

CMC-16 - Data Science Practices

Escolha do Projeto

Equipe:

Samir
Othon
Ulisses
Victor

Professor:

Filipe Verri

Multiclass Classification com target sendo o diagnóstico de doença cardíaca, indo de 0 a 4.



Heart Disease

Donated on 6/30/1988

4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach

Dataset Characteristics

Multivariate

Subject Area

Health and Medicine

Associated Tasks

Classification

Feature Type

Categorical, Integer, Real

Instances

303

Features

13

Dataset Information

Additional Information

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have ...

SHOW MORE

Variables Table



Variable Name	Role	Type	Demographic	Description	Units	Missing Values
age	Feature	Integer	Age		years	no
sex	Feature	Categorical	Sex			no
cp	Feature	Categorical				no
trestbps	Feature	Integer		resting blood pressure (on admission to the hospital)	mm Hg	no
chol	Feature	Integer		serum cholestoral	mg/dl	no
fbs	Feature	Categorical		fasting blood sugar > 120 mg/dl		no
restecg	Feature	Categorical				no
thalach	Feature	Integer		maximum heart rate achieved		no
exang	Feature	Categorical		exercise induced angina		no
oldpeak	Feature	Integer		ST depression induced by exercise relative to rest		no

- Nosso caso: conseguir diagnosticar aqueles com a doença. Queremos, assim, diminuir o máximo possível de falsos negativos.
- Isso nos leva a querer um alto macro average recall (devido ao class imbalance).
- Estimamos inicialmente $\geq 60\%$ como objetivo.
- Também é ideal uma accuracy (micro average recall) $\geq 75\%$, para estar de acordo com o benchmark.

