Classification and Prediction of Heart Diseases in Patients

Presenter: *Setare Hajarolasvadi*

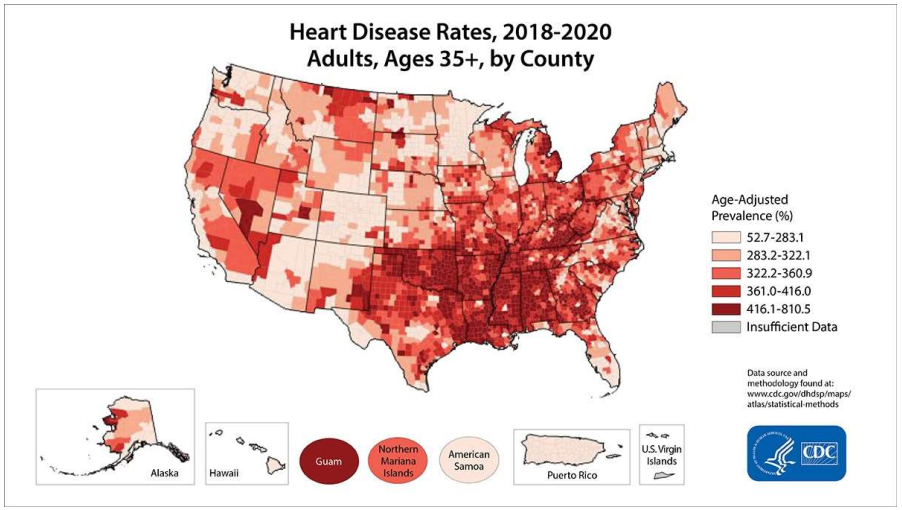
DS Flex Program: *Phase III Project*

Overview

- Objective
- Dataset
- Data Analysis
- Modeling Approach
- Results
- Recommendations
- Limitations

Objective

- One person dies *every 33 seconds* from cardiovascular disease.
- The **objective** of this study is to develop a machine learning model to predict the susceptibility to heart disease in patients.



Dataset

- Three subsets:
  - Cleveland
  - Switzerland
  - Hungary

Heart Disease

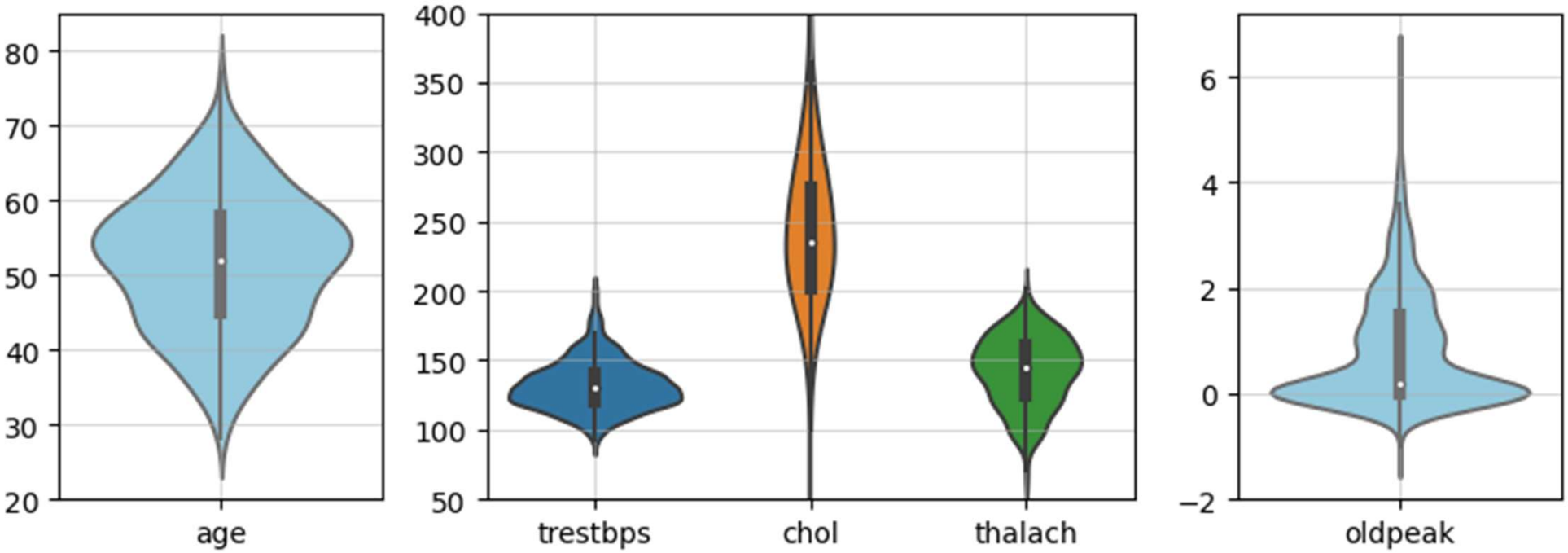
Donated on 6/30/1988

4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

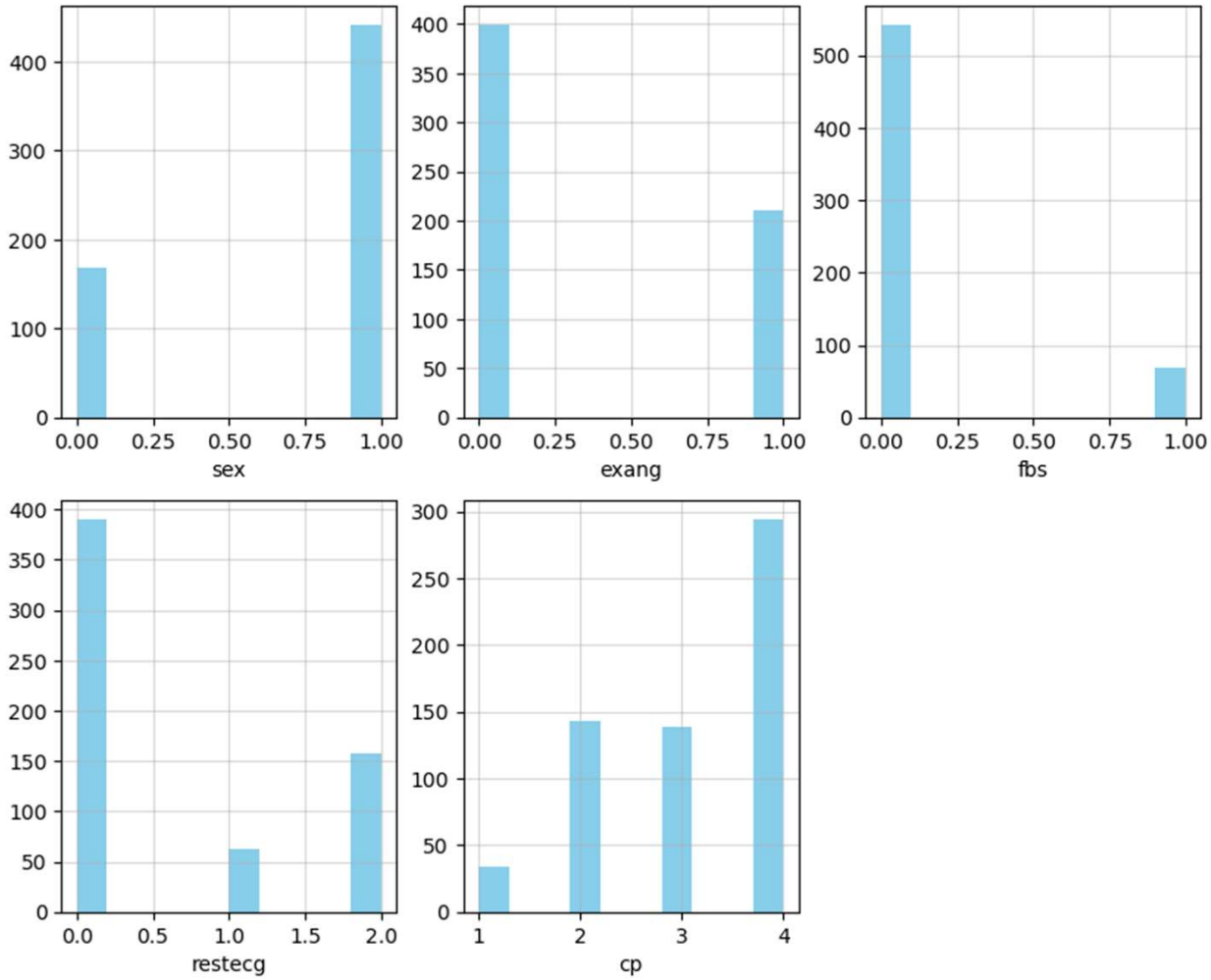
Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Health and Medicine	Classification

Feature Type	# Instances	# Features
Categorical, Integer, Real	~~303~~ **720**	~~13~~ **10**

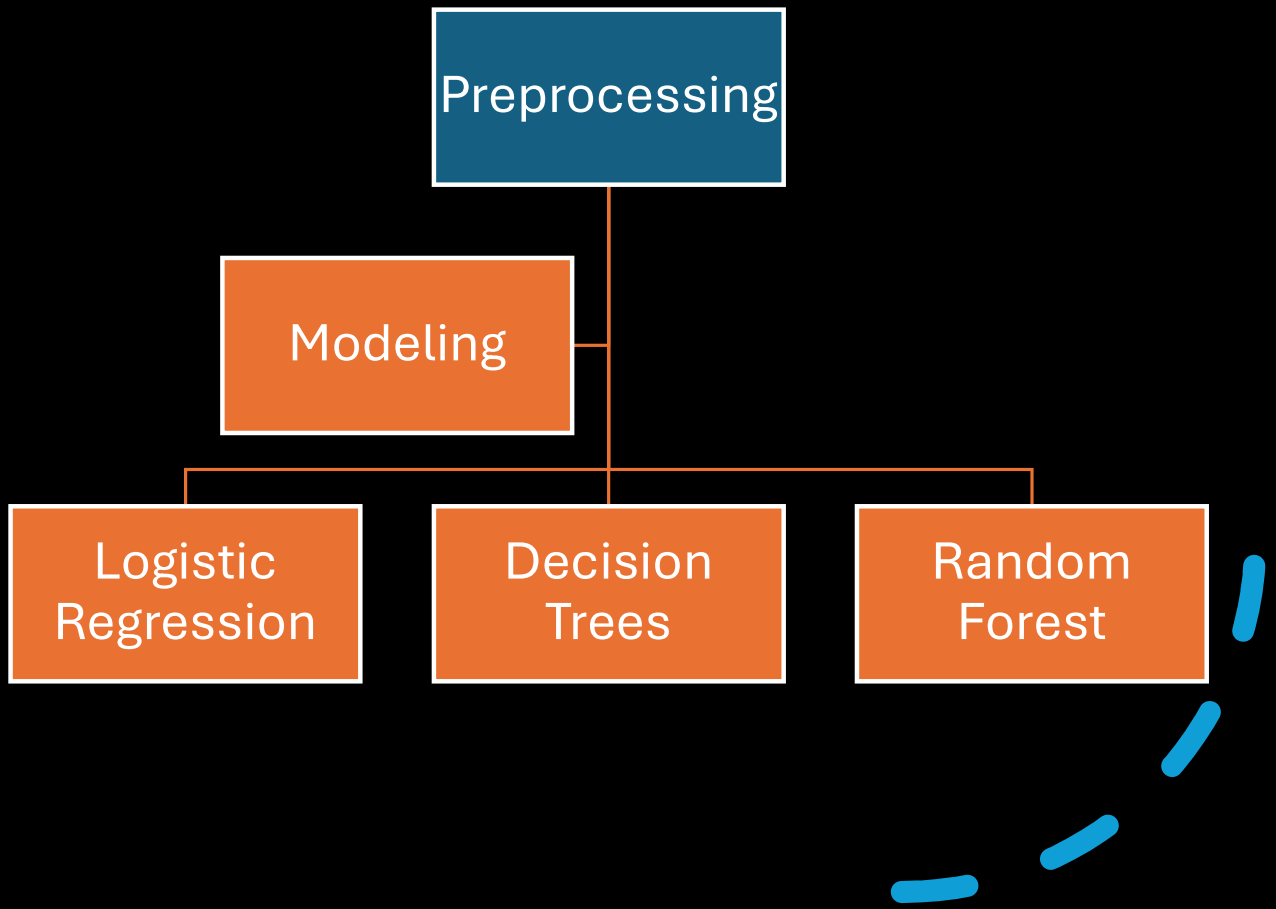
Data Analysis – Numerical Features



Data Analysis – Categorical Features



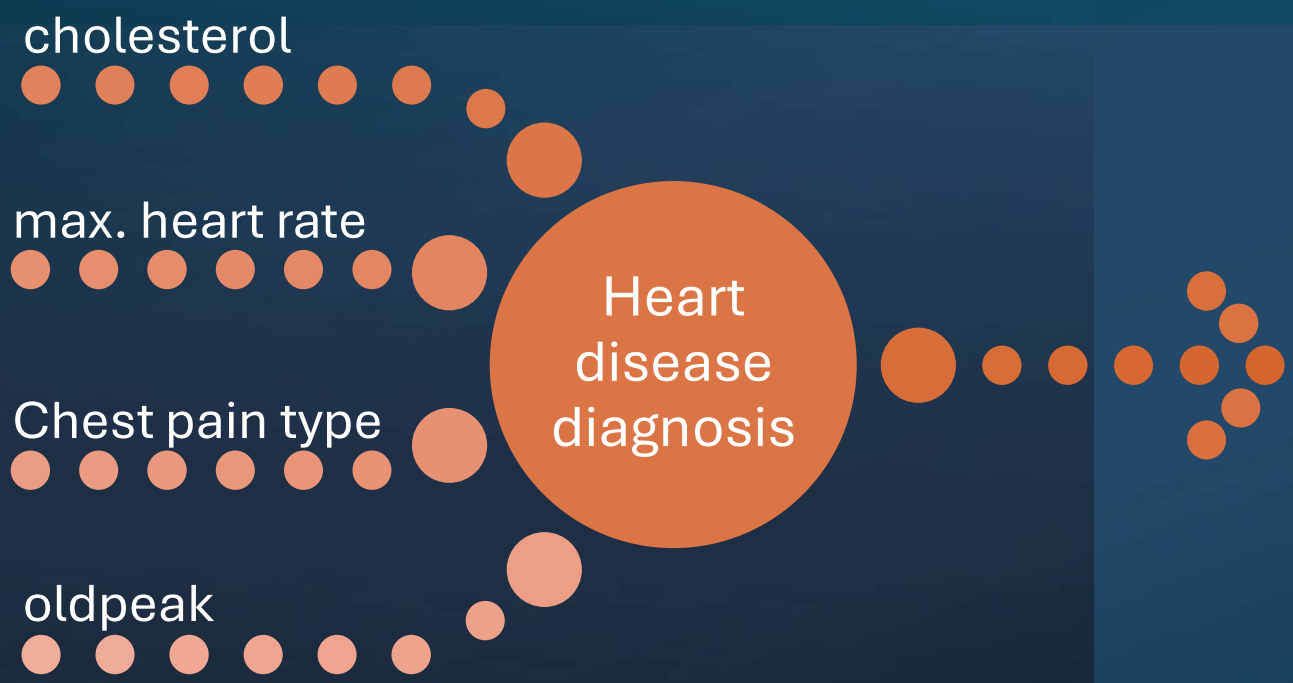
Modeling Approach



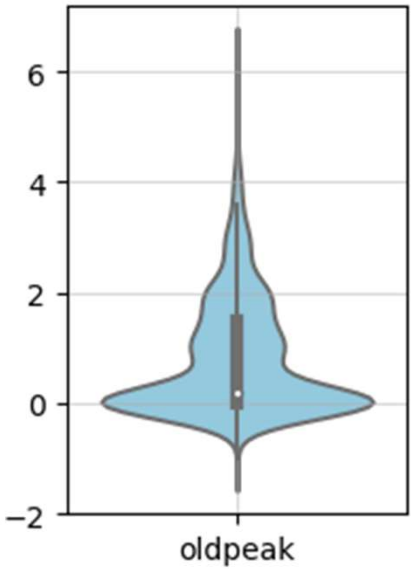
Results

Model	Accuracy	Precision	Recall	f1-score
Logistic Regression	0.78	0.85	0.71	0.77
Decision Tree	0.73	0.77	0.67	0.72
Random Forest	0.81	0.84	0.78	0.81

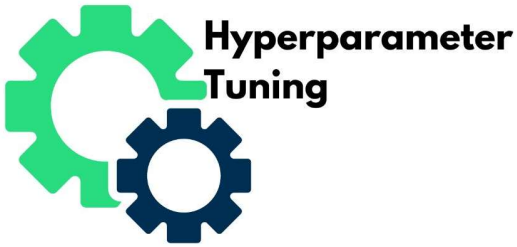
Recommendations



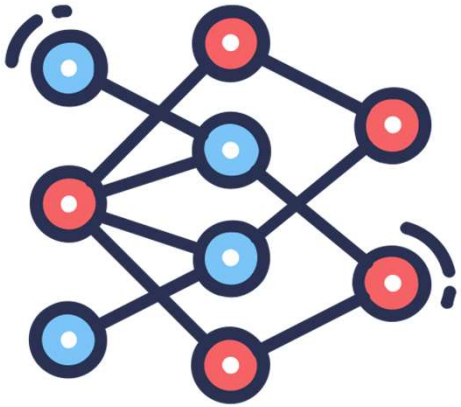
Limitations



Extremely right-skewed distribution



Optimize for recall



Neural Networks
K-Means Clustering