

Heart Disease

male	age	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	TenYe
1	39	0	0	0	0	0	0	195	106	70	26.97	80	77	0
0	46	0	0	0	0	0	0	250	121	81	28.73	95	76	0
1	48	1	20	0	0	0	0	245	127.5	80	25.34	75	70	0
0	61	1	30	0	0	1	0	225	150	95	28.58	65	103	1
0	46	1	23	0	0	0	0	285	130	84	23.1	85	85	0
0	43	0	0	0	0	1	0	228	180	110	30.3	77	99	0
0	63	0	0	0	0	0	0	205	138	71	33.11	60	85	1
0	45	1	20	0	0	0	0	313	100	71	21.68	79	78	0
1	52	0	0	0	0	1	0	260	141.5	89	26.36	76	79	0
1	43	1	30	0	0	1	0	225	162	107	23.61	93	88	0

Download CSV

About the Dataset

The World Health Organization reports 12 million annual deaths from heart diseases globally, with half in developed countries. Early prognosis can guide lifestyle changes in high-risk patients to reduce complications. This study uses logistic regression to identify key risk factors and predict heart disease risk. The dataset, sourced from Kaggle, stems from a cardiovascular study in Framingham, Massachusetts, comprising over 4,000 records and 15 attributes.

Ed Analysis and Insight Generation

Download the Colab File

Male: Gender (1 = Male, 0 = Female). Indicates the gender of the individual.	Age: Age in years. Represents the age of the individual.
CurrentSmoker: Currently smoking (1 = Yes, 0 = No). Indicates if the individual is a current smoker.	CigsPerDay: Cigarettes per day. Number of cigarettes the individual smokes per day.
BPMeds: Blood pressure medication (1 = Yes, 0 = No). Indicates if the individual is on blood pressure medication.	PrevalentStroke: History of stroke (1 = Yes, 0 = No). Indicates if the individual has a history of stroke.
PrevalentHyp: Prevalent hypertension (1 = Yes, 0 = No). Indicates if the individual has prevalent hypertension.	Diabetes: Diabetes (1 = Yes, 0 = No). Indicates if the individual has diabetes.
TotChol: Total cholesterol. Measures the total cholesterol level in the blood.	SysBP: Systolic blood pressure. Measures the systolic blood pressure.
DiaBP: Diastolic blood pressure. Measures the diastolic blood pressure.	BMI: Body mass index. A measure of body fat based on height and weight.
HeartRate: Heart rate. Measures the heart rate in beats per minute.	Glucose: Glucose level. Measures the glucose level in the blood.
TenYearCHD: Ten-year risk of coronary heart disease (1 = Yes, 0 = No). Indicates the ten-year risk of developing coronary heart disease.	

Knowledge Representation and Insight Generation

Contents

- Imports and data summary
- Dealing with null values
- Classification of discrete and continuous variables
- Analysis and Insight Generation
- PreProcessing for Model Training
- Different models and their Performance matrix
- Conclusion

Imports

```
In [ ]: import numpy as np
```

ML Models

Download Colab File

Model Training

Imports

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#Machine Learning Imports
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
from sklearn.metrics import accuracy_score, classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import GradientBoostingClassifier
from xgboost import XGBClassifier
from lightgbm import LGBMClassifier
```