



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: <https://doi.org/10.22214/ijraset.2022.42858>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Heart Disease Prediction Using ML Algorithm

Ritika Varshney¹, Himanshi Chaudhary², Mansi Bansal³, Mayank Goyal⁴

^{1, 2, 3, 4} Department of Electronic and Communication Engineering, KIET Group of Institutions, Ghaziabad, UP, India

Abstract: Cardiac disease could be a major international unhealthiness in fashionable drugs. The twenty-first century saying that by taking unusual or junk food cause cardio attack. The criticality of viscus diseases square measure a lot of crucial and may even result in vulnerable consequences if it is not detected at associate degree earlier stage. The techniques such as electronic health records, blood pressure measure machine are continuously monitoring human health condition through injection of major wearable devices. Since the information generated from the body space networks square measure continuous and tremendous in volume, the machine learning techniques square measure used for economical health information classification processes. This research paper is related to heart disease prediction. The outcomes are analyzed regarding factors such as accuracy, sensitivity, precision, and specificity. It has been found from the analysis that the proposed system provides comparatively better accuracy and prediction measures than the existing techniques.

Keywords: Heart disease, logistic regression, supervised learning

I. INTRODUCTION

Cardiac disease may be a major international unhealthiness in trendy drugs. The twenty first-century expression consummate proliferation in expectancy and a big transference within the causes of cardiopathy sorrowfulness throughout the planet. these days it's interpreted for roughly 30 % decrease across the world as well as approximately 40% within the high-income country and 28% in low and middle-income countries. Heart disease is the second leading disease in Asian country. Most of the time doctors alone cannot foresee if a patient is liable to a heart failure while not additional tests. This project aims to style a prediction model victimization machine learning technique to forecast the potential of a patient vulnerable to cardiovascular disease. Aim of this research is to develop heart disease prediction model using Machine Learning There are so many disease to diagnose data but this paper will focus on heart disease prediction. There area unit several factors associated with cardiopathy however we'll specialize in the smoking issue. We will specialize in Naïve Bayes model, MLP model and provision Regression. Grounded on the end of exploration, the significance of the study lies in assessing the prospects that initiates the habits of smoking. Alternately, the link between cardiovascular complaint and smoking can be estimated. Also, in field of healthcare, the exploration can work on chancing ways that in future can reduce the rates of conditions, especially cardiovascular conditions among the colonizing that practice events of smoking. In this chapter is introducing the topic of this report and has mentioned reasons why we research this topic in detail. Background problem related to this research topic has been discussed in this chapter of dissertation. On the other hand, aim, objectives and questions of this research are postulating the purpose of this research. This work is critical to review the problems associated with heart condition prediction and to seek out a far better solution by implementing an appropriate technique into the prediction model. Machine Learning, Classification, and Bayesian Network are some of the core concepts and terminologies covered in this chapter. Related and existing works on heart disease prediction using various machine learning techniques such as Naive Bayes, Logistic Regression, KNN, Decision Tree, SVM, and others are examined, with an emphasis on what was done, how it was done, the classification technique used, the data set used for implementation, the tools used, and the system's result and accuracy.

II. IMPORTANT COMPONENTS

A. Supervised Learning

Supervised learning is the process of training a data sample from a data source and then forecasting or predicting using a test dataset resulting from the data sample (Sathya & Abraham, 2013). In supervised learning, models are trained using a labelled dataset, where the model learns about each category of input. When the training phase is over, the model is evaluated using test data (a subset of the training set), and it then predicts the output.

1) Regression

A regression model attempts to forecast a continuous (real-valued) output. Trying to forecast the price of a specific house based on a variety of parameters is an example of a regression problem.

Some supervised learning regression algorithms are listed below:

- a) Linear Regression
- b) Regression Trees
- c) Non-Linear Regression
- d) Bayesian Linear Regression
- e) Polynomial Regression

2) Classification

A classifier model attempts to convert a discrete input space to a discrete output. A prediction of the existence of heart disease in a patient is an example of classification (Witten & Frank, Practical Machine Learning Tools and Techniques, 2005).

Some of the supervised learning classification algorithms are listed below:

- a) Random Forest
- b) Decision Trees
- c) Logistic Regression
- d) Support vector Machines

B. Unsupervised Learning

To anticipate the outcome from unlabeled datasets, unsupervised learning is utilized. The most frequent unsupervised learning technique is clustering.

Consider a scenario: the unsupervised system is given an input dataset which contains the photographs of various cats and dogs. Because the algorithm is never trained on the given dataset, it has no idea what its characteristics are. The purpose of the unsupervised learning algorithm is to recognise visual elements independently.

This task will be done by dividing the image collection into groups based on image similarity using an unsupervised learning method.

The following are some of the most important arguments for the relevance of unsupervised learning:

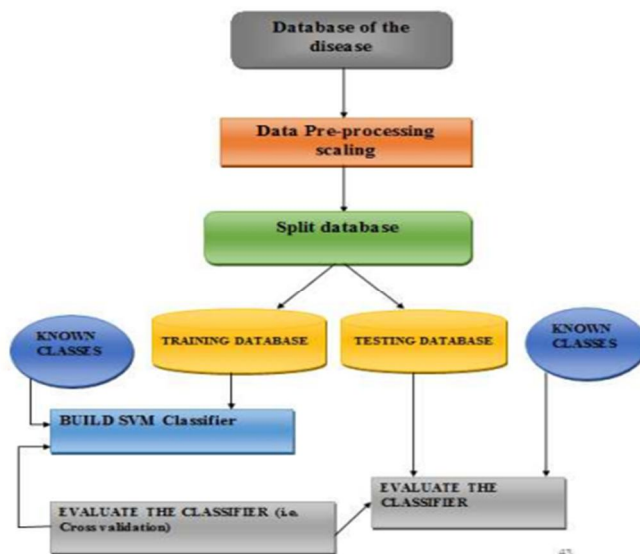
- 1) Unsupervised learning is beneficial for extracting relevant information from data.
- 2) Unsupervised learning is analogous to how a human learns to think via their own experiences, bringing it closer to true AI.
- 3) Because unsupervised learning works with unlabeled and uncategorized data, it is more important.
- 4) In the real world, we don't always have input data that corresponds to output, hence we require unsupervised learning to handle these problems.

C. Semisupervised Learning

Semi-supervised learning is similar to supervised learning in that it uses both labelled and unlabeled data. Because unlabeled data is less expensive and easier to obtain, it uses a little amount of labelled data with a large volume of unlabeled data. Classification, regression, and prediction can all be employed with this form of learning.

When the expense of labelling is too high to allow for a fully labelled training procedure, semi-supervised learning comes in handy. Identifying a person's face on a webcam is an early example of this. It would be nearly impossible to find a big number of tagged text documents in this case, therefore semi-supervised learning is suitable. This is due to the fact that having someone read through full text documents merely to assign a simple classification is inefficient. As a result, semi-supervised learning enables the algorithm to learn from a small number of labelled text documents while classifying a large number of unlabeled text documents in the training set.

III. METHODOLOGY



Project methodology is delineate as action to be taken to research a project’s problem. A strategy is outlined as a procedure, technique, and rules that may be used to understand the project. The technique are going to be accustomed determine, select, method and analyze the dataset to induce the knowledge for higher understanding the matter during this project. The effectiveness of the planned methodology are going to be recognized during this project. In order to fulfill the project objectives, many graphs, charts and applied mathematics knowledge have been accustomed perceive the project in a very correct manner. Because it has already been mentioned, secondary knowledge assortment methodology has been utilized in the study.

Attribute	Description	Domain of Values
Age	Age in years	29-79
Sex	Sex	0→ Female, 1→ Male
Cp	Chest pain type	1→Typical angina, 2→Atypical angina 3→Non-angina, 4→Asymptomatic
Trestbps	Resting blood sugar	94 to 200 mm Hg
Chol	Serum cholesterol	126 to 564 mg dL ⁻¹
Fbs	Fasting blood	sugar >120 mg dL ⁻¹ : 0→ False, 1→True
Restecg	Resting ECG result	0→Normal, 1→ST-T wave abnormality2→LV hypertrophy
Thalach	Maximum heart rate achieved	71 to 202
Exang	Exercise induced angina	0→No, 1→Yes
Oldpeak	ST depression induced by exercise relative to rest	0 to 6.2
Slope	Slope of peak exercise ST segment	1→ Upsloping, 2→Flat, 3→Downsloping
Ca	Number of major vessels colored by fluoroscopy	0-3


```

from sklearn.metrics import precision_score, recall_score, f1_score,
confusion_matrix

confusion_matrix(y_test_array, target_predict_values_array)

array([[28,  2],
       [ 7, 23]])

from sklearn import metrics

from sklearn.metrics import classification_report

print('accuracy %s' % metrics.accuracy_score(y_test_array,
target_predict_values_array))

print(classification_report(y_test_array, target_predict_values_array))

accuracy 0.85

      precision    recall  f1-score   support

0         0.80      0.93      0.86         30

1         0.92      0.77      0.84         30

avg / total         0.86      0.85      0.85         60

```

Fig Processed data

This is to show how well our model performed. Refer to section to get indepth knowledge on what each performance metrics means. The model was able to predict 51 out of the 60 test samples correctly thereby achieving an accuracy of 85%.

IV. CONCLUSION

The research work has made a model of the Bayesian network for predicting heart disease in person. This model is built using a hill climb algorithm installed in bnlearn packet in R. The main objective of the study was to demonstrate the effectiveness of the Bayesian classes in prediction of heart disease. We have used two different uses of the Bayesian classifier:

The Bayesian Belief Network has produced an accurate representation of the image. The model obtained helps us to easily identify the relationship of the possible cause conditional dependence and independence between attributes. The database was collected at the University of California, Irvine Machine Learning Center.

Data obtained from V.A. Medical Center, LBCC Foundation. It contains 303 cases of 14 cardiovascular disease each; 13 input number attributes and outcomes of one class. The effectiveness of the Bayesian classes is measured against each other with Bayesian Belief Networking Outperforming the logistic regression in Predicting Heart Disease. This research will help to think about heart disease, thus acting as a diagnostic tool to support physicians.

REFERENCES

- [1] Soni J, Ansari U, Sharma D and Soni S 2011 Predictive data mining for medical diagnosis: an overview of heart disease prediction International Journal of Computer Applications 17 43-8
- [2] Dangare C S and Apte S S 2012 Improved study of heart disease prediction system using data mining classification techniques International Journal of Computer Applications 47 44-8
- [3] Ordonez C 2006 Association rule discovery with the train and test approach for heart disease prediction IEEE Transactions on Information Technology in Biomedicine 10 334-43
- [4] Shinde R, Arjun S, Patil P and Waghmare J 2015 An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm International Journal of Computer Science and Information Technologies 6 637-
- [5] Bashir S, Qamar U and Javed M Y An ensemble-based decision support framework for intelligent heart disease diagnosis International Conference on Information Society (i-Society 2014) (IEEE) 259-6
- [6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W and Yun Y D 2014 A coronary heart disease prediction model: the Korean Heart Study BMJ open 4 e00502



- [7] Ganna A, Magnusson P K, Pedersen N L, De Faire U, Reilly M, Ärnlöv J and Ingelsson E 2013 Multilocus genetic risk scores for coronary heart disease prediction Arteriosclerosis, thrombosis, and vascular biology 33 2267-7
- [8] Jabbar M A, Deekshatulu B L and Chandra P Heart disease prediction using lazy associative classification 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s) (IEEE) 40-6
- [9] Dangare Chaitrali S and Apte Sulabha S 2012 Improved study of heart disease prediction system using data mining classification techniques International Journal of Computer Applications 47 44-8
- [10] Jyoti Soni 2011 Predictive data mining for medical diagnosis: An overview of heart disease prediction International Journal of Computer Applications 17 43-8
- [11] Chen A H, Huang S Y, Hong P S, Cheng C H and Lin E J HDPS: Heart disease prediction system 2011 Computing in Cardiology (IEEE) 557-60
- [12] Parthiban Latha and Subramanian R 2008 Intelligent heart disease prediction system using CANFIS and genetic algorithm International Journal of Biological, Biomedical and Medical Sciences 3
- [13] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E and Manefjord H 2016 Wireless body area network for heart attack detection [Education Corner] IEEE antennas and propagation magazine 58 84-92
- [14] Patel S and Chauhan Y 2014 Heart attack detection and medical attention using motion sensing device -kinect International Journal of Scientific and Research Publications 4 1-4
- [15] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A and Ghosh K Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design 2016 19th International Conference on Computer and Information Technology (ICCIT) (IEEE) 299-303



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)