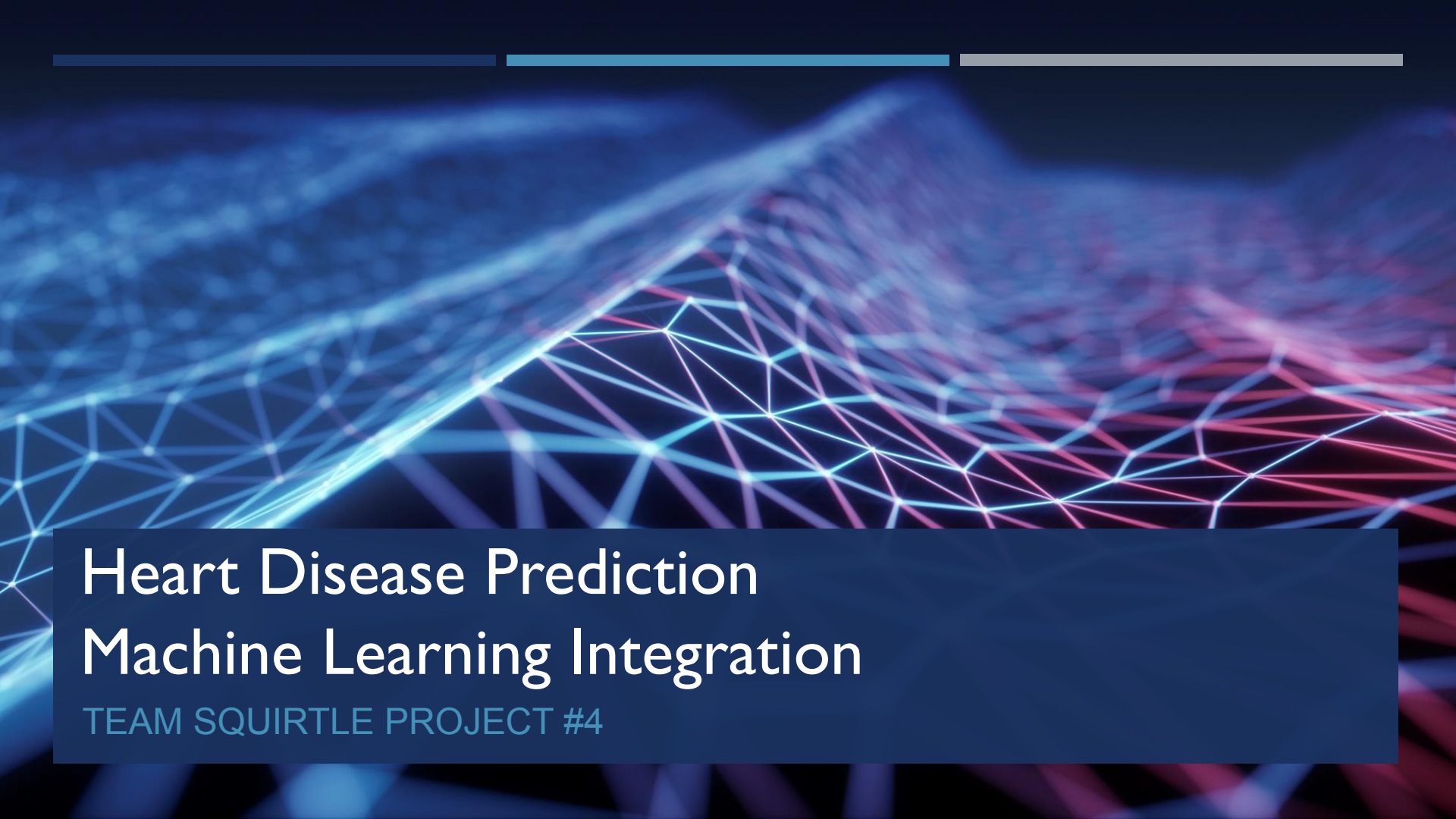


Heart Disease Prediction
Machine Learning Integration

TEAM SQUIRTLE PROJECT #4

Team Members

Randall Weaver

Rick Mora

AJ Lamar II

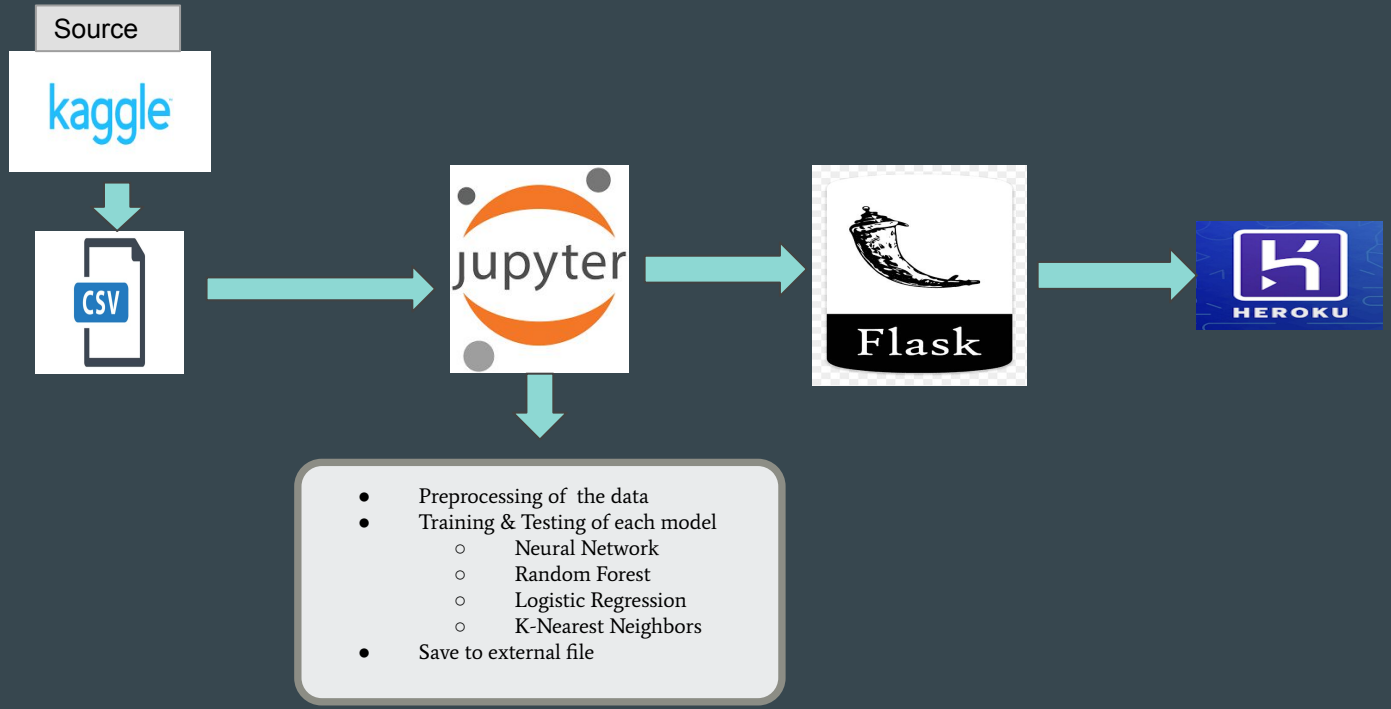
Evan Mickler

Anthony Ndungu

Yang Yu

Tamekia Phillips

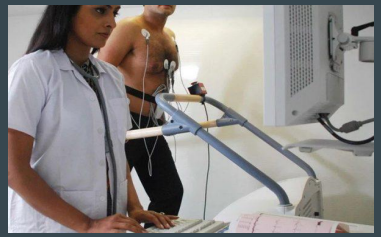
Process Overview:



Heart Disease: What is it?

- The term is often used interchangeably with the medical term “cardiovascular disease”
- It refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack.
- Heart disease describes a range of conditions that affect your heart. Under the heart disease umbrella include:
    - Blood vessel diseases
    - Coronary artery disease
    - Heart rhythm problems (arrhythmias)
    - Heart defects your born with (congenital heart defects , among others.
- Heart disease is the leading cause of death globally. Its estimated that 17.9 million people died from heart disease in 2019, representing 32% of global deaths. Of these deaths, 85% were due to heart attack and stroke.

Heart Disease Prediction Dataset:

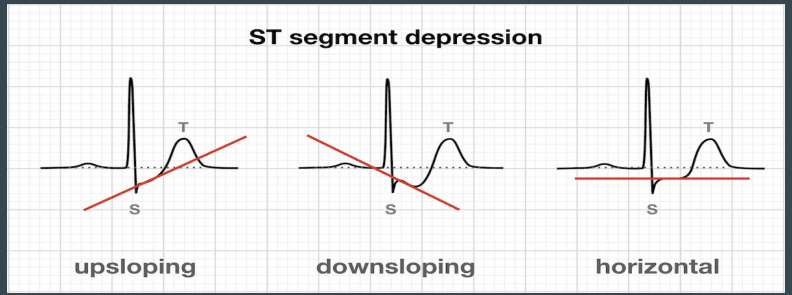


Is used to predict which patients are most likely to suffer from heart disease in the near future using 12 health attributes:

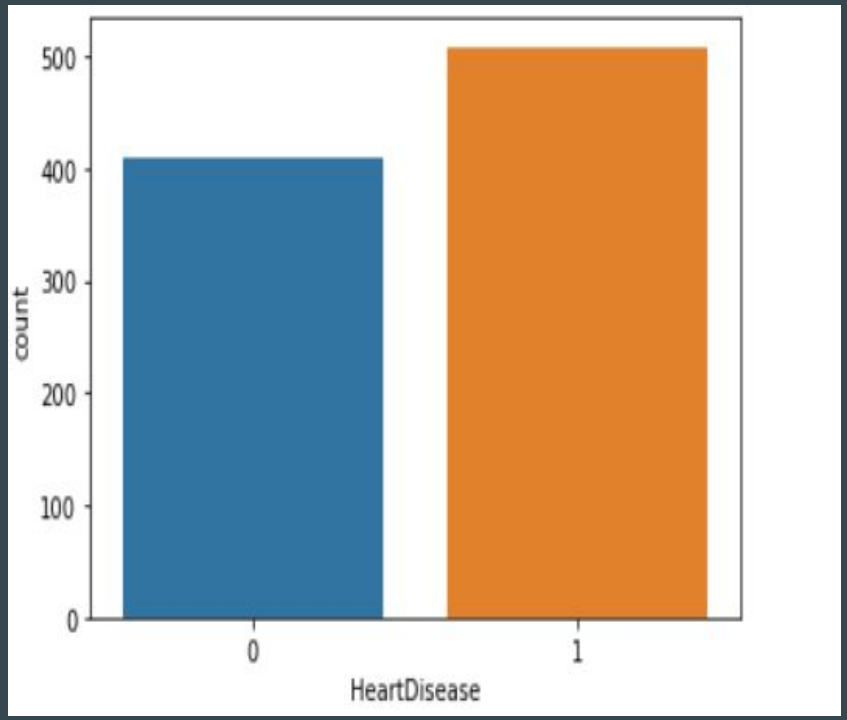
The information that is collected is usually collected through exercise electrocardiogram (stress test) and is commonly used because it's non-invasive and it's inexpensive method for detection of electrocardiogram changes.

- Age
- Sex
- Exercise Induced Angina [1= yes] [0 =No]
- Chest Pain Type:
  - Typical Angina
  - Atypical Angina
  - Non-Angina pain
  - Asymptomatic
- Resting blood pressure
- Cholesterol
- Fasting Blood Sugar
- Resting Electrocardiogram Results
- Max Heart Rate achieved
- Old Peak
  - ST slope depression(down)
    - ST depression induced by exercise relative to rest
- ST slope:
  - Elevation (up) Indication of some type (heart damage)
  - Depression(down) Indication of some type (heart damage)
  - Flat (normal) healthy heart
- Heart Disease
  - "0" No heart disease was present
  - "1" Yes, heart disease was present /heart

Table 1. Clinical classification of chest pain [8]	
Typical angina (definite)	Meets three of the following characteristics: • Substernal chest discomfort of characteristic quality and duration • Provoked by exertion or emotional stress • Relieved by rest and/or nitroglycerine
Atypical angina (probable)	Meets two of these characteristics
Non-cardiac chest pain	Meets one or none of the characteristics



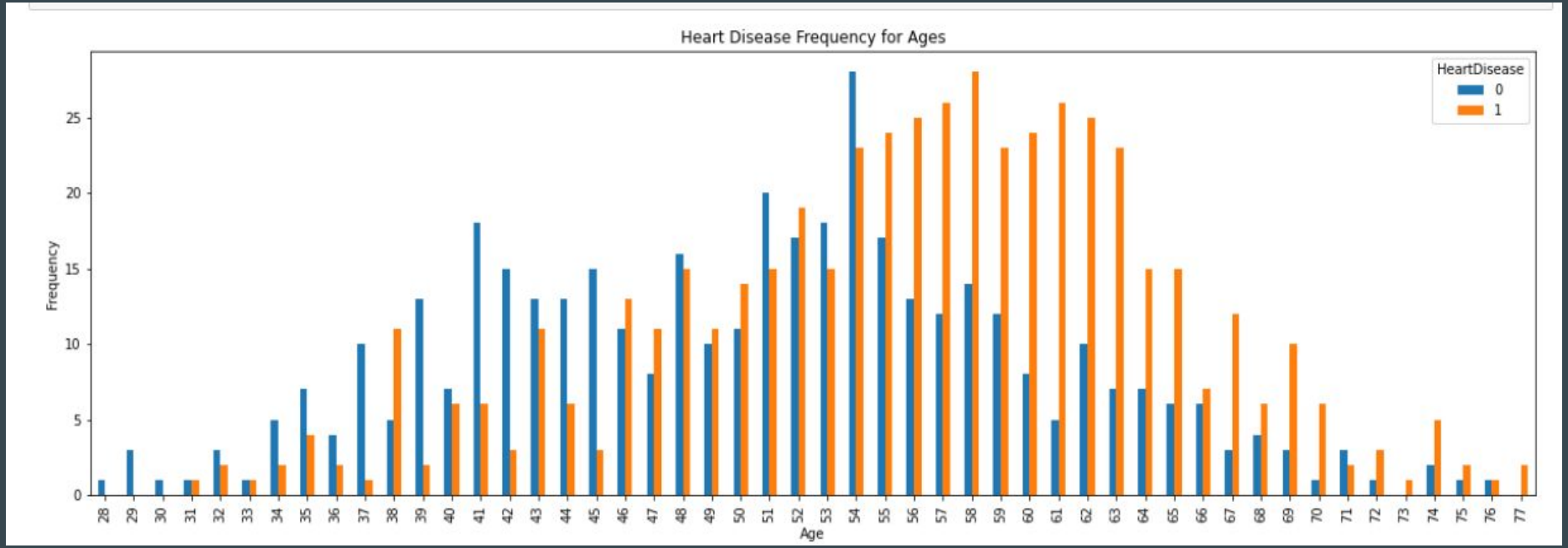
Percentage of Patient with or without Heart Disease in this Dataset



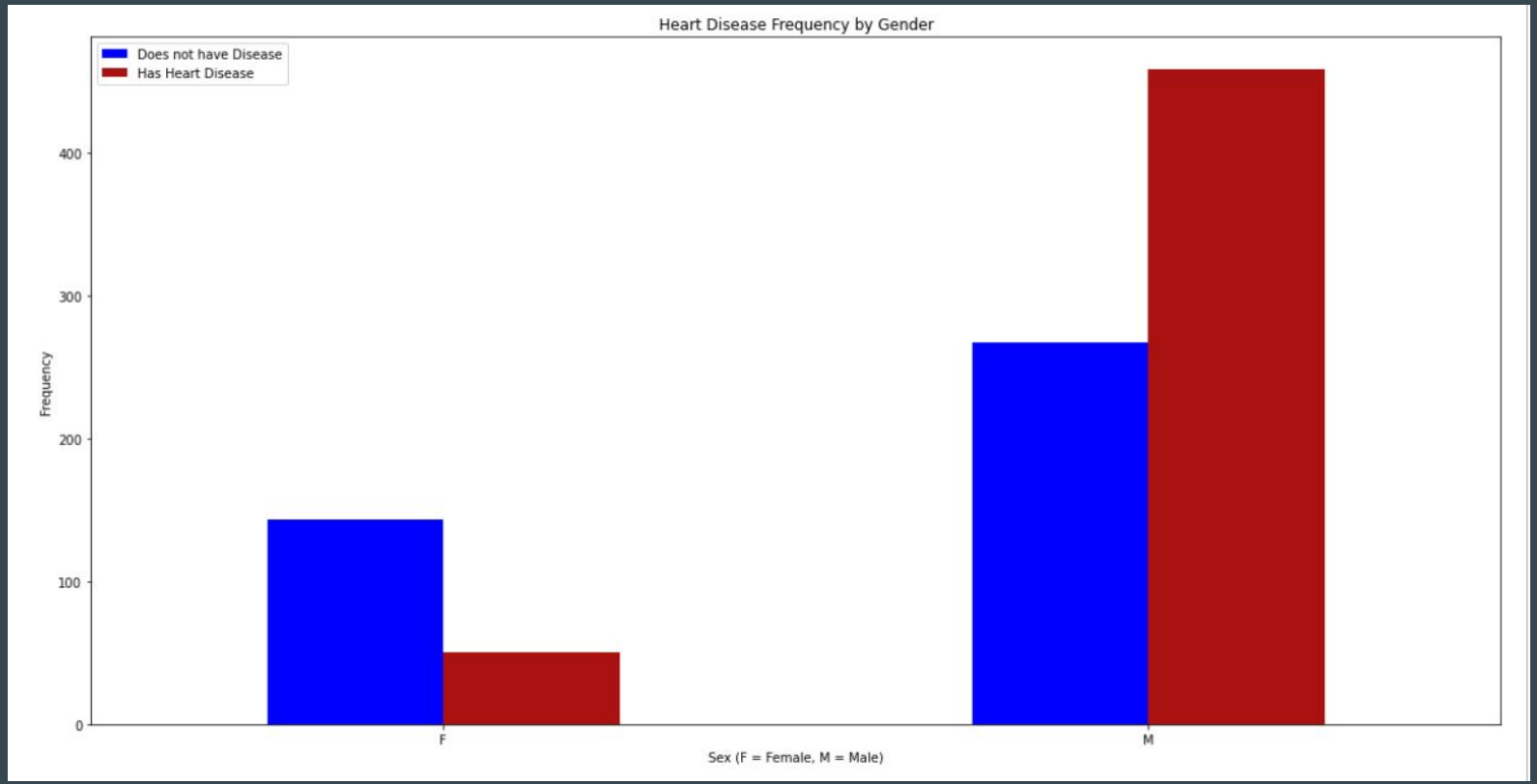
Percentage of patients without Heart Disease: 44.66%

Percentage of patients with Heart Disease: 55.34%

Heart Disease Frequency by Age:

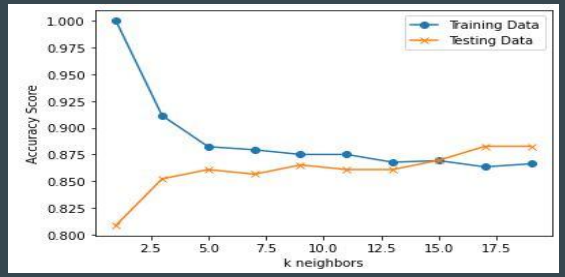


Heart Disease Frequency by Gender:



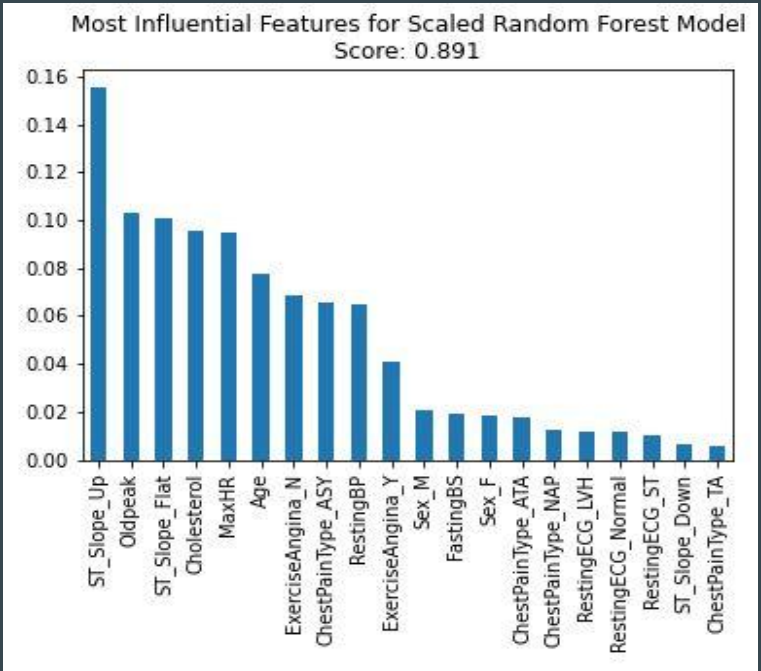
Models Trained and Tested:

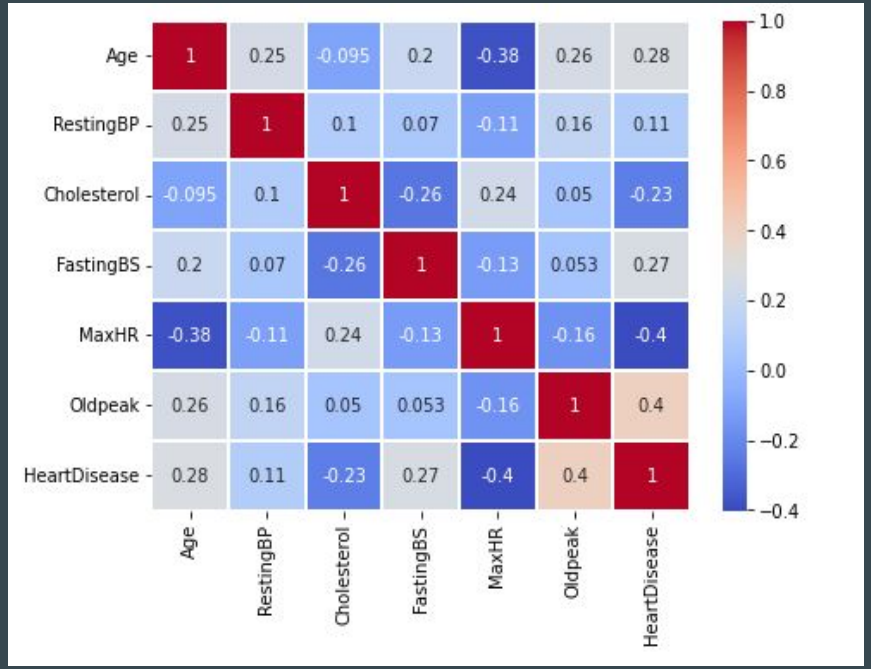
- Logistic Regression: 87% Accuracy
- K-Nearest Neighbors: 86% Accuracy
- Random Forest Classifier: 89% Accuracy
- Sequential Neural Network: 85% Accuracy
- Keras Tuned Neural Network: 96% Accuracy



Although the dataset that we used was already very clean, it still needed one-hot encoding and scaling prior to training and testing each model. Ultimately, we chose the Keras Tuned Neural Network as our production model to use in our application because of it's high testing accuracy and low data loss (1.2%). Also, it predicts the probability of a positive outcome rather than just a simple yes or no.

Most Influential Data Attributes





Challenges

- Data
  - About half of the people represented in our data set have heart disease. It would be interesting to use a more representative data set. According to the CDC, the rate of diagnosed heart disease in adults is 4.6%.
  - The health attribute names were not explained in our data set. It was quite difficult to find a definition for “old_peak” for example.
- Heroku deployment
  - What is a slug, and why was it so massive? It’s all about our required python libraries that heroku must import.
  - It turns out tensorflow is an ENORMOUS library!

Demo