Heart Attack Prediction

Kelompok 8

Allin Setiawan - 2602191301 Callista Serena E. - 2602197305 Janice Ashley K. - 2602203794

1 Problem Statement

Heart attacks are the number one cause of death worldwide. According to WHO data there are 17 million people in the world who die from heart disease. Many factors make a person suffer from a heart attack such as obesity, cholesterol etc. From the dataset that we have taken from Kaggle, namely the heart attack dataset, we want to predict how many patients get the heart attack using several machine learning models what is good to use in this dataset.

2 Data Overview

Key attributes of the data:

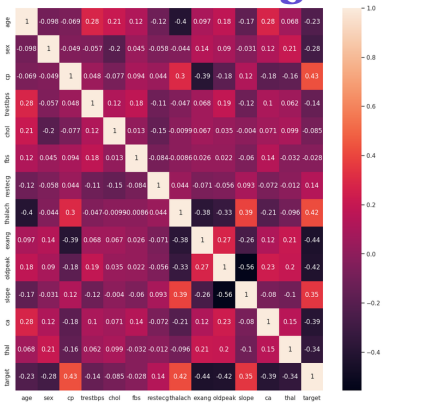
- age : int
- sex : int
- chest pain type (cp) : int
- resting blood pressure : int
- serum cholesterol in mg/dl: int
- fasting blood sugar : int
- restecg : int
- thalach : int
- exang : int
- oldpeak : float
- ca: int
- thal : int
- target : int

The data has 303 rows and no null values.

3 Objective

We're trying to predict the 'target' value from the dataset, which indicates whether a person has a heart attack or not. We predict the 'target' value to help someone know the possibility of having a heart attack and determine whether they need a treatment or not

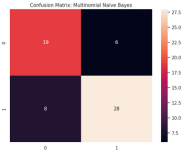
4 Data Understanding



For your information none of the variables in this dataset are null.

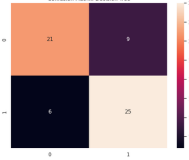
5 Machine Learning Modelling

Naive Bayes



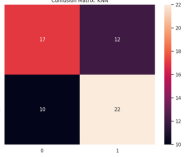
Accuracy score: 0.7704918032786885
F1-Score: 0.7999999999999999

Decision Tree



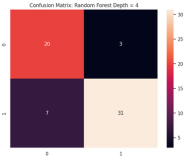
Accuracy score: 0.7540983606557377
F1-Score: 0.7692307692307693

K Nearest Neighbors



Accuracy score: 0.639344262295082
F1-Score: 0.6666666666666667

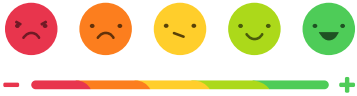
Random Forest



Accuracy score: 0.8360655737704918
F1-Score: 0.861111111111111

6 Evaluation

Of the four multi-classification methods that we have done. We get the results of the classification report of each method. As can be seen from the results of the accuracy value and F-1 score of each method. Random forests had the highest yields, meaning that they performed better at predicting someone would have a heart attack than other models. However, we still need to know whether the random forest model is really consistent in doing good predictability on new data that has not yet been seen



7 Conclusion

Classification Report

	precision	recall	f1-score	support
0	0.87	0.74	0.80	27
1	0.82	0.91	0.86	34
accuracy			0.84	61
macro avg	0.84	0.83	0.83	61
weighted avg	0.84	0.84	0.83	61

Based on the results of a more detailed classification report by obtaining precision, recall, f1-score and support values from the four models. We can conclude again that Random Forest performs better than other models because it shows a good balance between precision and recall for the class.