

# DAV Project

## 1) Project Name:

- Data Analysis and Visualization for Cardiovascular Disease Prediction

## 2) Team Members:

- Nithish Chouti (21bcs074)
- Udayini Vedantham (21bcs130)
- Karan R Naik (21bcs051)

## 3) Datasets and Inferences:

- <https://www.openml.org/search?type=data&status=active&id=43154&sort=runs> - 4
- <https://www.openml.org/search?type=data&sort=runs&id=43823&status=active> - 3
- <https://www.openml.org/search?type=data&status=active&id=45547&sort=runs> - 2
- <https://www.openml.org/search?type=data&status=active&id=43672&sort=runs> - 1

## 4) Finalized Dataset:

- **Name** - Heart-Disease-Dataset-(Comprehensive)
- Heart disease is also known as Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year which is about 32 of all deaths globally. CVDs are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions. Four out of 5 CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age.
- The five datasets used for its curation are database of instances:
  - Cleveland: 303
  - Hungarian: 294
  - Switzerland: 123
  - Long Beach VA: 200
  - Stalog (Heart) Data Set: 270
- **Source** -

- OpenML:  
(<https://www.openml.org/search?type=data&status=active&id=43672&sort=runs>)
- IEEE:  
(<https://ieee-dataport.org/open-access/heart-disease-data-set-comprehensive>)
- **Instances** - 1190
- **Features** - 12

#### 5) Base Research Papers Links:

- **Name** - Predictive Modeling for Heart Disease Detection with Machine Learning.
- **Link** - [<https://ieeexplore.ieee.org/abstract/document/10402340>]
- **Name** - Automated Heart Disease Prediction System using Machine Learning Approaches.
- **Link** - [<https://ieeexplore.ieee.org/abstract/document/10407008>]
- **Name** - A Study on Heart Disease Prediction using Different Classification Models based on Cross Validation Method
- **Link** - [<https://www.ijert.org/a-study-on-heart-disease-prediction-using-different-classification-models-based-on-cross-validation-method>]
- **Name** - AI Models for Early Detection and Mortality Prediction in Cardiovascular Diseases.
- **Link** - [<https://www.techrxiv.org/doi/full/10.36227/techrxiv.24248827.v1>]

#### 6) Project Scope:

- Provide an overview of the project, including the significance of predicting cardiovascular diseases (CVDs) and the importance of data analysis and visualization in healthcare.
- Analyzing historical patient data to identify risk factors associated with CVDs.
- Developing predictive models based on the identified risk factors.
- Describing the sources of data, which may include electronic health records, medical databases, or research datasets.
- Performing statistical analyses, including summary statistics, correlation analysis, and hypothesis testing to identify significant factors related to CVDs.

- Engineering new features if necessary to enhance the predictive power of the models.
- Machine Learning algorithms used:
  1. Logistic Regression (Scikit-learn)
  2. Naive Bayes (Scikit-learn)
  3. Support Vector Machine (Linear) (Scikit-learn)
  4. K-Nearest Neighbours (Scikit-learn)
  5. Decision Tree (Scikit-learn)
  6. Random Forest (Scikit-learn)
  7. XGBoost (Scikit-learn)
  8. Artificial Neural Network with 3 Hidden layers (Keras)

## 7) Code Links:

1. AI Models for Early Detection and Mortality Prediction in Cardiovascular Diseases -  
[\[https://github.com/datascintist-abusufian/AI-Models-for-Early-Cardiovascular-Diseases-Detection/tree/main\]](https://github.com/datascintist-abusufian/AI-Models-for-Early-Cardiovascular-Diseases-Detection/tree/main)
2. Heart Disease Prediction using Machine Learning -  
[\[https://github.com/g-shreekant/Heart-Disease-Prediction-using-Machine-Learning\]](https://github.com/g-shreekant/Heart-Disease-Prediction-using-Machine-Learning)

\*\*\* NEW Code:

[https://colab.research.google.com/drive/16Srj\\_YT2MVSrrU1TMIALnywxJESyxV67?usp=sharing](https://colab.research.google.com/drive/16Srj_YT2MVSrrU1TMIALnywxJESyxV67?usp=sharing)

\*\*\* OLD Code:

[https://colab.research.google.com/drive/1QXF2jGzs1VkF8msgVHjTpFSf0rBoODT\\_?usp=sharing](https://colab.research.google.com/drive/1QXF2jGzs1VkF8msgVHjTpFSf0rBoODT_?usp=sharing)

\*\*\* OLD Code:

<https://colab.research.google.com/drive/17DiSfWLE0jbYaKcW0sCTbk4aCRYH-uX4?usp=sharing>

## 8) Any Other Updates or Information:

## 4. Project Overview:

Title: "Cardiovascular Disease Prediction Project"

Objective: Clearly state the objective of your project, i.e., to develop a predictive model for cardiovascular disease.

**Data Sources:** Mention the sources of data used for the project.

**Features Used:** List the features or variables utilized in the predictive model, such as age, gender, blood pressure, cholesterol levels, etc.

**Models Implemented:** Outline the machine learning models used for prediction (e.g., logistic regression, random forest, XGBoost).

**Evaluation Metrics:** Specify the evaluation metrics used to assess the performance of the models (e.g., accuracy, precision, recall, F1-score).