

Cardiovascular Disease Prediction using Machine Learning

Vandana Joshi¹, Varshitha B², Mallikarjuna M³, Shruthi R⁴, Devarasetty Vivek Kumar⁵

^{1,2,3,4,5}School of Computing and Information Technology, Reva University, Bengaluru, India

¹vandanaj410@gmail.com, ²varshithab864@gmail.com, ³mathada.mass@gmail.com,

⁴shrurhiraj39672@gmail.com, ⁵vivekkumar.devarasetty@gmail.com

Article Info

Volume 83

Page Number: 4454-4457

Publication Issue:

May - June 2020

Abstract

Throughout world wide the main cause of death is happening due to heart attacks and this costs millions of deaths and threat to life. Data mining techniques have the advantage of less human intervention and it is also cost effective while the clinical parameters has to undergo many procedures. The significance of using data mining or machine learning techniques is that they give precise output. The study is made under supervised classification techniques Bayesian and logistic regression. The proposed web application is user interactive and helps in the medical field mainly in laboratories which incorporates the developed model and predicts the threat of heart disease in an easier and cost effective way. The outcome of the proposed model predicts the accuracy of the classifier and application proposed will predict if a person is chronic to Heart disease.

Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 12 May 2020

Keywords: Machine Learning; Supervised Learning; naive Bayes; Logistic Regression; Web Application.

1. Introduction

The most common type of heart disease is the coronary heart disease. About 18.2 million adults aged 20 and older have Cardiovascular Disease. To predict whether the person is suffering from heart disease, many approaches are to be taken. Machine Learning is one of the best techniques possible to give an accurate result. Supervised Classification Machine learning algorithms namely Naïve Bayes and Logistic regression techniques are taken and the comparative study between them is made. The one with the best accuracy is selected and implemented for the application that is been implemented later. The application is developed which will be user convenient where the section provided will be filled with the details including the clinical parameters such as name, age, sex, cholesterol, bp etc which are processed further.

In section III, the proposed architecture is shown and briefly described the algorithms of Naïve Bayes and Logistic Regression. The comparative study is made between these two algorithms and the accuracy graph is shown. The concept of web application serves as a user interface through which the user provides the details pertaining to the heart and is treated as the input to the model.

Section IV, is based on the output of the proposed model which makes the comparative study between Naïve bayes and Logistic Regression. The accuracy is compared between these techniques for both the training and testing set.

2. Literature Survey

Santhana Krishna.J. and Geetha.S. recommend a system where the heart disease is predicted using the Naïve Bayes and Decision Tree algorithms. These two supervised classification algorithms are applied on the same dataset. The accuracy of the Decision Tree and Naïve Bayes algorithms are evaluated. The classification model is generated for both the training and test data using these machine learning algorithms.

Anjan Nikhil Repaka .et.al., propose a system which makes use of Naïve Bayes algorithm to predict heart disease. The proposed system involves a procedure, wherein the user has to login through a mobile application. The data associated with the heart disease is collected. The collected data is then classified using Naïve Bayes and the disease prediction is made. It's revealed that with respect to accuracy, Naïve Bayes has 89.77% accuracy. The result is then encrypted into the database

using the Advanced Standard Encryption(AES) algorithm.It is observed that,Parallel Homomorphic Encryption Algorithm(PHEA) is less secured when compared to Advanced Standard Encryption(AES).[2]

Aditi Gavhane .et.al, propose a work which utilizes the concept of neural networks for heart disease prediction.The training and testing of data is performed using the Multilayer perceptron algorithm.MLP is a supervised neural network algorithm.The MLP algorithm has an input layer,one or more hidden layers,followed by an output layer.The weighted inputs are fed to the activation function and the algorithm nearly gives a reliable output.

K.G. Dinesh .et.al., proposes a cardiovascular disease prediction system using machine learning algorithms.Various machine learning algorithms like Gradient Boosting,Logistic Regression,Support Vector Machine,Naive Bayes and Random forest are applied on the dataset that is collected from the UCI Machine Learning Repository.The accuracy of these algorithms are calculated.Logistic regression yields the high overall accuracy of 86.51%.

Rahma Atallah and Amjed Al-Mousa recommend a Majority Voting Ensemble method which predicts the presence of heart disease.The various machine learning classifiers like Stochastic Gradient Descent(SGD), K-Nearest Neighbour Classifier, Random Forest Classifier,Logistic Regression Classifiers are applied.In an ensemble approach the above mentioned classifiers are combined.It is observed that an ensemble mechanism yields high accuracy when compared to the individual implementation of the algorithms.

3. Methodology

A. Proposed Mechanism:

In the figure 1 shown below initially the dataset has been taken from the UCI repository to train the model. In this system, naïve bayes and Logistic regression algorithms are employed.UCI repository is the database which provides the collection of all the static datasets.

The comparative study is made between the two algorithms.The one with the highest accuracy is selected for further process and implemented in the application.The user enters the data manually which includes the attributes such as sex, age, cholesterol etc.

The data collected from the user through the website is further processed using the Logistic Regression algorithm.The prediction for the individual instance of data is done.The predicted result is displayed on the webpage.

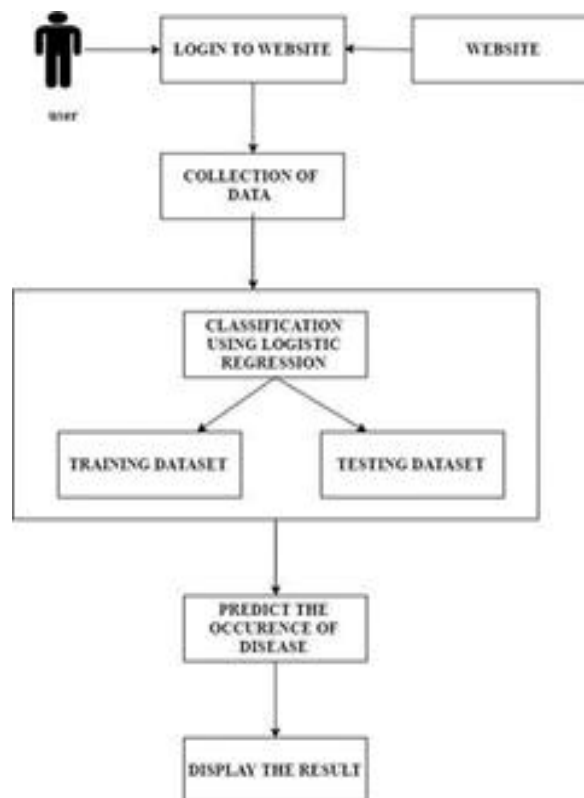


Figure 1: Proposed Architecture

Figure 1 shows the proposed Architecture for the system. The user provides the clinical details through website.The prediction is made using logistic regression and the predicted result is displayed on the webpage.

B. Description of Algorithms

The two main algorithms that are used in this system are

- Logistic Regression
- Naive Bayes

Logistic Regression:

Logistic Regression comes under the classification category of the supervised Machine Learning.The logistic regression is appropriate when the output of the model to be returned in binary, that is it is mainly derived as positive class or negative class.

For instance to evaluate the person as cheat or not cheat or to predict is the person is vulnerable to attack:

$$1 / (1 + e^{-\text{value}})$$

In the above given formula e represents the base of the natural logarithms ,value is actual numerical value that needs to be transformed.

Naive Bayes:

Naïve Bayes algorithm is based on a probabilistic approach where the posterior probability is calculated with the aid of prior probability.

There are different types of Naïve Bayes Classifier:

1. Gaussian Naïve Bayes Classifier
2. Bernoulli Naïve Bayes Classifier
3. Multinomial Naïve Bayes Classifier

Mathematical representation of bayes theorem:

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)}$$

$P(M)$ => independent probability of M (prior probability)

$P(N)$ => independent probability of N (evidence)

$P(N|M)$ => conditional probability of N given M (likelihood)

$P(M|N)$ => conditional probability of M given N (posterior probability)

4. Result & Discussion

The output of the proposed model which makes the comparative study between Naïve bayes and Logistic Regression. The accuracy is compared between these techniques for both the training and testing set. After the process, accuracy check it is seen that the Logistic Regression is having accuracy of around 80 percent and it is efficient. The web application proposed in this system predicts the occurrence of heart disease for an individual based on the clinical parameters provided to the system and the prediction is made using logistic regression algorithm.

Table 1: Comparison of Accuracies

ALGORITHM NAME	TRAINING SET ACCURACY	TEST SET ACCURACY
Naive Bayes	0.8677685950413223	0.7868852459016393
Logistic regression	0.8636363636363636	0.8032786885245902

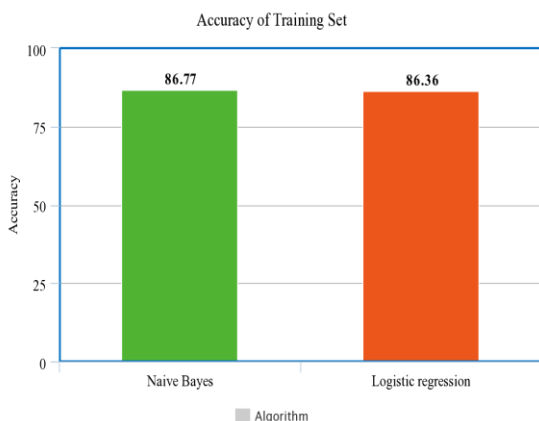


Figure 2: Comparison of training set accuracy

Figure 2 shows the bar graphs plotted for the efficiency of the algorithms naïve bayes and logistic regression by making comparative study and proves Logistic Regression is more accurate.

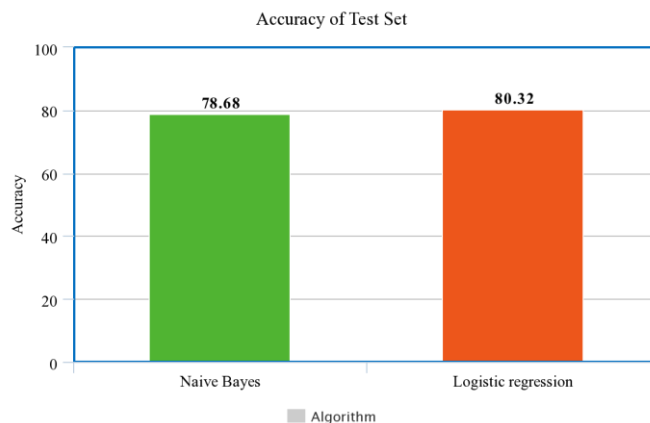


Figure 3: Comparison of test set accuracy

Figure 3 shows the accuracy of the Testing dataset when the comparative study is made using naïve bayes and logistic regression algorithms. It shows Logistic Regression has an accuracy of 80 percent.

5. Conclusion

In the proposed system, two supervised algorithms are applied on the same dataset. The accuracies of the two supervised algorithms are reviewed. Logistic regression is more accurate when compared to Naive Bayes. The diagnosis of a patient is performed using a Logistic regression algorithm because predicted output is reliable.

6. Future Work

Many web applications can be developed which resemble the proposed system for the prediction of other types of diseases. The machine learning hybrid algorithmic approach can be deployed. The machine learning ensemble method that combines multiple machine learning algorithms are perceived to provide better accuracy.

References

- [1] Santhana Krishna.J. and Geetha.s., "Prediction of Heart Disease Using Machine Learning Algorithms", 1st International Conference on Innovations in Information and Communication Technology (ICIICT), pp. 1-5, 2019.
- [2] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin "Design and Implementing Heart Disease Prediction Using Naives Bayesian", Proceeding of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), pp. 292-297.
- [3] Aditi Gavhane, Gouthami Kukkula, Ish Pandya, Prof. Kailas Deavadka, "Prediction of Heart

- Disease Using Machine Learning”, Proceeding of 2nd International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), pp. 1275-1278.
- [4] K.G. Dinesh, K. Arumugaraj, K.D. Santosh and V. Mareeswari, ”Prediction of Cardiovascular Disease Using Machine Learning Algorithms”, Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, pp. 1-7, 2018.
- [5] Rahma Atallah and Amjed Al-Mousa, ”Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method”, 2nd International Conference on new Trends in Computing Sciences (ICTCS 2019), pp. 1-6.
- [6] A. Ed-Dauody and K. Maalmi, ”Real-time machine learning for early detection of heart disease using big data approach”, 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, pp. 1-5.