SYNOPSIS

Project Team No: 21SOCU1032

Register No: 122003059 **Name:** Darshan Kumar

Project Title: Early prediction of heart diseases using ML based Models , their comparison

Name Of Guide: Professor Ezhilarasie R.

Abstract

India witnesses 54.5 million cases of CVD(cardio vascular diseases), every year. Most of them are identified, only after reaching advanced Stage, from which it could not be cured. Tests such as treadmill Test, CT scans, etc. which are used in western countries, for identifying CVD, are out of reach of Rural parts of India. A ML based Prediction model, is proposed, as a reliable, accessible, affordable solution. 5 state of the art algorithms are proposed: 1)KNN 2) Naïve bayes 3)logistic regression. Other hybrid algorithms implemented are: 4)Random Forest 5) Adaptive Boosting. Useful attributes of dataset were selected, using 1)T-test, 2)co-variance test 3) correlation Test.70% of dataset were used for Training, the rest 30% for testing. Highest accuracy is given by Logistic Regression, highest sensitivity is given by Naïve Bayes classifier, whereas Random Forest , shows the highest specificity . Hyper Tuning of Parameters is done , to get highest score, in each model. Additional algorithms, like Decision Tree, was implemented, to show, how it is likely to overfit, thus not suitable, for this case. Thus hybrid algorithms like Adaboost, RF were used(decision tree as weak learner), to achieve better performance. The Trained models were saved in ".pkl" format using "pickle" library. These files are used, to directly employ the trained model in deployment. The models were deployed using Flask Framework website, for trial. It can be hosted in cloud, for easy reach, of all primary health centers of rural areas.

Specific Contribution:

Subroutines of algorithms like KNN, Naïve Bayes, Logistic regression, decision trees ,were implemented from scratch, without using any pre-defined library functions.

Specific Learning:

Old age, History of Hyper Tension, presence of high BP, were seen as the most contributing factors for CVD

Technical Limitations & challenges faced

The accuracy of models, depends on the kind of dataset used.

Keywords: CVD, ML, KNN, RF , Adaboost , affordable , accessible , reliable

Darshan Kumar R. Ezhi

Name & Signature of Student **Signature of Guide**

Darshan Kumar (122003059)

Date: 9 Jan 2022