

Dataset Heart

What dataset ?

- This dataset gives us informations about things related to the heart, and we use this dataset to make a classification between healthy and sick people according to certain information.
- We have used this dataset to make more than 5 machine learning algorithms to see which algorithm achieves the highest accuracy

features

- **Age:-** age of person
- **Sex:-** the gender of person
- **Cp:-** Constrictive pericarditis (CP) is a form of diastolic heart failure that arises because an inelastic pericardium inhibits cardiac filling.
- **Trestbps:-** The person's resting blood pressure (mm Hg on admission to the hospital) **Chol:** The person's cholesterol measurement in mg/dl. **fbs:** The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false) **restecg:** resting electrocardiographic results.
- **Chol:** - Cholesterol
- **Fbs:** -Fanconi Bickel syndrome (FBS) is a rare condition characterized by the accumulation of a substance called glycogen in different parts of the body.

features

- **Restecg**:-Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria) • thalach: The person's maximum heart rate achieved.
- **Thalach**:- **The person's maximum heart rate achieved.**
- **Exang**:- ST depression induced by. exercise relative to rest(old peak), the slope of the peak. exercise ST segment(slope), number of major vessels. colored by fluoroscopy(ca) and thalassemia(thal) for. prediction of heart diseases
- **Old peak**:- **ST depression induced by exercise relative to rest**
- **Slope**:- **The ST segment shift relative to exercise-induced increments in heart rate**
- **Ca**:- Calcium in blood

Features

- **Old peak:-** *ST depression induced by exercise relative to rest*
- **Slope:-** *The ST segment shift relative to exercise-induced increments in heart rate*
- **Ca:-** *Calcium in blood*
- **Thal:-** *Thalassemia is an inherited (i.e., passed from parents to children through genes) blood disorder caused when the body doesn't make enough of a protein called hemoglobin, an important part of red blood cells.*
-

machine learning algorithm we used

- **Logistic Regression**
- **- K-Nearest Neighbors' Classifier**
- **- Support Vector Machine**
- **- Decision Tree Classifier**
- **- Random Forest Classifier**
- **- XGBoost Classifier**
- **- MLP Classifier**

Accuracy in train and test

[35]:

	Model	Training Accuracy %	Testing Accuracy %
0	Logistic Regression	86.792453	91.208791
1	Decision Tree Classifier	100.000000	71.428571
2	Support Vector Machine	91.981132	89.010989
3	K-nearest neighbors	86.320755	82.417582
4	XGBoost Classifier	100.000000	82.417582
5	Random Forest Classifier	100.000000	89.010989
6	MLP Classifier	95.754717	86.813187
7	Gaussian Naive Bias Classifier	81.603774	86.813187