**Effective Heart Disease Prediction Using Hybrid**

**Machine Learning Techniques**

**Alternate title:** Improving the accuracy in the prediction of heart disease using machine learning algorithms

**Aim**:

To apply machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease

**Synopsis:**

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this paper, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques.

**Existing System:**

The obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 86:8% for *F*-measure, competing with the other existing methods.

**Proposed System**

We have used python and pandas operations to perform heart disease classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation, and the results with improved accuracy.

**Module Description:**

* Data Pre-Processing
* Feature Selection And Reduction
* Classification Modeling
* Performance Measures

**Data Pre-Processing**

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing.

**Feature Selection and Reduction**

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease.

**Classification Modeling**

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

* Decision Trees Classifier
* Support Vector Classifier
* Random Forest Classifier
* K- Nearest Neighbour

**Performance Measures:**

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model.

**Software Requirements:**

* Operating System : Windows 7 , 8, 10 (64 bit)
* Software : Python 3.7
* Tools : Anaconda (Jupyter Note Book IDE)

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Architecture Diagram:**

