
Welcome

HEART DISEASE AND FAILURE *PREDICTION USING TEST RESULTS AND CLINICAL RECORDS*

Disclosure: **Heart disease and failure prediction using Test results and Clinical records during these lectures is only taken as the general example to show, how one can do data analysis using pandas (in python).** Data in these records is idealized to meet the machine model requirements. In real-life scenarios, these predictions may not be applicable. Please consider the specialist / doctors before application to the real-life scenarios of this type of dataset. I would be not responsible for any kind of harm/loss to you.

NOTE: Data is available under education license only. Don’t use dataset other than educational purposes.

мξƏДмαĉĦĮŋΞ

About Dataset – Heart Disease Dataset

Abstract: 4 databases – Cleveland, Hungary, Switzerland, and the VA Long Beach. 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 this date.

Data Set Characteristics:	Multivariate	Number of Instances:	303	Area:	Life
Attribute Characteristics:	Categorical, Integer, Real	Number of Attributes:	75	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	1585441

Dataset source (ics.uci): https://archive.ics.uci.edu/ml/datasets/Heart+Disease (full dataset)

Dataset download links (dataset description, processed.cleveland.data, processed.hungarian.data, processed.switzerland.data, processed.va.data), drive link: direct_drive_link (full dataset)

Note: Only processed data files are considered for this tutorial.

NOTE: Data is available under education license only. Don't use dataset other than educational purposes.

About Dataset – Heart Disease Dataset(Conti.)

Data Set 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 this 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 concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this directory.

To see Test Costs (donated by Peter Turney), please see the folder "Costs“.

About Dataset – Heart Disease Dataset(Conti.)

Total number of rows – 303 (Cleveland) | 294 (Hungarian) | 123 (Switzerland) | 200 (VA)

Missing values – Yes (Marked by – '?').

Total number of columns – 14 – as – >

- age – Column 0: age in years (numerical).
- sex – Column 1: sex (1 = male; 0 = female) (categorical).
- cp – Column 2: chest pain type (categorical) –
  - Value 1: typical angina.
  - Value 2: atypical angina.
  - Value 3: non - anginal pain.
  - Value 4: asymptomatic.
- trestbps – Column 3: resting blood pressure (in mm Hg on admission to the hospital) (numerical)

About Dataset – Heart Disease Dataset(Conti.)

- chol – Column 4: serum cholestoral in mg/dl (numerical).
- fbs – Column 5: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) (categorical).
- restecg – Column 6: resting electrocardiographic results (categorical) –
  - Value 0: normal
  - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalach – Column 7: maximum heart rate achieved (numerical).
- exang – Column 8: exercise induced angina (1 = yes; 0 = no) (Categorical).
- oldpeak – Column 9: ST depression induced by exercise relative to rest (numerical).

About Dataset – Heart Disease Dataset(Conti.)

- slope – Column 10: the slope of the peak exercise ST segment (categorical) –
  - Value 1: up sloping
  - Value 2: flat
  - Value 3: down sloping
- ca – Column 11: number of major vessels (0-3) colored by fluoroscopy (categorical - numerical).
- thal – Column 12: 3 = normal, 6 = fixed defect, 7 = reversible defect (categorical - numerical)
- num – Column 13: y (to predicted) diagnosis of heart disease (angiographic disease status) –
  - Value 0: < 50% diameter narrowing
  - Value 1: > 50% diameter narrowing

About Dataset – Heart Disease Dataset (Conti.)

Source: - Creators:

- Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
- University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
- University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
- V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Donor: David W. Aha (aha '@' ics.uci.edu) (714) 856-8779.

Citation Request: The authors of the databases have requested that any publications resulting from the use of the data include the names of the principal investigator responsible for the data collection at each institution. They would be:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

NOTE: Data is available under education license only. Don’t use dataset other than educational purposes.

About Dataset – Heart Disease Dataset (Conti.)

Relevant Papers:

Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64,304--310.
[Web Link].

David W. Aha & Dennis Kibler. "Instance-based prediction of heart-disease presence with the Cleveland database."
[Web Link].

Gennari, J.H., Langley, P, & Fisher, D. (1989). Models of incremental concept formation. Artificial Intelligence, 40, 11--61.
[Web Link].

About Dataset – Heart Failure Clinical Records Dataset

Abstract: This dataset contains the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 13 clinical features.

Data Set Characteristics:	Multivariate	Number of Instances:	299	Area:	Life
Attribute Characteristics:	Integer, Real	Number of Attributes:	13	Date Donated	2020-02-05
Associated Tasks:	Classification, Regression, Clustering	Missing Values?	N/A	Number of Web Hits:	75963

Dataset source (ics.uci): https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records

Dataset download links (heart_failure_clinical_records_dataset.csv), drive link: direct_drive_link

About Dataset – Heart Failure Clinical Records Dataset (Conti.)

Source:

Provide the names, email addresses, institutions, and other contact information of the donors and creators of the data set.The original dataset version was collected by Tanvir Ahmad, Assia Munir, Sajjad Haider Bhatti, Muhammad Aftab, and Muhammad Ali Raza (Government College University, Faisalabad, Pakistan) and made available by them on FigShare under the Attribution 4.0 International (CC BY 4.0: freedom to share and adapt the material) copyright in July 2017. The current version of the dataset was elaborated by Davide Chicco (Krembil Research Institute, Toronto, Canada) and donated to the University of California Irvine Machine Learning Repository under the same Attribution 4.0 International (CC BY 4.0) copyright in January 2020. Davide Chicco can be reached at <davidechicco '@' davidechicco.it>

Data Set Information:

A detailed description of the dataset can be found in the Dataset section of the following paper: Davide Chicco, Giuseppe Jurman: "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". BMC Medical Informatics and Decision Making 20, 16 (2020). [Web Link]

NOTE: Data is available under education license only. Don’t use dataset other than educational purposes.

About Dataset – Heart Failure Clinical Records Dataset (Conti.)

Total number of rows – Total 299 entries (rows) with no null values.

Total number of columns – Thirteen (13) clinical features:

- age – Column 0: age of the patient (years).
- anemia – Column 1: decrease of red blood cells or hemoglobin (Boolean).
- high blood pressure – Column 2: if the patient has hypertension (Boolean).
- creatinine phosphokinase (CPK) – Column 3: level of the CPK enzyme in the blood (mcg/L).
- diabetes – Column 4: if the patient has diabetes (Boolean).
- ejection fraction – Column 5: percentage of blood leaving the heart at each contraction (percentage).
- platelets – Column 6: platelets in the blood (kilo platelets/mL).
- sex – Column 7: woman or man (binary).

About Dataset – Heart Failure Clinical Records Dataset (Conti.)

- serum creatinine: level of serum creatinine in the blood (mg/dL).
- serum sodium: level of serum sodium in the blood (mEq/L).
- smoking: if the patient smokes or not (Boolean).
- time: follow-up period (days).
- [target] death event: if the patient deceased during the follow-up period (Boolean)

For more information, please check Table 1, Table 2, and Table 3 of the following paper:

Davide Chicco, Giuseppe Jurman: "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". BMC Medical Informatics and Decision Making 20, 16 (2020). [Web Link]

- Table 1: Meanings, measurement units, and intervals of each feature of the dataset.
- Table 2: Statistical quantitative description of the category features.
- Table 3: Statistical quantitative description of the numeric features.

NOTE: Data is available under education license only. Don’t use dataset other than educational purposes.

About Dataset – Heart Failure Clinical Records Dataset (Conti.)

Citation Request:

Davide Chicco, Giuseppe Jurman: "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". BMC Medical Informatics and Decision Making 20, 16 (2020). [Web Link]

Relevant Papers:

Original dataset version:

Tanvir Ahmad, Assia Munir, Sajjad Haider Bhatti, Muhammad Aftab, and Muhammad Ali Raza: "Survival analysis of heart failure patients: a case study". PLoS ONE 12(7), 0181001 (2017). [Web Link]

Current dataset version on the UCI ML Repository:

Davide Chicco, Giuseppe Jurman: "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". BMC Medical Informatics and Decision Making 20, 16 (2020). [Web Link]

References

- Lecture drive link:
- https://archive.ics.uci.edu/ml/index.php
- https://archive.ics.uci.edu/ml/datasets/Heart+Disease
- https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.names
- https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data
- https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.hungarian.data
- https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.switzerland.data
- https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.va.data
- https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records
- https://archive.ics.uci.edu/ml/machine-learning-databases/00519/heart_failure_clinical_records_dataset.csv
- https://doi.org/10.1186/s12911-020-1023-5
- https://doi.org/10.1371/journal.pone.0181001
- https://doi.org/10.1186/s12911-020-1023-5

THANKS FOR UR PRECIOUS TIME! ☺

- Questions? 🗣 💬

βy мξƟДмαĉĦ¡ŋΞ

Thank you