

CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES.

K.Tejaswini¹, G.Vaishnavi², C.Upanya³, P.Niharika⁴

1,2,3 Junior Research Scholar, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India,

4 Junior Research Scholar, Department of Computer Science, Gitam University, Vizag, India.

Abstract: Over the last decade cardiovascular disease is the main reason for deaths in the world. Even if cardiovascular diseases is found as the important source of death in world in ancient years, these have been announced as the most avoidable and manageable diseases. The main idea behind this work is to study diverse prediction models for cardiovascular disease and selecting important cardiovascular disease features using various algorithms such as Stochastic Gradient descent, Random forest, Logistic Regression, Decision Tree and Support Vector Machine.

I. INTRODUCTION

Even though cardiovascular disease is acknowledged as the supreme chronic sort of disease in the world, it can be most avoidable one also at the same time. A healthy way of life (main prevention) and timely analysis (inferior prevention) are the two major origins of cardiovascular disease director. Conducting steady check-ups (inferior prevention) shows outstanding role in the judgment and early prevention of cardiovascular disease difficulties. Several tests comprising of angiography, chest X-rays, echocardiography and exercise tolerance test support to this significant issue. Nevertheless, these tests are expensive and involve availability of accurate medical equipment. Heart expert's create a good and huge record of patient's database and store them. Statistical analysis have acknowledged the count of risk factors for heart diseases counting age, blood pressure, diabetes, total cholesterol, and hypertension, heart disease training in family, obesity and lack of exercise.

II. LITERATURE SURVEY

Franck Le Duff (2004), made the decision tree rapidly by taking some help from the doctors and administrators. He also mentions that not much mining information actually helps patients.

Boleslaw Szymanski, et. al. (2006), performed to check the fitness of computation of a rare piece in SUPERNOVA. He used this method on a standard Boston lodging retail dataset for finding heart sicknesses, An estimation of heart exercises and desire for heart illnesses which discovered 83.7% to be right and estimation was carried out using the help vector machine.

Niti Guru, et. al. (2007), used the neural framework which had 13 kinds such as pulse, period, and so on as an advice for anticipating coronary illness, Blood Stress and Sugar.

Sellappan Palaniappan (2008), used methods for information mining calculation industrializing IHDPS- Intelligent Coronary disease figure System where every procedure has its own way to get the right outcomes.

III. PROPOSED WORK

In this paper, comparison of various machine learning methods is done for predicting the risk of cardiovascular disease of the patients from their medical data. We have tried prediction and analysis of cardiovascular disease by considering the parameters like age, gender, blood pressure, heart rate, diabetes. Since numerous factors are involved, the prediction of this disease is challenging. We have created a user interface where a patient having an internet access can easily enter their readings and get the output that predicts the possibility of the disease.

IV. MODULES INVOLVED

The entire work of this task is divided into four modules.

4.1 Real time information Pre processing

This module contains all the pre-preparing capacities that are required to process all information reports and messages. Firstly to preuse the train, test and approval information records at that point played out some preprocessing like slicing and dicing and so forth. There is some exploratory information investigation is performed to check the quality like invalid or missing qualities.

4.2 Feature selection

In this record we have calculated highlight extraction and determination strategies through sci-unit various libraries of python. For highlight choice, include change we have utilized different strategies like examination between the properties and finding the best-included qualities present in the dataset. Some qualities are additionally made for the factor Standardizing dataset after factors creation to bring each component at the same scale.

4.3 Classification

Importing Machine Learning libraries and evolution matrix to check the presentation of the model. Here we have built all the classifiers for cardio vascular disease detection. The features that are extracted are build using various classifiers. There are various Machine Learning methodologies compared. The model with the highest accuracy is considered.

4.4 Prediction

Once after the prediction of the most accurate algorithm. Then comes to Analysis the Result using various result analysis models and then the process of data Visualization. At the end the patient can enter the readings and prediction can be done.

V. ALGORITHMS

5.1 Random Forest

The random forest is a classification algorithm which consists of many decision trees. It used for bagging and feature randomness where it builds each individual tree to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of an individual tree.

5.2 Logistic Regression

This is the algorithm defined to classify which allocates observations to discrete classes. It uses sigmoid functions for the transformation of output. The values in this algorithm are not as continuous as a precise lapse. The obtained probability value in the sigmoid function is matched with various discrete classes.

5.3 Stochastic Gradient Descent

Stochastic Gradient descent is an iterative algorithm, that which starts from a random point on a function and travels down its slope in steps until it reaches the lowest point of that function. Subsequently, in Stochastic Gradient Descent, a few tests are taken arbitrarily rather than the entire informational index for every emphasis. While playing out the Gradient Descent, and it ought to be accomplished for each emphasis until the minima are reached. Consequently, it turns out to be computationally over the top expensive to perform.

5.4 Decision Tree Classifier

It is utilized to explain both the relapse and order issues. The choice tree utilizes the tree portrayal to take care of the issue in which each leaf hub compares to class marks and qualities are spoken to on the inward hub of the tree.

5.5 Support Vector Machine

Separation of Hyperplane is formally considered in this algorithm the information that trained is given as input. The results of this method depicts the hyperplane. In a 2D plane, it is shown as a line that divides the plane into two parts where each class lies on an either plane.

5.6 Result

The machine learning models is evaluated using the AUC-ROC metric. This can be used to understand the model performance. The Stochastic Gradient Descent is more efficient.

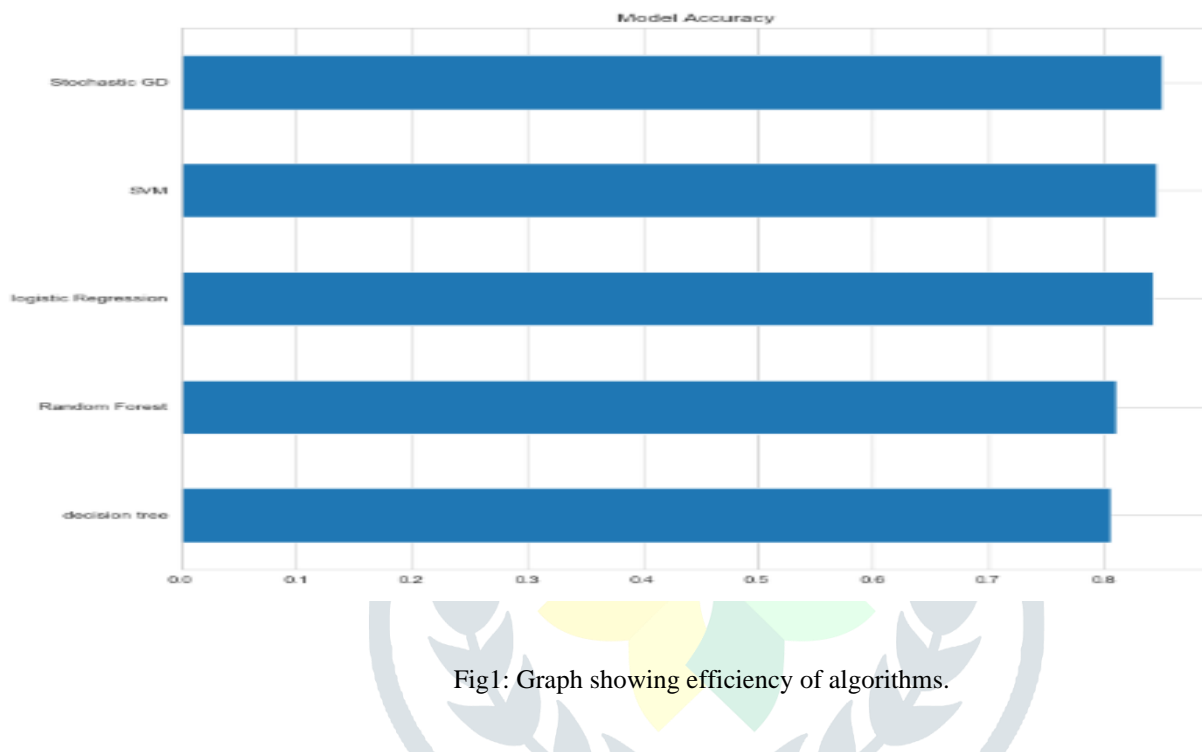


Fig1: Graph showing efficiency of algorithms.

VI. USER INTERFACE

Running the web page on the local host port 5000. Copying the URL <http://127.0.0.1:5000> and running on the web browser results in opening the Web page where the input of the patient is given.

```
(base) C:\Users\ganva\OneDrive\Desktop\cardio vascular diseases prediction using machine learning>py r.py
* Serving Flask app "r" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
* Debugger is active!
* Debugger PIN: 142-977-682
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

Fig2: Anaconda prompt showing the local host URL

Designed UI to give the USER INPUT .

The screenshot shows a web browser window with the address bar displaying '127.0.0.1:5000'. The page contains a form with the following fields and labels:

- AGE: Enter Age
- Gender: Enter Gender
- Cp: Enter Chestpain
- Testbps: Enter Person's Resting I
- Cholestrol: Enter Persons Cholestrc
- fbs: Enter fasting blood suga
- Restecg: Enter Resting ECG
- Thalach: Enter Persons maximum
- Exang: Enter Exercise induced
- Oldpeak: Enter ST depression ind
- Slope: Enter the slope of the pe
- Ca: Enter The number of mg
- Thal: Enter A blood disorder c
- submit

Fig3: User Interface for giving the inputs for various attributes

VII. RESULT:

Result Screen 1 :

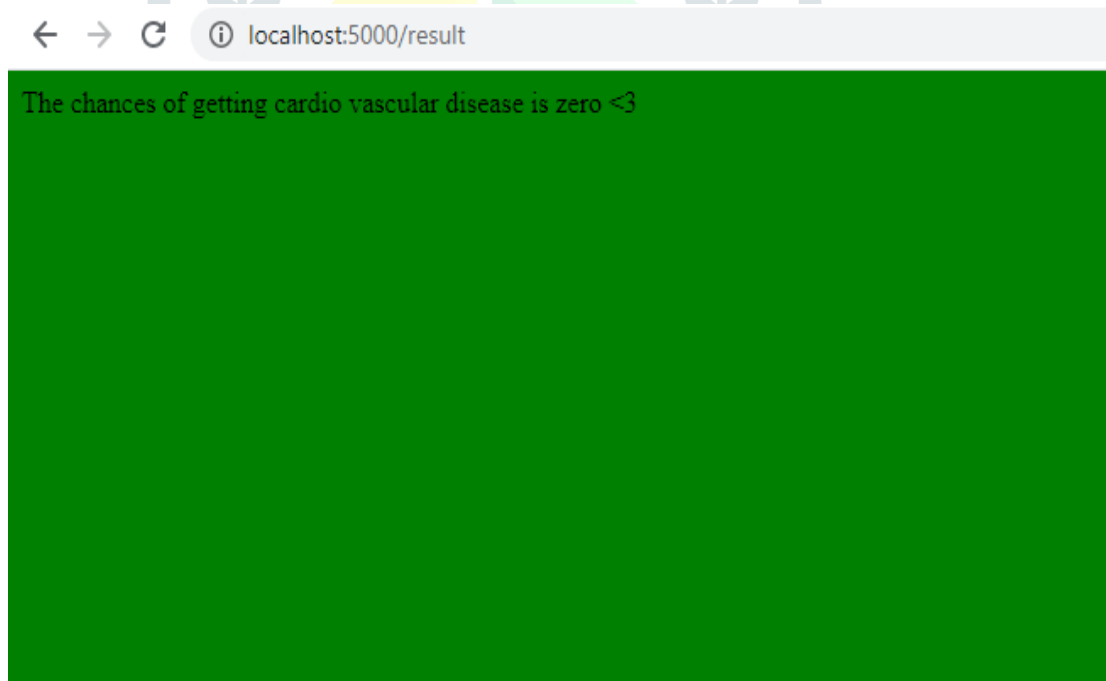


Fig4: The result showing that the chances of getting cardio vascular disease is zero.

Result Screen 2 :

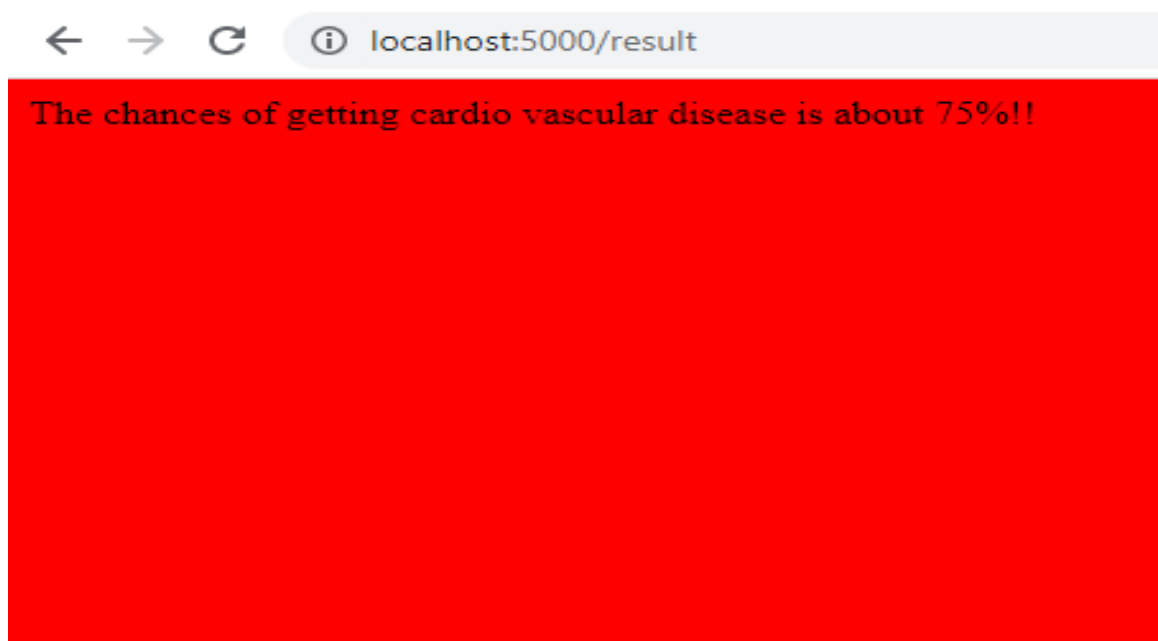


Fig.5: The result showing that the chances of getting the cardio vascular disease is about 75%.

VIII. CONCLUSION

Various Classifier techniques have introduced in coronary illness. After a comparison of all the methodologies Stochastic Gradient Descent gives the highest accuracy after training and verification of information that is present in the dataset. As illustrated before the system can be identified as a scientific assistant for any clinicians. The disease prediction through the risk factors can be organized online and hence any internet, users can go through the system through a web engine and understand the risk of cardio vascular disease. The scheduled model can be implemented for any real-time application. This helps in identifying the disease in the early stages. The proposed model has a wide area of applications like grid computing, cloud computing, robotic modeling, etc. The development of Web or Android- based application helps for the detection of disease.

REFERENCES

- [1] P .K. Anooj, —Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules!; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40. Computer Science & Information Technology (CS & IT) 59
- [2] Nidhi Bhatla, Kiran Jyoti”An Analysis of Heart Disease Prediction using Different Data Mining Techniques”.International Journal of Engineering Research & Technology
- [3] Jyoti Soni Ujma Ansari Dipesh Sharma, Sunita Soni. “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”.
- [4] Chaitrali S. Dangare Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques” International Journal of Computer Applications (0975 – 888.
- [5] Dane Bertram, Amy Volda, Saul Greenberg, Robert Walker, “Communication, Collaboration, and Bugs: The Social Nature of Issue Tracking in Small, Collocated Teams.