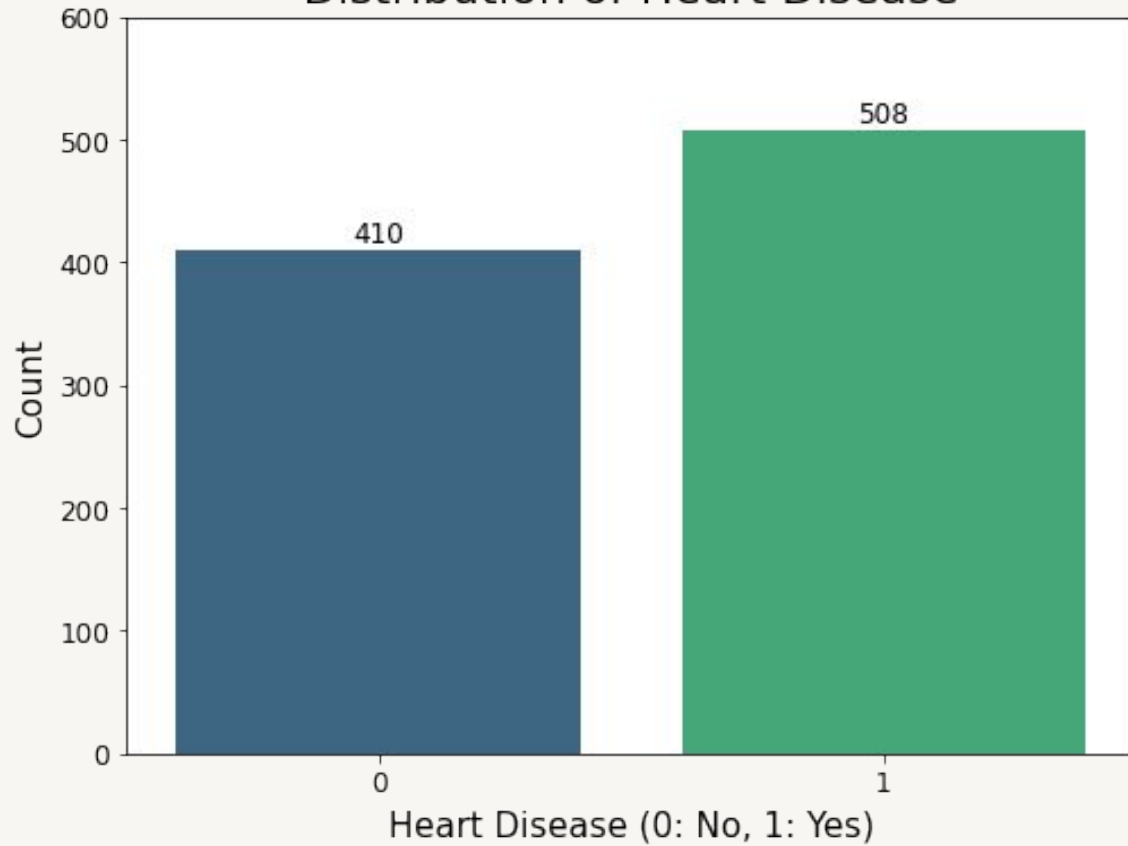
Methods

Cleaning

- Appropriate Data Types
- Numerically Encode Categories

Normalization

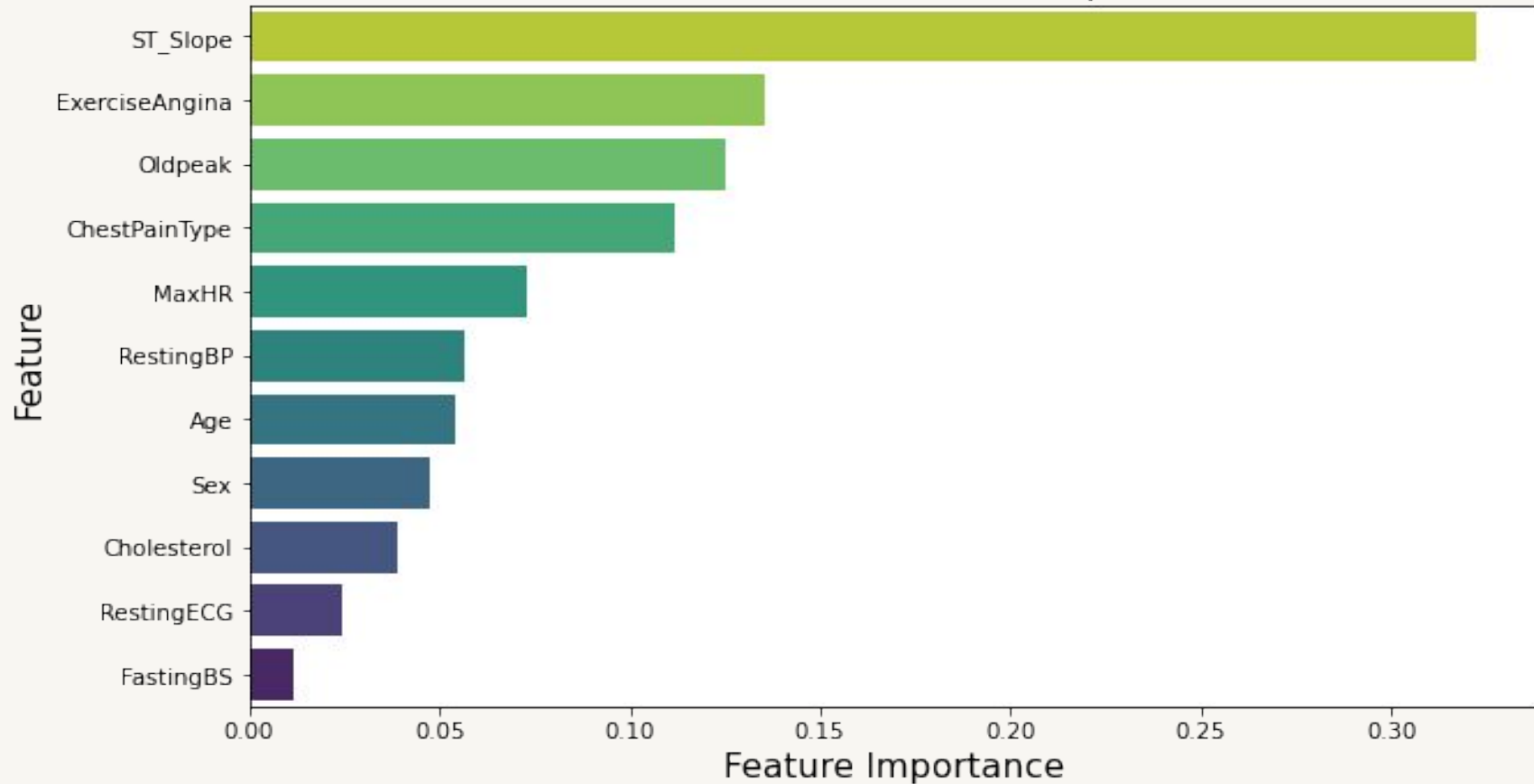- Prepare for modeling
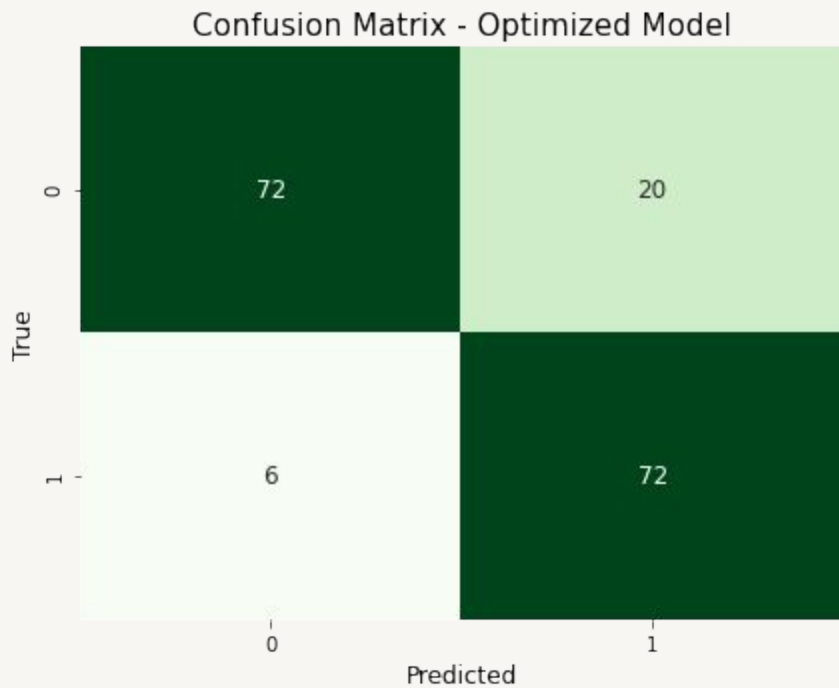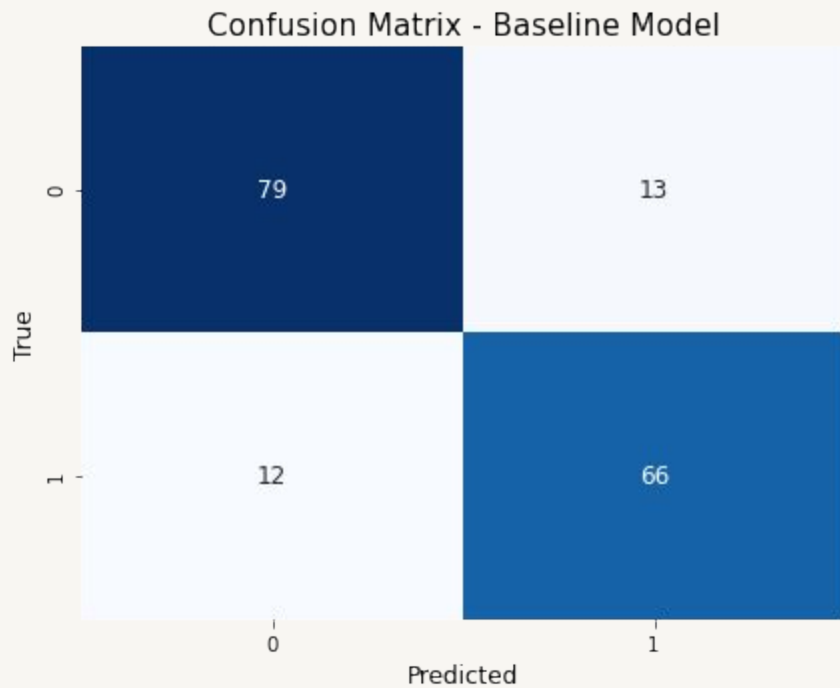- All continuous features are on the same scale

Baseline Models

- Baseline vs Optimized
- Scoring Metric: Recall (Sensitivity)

Optimized Models

Results

Random Forest Feature Importances

Confusion Matrix - Baseline Model

Confusion Matrix - Optimized Model

Evaluation

1. Incorporate the Random Forest model into diagnostic medical device software for the purposes of monitoring Cardiovascular related symptoms for Cardiovascular Diseases.

2. I recommend to develop medical devices geared towards monitoring the slope of the S–T segment of the Electrocardiogram, Exercise–induced angina, and measuring the S–T segment depression.

Thank You

Questions?

 github.com/cristopher-d-delgado

 linkedin.com/cristopher-d-delgado/