International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Heart Diseases Prediction using Machine Learning Algorithms

[1] P Sampath [2] A Riyaz [3] G Narendra [4] G Durvasi
[1,2,3] Department of Information Technology, Andhra Loyola Institute of Engineering and Technology
[4] Assistant Profeser, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology

1. ABSTRACT

Heart related diseases are the major reason for deaths worldwide. Increasing Deaths due to heart related issues in the last few years is an alarming situation. The leading factors which causes cardiac arrest and remaining heart problems are high blood pressure, diabetes, smoking and unhealthy diet. In order to reduce the death rate due to cardiovascular disease time to time diagnosis and effective treatment is obvious. Chest X-ray, Electrocardiogram (ECG) and coronary algorithms are Traditional cardiovascular disease diagnosis methods. These methods would require a highly qualified professionals and takes a lot of time and sometimes may lead to misdiagnosis. n this paper we are proposing machine learning algorithms which includes Logistic Regression, Decision Tree, Naïve Baye's, SVM for automating the heart diseases prediction. We have used the medical profiles of patients like age, gender, previous chest pain data some other medical profiles in order to predict the odds of getting a heart disease. Depending upon our model prediction, patients can consult doctor.

2. INTRODUCTION

The saying "Health is Wealth" is well-known throughout the world. The ability of the body to function well overall is referred to as being in good health. The heart is an essential component of the human body. If the heart is not functioning properly, it will also impact the other body parts. The heart serves as the body's operating system. Heart illnesses are becoming the leading cause of death in humans. A recent survey found that one in three persons globally have high blood pressure, which accounts for around half of all stroke and heart disease deaths. Cardiovascular disease is another name for heart disease (CVD).

Here, using machine learning methods, we create a smart system that accurately forecast cardiac problems. we used Cleveland data that was already accessible to train the dataset. The patient must register in the application by supplying the necessary information if they wish to anticipate their cardiac illness. After the user has successfully registered, a medical form that asks them to provide information to anticipate their heart illness must be filled out. This form contains 13 distinct medical characteristics that help us estimate the likelihood of developing a heart condition. In order to predict cardiac disease, this study uses machine learning methods including Logistic Regression, Decision Tree, Naive Baye's, and Support Vector Machine.

3. LITERATURE SURVEY

It has been discovered through research that supervised machine learning algorithms can be used to forecast cardiovascular disease. On this subject, numerous study papers have been prepared. A survey that evaluates the effectiveness of various models built using machine learning algorithms and methodologies has been published as a paper [4]. In one of the publications, effort has been done to develop a Graphical User Interface (GUI) that uses a Weighted Association rule-based Classifier to predict whether or not a person has heart disease [5]. A novel strategy for the prediction of cardiac illness has been introduced in another study [6] and is based on the coactive neuron-fuzzy interference system (CANFIS). One of the studies [7] provides an overview of the methods frequently employed for heart disease prediction and their difficulties. One of the publications showed how Naive Baye's can be used for classification purposes and proposed a classifier technique for the diagnosis of heart disease [8]. In one of the publications, a survey is conducted that contains various papers in which one or more data mining algorithms have been utilised to predict cardiac disease [9].

4. METHODOLOGY

Supervised learning:

Supervised learning, often known as "task learning," is the approach to machine learning that is most frequently employed. Generally, it is used to predict the following values. A driven mechanism. It is well-liked since it is the simplest and most straightforward technique to adopt. In this case, we supply a dataset for training and prediction. Based on the supplied data, the algorithm predicts and consistently produces an output. To acquire the desired projected output over time, we must train it back by providing the feedback as an input.

Logistic Regression:

It is one of the supervised learning category's most widely used ML algorithms. The various probability of a necessary variable or a target are predicted using it as target element. There would only be two classes that may exist for a target's qualities.Generally speaking, logistic regression is binary in nature; the data are coded as either 1 or 0, where 1 denotes success and 0 denotes failure. This model predicts Predict(Y=1) as a function of X mathematically. Owing to its crude predictability of a target variable

Decision Tree:

It is a technique for decision prediction modelling that can be used in various contexts. It can be used with a decision-building strategy that quickly separates theutilising a dataset in many logical ways to make predictions about potential scenarios or conditions. The phrase "one of the most powerful algorithms" is also used to describe this algorithm, which belongs to the class of supervised learning models.Classification tasks and Regression tasks are the additional classifications for the decision tree algorithm.In the modelling process, these components of a decision tree are known as decision nodes, where the data is departing and dividing in different directions dependent on conditions.

Naive Baye's :

It is well recognised as a classification method that relies on the Baye's theorem, which makes the strong assumption that all predictors or algorithms are apart from one another. The posterior probabilities are the most common Bayesian classification. It is stated that the Nave Baye's method can be used to demonstrate this feature in a quantitative structure for a label given with some observable nature.

Predict(L| f) = predict(L)(f | L) / Predict(f)

Support Vector Machine :

It is one of the most well-known supervised learning algorithms and is applied to both classification and regression issues. SVM is employed in the majority of scenarios. As a machine learning approach to solving categorization issues. The SVM's formative function is to display the best line or decision boundary that is divided into classes in n-dimensional space. As the accuracy is time-lined, we can later place a new data point in the appropriate sector. And this choice limit is referred to as a "hyper plane." To build a hyper plane, this technique selects the extreme vectors or data points. Support vectors are these classes, and support vector machines are easily available. Utilised for text classification, image classification, and face detection.

ARCHITECTURE:

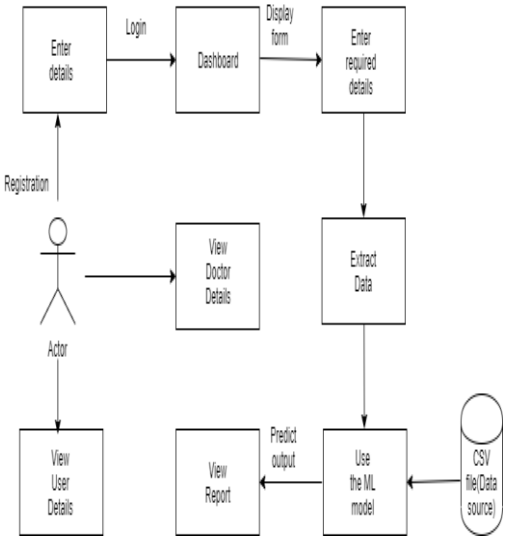


figure1:architecture

Registration: The user should register by providing the information needed to forecast cardiac illness.

Dashboard: After creation of an account, the user needs to login himself/herself with their respective login credentials so that it will directly redirect to the dashboard.

Enter data: After successfully logging in, the individual should provide information about their heart health.

Data extraction: The user's details will be used to extract the secret data.

Using ML model: By using ML model, the hidden data will be extracted.

View report: See the report The user can view their prognosis, which indicates the possibility of a heart attack, by utilising the ML model to predict the outcome of heart illness.

Algorithms implementation:

Input: Information entered into the form.

Output: Prediction of the class as label.

We initialize the classifier which is to be used.

Classifier training classifiers in scikit-learn uses the fit(x,y) method. Then we predict target and evaluates the classifier model.

EXPERIMENTAL RESULT

The application's operational functionality is described in the steps that follow. As follows:

1. A computer with Jupyter, a Python programme, is first required. jupyter notebook with django and anaconda installations inside.
2. We require a dataset with various attributes pertaining to heart illnesses included in it.
3. Using machine learning techniques like Logistic Regression, Decision Tree, Naive Baye's, and SVM, we should train and test the dataset. These methods should be saved as logistic regression.py, decision tree.py, naive bayes.py, and scv.py, respectively.
4. Next, we must run the server at the Anaconda prompt, apply the migrations, and launch the application.
5. On the login screen, if the user is legitimate, they will be taken to their profile.
6. The user can then fill out the form with their medical information and click the forecast button to see their outcome of coronary disease.
7. The user may finally see the heart disease prediction.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0

Figure 2: Prediction dataset (HealthData.csv)

Figure 3: Testing Accuracy for Logistic Regression

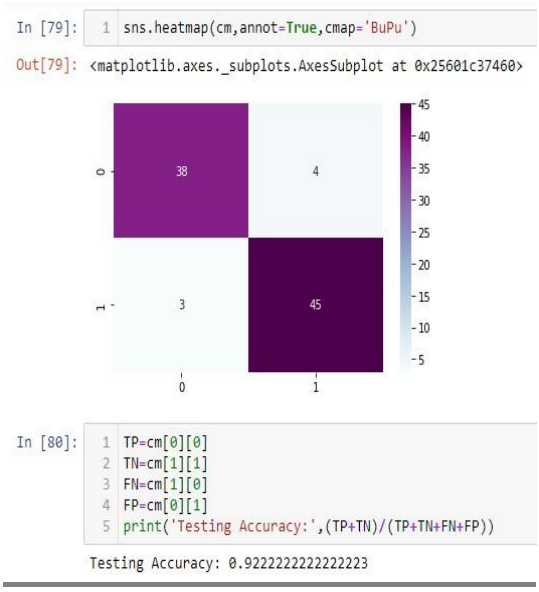


Figure 4: Testing Accuracy for Decision Tree

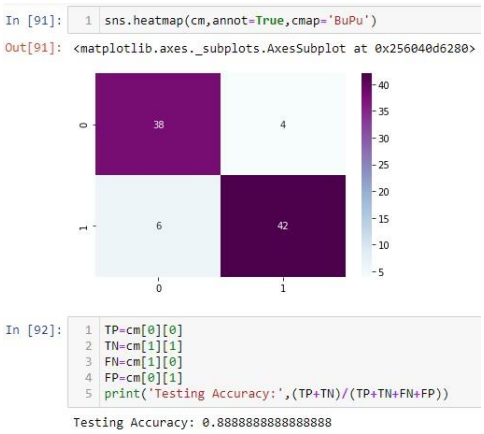
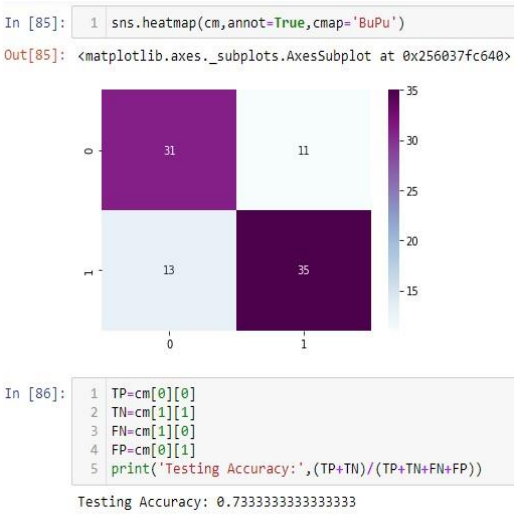
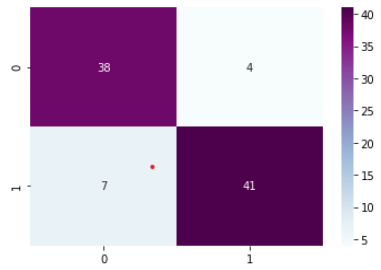


Figure 5: Testing Accuracy for Naïve Bayes



```
In [97]: 1 sns.heatmap(cm,annot=True,cmap='BuPu')
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x25601ba2430>
```




```
In [98]: 1 TP=cm[0][0]
         2 TN=cm[1][1]
         3 FN=cm[1][0]
         4 FP=cm[0][1]
         5 print('Testing Accuracy:',(TP+TN)/(TP+TN+FN+FP))
Testing Accuracy: 0.8777777777777778
```

Figure 6: Testing Accuracy for SVM

CONCLUSION:

In order to anticipate heart attacks and assess the most effective treatment options, the study conducted an experiment using a variety of machine learning algorithms. The research findings do show a significant variation in prediction when utilising various ML classification methods. The experiment can be a useful tool for doctors to identify high-risk situations and provide appropriate advice. The classification model will be able to respond to more complicated questions regarding heart attack illness prediction. According to regression, Naive-bayes, and SVM algorithms' predictions, the metrics employed to predict the existence of cardiac disorders are trustworthy indicators. This study primarily focuses on developing algorithms with high accuracy for predicting cardiac disease, and the dataset offers a wealth of information. superior accuracy to one with missing data that is unaltered. So, using sophisticated and effective methods for data cleaning together with precise and conventional machine learning algorithms will aid in the creation of prediction systems that yield effective and consistent Outcomes.

REFERENCES

1. Andrea D'Souza, ,Heart Disease Prediction, IJRES, Vol 3, No 3, April-2015.

2. Dhanashree S. Medhekar, Mayur P. Bote, Shruti D.Deshmukh, ,Heart Disease Prediction System using Naïve Bayes', IJERSTE, Vol 2, No 3, March-2013.

3.Jyoti Soni,Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction', International Journal Computer Applications, Vol17, No 8, March-2011.

4.G. Purusothaman and P. Krishnakumari, ,A Survey on Risk Prediction: Heart Disease', Indian Journal of Science, Vol 8, No 12, 2015. 5.https://researchdesignlab.com/smart-health-monitoring-s ystem.html

6.https://pdfs.semanticscholar.org/8610/4041dd610f78b6a5718e7ebe3cf6affb6f28.pdf

7.https://www.ijrte.org/wpcontent/uploads/papers/v8i6/F9241038620.pdf

8.https://pdfs.semanticscholar.org/caf8/b4d1b1cabe259aa84d12e44c6cdbdf908c81.pdf

9.https://www.ijana.in/Special%20Issue/TPID22.pdf

10.https://towardsdatascience.com/what-are-the-types-of-machine-learning e2b9e5d1756f

11.https://www.ripublication.com/acst17/acstv10n7_13.pdf

12.ShantakumarB.Patil,Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network', European Journal of Scientific Research, ISSN 1450-216X, Vol 31, No 4, 2009.

13. Syed Immamul Ansarullah, ,Heart Disease Prediction System', IRJET.

14. S.B.Bhalerao1, Dr. B.L.Gunjal2, ,Survey Of Heart Disease Prediction Based On Algorithms', IJARIIE vol-2 Issue-02, 2016.