Heart Disease Prediction Using Machine Learning

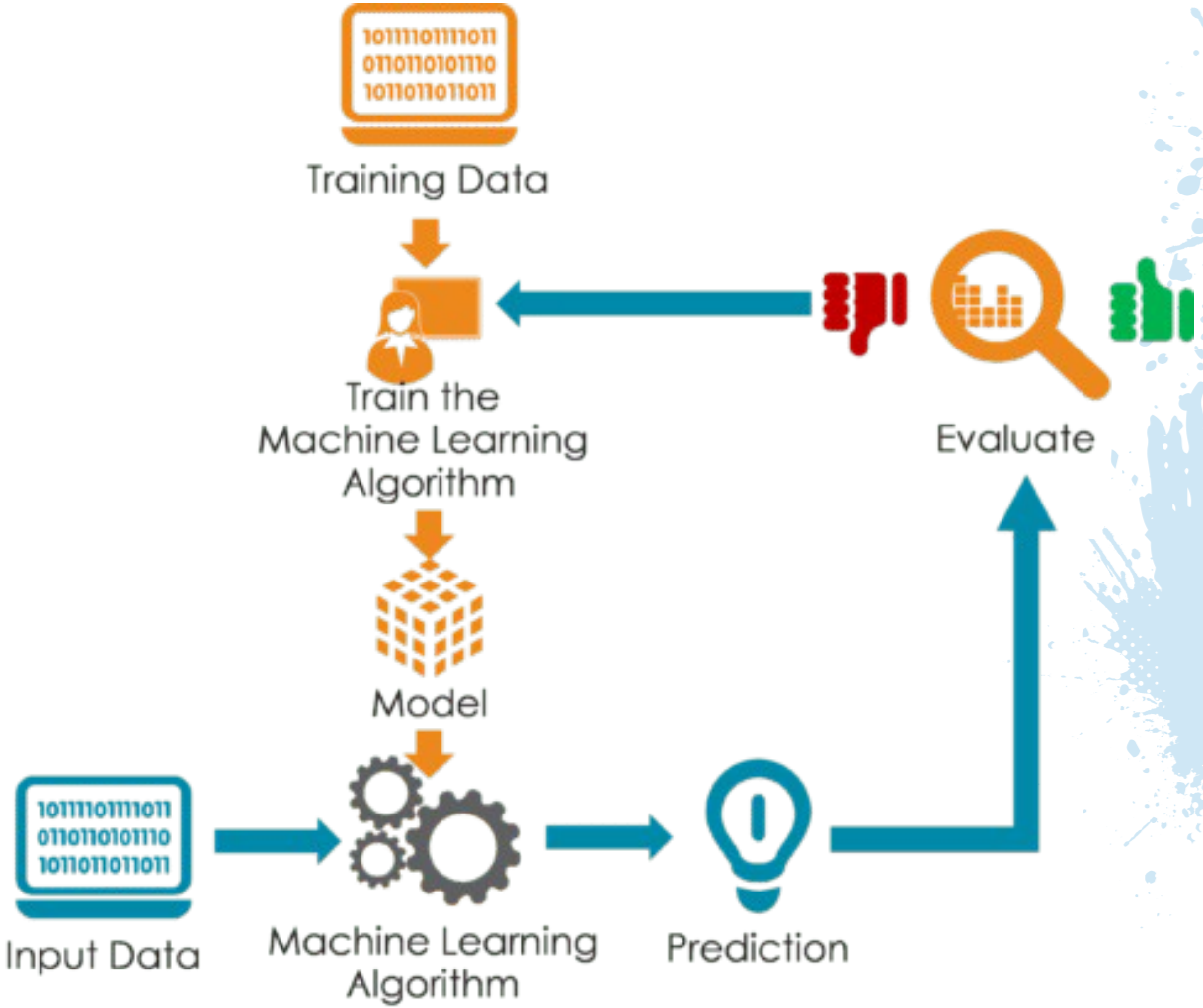
By

SHIVAM JOHARI

ABSTRACT

Nowadays, health disease are increasing day by day due to lifestyle, hereditary. Especially, heart disease has become more common these days, i.e. life of people is at risk. Each individual has different values for Blood pressure, cholesterol and pulse rate. But according to medically proven results the normal values of Blood pressure is 120/90, Cholesterol is 100-129 mg/dL ,Pulse rate is 72, Fasting Blood Sugar level is 100 mg/dL ,Heart rate is 60-100 bpm, ECG is normal, Width of major vessels is 25 mm (1 inch) in the aorta to only 8 μm in the capillaries. This paper gives the survey about different classification techniques used for predicting the risk level of each person based on age, gender, Blood pressure, cholesterol, pulse rate.

“Disease Prediction” system based on predictive modeling predicts the disease of the user on the basis of the symptoms that user provides as an input to the system. The system analyzes thesymptoms provided by the user as input and gives the probability of the disease as an output. Disease Prediction is done by implementing 5 techniques such as Naïve Bayes, KNN, Decision Tree, Linear Regression and Random Forest Algorithms. These techniques calculate the probability of the disease.


Training Data
Train the Machine Learning Algorithm
Evaluate
Model
Input Data
Machine Learning Algorithm
Prediction

TECHNOLOGY STACK USED

Heart Disease prediction
using 5 algorithms

- Logistic regression,

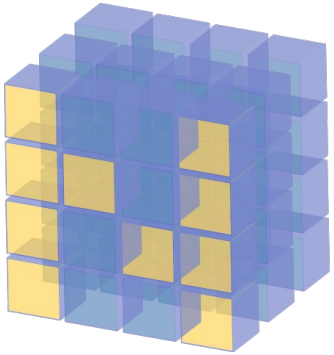
- Random forest,

- Naive Bayes,

- KNN(K Nearest Neighbors),

- Decision Tree

LIBRARIES USED IN THIS ML MODEL

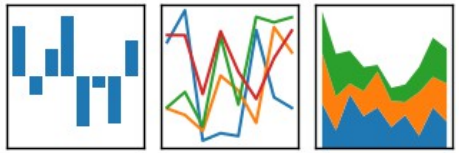


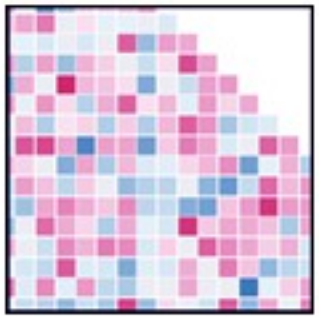
scikit learn

NumPy

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$





Seaborn

matplotlib

EXPLAINATION OF THE PROJECT

The main goal of the project is to predict and determine whether the patient has an chance of having heart disease in future or not on the basis of dataset with Recorded details. For this, I extracted the dataset from kaggle website. This dataset consisted of various columns such as age,sex,cp,trestbps,chol,fbs,restecg,thalach, Exang,oldpeak,slope,ca,thal,target. In each column of the dataset we have 304 entries which means that we have entries of 304 patients with the record of there various Body tests listed in the dataset.

The proposed work predicts heart disease by exploring the above mentioned four classification algorithms and does performance analysis. The objective of this study is to effectively predict if the patient suffers from heart disease. The health professional enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart disease.

●Conclusion

With the increasing number of deaths due to heart diseases, it has become mandatory to develop a system to predict heart diseases effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart disease using dataset.