

Etapas do processo

As fontes utilizadas foram obtidas no site Kaggle. O nome do conjunto de dados é "Prever a sobrevivência de pacientes com insuficiência cardíaca".

As variáveis utilizadas, ou seja, os nomes das colunas, foram: age, anaemia, creatinine_phosphokinase, diabetes e ejection_fraction.

Escolhi esse algoritmo por sua facilidade de uso e porque foi explicado na aula. Era o que eu conhecia para colocar em prática.

Utilizei a biblioteca Scikit-learn porque é conhecida e tem uma variedade de ferramentas que facilitam para quem está começando a entender e colocar em prática. Com ela, é possível criar modelos e realizar treinamentos.

Utilizei o Grid Search para otimizar o modelo, pois nos vídeos que assisti e nas pesquisas que fiz, foi o método recomendado e que eu entendi melhor.

Para validar o modelo, utilizei a validação cruzada com 5 folds, ou seja, o modelo foi treinado 5 vezes para teste, proporcionando uma avaliação mais completa do seu desempenho.

Os hiperparâmetros foram otimizados utilizando o Grid Search.

As métricas de avaliação utilizadas foram o MSE (Erro Quadrático Médio) e R2, entre outras informações que indicaram um bom ajuste no modelo.