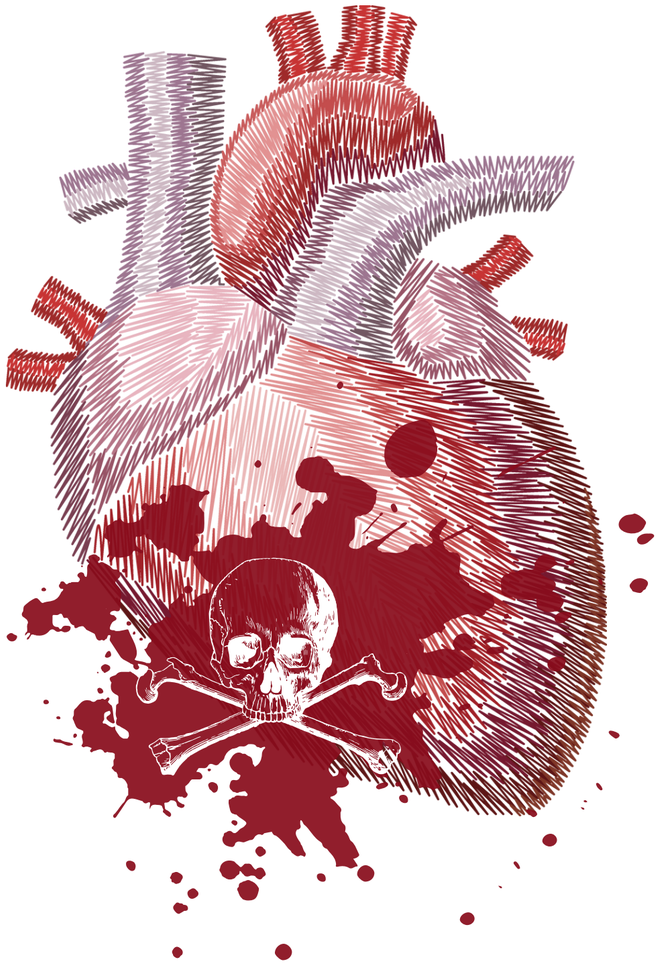


Heart Failure Prediction

Stefanus Hermawan Sebastian 2301878605
Fernando Asikin 2301870142
Matthew Lim 2301873182

Introduction

WHAT?

Prediction of Heart Failure.

WHY?

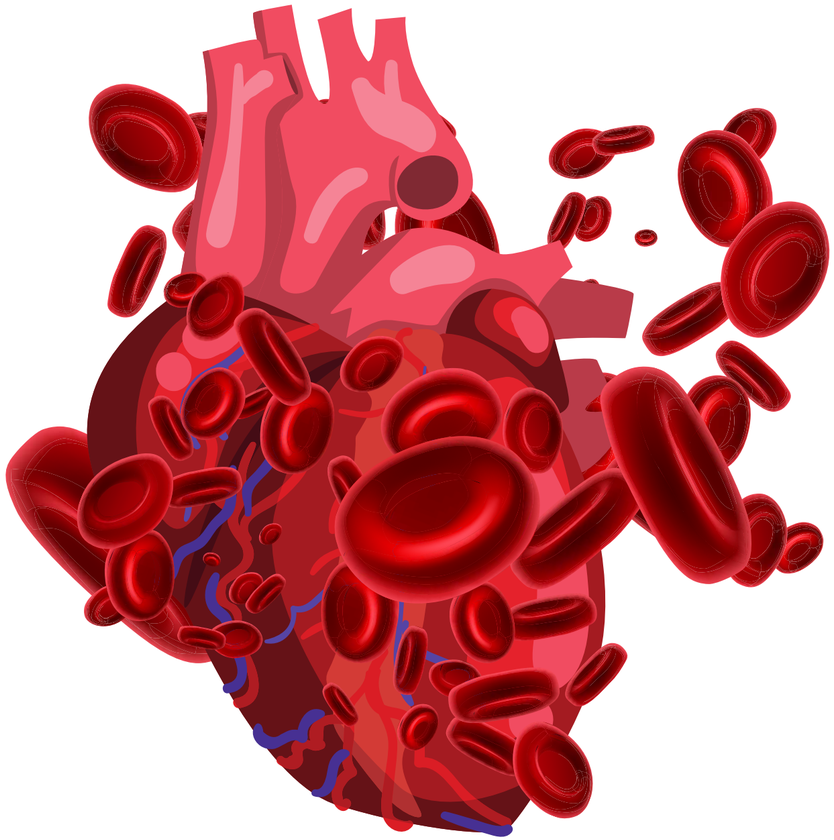
Because we want to prevent the death risk of heart failure.

HOW?

From the data, our machine learning can predict if people will have a probability of having a heart failure using supervised learning.

WHO?

people who have a potential heart failure



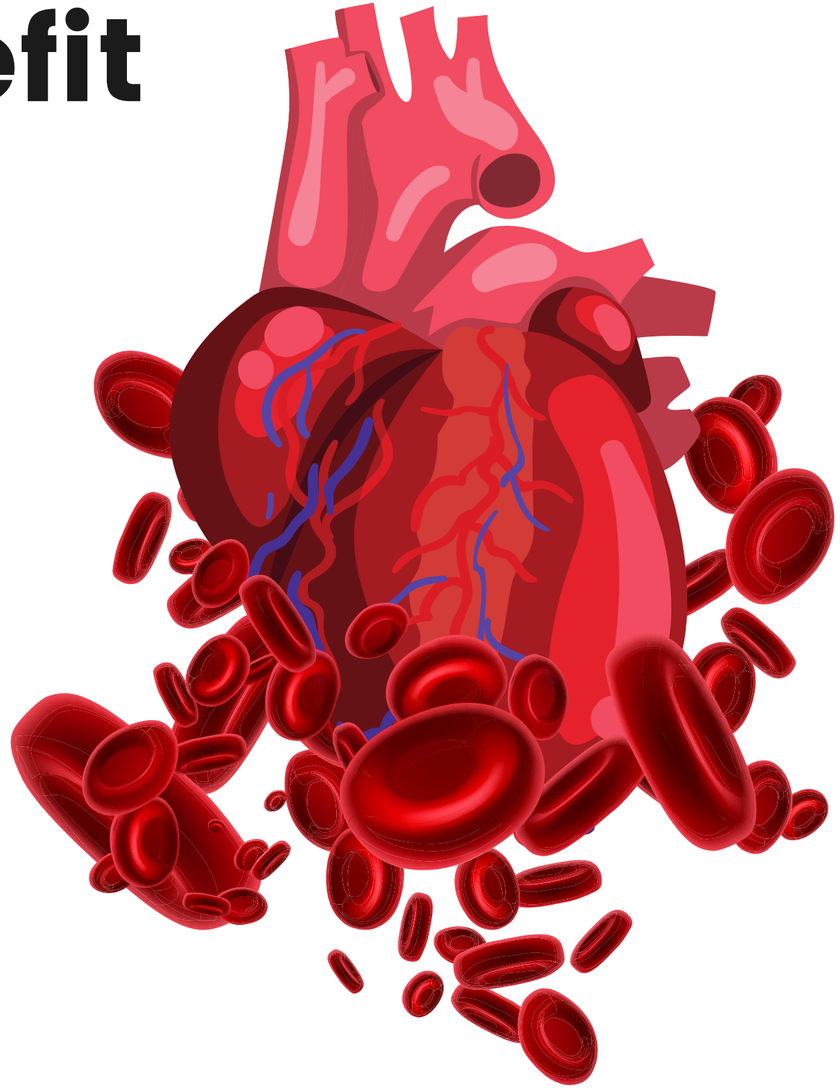
Purpose and Benefit

PURPOSE

To reduce a person's potential for death caused by heart failure

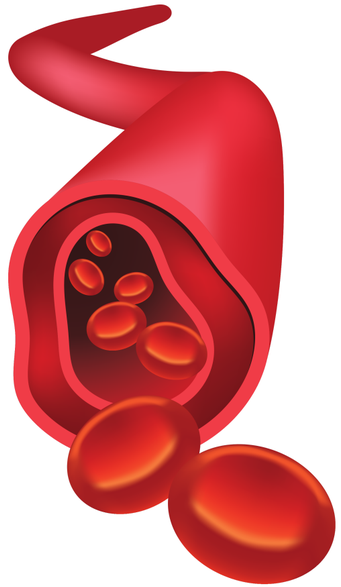
BENEFIT

Users can predict whether they have the potential for heart failure or not



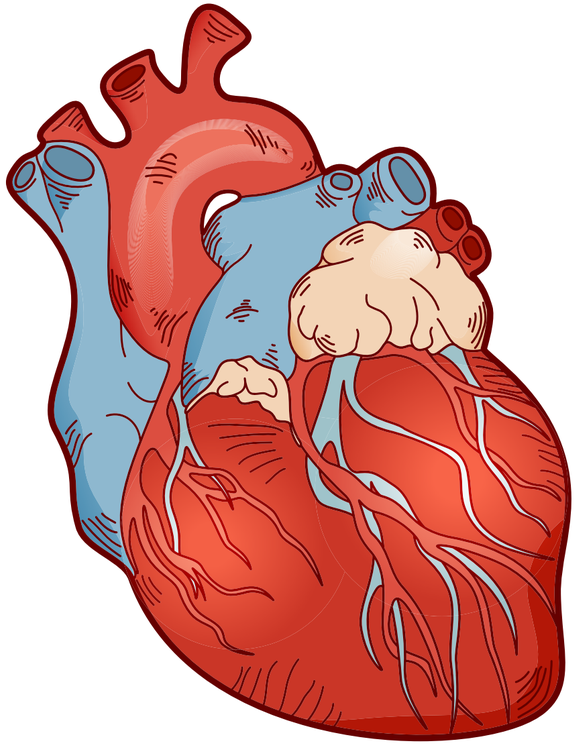
Resources

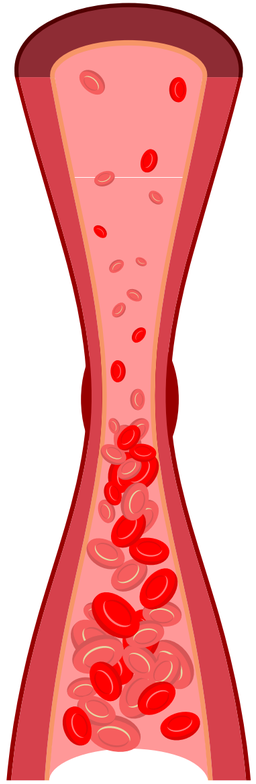
https://www.kaggle.com/andrewmvd/heart-failure-clinical-data



Dataset

- age
- anaemia
- creatinine_phosphokinase
- diabetes
- ejection_fraction
- high_blood_pressure
- platelets
- serum_creatinine
- serum_sodium
- sex
- smoking
- time
- DEATH_EVENT





//05

How can we predict?

We will using several classification methods for classify the probability of people getting the heart failure.

Why?

Classification is used to identify which people have a risk of getting heart failure.

In this experiment we use several classification to compare which of the method have the highest accuracy.

//06

source

Heart Failure Prediction Using Machine Learning Techniques

Prasanta Kumar Sahoo
Sreenidhi Institute of Science and Technology
Pravalika Jeripothula
Sreenidhi Institute of Science & Technology (SNIST)

December 15, 2020

Literature Review

- based on a paper written by Prasanta Kumar Sahoo and his friends entitled "Heart Failure Prediction Using Machine Learning Techniques" proves that SVM has good accuracy for classifying heart disease.
- Prasanta Kumar Sahoo and friends implemented many algorithms such as SVM, Naïve Bayes, Logistic Regression, Decision Tree, and KNN. It was found that SVM gave the best results with an accuracy of up to 85.2%.
- Prasanta Kumar Sahoo and friends were using 3 types of python libraries: Numpy, Pandas, and Sklearn

//07

What Have Been Done?

- Searched for the dataset
- Decided the method
- Set the background
- Analyze dataset
- Doing research
- Creating report

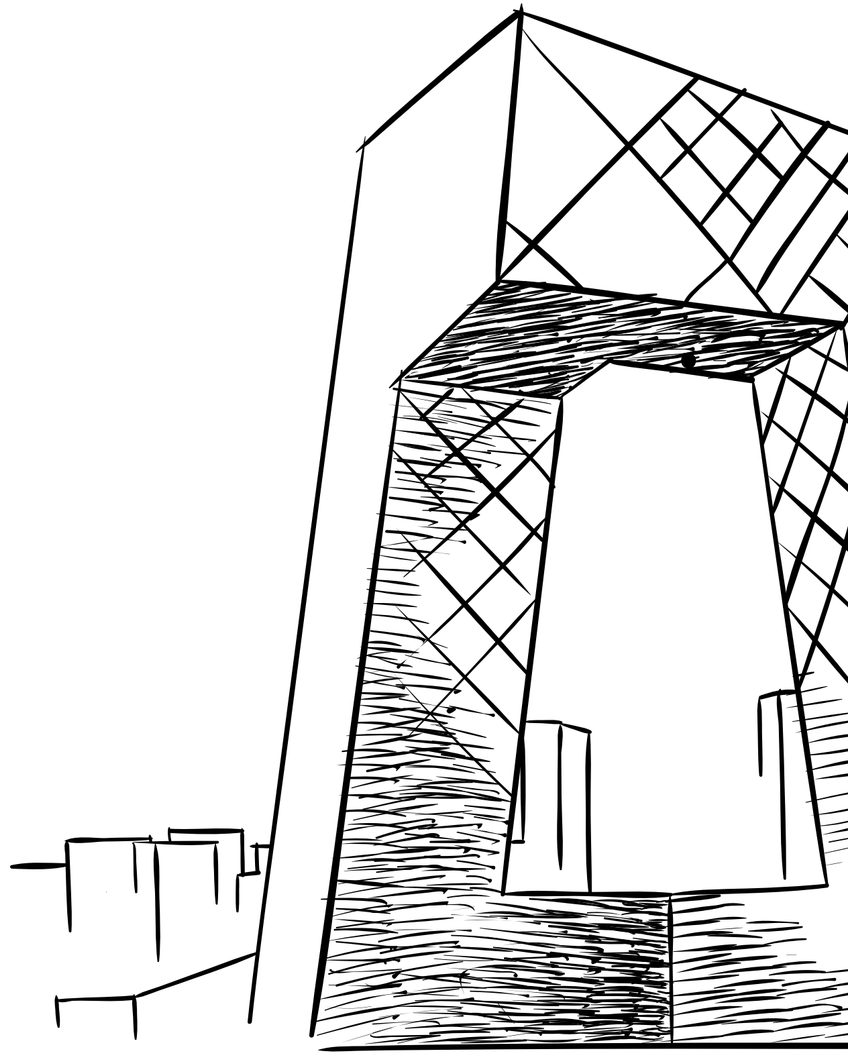
//08

Details of future works

- Can improve the predictive outcome of heart failure.
- Can use less features to predict outcome of heart failure.
- Can be implemented in real life to help doctors or nurses warn user to become more aware of their health condition.

//09

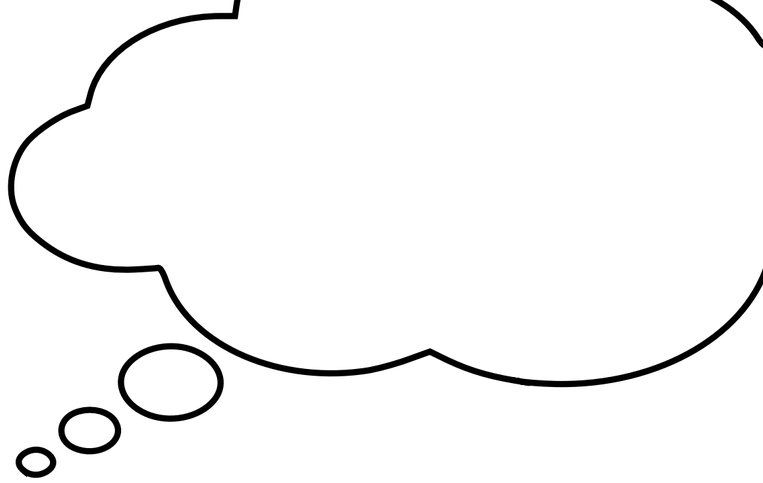
Architecture System



HEART FAILURE | Machine Learning

Flow of the experiment

Start



- Start to determine what topics we will discuss, the problems and ways to classify them

//12

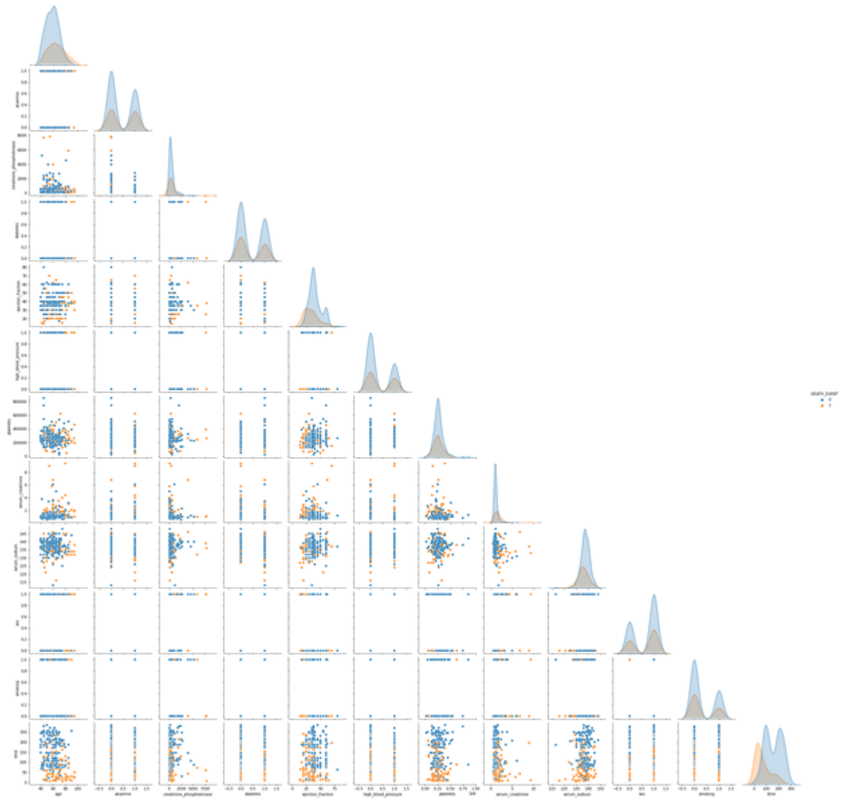
Dataset

- Determine the dataset to be used in conducting experiments
- Searching the dataset on the internet (Kaggle)
- Set the dataset into python

Preprocessing Data

- Make the pairplot and heatmap of the feature to see which have a great correlation with DEATH_EVENT.
- Determine the rank of 5 best features using heatmap and pairplot
- Make the pairplot of the best 5 features to make sure it have a great correlation
- Clean out the existing outliers on the 5 selected data features
- Split data into 70% training data, and 30% testing data

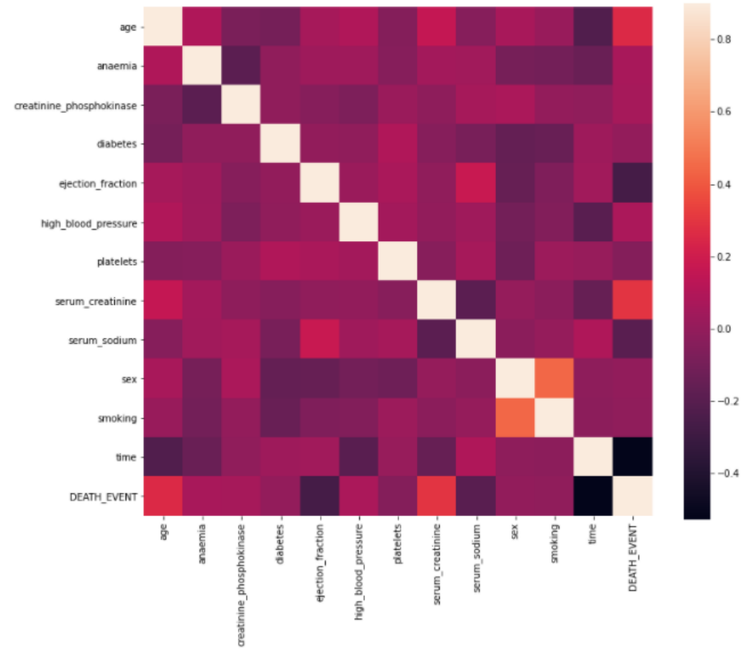
Preprocessing Data



here is the "pairplot" for all the fatures on the existing dataset

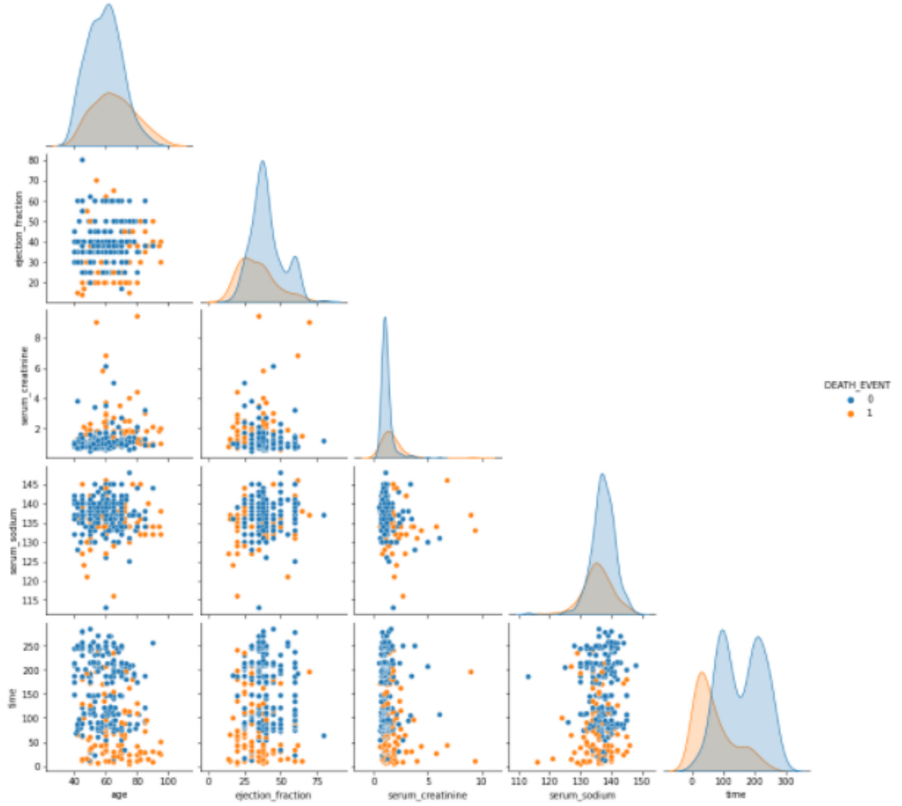
Preprocessing Data

```
age                         0.253729
anaemia                     0.066270
creatinine_phosphokinase    0.062728
diabetes                   -0.001943
ejection_fraction          -0.268603
high_blood_pressure         0.079351
platelets                  -0.049139
serum_creatinine            0.294278
serum_sodium               -0.195204
sex                        -0.004316
smoking                    -0.012623
time                       -0.526964
DEATH_EVENT                 1.000000
Name: DEATH_EVENT, dtype: float64
```



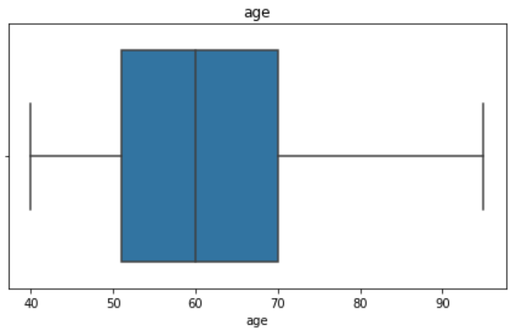
Here is a heatmap to define 5 best features

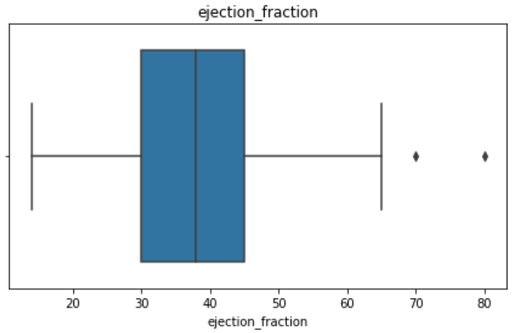
Preprocessing Data

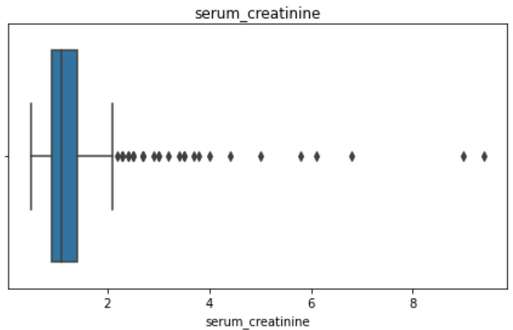


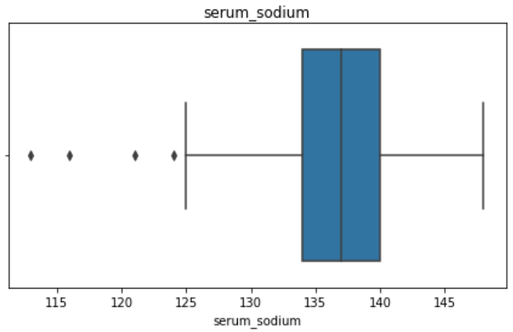
Here is the "pairplot" for 5 best features

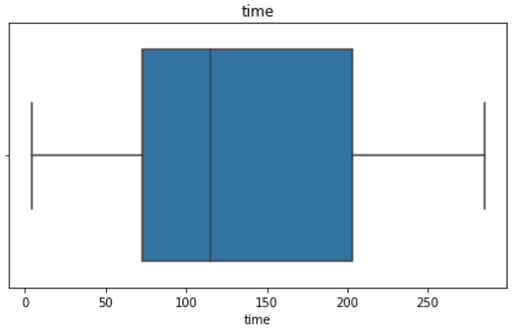
Preprocessing Data

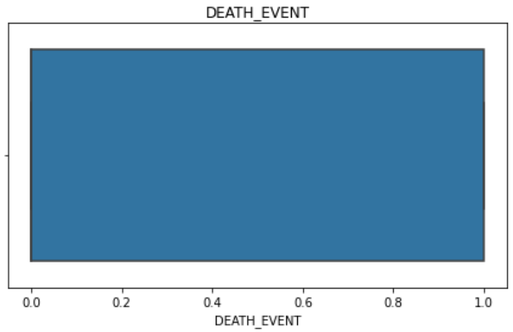






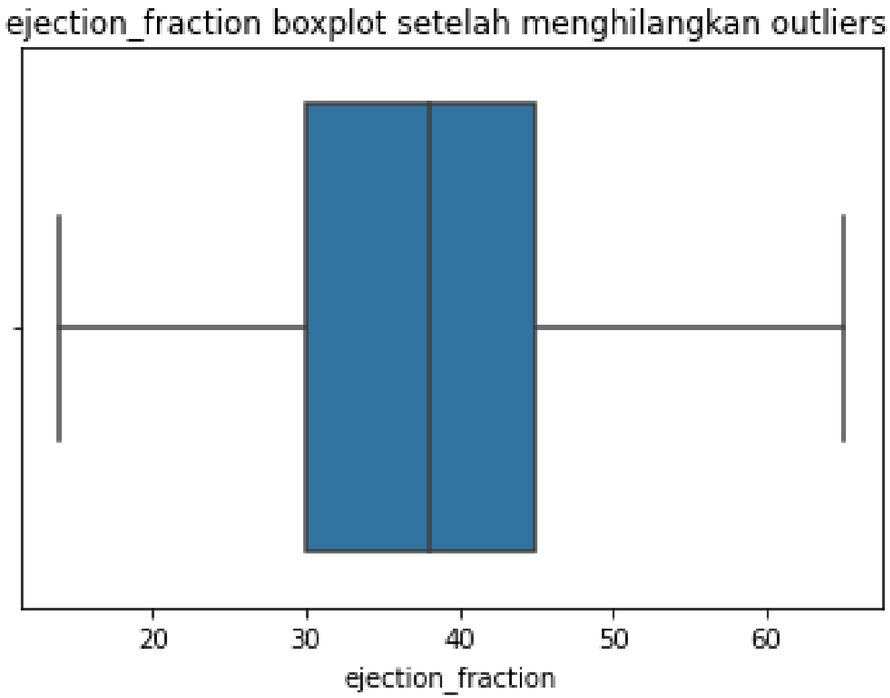






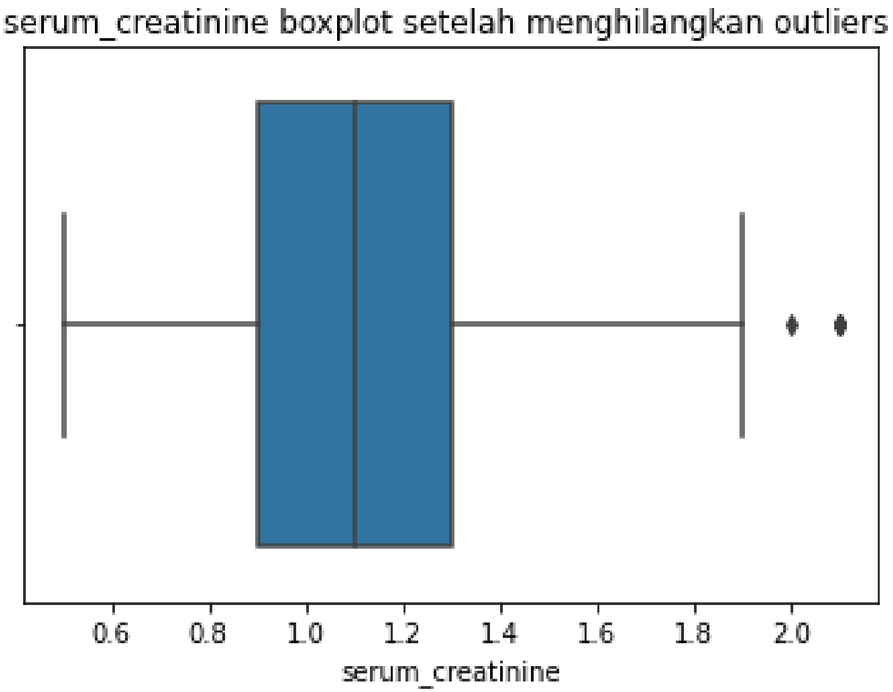
determine which features to clean

Preprocessing Data



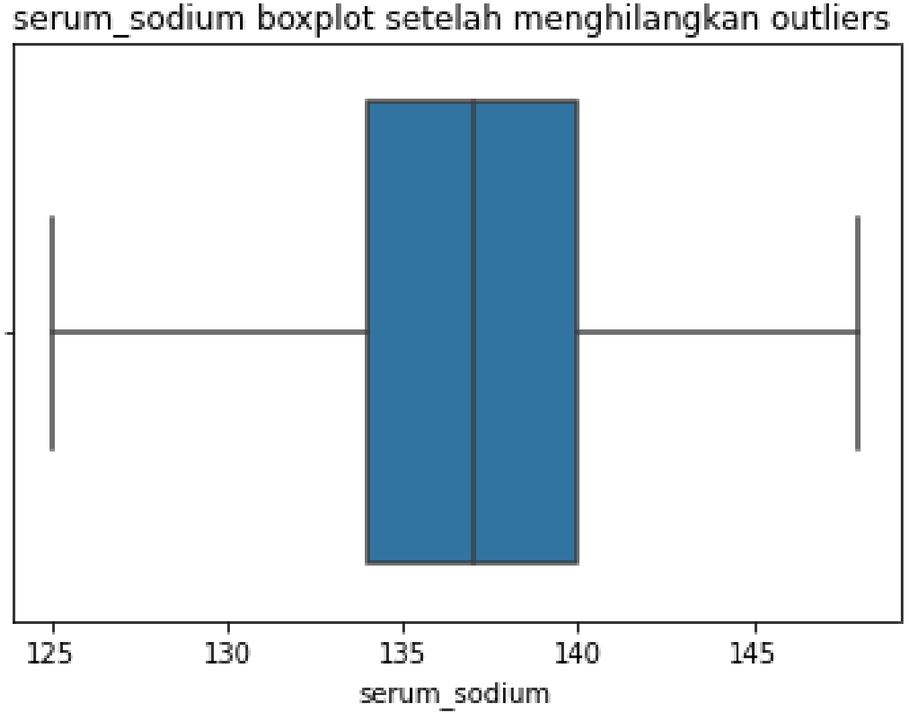
cleaning
ejection_fraction feature

Preprocessing Data



cleaning
serum_creatinine
feature

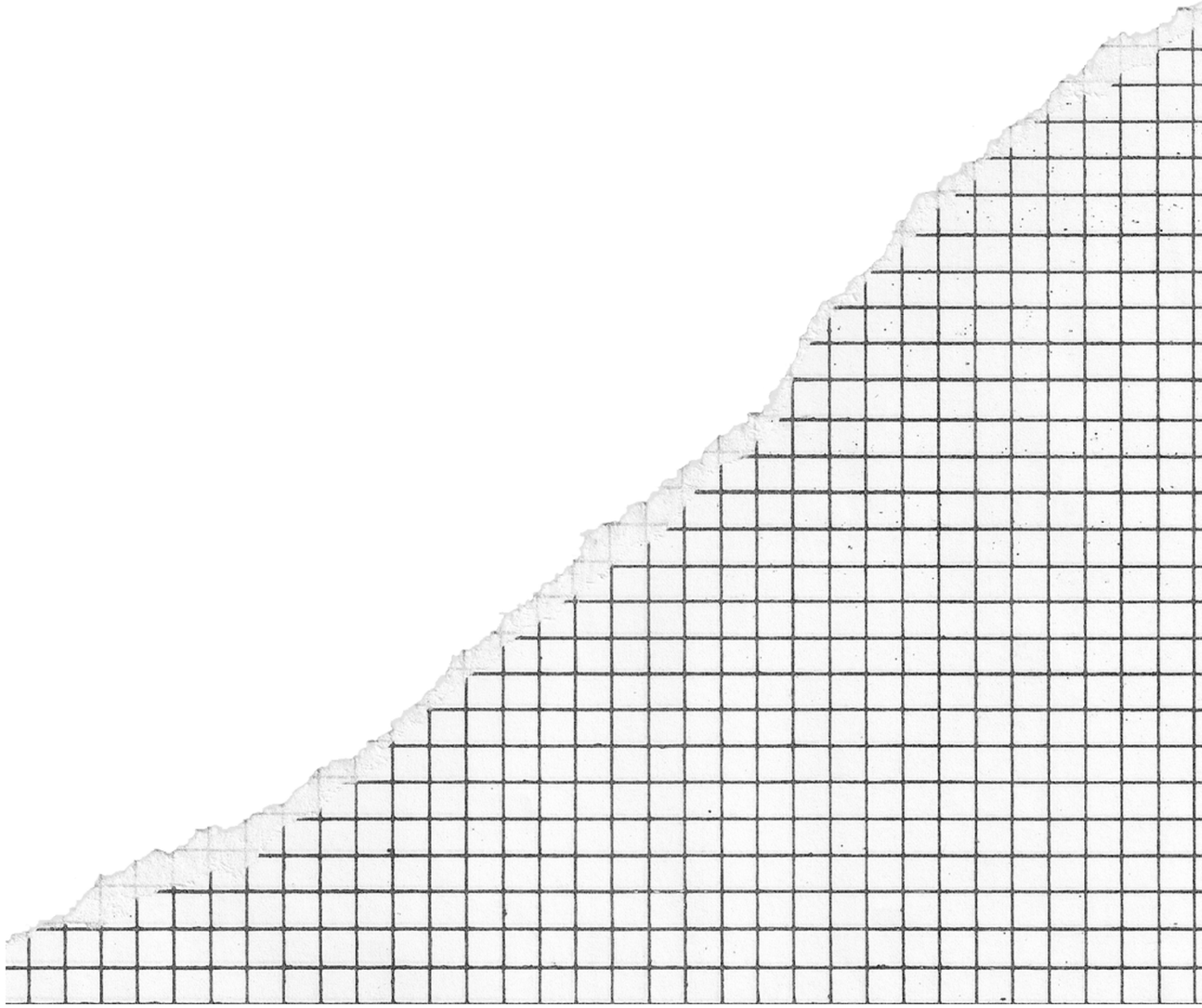
Preprocessing Data



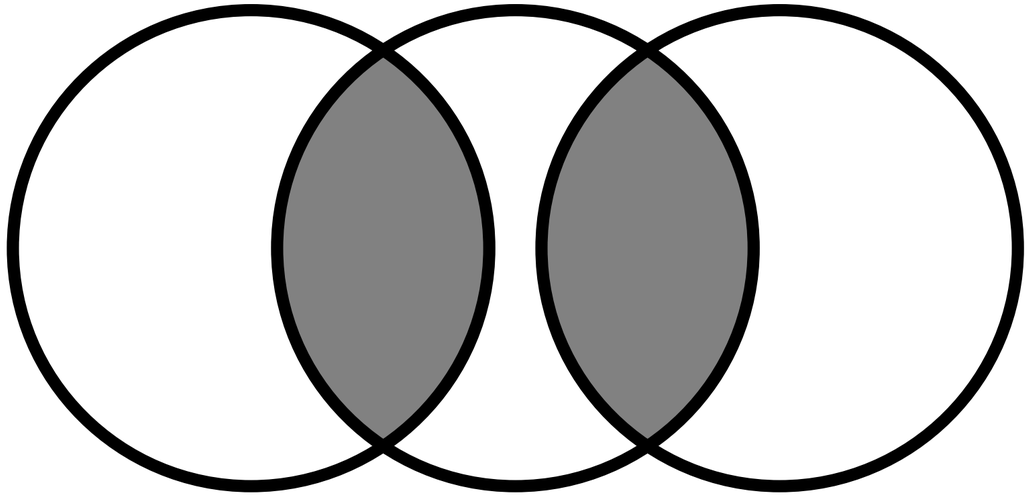
cleaning serum_sodium feature

Model Building

- Modeling :
  SVM, KNN, Logistic Regression,
  Random Forest Classifier



Testing



- Testing data using 4 classifier methods :
  SVM, KNN, Logistic Regression, Random Forest Classifier

Result (Output)

LOGISTIC REGRESSION

ACCURACY : 0.8875

CONFUSION MATRIX :

	T	F
T	61	0
F	9	10

SUPPORT VECTOR MACHINE

ACCURACY : 0.8625

CONFUSION MATRIX :

	T	F
T	59	2
F	9	10

K NEAREST NEIGHBOUR

ACCURACY : 0.8625

CONFUSION MATRIX :

	T	F
T	56	5
F	6	13

RANDOM FOREST ACCURACY

ACCURACY : 0.8625

CONFUSION MATRIX :

	T	F
T	57	4
F	7	12

Thank you