

# Dataset Documentation: heart.csv

## Title: **AI Models for Early Detection and Mortality Prediction in Cardiovascular Diseases**

### 1. Overview:

The `heart.csv` dataset contains medical attributes of individuals, providing essential information used for the prediction and early detection of cardiovascular diseases. The dataset is pivotal for developing AI models aimed at mortality prediction and early detection in cardiovascular diseases.

### 2. Variables Description:

- age:
  - type: Numerical
  - Description: Age of the individual.
  - Range: [29, 77]
- sex:
  - Type: Categorical (Binary)
  - Description: Gender of the individual.
  - Categories:
    - `0`: Female
    - `1`: Male
- cp:
  - Type: Categorical
  - Description: Chest pain type experienced by the individual.
  - Categories:[0, 1, 2, 3]
- trestbps:
  - Type: Numerical
  - Description: Resting blood pressure (in mm Hg) upon admission to the hospital.
- chol:
  - Type: Numerical
  - Description: Serum cholesterol level in mg/dl.
- fbs:
  - Type: Categorical (Binary)
  - Description: Fasting blood sugar.
  - Categories:
    - `0`: < 120 mg/dl
    - `1`: > 120 mg/dl

- restecg:
  - Type: Categorical
  - Description: Resting electrocardiographic results.
  - Categories: [0, 1, 2]
- thalach:
  - Type: Numerical
  - Description: Maximum heart rate achieved during the Thallium stress test.
- exang:
  - Type: Categorical (Binary)
  - Description: Exercise-induced angina.
  - Categories:
    - `0`: No
    - `1`: Yes
- oldpeak:
  - Type: Numerical
  - Description: ST depression induced by exercise relative to rest.
- slope:
  - Type: Categorical
  - Description: Slope of the peak exercise ST segment.
  - Categories: [0, 1, 2]
- ca:
  - Type: Numerical
  - Description: Number of major vessels colored by fluoroscopy.
- thal:
  - Type: Categorical
  - Description: Thalassemia type.
  - Categories: [1, 2, 3]
- target:
  - Type: Categorical (Binary)
  - Description: Diagnosis of heart disease.
  - Categories:
    - `0`: Absence of heart disease
    - `1`: Presence of heart disease

### 3. Usage:

The dataset is intended for researchers and practitioners in the healthcare and medical domain for developing and validating AI and machine learning models aimed at predicting cardiovascular diseases. It serves as a resource for the exploration of feature importance,

model explainability, and the development of interpretable models in the cardiovascular health domain.

#### **4. Limitations and Ethical Considerations:**

Researchers should approach the dataset considering the ethical aspects of patient data, even if anonymized. Understanding the limitations in terms of representativeness, possible biases in data collection, and the need for external validation is crucial.

#### **5. Acknowledgments:**

Researchers utilizing this dataset should acknowledge the source, and if applicable, the funding agencies and institutions supporting the work related to the dataset's creation and distribution.