Characterization of Binary Machine Learning Classifier for Robust Heart Disease Prediction

Ishita, Ankit Sharma, Parag VijayVargiya , Aishwarya Soman
Dr.HirenThakkar

Abstract

Large number of patients related data is stored and maintained in the health industry. Heart disease is the most common one nowadays. The different ways of predicting it are Electrocardiogram (ECG), stress test, and Heart MRI. Here, the proposed model uses 13 parameters for the prediction of heart disease that includes heart rate, chest pain, cholesterol level, blood pressure, Age etc. The aim of this model is to predict whether heart disease is present or not using the various machine learning models such as Decision Tree, Random Forest, Logistic Regression, Naïve Bayes. We have achieved 0.3312 log loss using the Logistic Regression.

Introduction

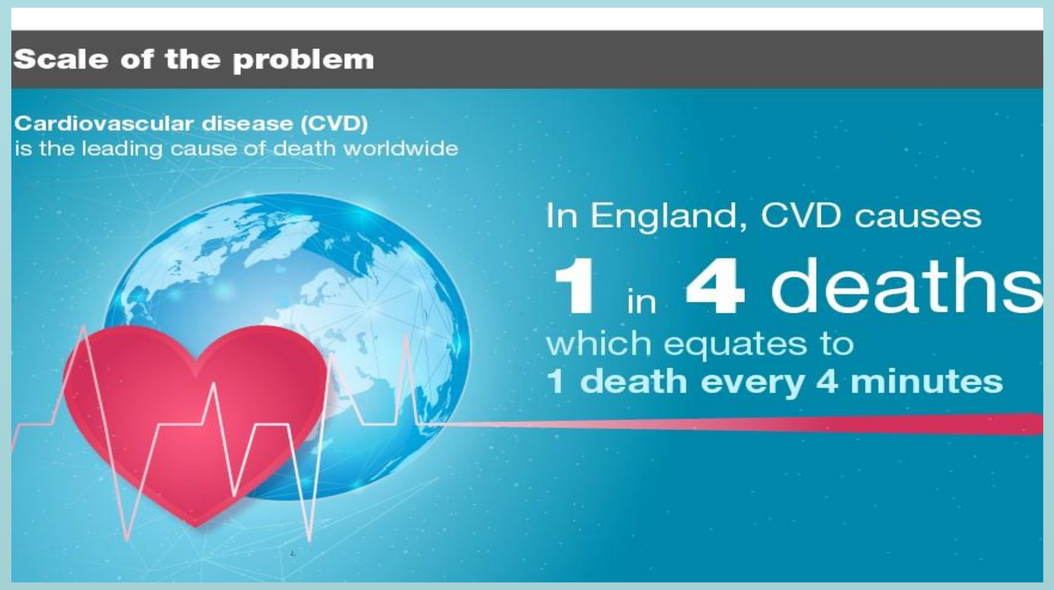

Fig 1:Cardiovascular diseases [2].

- Heart is one of the most important organs in the human body.
- Heart disease is the biggest killer all around the world, prevention of heart diseases is most important one [3, 4].
- Person with high cardiovascular(which means the symptom or factors are hypertension, Diabetes). So we need to early detect the diseases and cure them using appropriate medicine.
- To achieve that goal by using different machine learning algorithm.

Proposed Method

The architecture of the model is shown in Fig 2. The dataset is divided into two such as Training and testing. We have pre-processed the data and analysed through various Machine Learning algorithms (Logistic Regression [1], Decision Tree etc.). Then we have recorded the Log Loss.

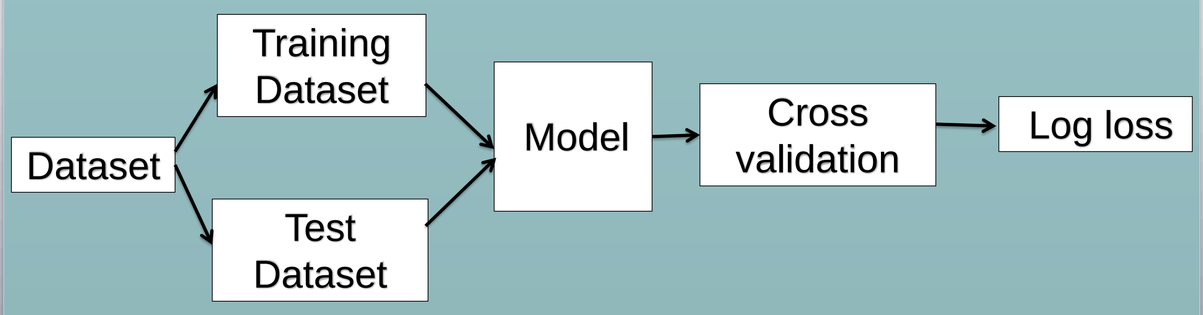

Fig 2:Architecture of the model

Experimental Results and Discussion

- The dataset has 13 features and 181 instances. These features are independent on each other.
- The 13 features include:
  slope_of_peak_excercise_st_segment, thal, resting_blood_pressure, chest_pain_type, num_major_vessels, fasting_blood_sugar_gt_120_mg_per_dl, resting_ekg_results, serum_cholesterol_mg_per_dl, oldpeak_eq_st_depression, sex, age, max_heart_rate_achieved, exercise_induced_angina**.**

In order to predict we use the data from Driven data Competition. Here the metric used is logarithmic loss:

Log loss=- $\frac{1}{n}\sum_{i=1}^{n}$ *[y log(y` )+(1−y )log(1−y`)]*

Here, ‘y’ is the probability that y=1. Log loss provides a steep penalty for prediction that are both confident and wrong.

The goal of our machine learning algorithm is to minimize this value(Log loss).

- Testing procedure involved Dataset is divided in to two such as Training data and Testing data. Then passing the pre-processed data through the model, and recorded Log Loss and repeating it for the various models.
- The model with Least Log Loss value give the best result.

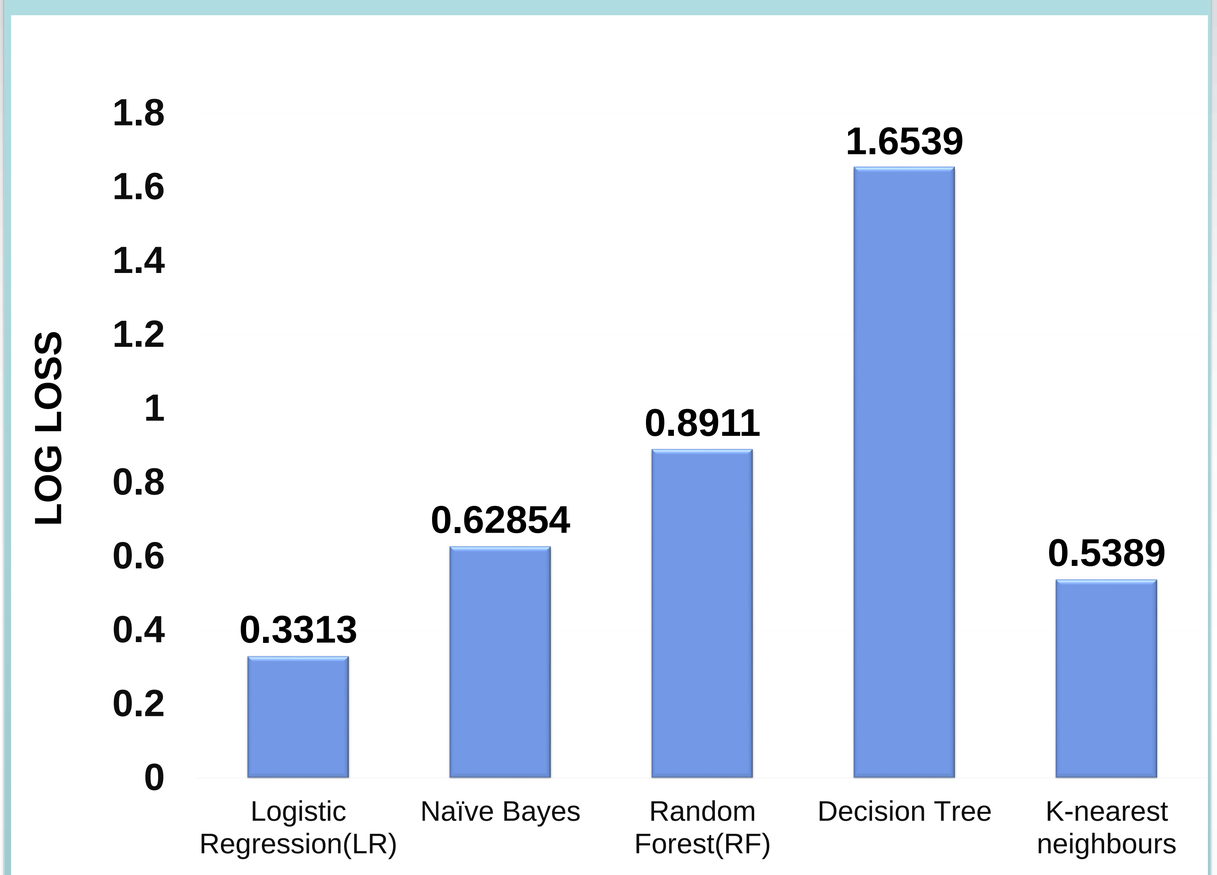

Fig 3: Log loss Results

The graph in Fig 3 shows the Log loss result that we have gotten from our models. Logistic regression has least value(0.3313), then K-nearest neighbours (0.5389) and so on.

Conclusions

In this experiment we get deep knowledge and understanding into different Machine learning techniques for the binary classification of heart diseases. After applying different models and defining Log loss as the evaluation metric. We used different algorithms like Logistic Regression , Decision Tree, Random Forest, Naïve Bayes are used in the evaluation of models. The main aim is to minimize the Log loss. The least Log loss will provide the best result.

References

1. https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
2. https://www.gov.uk/government/publications/health-matters-preventing-cardiovascular-disease/health-matters-preventing-cardiovascular-disease.
3. Sahoo, P., Thakkar, H., Lin, W. Y., Chang, P. C., & Lee, M. Y. (2018). On the design of an efficient cardiac health monitoring system through combined analysis of ecg and scg signals. Sensors, 18(2), 379.
4. Sahoo, P., Thakkar, H., & Lee, M. Y. (2017). A cardiac early warning system with multi channel SCG and ECG monitoring for mobile health. Sensors, 17(4), 711.

Team No. - X