Heartrate failure classification

1) pre-process data (find outliers with box-plot)

iqr = (75% quartile) - (25% quartile)

uw = q3 + 1.5 * iqr
lw = q1 + 1.5 * iqr

2) i) ~~how~~ fixing outliers

make value above uw = uw
values below lw = lw

ii) normalize data (max - min)

df [' data '] = (df ['data']
- min_value) /
(max_value - min_value)

↑

Use if columns have large values

2) ML model

→ oversampling is used to make sure the target data points is are ~~un~~balanced.
(This generates new data-points)

Date:

x_train, x_test, y_train, y_test
= test_train_split(x, y,
random_state = 4,
stratify = y, test_size = 0.1)

go 10% of og data

makes sure equal category of target data is taken
(No new data points generated)

3) Algorithm (KNN (Nearest neighbours)

checks distance of (example 10) nearest points
(finds avg of each category and takes which has minimum one)

knn_model = KNeighboursClassifier
(n_neighbours = 10)

4) Confusion matrix