
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

Recommended to Read Before Starting

About python: Python is powerful... and fast; plays well with others; runs everywhere; is friendly & easy to learn; is Open –> https://www.python.org/about/.

Python docs: https://docs.python.org/3/ (all documentation); https://docs.python.org/3.7/ **(Recommended version – 3.7)**.

The Python Tutorial (python3.7): https://docs.python.org/3.7/tutorial/index.html

The Python Language Reference: https://docs.python.org/3.7/reference/index.html#reference-index

Familiar with python notebook environment: The Jupyter Notebook (Formerly known as the IPython Notebook): Documentation.

Google Colab (Our working environment): Get started with Google Colaboratory (Coding TensorFlow) (Video tutorial); Tutorial: CS231n Python Tutorial With Google Colab (**Must have a go-through**, ".ipynb" file).

Recommended to Read Before Starting (Conti.)

NumPy : The fundamental package for scientific computing with Python. NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

About: https://numpy.org/

Installation: https://numpy.org/install/

Docs: https://numpy.org/doc/stable/

NumPy quick-start: https://numpy.org/doc/stable/user/quickstart.html

Recommended to Read Before Starting (Conti.)

Pandas: pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

About: https://pandas.pydata.org/

Docs: https://pandas.pydata.org/docs/

Installation: https://pandas.pydata.org/getting_started.html

Getting started: https://pandas.pydata.org/docs/getting_started/index.html

User Guide: https://pandas.pydata.org/docs/user_guide/index.html#user-guide

Download documentation: PDF Version | Zipped HTML

Recommended to Read Before Starting (Conti.)

Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

About: https://matplotlib.org/

Installation: https://matplotlib.org/stable/users/installing.html

Docs: https://matplotlib.org/stable/contents.html

Tutorials: https://matplotlib.org/stable/tutorials/index.html

Examples: https://matplotlib.org/stable/gallery/index.html

Recommended to Read Before Starting (Conti.)

Scikit-Learn: Machine Learning in Python. Simple and efficient tools for predictive data analysis. Built on NumPy, SciPy, and matplotlib.

About: https://scikit-learn.org/stable/

Installation: https://scikit-learn.org/stable/install.html

Getting started: https://scikit-learn.org/stable/getting_started.html

Docs: https://scikit-learn.org/stable/user_guide.html

Tutorials: https://scikit-learn.org/stable/tutorial/index.html

Examples: https://scikit-learn.org/stable/auto_examples/index.html

API references: https://scikit-learn.org/stable/modules/classes.html

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.)

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.)

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://www.python.org/
- https://www.python.org/about/.
- https://docs.python.org/3/
- https://docs.python.org/3.7/
- https://docs.python.org/3.7/tutorial/index.html
- https://docs.python.org/3.7/reference/index.html#reference-index
- https://ipython.org/notebook.html
- https://ipython.org/documentation.html
- https://colab.research.google.com/notebooks/intro.ipynb?utm_source=scs-index#recent=true
- https://youtu.be/inN8seMm7UI
- https://numpy.org/
- https://numpy.org/install/
- https://numpy.org/doc/stable/
- https://numpy.org/doc/stable/user/quickstart.html
- https://pandas.pydata.org/
- https://pandas.pydata.org/getting_started.html
- https://pandas.pydata.org/docs/
- https://pandas.pydata.org/docs/getting_started/index.html
- https://pandas.pydata.org/docs/user_guide/index.html#user-guide

THANKS FOR UR PRECIOUS TIME! ☺

- Questions? 🗣 💬

βy мξƟДмαĉĦ¡ŋΞ

Thank you