Predicting Death by Heart Failure

Introduction

Disease of the heart is considered the highest cause of death worldwide. One of the most frequent ways it can cause death is for the heart disease to cause a heart failure event.

Data

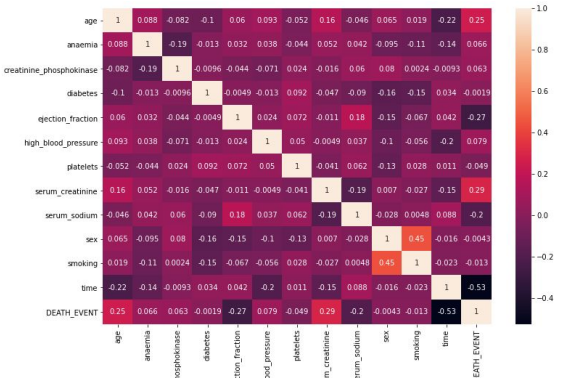
The url for the data is:
https://www.kaggle.com/andrewmvd/heart-failure-clinicaldata.

The data contains the following columns:

Age	High Blood Pressure	Smoking
Anaemia	Platelets	Time between checks and follow-up
Creatinine Phosphokinase	Serum Creatinine	Death event
Diabetes	Serum Sodium	
Ejection Fraction	Sex	

Preprocessing Data

- No missing data after check
- Data types are valid
- Numerical data normalized to a range of (0,1)
- Correlation heat map produced to determine best features:



Feature Selection

Eliminated:

- Sex
- Diabetes

Modeling Algorithm and Evaluation

- Problem is Classification problem
- Decision Trees will be used to model and predict the classification of death event.
- Results:
  - 78% accuracy

Conclusion

The model produced could be implemented to all heart failure patients after medical tests to detect whether further intervention is needed to avoid a possible death.