FDS: Heart disease prediction

Second presentation

Elios Buzo, Anthony Giusti, Laurentiu Adrian Cristurean, Ludovico Lentini, Michele Spina

SAPIENZA
UNIVERSITÀ DI ROMA

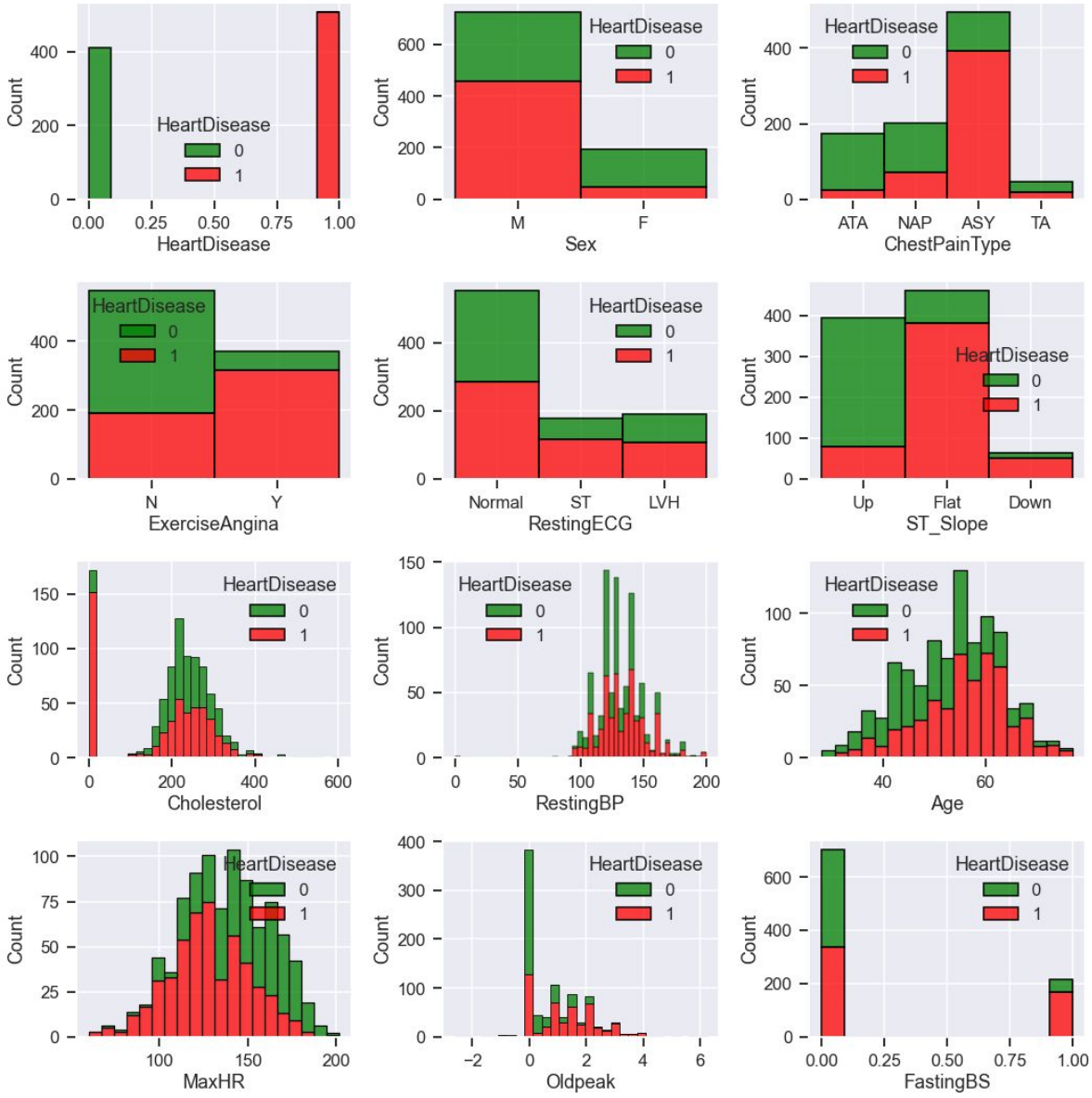
Project for:
Fundamentals of Data Science
2021/22

Prof: Fabio Galasso

SAPIENZA
UNIVERSITÀ DI ROMA

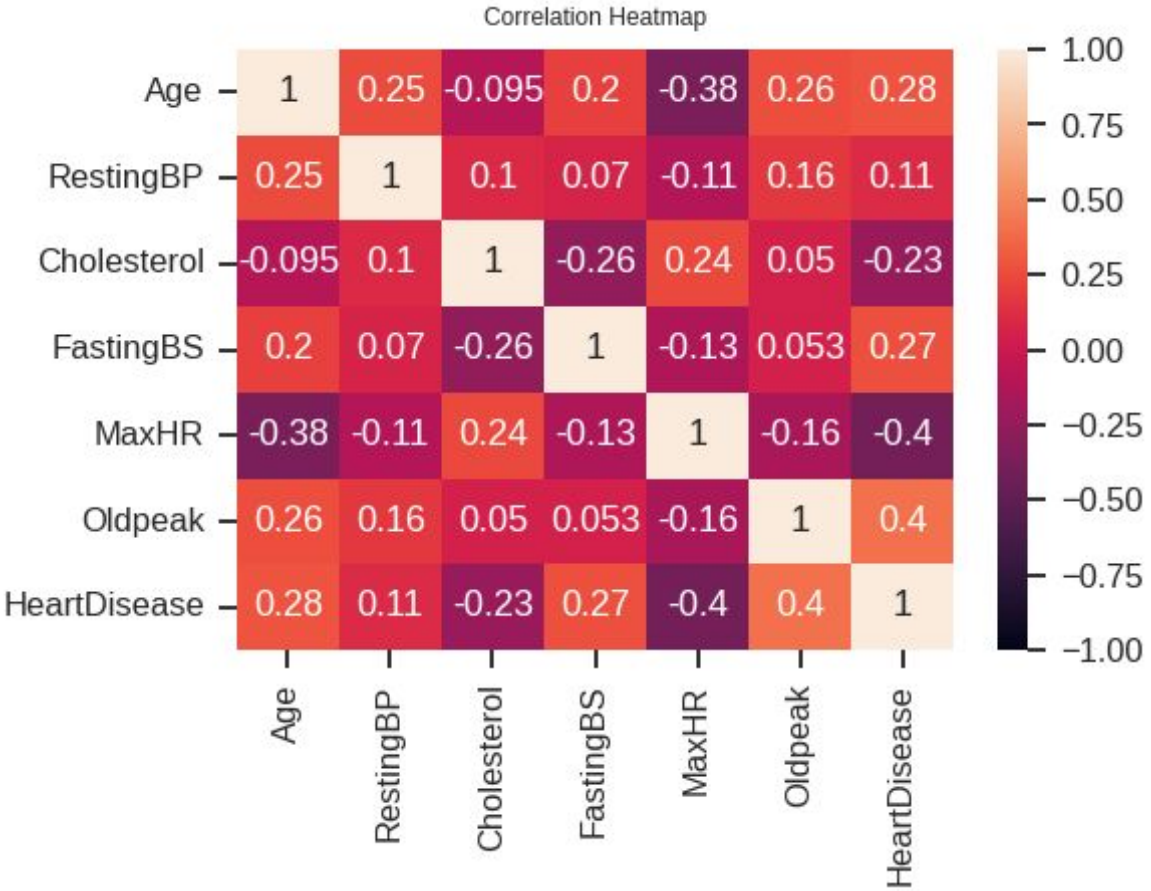
Exploratory Data Analysis

- Numerical Variables normally distributed



SAPIENZA
UNIVERSITÀ DI ROMA

EDA - Correlation HeatMap



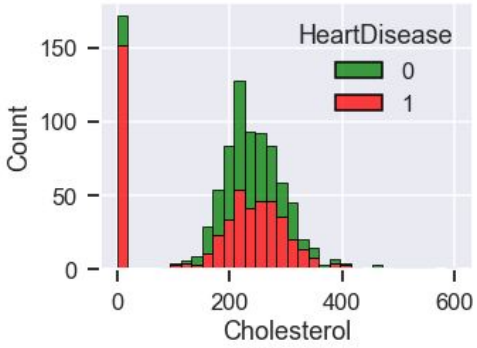
- No strong correlation between variables
- There is no redundancy in the feature variability

SAPIENZA
UNIVERSITÀ DI ROMA

Exploratory Data Analysis - Preprocessing

Manage categorical values: use one-hot encoding

Manage inconsistent values in Cholesterol



1. Delete cols
2. Delete rows
3. Substitute with mean
4. **Substitute with mean and add a column isVirtualValue**
5. Apply linear regression

Normalize values

SAPIENZA
UNIVERSITÀ DI ROMA

Models

- Logistic regression
- Logistic regression with stochastic gradient ascent
- Gaussian Discriminant Analysis
- Naive Bayes:
  - GaussianNB
  - BernoulliNB

SAPIENZA
UNIVERSITÀ DI ROMA

First Analysis

Model	Accuracy	AUC	AP
Logistic regression	0.869	0.92	0.92
LR with Gradient	0.86	0.92	0.92
GDA	0.86	0.93	0.93
GaussianNB	0.87	0.92	0.92
BernoulliNB	0.86	0.91	0.91

All the models are performing well

SAPIENZA
UNIVERSITÀ DI ROMA

Best Results so far: Tuning hyperparameters and cross-validation

Model	Tuning Parameters	Mean accuracy	Mean Roc_Auc
Logistic regression	Max_iter = 1000, C=1 Degree = 1	**0.872**	**0.928**
GDA	solver = svd, lsqr	**0.870**	**0.927**
GaussianNB	var_smooth = 1.23e-05	**0.863**	0.925
GaussianNB	var_smooth = 2.3e-05	0.860	**0.926**

Work in progress

Continue tuning hyperparameters

FDS: Heart disease prediction

Second presentation

Elios Buzo, Anthony Giusti, Laurentiu Adrian Cristurean, Ludovico Lentini, Michele Spina

Thank you for your attention 😊

SAPIENZA
UNIVERSITÀ DI ROMA

Project for:
Fundamentals of Data Science
2021/22

Prof: Fabio Galasso