How to leverage data to reduce costs for inpatients visits of patients with cardiovascular diseases?

Shadi Mahdiani
October 2020

Contents

Heart Dataset

303 rows, 14 columns
with only one duplicate

Description	Values	Type
Age	29-77	Numerical
Sex	0, 1	Categorical
Chest pain type	0, 1, 2, 3	Categorical
Resting blood pressure	94-200	Numerical
Cholesterol	126-564	Numerical
Fasting blood sugar	0, 1	Binary
Rest ECG	0, 1, 2	Categorical
Max HR	71-202	Numerical
Exercise included angina	0, 1	Binary
ST depression	0-6.2	Numerical
ST segment slope	0, 1, 2	Categorical
Number major vessels	0, 1, 2, 3, 4	Ordinal
Thallium stress test	0, 1, 2, 3	Categorical
Target	0, 1	Boolean

Some of parameters are correlated with target in the opposite direction. HOW?

Expectation: patients with heart diseases

- Are older
- Have higher ST depression
- Have higher resting BP
- Have higher Cholesterol
- Their maximum achieved HR is smaller
- Have more number of vessels recognised in fluoroscopy

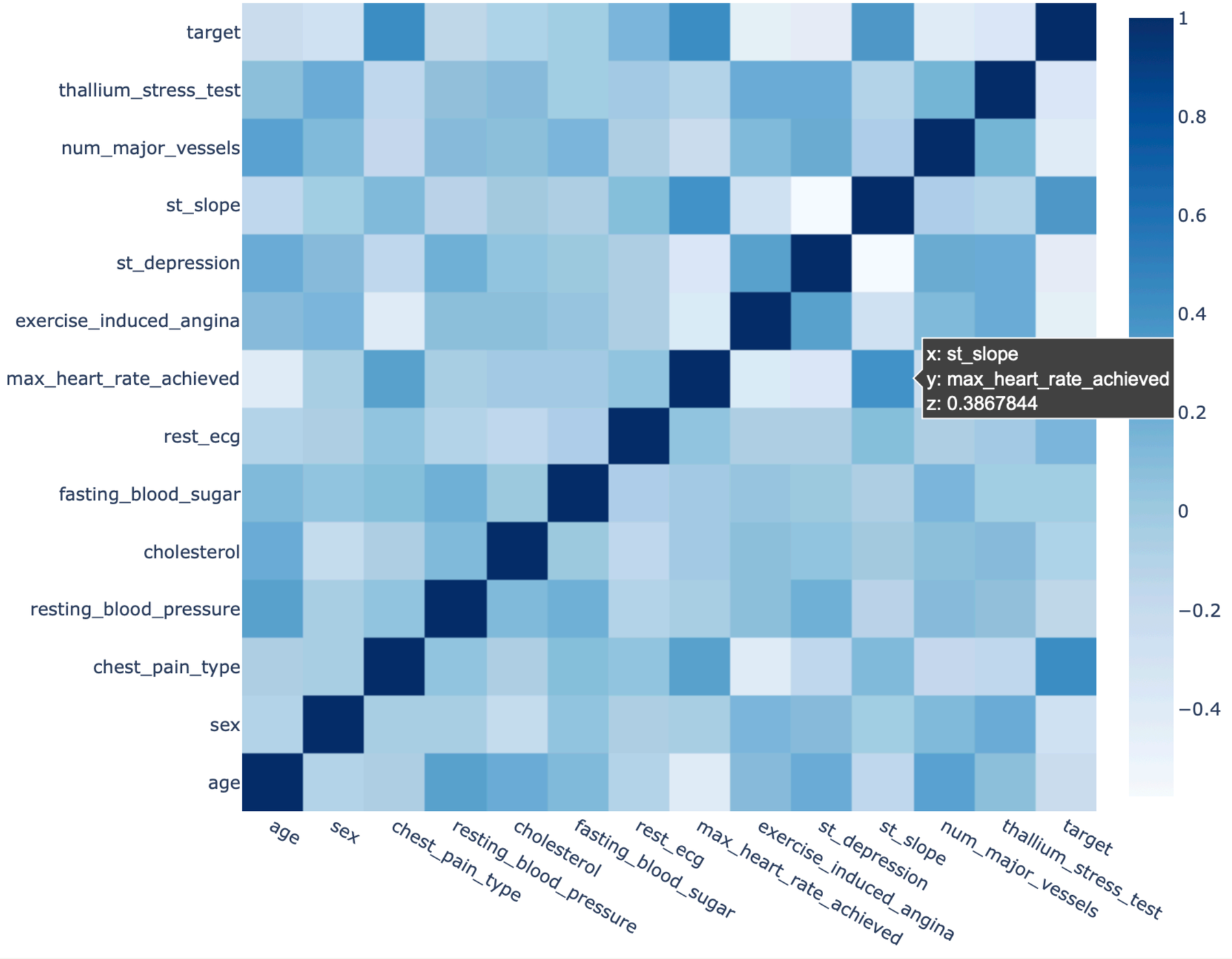
Target = 0	Age	ST Dep	Rest BP	Chol	Max HR	Num Vess
Mean	56.6	1.58	134.39	251.08	139.01	1.16
Std	7.96	1.3	18.72	49.45	22.59	1.04

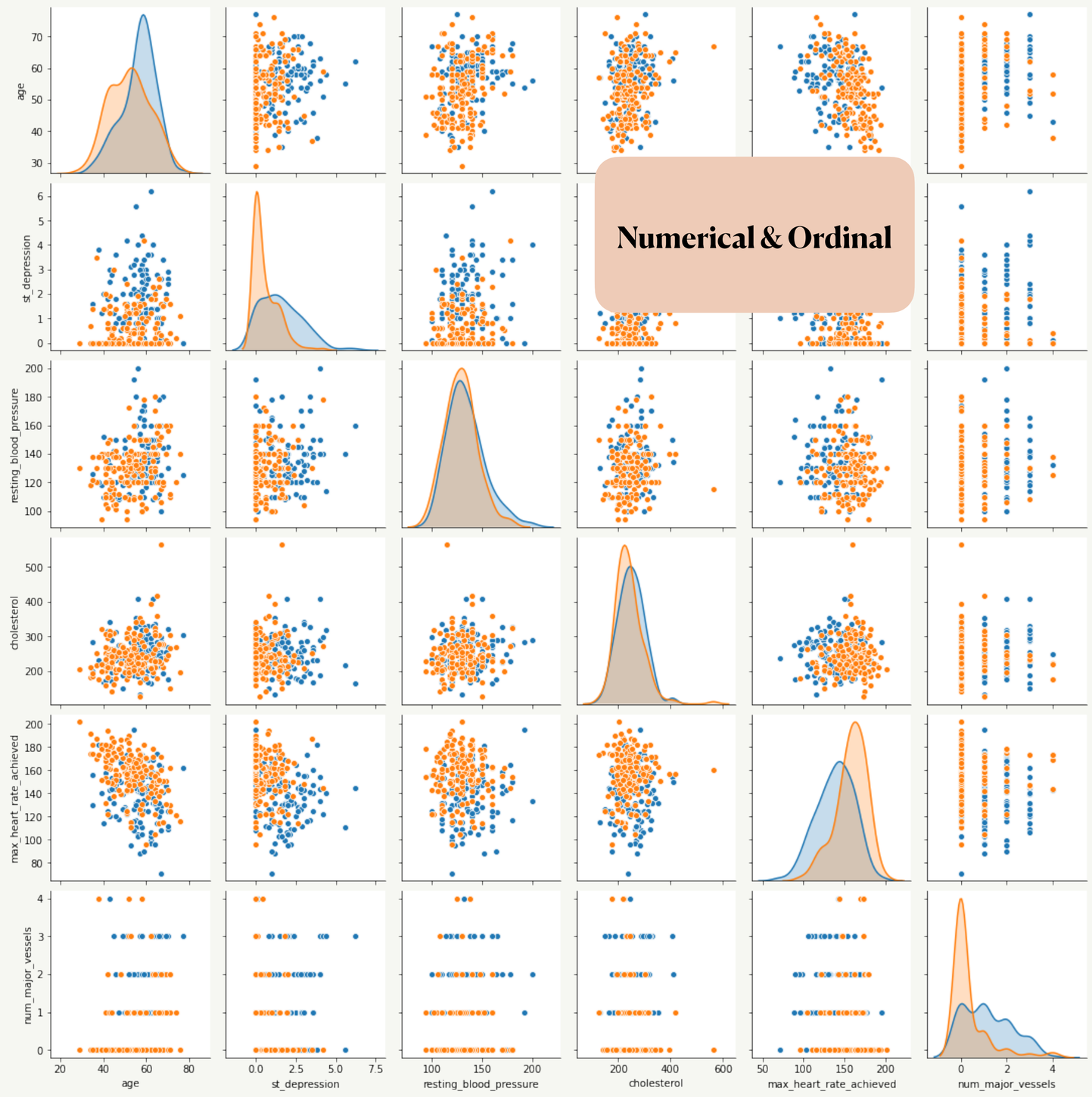
Target = 1	Age	ST Dep	Rest BP	Chol	Max HR	Num Vess
Mean	52.49	0.58	129.3	242.23	158.46	0.36
Std	9.55	0.78	16.16	53.55	19.17	0.84

HeatMap

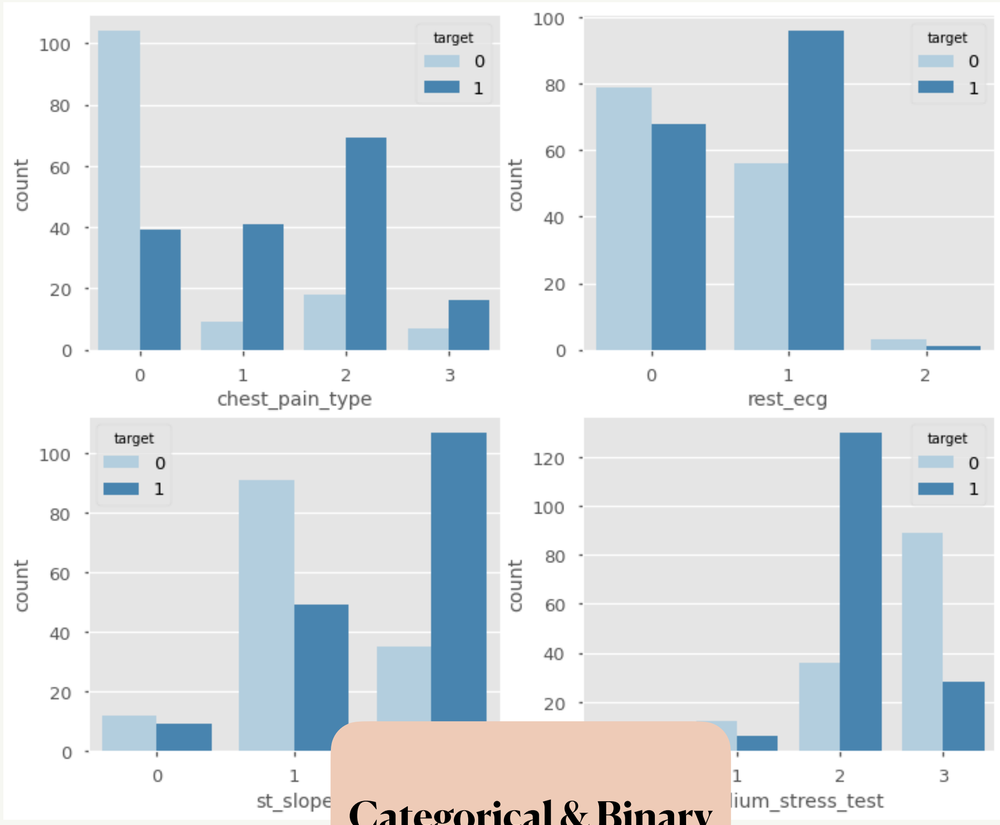
Correlation between all variables

- Parameters seem to be poorly correlated with one another
- No need for eliminating highly correlated features or applying PCA

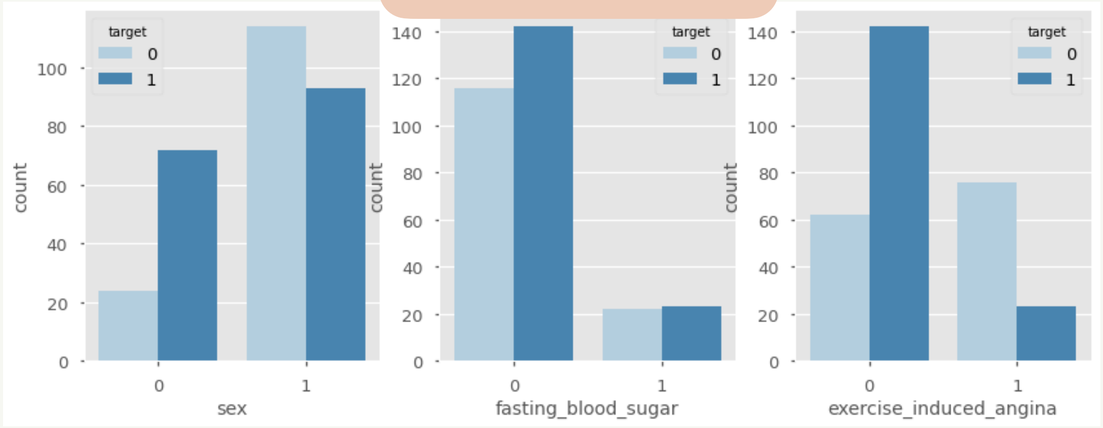




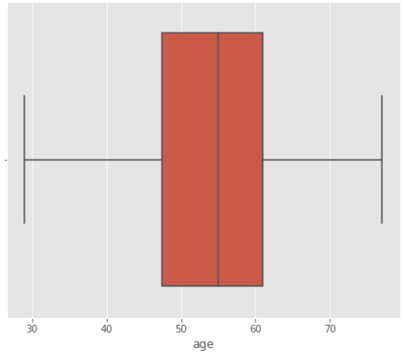
Numerical & Ordinal

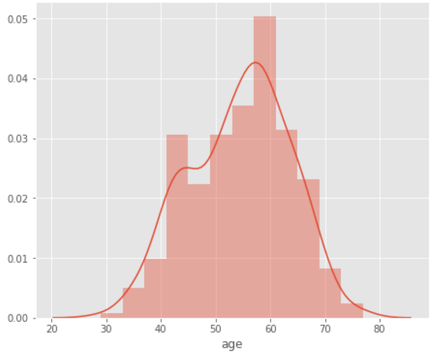


Categorical & Binary

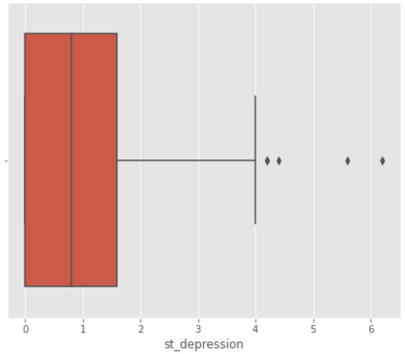


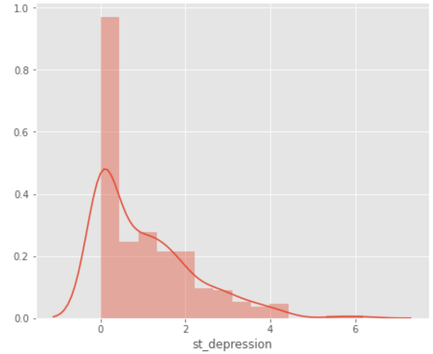
Age



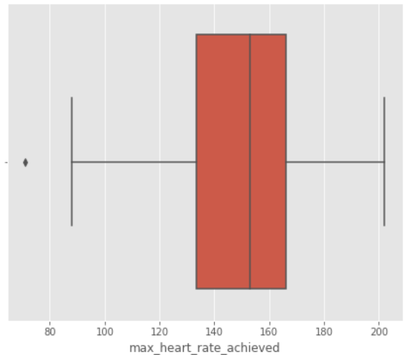


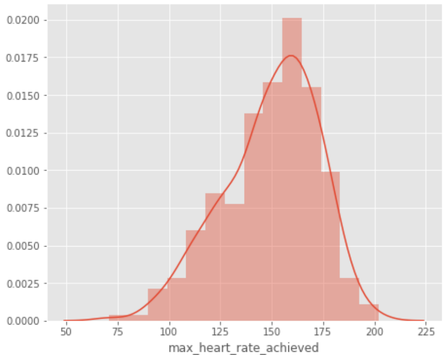
ST Depression



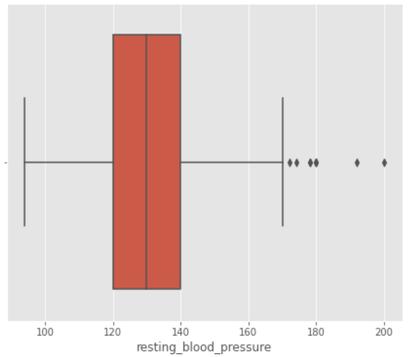


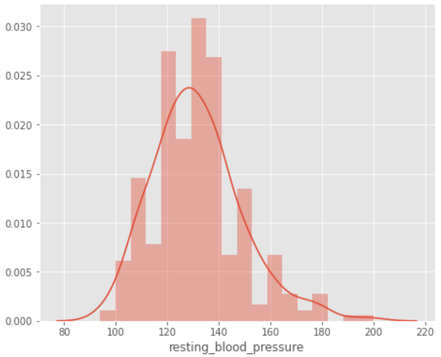
Max HR

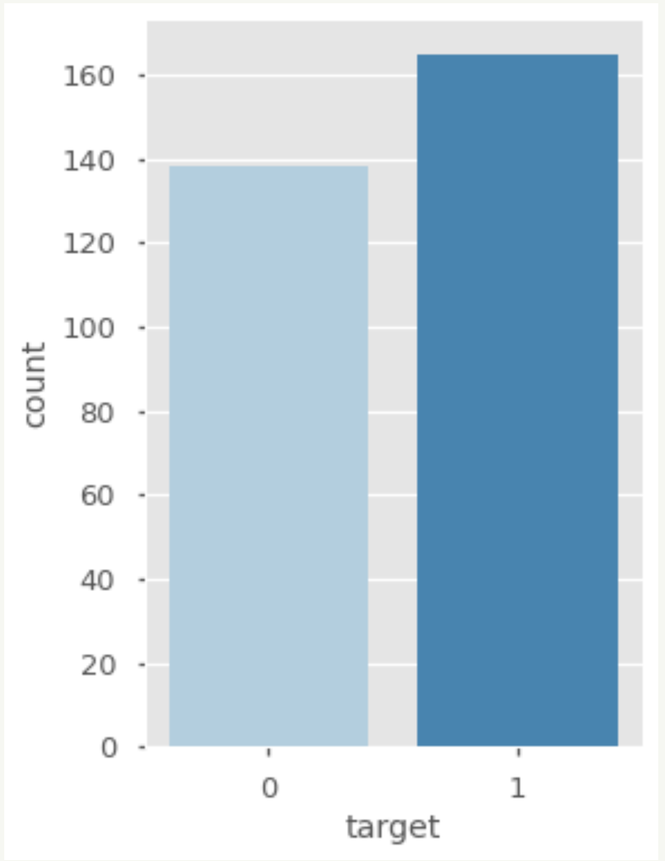




Rest BP







Target (healthy/patient)

Balanced dataset

Feature Encoding & Normalization

- Categorical features: One hot encoding (dummy variables, dropping first column)
- Numerical features: Standard Scaling (mean 0 and std 1)

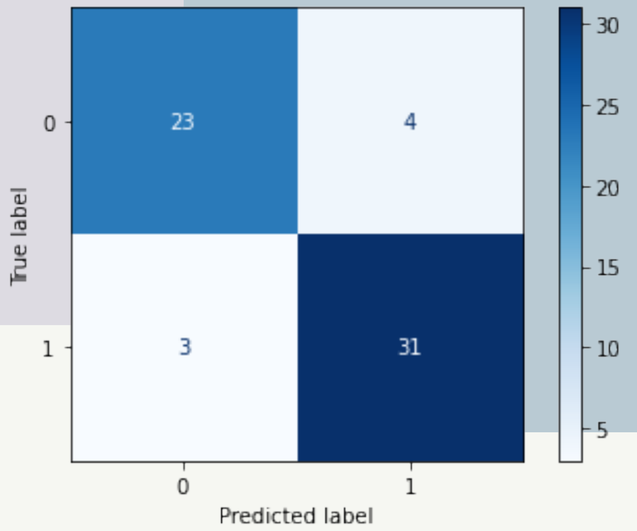
Model training

Split data —> 80% training, 20% validation

Model —> Random Forest

```
Training Accuracy : 0.8966942148760331
Testing Accuracy : 0.8852459016393442
=====================================================
              precision    recall  f1-score   support

           0       0.88      0.85      0.87        27
           1       0.89      0.91      0.90        34
```

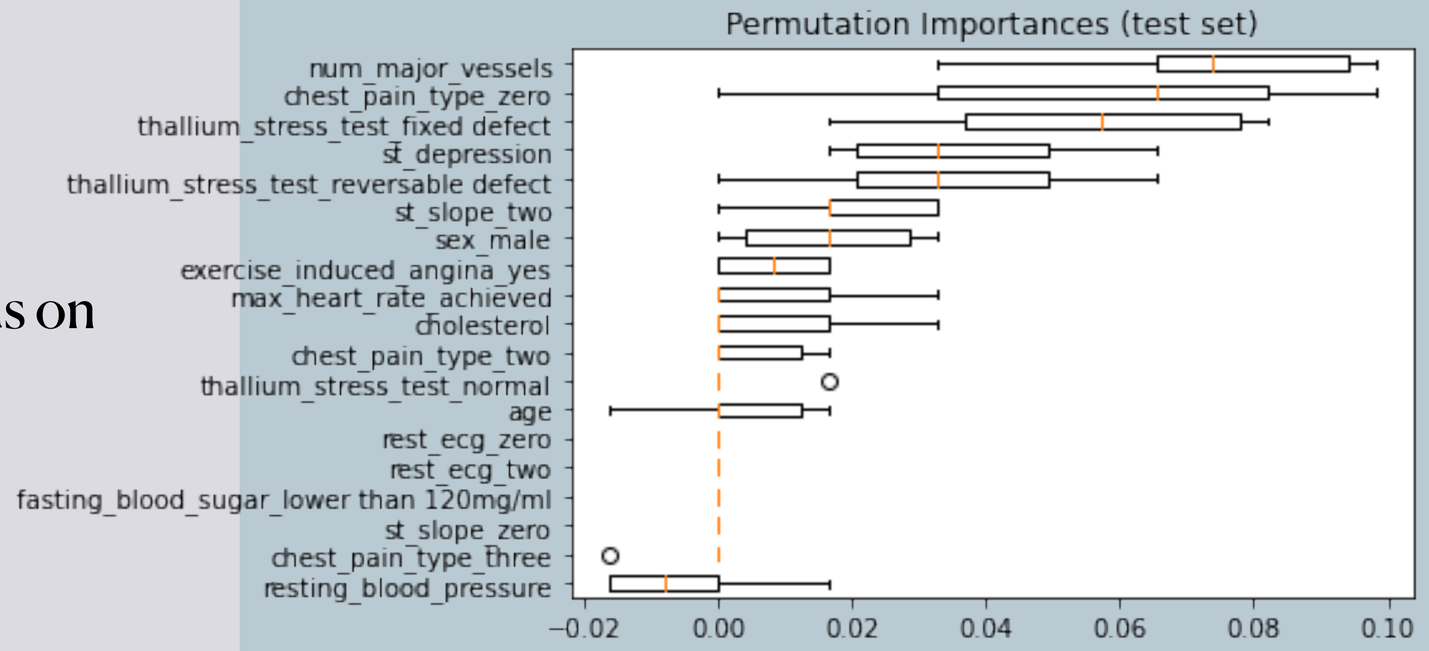


Sensitivity : 0.88, Specificity: 0.87

In medical diagnosis, **sensitivity** is the ability to correctly identify those with the disease (true positive rate), whereas **specificity** is the ability to correctly identify those without the disease (true negative rate).

Permutation Feature Importance

- Computed on a held out test set
- Shows how much the model depends on the feature



HOW AI-driven tool can improve patient safety, avoid unnecessary spending, and enable medical experts to deliver the best possible care in the most efficient manner?

- ✦Identify unnecessary treatments steps
- ✦Identify the best treatment methodologies based on historical data

- ✦Identify patients at risk for death or congestive heart failure
- ✦Identify patients with possible rehospitalization after initial treatment in near future