Heart Disease Project

Avedis Abakelian
Thomas Habeshian

TABLE OF CONTENTS

01

Introduction

Focus

Heart Disease Dataset from Kaggle

Objective

Heart Disease Prediction and Risk Assessment

Analysis

patient characteristics and medical measurements

Structure

Comprehensive structure conducive to predictive modeling

Goal

Comprehensive structure conducive to predictive modeling

Data Preparation 02

Age	Sex	Chest-pain	Resting Blood Pressure
Individual's age	1=male 0=female	1= Typical angina 2 =Atypical angina 3= Non-anginal 4= Asymptotic	mmHg
Cholesterol	Fasting Blood Sugar	Resting ECG	Max Heart Rate
mg/dl	>120 mg/dl 1 = true 0=false	0 = Normal 1 = Wave abnormality 2 = Ventricular hypertrophy	bps

Exercise Induced Angina	Oldpeak	Peak Exercise ST Segment
1=yes 0=no	ST depression induced by exercise relative to rest	1= Upsloping 2= Flat 3 = Downsloping
Major Vessels	Thalassemia	Diagnosis of Heart Disease
0->3 Colored by Fluoroscopy	3 = Normal 6 = Fixed defect 7 = Reversible defect	0 = absence 1, 2, 3, 4 = present

Why these parameters

03

Age	Sex	Chest Pain	Resting Blood Pressure
Most crucial risk factor for cardiovascular diseases	Men have a higher risk of heart disease than premenopausal women.	Chest pain or discomfort due to inadequate oxygen-rich blood to the heart muscle.	High blood pressure damages heart-feeding arteries over time.
Risk triples with each decade of life.	women's risk may be similar to men, but recent data disputes this.	Sensation of pressure or squeezing in the chest.	Elevated risk when combined with obesity, high cholesterol, or diabetes.
Coronary fatty streaks can start forming in adolescence.	Women with diabetes have a higher likelihood of developing heart disease than men with diabetes.	Pain may radiate to shoulders, arms, neck, jaw, or back.	-
82% of coronary heart disease deaths occur in individuals aged 65 and older	-	Can mimic indigestion.	-
Stroke risk doubles every decade after age 55.	-	-	-

Cholesterol	Fasting Blood Sugar	Resting ECG	Max Heart Rate
High LDL cholesterol narrows arteries	Insufficient insulin production or response leads to elevated blood sugar	USPSTF recommends against screening with resting or exercise ECG for low-risk individuals	Acceleration of heart rate comparable to high blood pressure in increasing cardiovascular risk.
Elevated triglycerides increase heart attack risk	Increased risk of heart attack	Insufficient evidence for assessing benefits and harms in intermediate to high-risk individuals	10 bpm increase associated with at least a 20% rise in cardiac death risk.
High HDL cholesterol lowers heart attack risk	-	-	Similar risk increase to a 10 mm Hg rise in systolic blood pressure.
-	-	-	-
-	-	-	-

Exercise Induced Angina	Peak Exercise ST Segment
Angina characterized by tight, gripping, or squeezing pain or discomfort.	Abnormal test: Horizontal or downsloping ST-segment depression ⩾ 1 mm at 60–80 ms after J point.
Pain usually centered in the chest, may spread to shoulders, back, neck, jaw, arm, or hands	Up-sloping ST-segment depressions considered ‘equivocal’
Types: Stable Angina, Unstable Angina, Variant (Prinzmetal) Angina, Microvascular Angina	Worse prognosis with lower workload or heart rate showing ST-segment depression.
-	Prolonged recovery after peak stress indicates a positive test.
-	- ST-segment elevation > 1 mm, suggesting transmural ischemia, often prompts urgent coronary angiography referral

Analysis

04

Data Cleaning and Exploration

- Checked for missing values in each attribute
- Found 2 null values in 'thal' and 4 in 'ca' columns
- Eliminated cases with missing values to preserve data integrity.
- Used 'describe()' method for descriptive statistics, revealing data distribution properties.

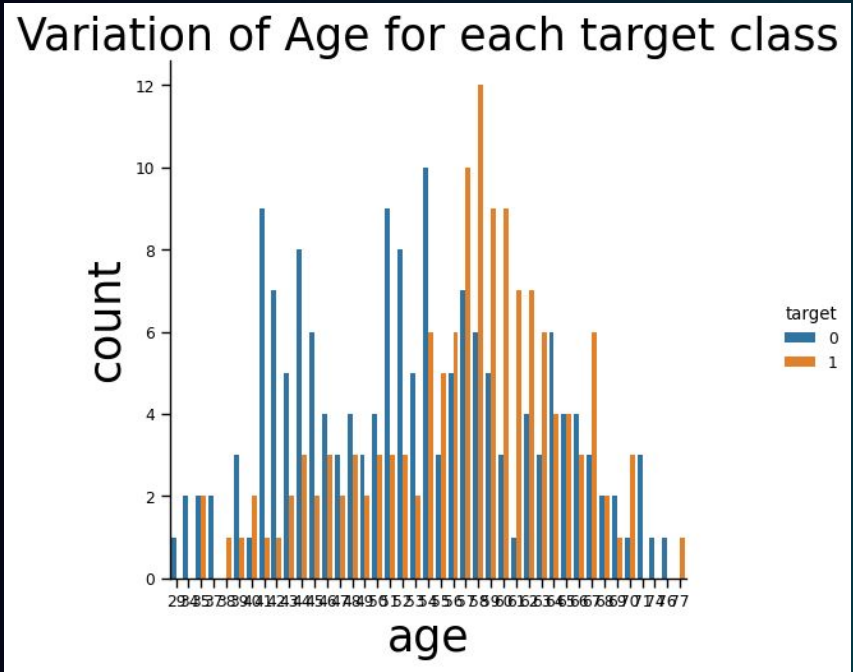
Data Types

- 3 features of type float64, 11 features of type int64

Duplicate Entries

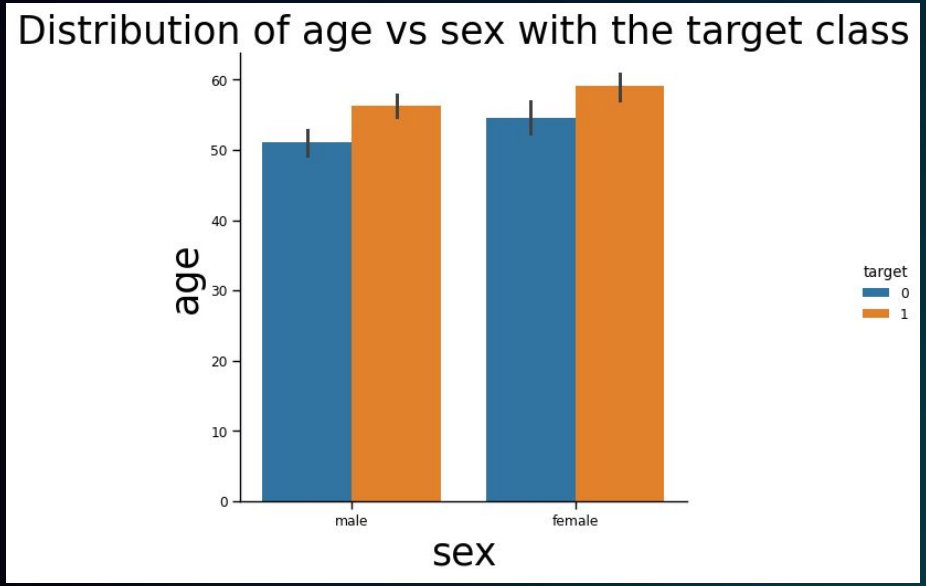
- No duplicate items found in the dataset

Age Distribution



- The concentration of individuals with heart disease is notably centered around ages 58 and 57.
- There is a significant prevalence of heart disease among individuals aged 50 and above.
- The dataset showcases a higher occurrence of heart disease in individuals within this older age bracket, particularly around 50, 57, and 58 years old.

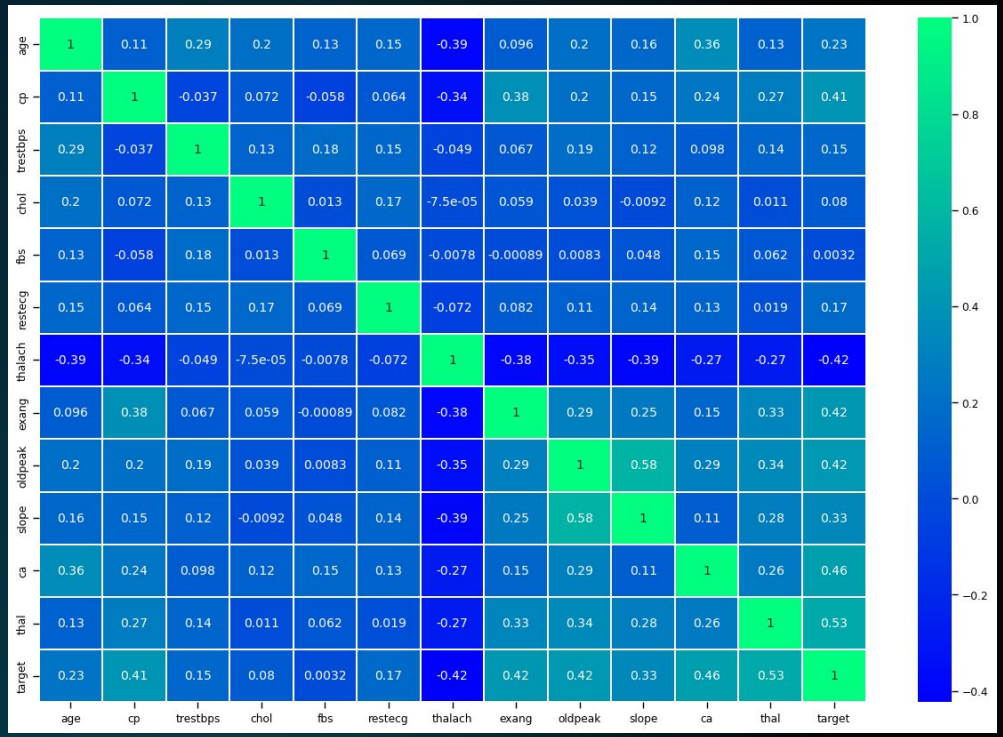
Gender-Based Analysis



Females with heart disease tend to be older than males, indicating age-related gender disparity

Correlation Analysis

- Heatmap showed significant associations with heart disease ('target').
- Positive correlations with chest pain ('cp'), maximum heart rate ('thalach'), and peak exercise ST segment slope ('slope').
- Negative correlations with exercise-induced angina ('exang'), number of major vessels ('ca'), ST depression ('oldpeak'), and thalassemia ('thal').



Supervised Learning Models

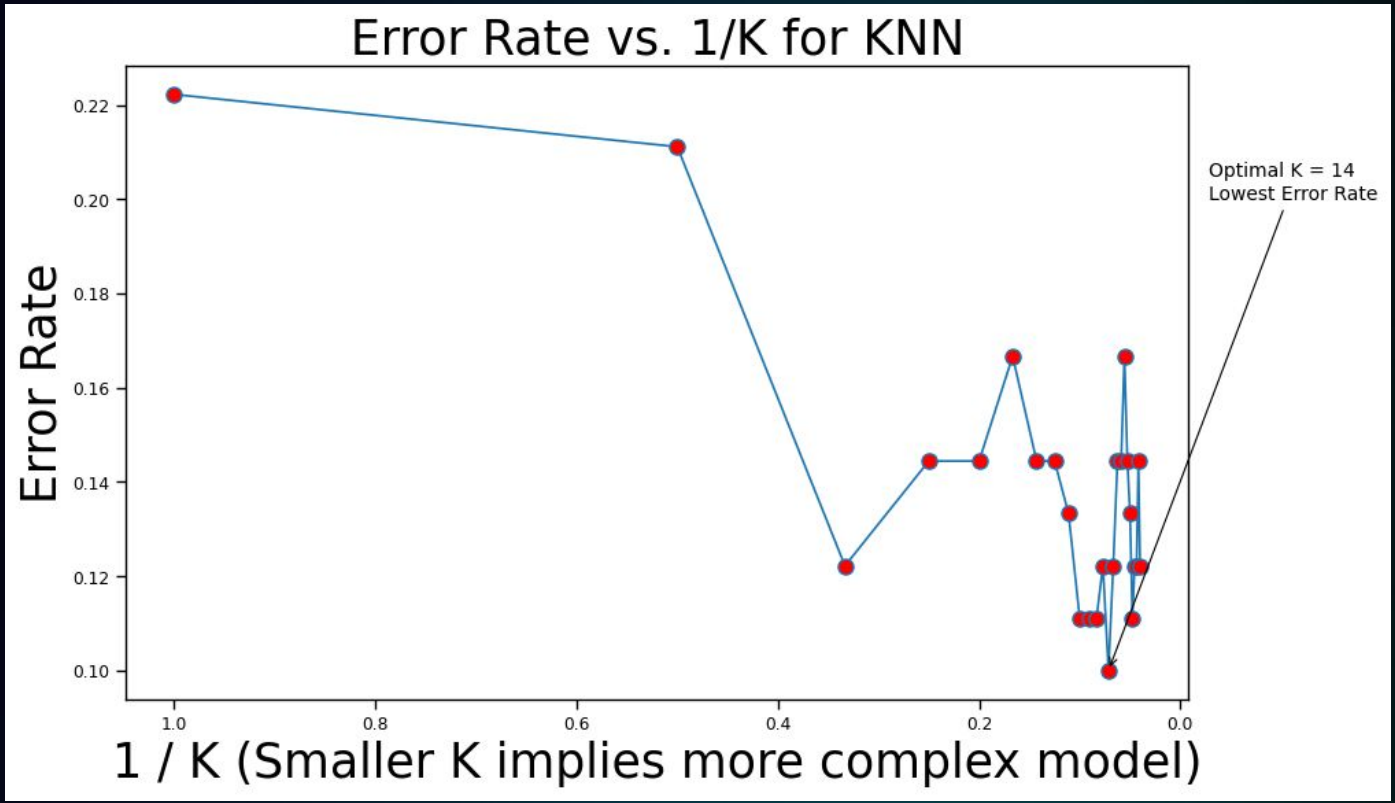
- Decision Tree, Random Forest, Naive Bayes, Logistic Regression, SVM, and KNN used for classification.
- Data divided into 80:20 segments for training and testing.
- Evaluation metrics: accuracy, precision, recall, and F1 score.

Best Model

- Naive Bayes performed best overall with exceptional recall, accuracy, and precision.

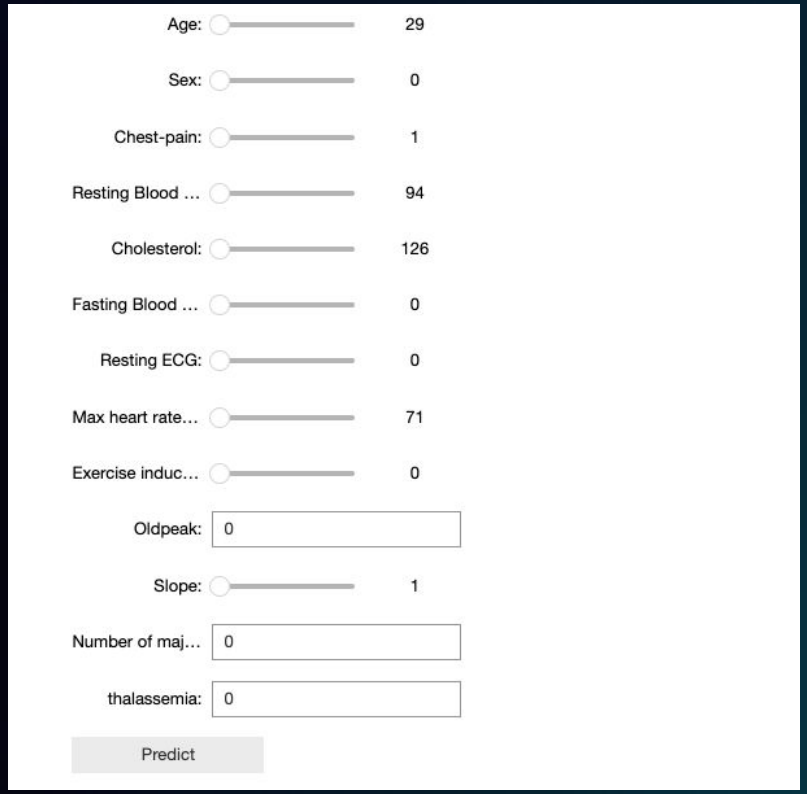
KNN Analysis

- Optimal 'K' value determined as 14 for superior predictive accuracy.



Hyperparameter Tuning

- Accuracy on the test set after tuning: 0.8778.
- Concerns about potential overfitting with a training set accuracy of 0.8164



Model Evaluation

- Caution against relying solely on accuracy for model success evaluation.
- Emphasis on understanding the total performance beyond accuracy.
- Overall promising results in the model's ability to predict heart disease.

Thank you!