

# **Report - 2 : Heart disease detection**

**CS21B2039 M.VARUN REDDY**

In our project on heart disease detection we have collect the data from Kaggle we have searched for data from various sources but among all Kaggle has a around 918 individual data sets with no null values and covering different age groups and 11 constraints for each data set so that is the reason we choose Kaggle data for our project .

In 5 data constraints( sex, chestPainType, RestingECG, ExerciseAngina, ST\_slope) are not in numerical for so we encoded them to numerical form (0,1,2,...).

Later we plotted graphs between each constraints to get a clear view of the data.

After preparing the data we have spliced dataset into training and test set. We use the training set to create the model, which is subsequently used on the test set. We can assess the effectiveness of our model in this way.

Now we have feature scaled our data. A technique for normalising the variety of independent variables or features in data is called feature scaling. It is typically carried out during the data preprocessing step and is sometimes referred to as data normalisation in the context of data processing.

Now after processing the entire data we trained our dataset using different type of machine learning algorithms such as:

- Random Forest classifier
- Logistic Regression
- KNN Classifier (KNeighbors Classifier)
- Decision Tree classifier
- XGB Classifier(XG boost)
- SVC (Support vector classifier)
- Naive Bayes Classifier

After training the data set with various M.L. Algorithms and changing the test size of the data set we found that Naive Bayes classifier later SVM and random forest classifier has been the consistent throughout the process of changing the test size and running the model and decision tree classifier has the least accuracy every time.

As the test size decreases the accuracy of the model is increasing. Therefore , Naives Bayes Classifier gives the best accuracy.

Here are the results of our model accuracy at various test sizes for Naive bayes classifiers and other classifiers

Test size	Highest Accuracy	Algo name	Naive Bayes Accuracy
0.5	0.862745098039216	KNN	0.860566448801743
0.33	0.861386138613861	SVM	0.854785478547855
0.25	0.856521739130435	SVM	0.852173913043478
0.15	0.869565217391304	Naive's	0.869565217391304
0.1	0.923913043478261	XGBoost	0.891304347826087
0.9	0.84764207980653	Naive's	0.84764207980653