**PROBLEM STATEMENT:**

The Goal of this project is to develop an accurate and efficient **HEART DISEASE PREDICTION** using machine learning techniques.

**APPROACH:**

* **Algorithm**: Logistic regression
* **Programming language:** Python
* **Tools and Libraries:** Jupyter Notebook, scikit-learn, Numpy, Pandas, matplotlib, Sklearn.
* **Dataset**: framingham.csv(from **Kaggle** website)

**ABSTRACT:**

Logistic regression models predicts the risk of suffering from heart disease among the elderly by exploring the feasibility of using logistic regression models. Through the technology of data mining, the main pathogenic factors of heart disease were found, and the incidence of heart disease was predicted by using the regression model. The accuracy of logistic regression model was compared with other explored algorithms, and I found that the logistic regression model was worthy of research in the field of heart disease prediction.

1. **ALGORITHM:**

**LOGISTIC REGRESSION ALGORITHM:**

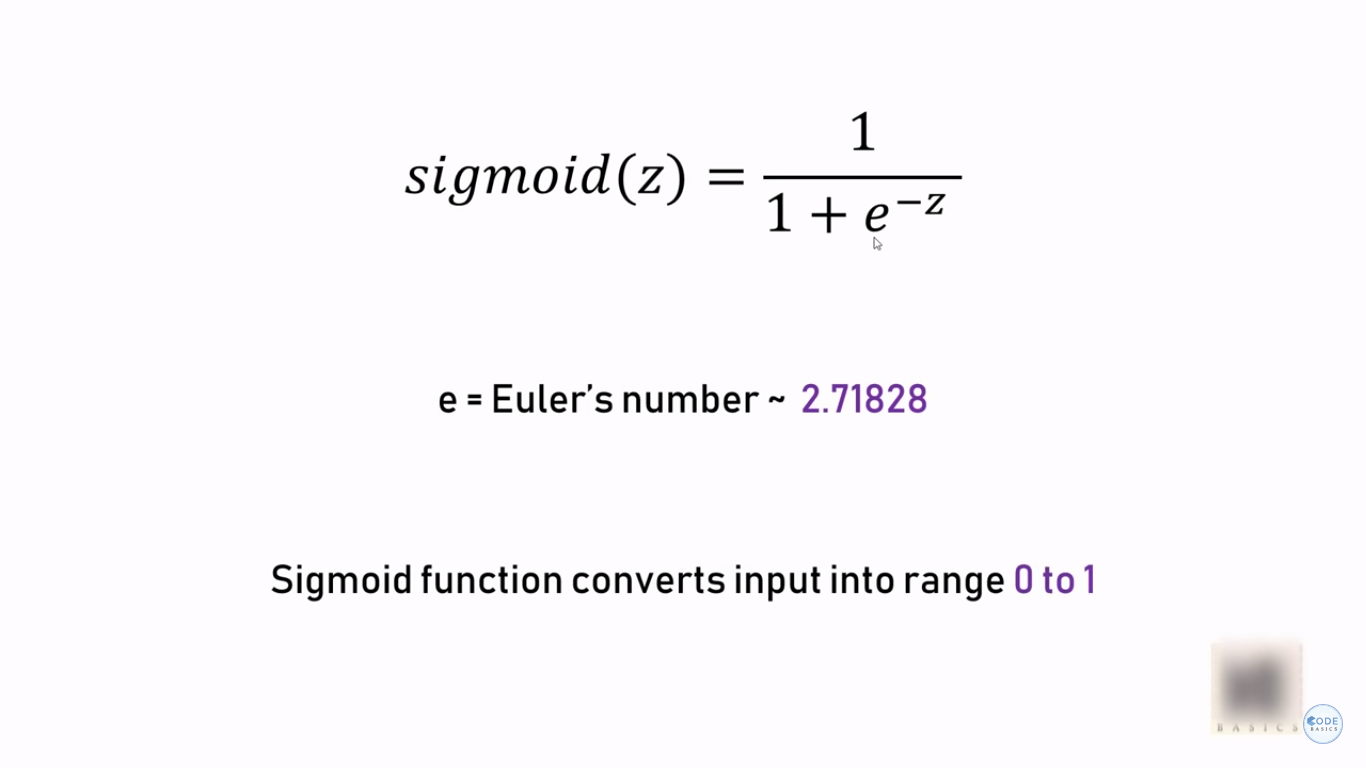
Logistic Regression is a popular and widely used classification algorithm in machine learning. It is primarily used for binary classification tasks, where the target variable is categorical value (yes/no ,0/1,true/false,fraud/not fraud). Logistic Regression algorithm uses sigmoid function to map the predicted values to the probability of belonging to a particular class.

**Logistic Function (Sigmoid Function):**

* The sigmoid function is a mathematical function used to map the predicted values to probabilities.
* It maps any real value into another value within a range of 0 and 1.
* The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the “S” form. The S-form curve is called the Sigmoid function or the logistic function.
* In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

**Assumptions for Logistic Regression:**

* The dependent variable must be categorical in nature.
* The independent variable should not have multi-collinearity.



1. **TOOLS AND LIBRARIES USED:**

HEART DISEASE PREDICTION data in real-time. To accomplish this task efficiently, various tools and libraries are commonly used. Below are some of the key tools and libraries employed in HEART DISEASE PREDICTION projects:

1. **Python**: Python is a widely used programming language for data science and machine learning tasks. It offers an extensive range of libraries that make it a popular choice for credit card fraud detection projects.
2. **Pandas**: Pandas is a powerful library in Python used for data manipulation and analysis. It helps with tasks like data preprocessing, cleaning, and feature engineering, which are crucial steps in preparing the data for model training.
3. **NumPy**: NumPy is another fundamental library in Python that provides support for large, multi-dimensional arrays and matrices. It is extensively used for numerical computations and vectorized operations.
4. **Scikit-learn**: Scikit-learn is a comprehensive machine learning library in Python that offers various algorithms for classification, regression, clustering, and more. It includes implementations of logistic regression and other classification models used in credit card fraud detection.
5. **Matplotlib / Seaborn**: These visualization libraries in Python are used to create informative plots and charts, aiding in data exploration and understanding the distribution of data.
6. **DATA COLLECTION:**

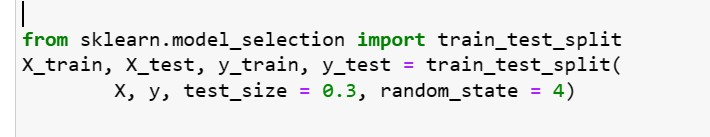
The dataset contains Logistic regression models predicts the risk of suffering from heart disease among the elderly by exploring the feasibility of using logistic regression models. Through the technology of data mining, the main pathogenic factors of heart disease were found, and the incidence of heart disease was predicted by using the regression model. The accuracy of logistic regression model was compared with other explored algorithms, and I found that the logistic regression model was worthy of research in the field of heart disease prediction

1. **DATA PREPROCESSING:**

The data reduction is the most used preprocessing task in cardiology, followed by data cleaning. In general, preprocessing either maintained or improved the performance of heart disease classifiers. Some combinations such as (ANN + PCA), (ANN + CHI) and (SVM + PCA) are promising terms of accuracy. the deployment of these models in real-world diagnosis decision support systems is subject to several risks and limitations due to the lack of interpretation.

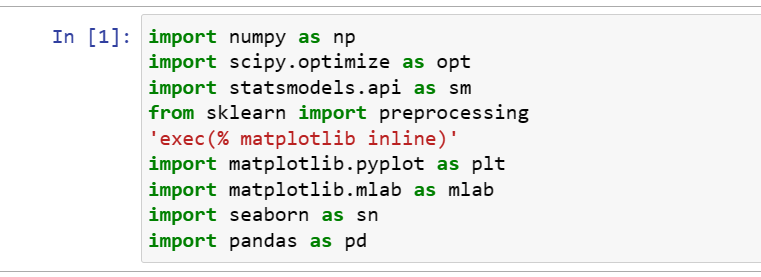
1. **SPLITTING THE DATA:**

Splitting a dataset into training and test data is a fundamental step in machine learning and data analysis. The purpose of this split is to train a model on the training data and then evaluate its performance on unseen data, represented by the test data, to estimate how well the model generalizes to new, unseen examples.



1. **MODEL TRAINING:**

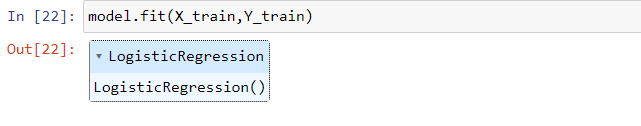
(Using Logistic Regression)



**Choose ML model:**

****

**Training the ML model:**



**7.MODEL EVALUATION:**

**Predict model:**

