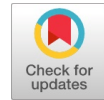


# Heart Disease Prediction Using Machine Learning Algorithm



Praveen Kumar Reddy M, T Sunil Kumar Reddy, S.Balakrishnan, Syed Muzamil Basha  
Ravi Kumar Poluru

**Abstract:** Heart disease is a common problem which can be very severe in old ages and also in people not having a healthy lifestyle. With regular check-up and diagnosis in addition to maintaining a decent eating habit can prevent it to some extent. In this paper we have tried to implement the most sought after and important machine learning algorithm to predict the heart disease in a patient. The decision tree classifier is implemented based on the symptoms which are specifically the attributes required for the purpose of prediction. Using the decision tree algorithm, we will be able to identify those attributes which are the best one that will lead us to a better prediction of the datasets. The decision tree algorithm works in a way where it tries to solve the problem by the help of tree representation. Here each internal node of the tree represents an attribute, and each leaf node corresponds to a class label. The support vector machine algorithm helps us to classify the datasets on the basis of kernel and it also groups the dataset using hyperplane. The main objective of this project is to try and reduce the number of occurrences of the heart diseases in patients.

**Keywords :** About four key words or phrases in alphabetical order, separated by commas.

## I. INTRODUCTION

Machine learning is one of the most disruptive technologies of this generation. It is a part of data science wherein the computer systems are made to learn from the different data sets on the basis of patterns generated from the datasets. It basically enables a machine to learn on its own based on some useful algorithms specifically developed for it. Machine learning is being heavily used in all the sectors including manufacturing, healthcare, Research and Development etc. It is also a recent trend in the world and is forming a major branch of computer science due to its power to automate things. Some notable examples of machine learning implementation comprises of the online chess board game or the online music player which works on a particular mechanism that is it learns the patterns in which we provide it with the input for example when we are listening to a particular song every day at a specific time then after some

Manuscript published on 30 August 2019.

\*Correspondence Author(s)

**Praveen Kumar Reddy M**, Assistant Professor in School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India.

**T Sunil Kumar Reddy**, Professor in the Department of Computer Science Engineering, Sri Venkateswara College of Engineering and Technology, Chittoor.

**S.Balakrishnan**, Professor in Computer Science and Business Systems, Sri Krishna college of Engineering and Technology

**Syed Muzamil Basha**, Ph.D. degree in School of computer science and Engineering from the VIT (Deemed by university), Vellore, India.

**Ravi Kumar Poluru**, Research Scholar in the School of Computer Science & Engineering, Vellore Institute of Technology, Vellore

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

days the music player learns the pattern and automatically it displays us the song which we would like to be played. This is actually a more detailed concept based on machine learning which is called deep learning. Similarly machine learning can be used to predict some useful information about large volumes of data to ensure the occurrence of a particular outcome. We can use machine learning algorithms to predict the result of a student based on some attributes given, we can even predict the most likely occurrences of road accidents in a particular area and so on. In this work we are predicting the heart disease occurrence in a patient based on some important characteristics which are best suited based on our data set that we have collected. The famous decision tree model is being implemented in our project based on the symptoms that are specifically the attributes required to perform the prediction. Using the decision trees we can easily find out which attributes are the best one which will lead us to a better prediction of the datasets. The decision tree is created based on the highest information gain of the attribute and also the Gini index is calculated and based on the best split the attribute with the best split is taken into consideration. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label. We are also implementing the support vector machine to classify the datasets based on the class label. The support vector machine works on kernel and uses hyper plane to group datasets.

## II. LITERATURE SURVEY

Achieves reduction of time complexity by avoiding finding frequent action sets [3]. To obtain near-optimal solutions in a reasonable amount of time [4]. To improve the accuracy of the classification result of heart attack prediction [5]. Gives a brief explanation about different classification algorithm and predict the risk level of a person for heart problem [6]. To remove the use of AI devices for order and expectation of heart disease [7]. On the basis of mining algorithm and the given inputs they are trying to find the issues of prediction of heart disease [8].

## III. PROPOSED WORK

We are using certain PYTHON libraries to carry out all the manipulations in the dataset that we take for our work. We use PANDAS to extract the data frame to carry out the data manipulations there are some attributes in our dataset like age, occurrence of chest pain and its types, bp, cholesterol levels etc.



# Heart Disease Prediction Using Machine Learning Algorithm

We implement the machine learning algorithms like Decision Tree and SVM using the SCIKIT LEARN library in PYTHON. We visualized the outcomes of our implementation using MATPLOTLIB library in PYTHON as well as WEKA. After this we compare between the two algorithms to find out which one is the best based on the accuracy of both.

## IV. IMPLEMENTATION OF MODULES

We are implementing this project using WEKA tool and also using PYTHON programming language. Weka is an inbuilt tool used for machine learning purpose which is created in JAVA environment and consists of some tabs like PREPROCESS, CLASSIFY, CLUSTER etc. These tabs are used to perform the data manipulations on a specific data set. With the help of the Pre-process tab we firstly import the dataset into our WEKA tool. After this we clean the data that is we remove all the data dirty data or inconsistent data and also fill up all the missing values in the data using the filter option. Then by using the classify tab we incorporate the algorithms which we want to apply in our dataset to make the necessary predictions. We can also visualize the class label with the help of a graph. We save the predictive model based on the training dataset that we have used and apply the same on the test data created by us and then perform the prediction. Similarly using PYTHON programming language we are using some specific libraries to carry out the data manipulations. The libraries that we are using in PYTHON are pandas for data pre-processing, Mat plot lib for data visualization and Sci-kit learn for invoking the machine learning algorithm on our dataset. The use of panda's library is to import the data sets and perform all sorts of pre-processing techniques like data cleaning, data mugging, filling in the missing values and so on. Mat plot lib can be used for visualizing the dataset in terms of graphs.

## V. ARCHITECTURE

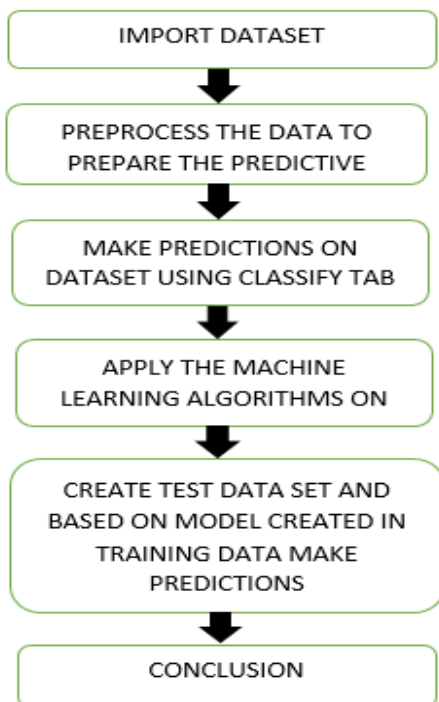


Fig.1. Architecture of the Proposed Model

## VI. RESULTS

We implemented the decision tree algorithm on our dataset and the results that are produced is given in the form of confusion matrix which shows the accuracy of the particular model in the form of true positive and true negative values as given below:

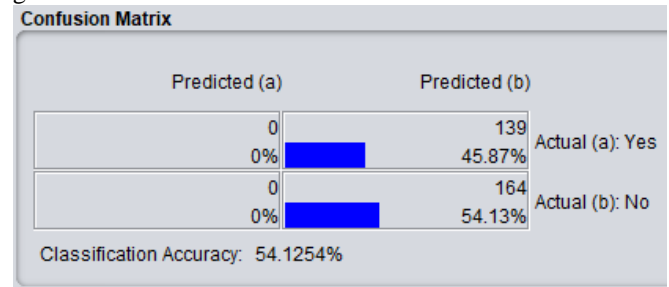


Fig.2. Confusion Matrix of the Proposed Model

The threshold curve for the heart disease data set based on our predictions using the training model is given below:-

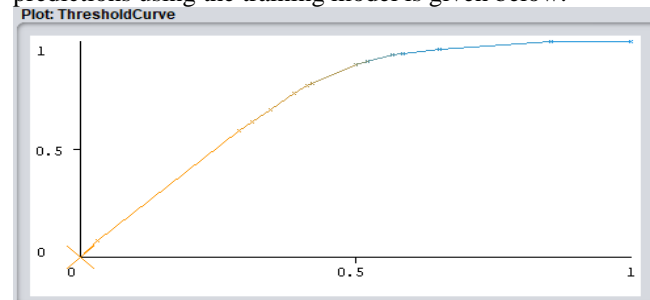


Fig.3. Threshold Curve For Heart Disease Data Set

We implemented both the decision tree algorithm as well as the Support vector machine and compared between both of them in PYTHON programming language also. After comparison we found out that the decision Tree works based for the attributes in our dataset because it shows 100% accuracy whereas the support vector machine shows just 55% accuracy. Thus we deduce that the decision tree algorithm works best for our dataset. We visualized the graph for both of the models along with their accuracy. The figure for the same is given below:-

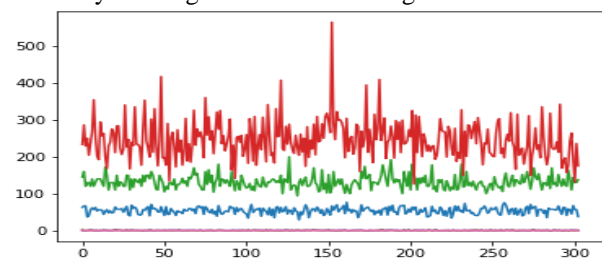


Fig.4. The Normalized Data Set

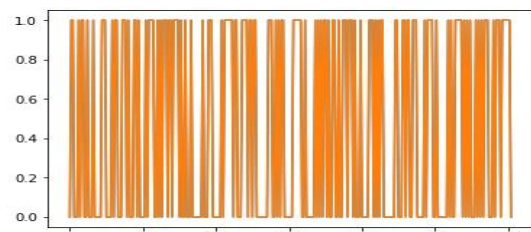
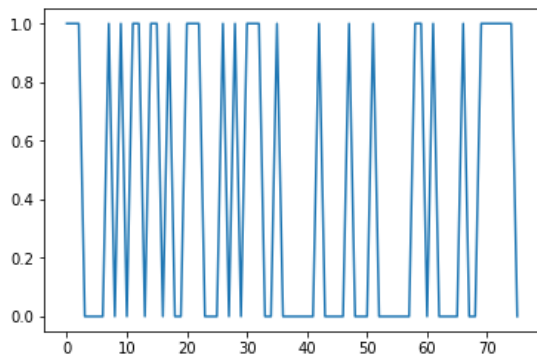
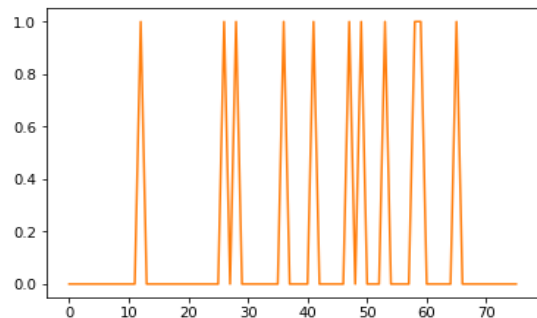


Fig. 5. Data Visualization Of Class Labels



**Fig.6. Data Visualization For Decision Tree On Test Dataset**



**Fig.7. Data Visualization For Support Vector Machine On Test Dataset**

**VII. CONCLUSION**

We can thus conclude that depending on the data set we are using for our work we can incorporate the machine learning algorithms on it to predict the specific outcomes. The decision tree model usually works best for categorical data and especially when the datasets are large. The different attributes in our dataset represents the symptoms most favourable leading to heart disease in patients. Selecting the best attribute becomes a challenge because based on this attribute only we will successfully be able to generate our results. The class label which is categorical in our data set also will be used in the test dataset. In case of scikit learn the nominal values in the class label has to be converted into binary values 1 and 0 so we imported label encoder to perform this operation and then we applied the cross validation to split the original dataset into training set and test set in the ratio of 75% and 25 % by default. Then we applied the decision tree and the support vector machine to compute the result and we found that decision tree worked well as compared to SVM as the accuracy for prediction on the test set in case of decision tree was 100% and for SVM was 55%.

**REFERENCES**

1. M. Amiri and G. Armano, "Early diagnosis of heart disease using classification and regression trees," *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-4. doi: 10.1109/IJCNN.2013.6707080
2. S. Ekiz and P. Erdoğan, "Comparative study of heart disease classification," *2017 Electric Electronics, Computer Science, Biomedical Engineering's Meeting (EBBT)*, Istanbul, 2017, pp. 1-4. doi: 10.1109/EBBT.2017.7956761
3. P. Su, J. Yang, Z. Li and Y. Liu, "Mining Actionable Behavioral Rules Based on Decision Tree Classifier," *2017 13th International Conference on Semantics, Knowledge and Grids (SKG)*, Beijing, 2017, pp. 139-143. doi: 10.1109/SKG.2017.00030

4. D. Bertsimas, J. Dunn and A. Paschalidis, "Regression and classification using optimal decision trees," *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*, Cambridge, MA, 2017, pp. 1-4. doi: 10.1109/URTC.2017.8284195
5. P. Nag, S. Mondal, F. Ahmed, A. More and M. Raihan, "A simple acute myocardial infarction (Heart Attack) prediction system using clinical data and data mining techniques," *2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, 2017, pp. 1-6. doi: 10.1109/ICCITECHN.2017.8281809.
6. J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Nagercoil, 2016, pp. 1-5. doi: 10.1109/ICCPCT.2016.7530265
7. S. Radhimeenakshi, "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 3107-3111.
8. M. Sultana, A. Haider and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, Dhaka, 2016, pp. 1-5. doi: 10.1109/CEEICT.2016.7873142

**AUTHORS PROFILE**



**Dr. M. Praveen Kumar Reddy** is currently working as Assistant Professor in School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India. He obtained his B. Tech in CSE from JNT University, A.P and M.Tech. in C.S.E. from VIT University, Vellore, Tamil Nadu, India. He produced more than 15 international/national publications. Currently, Dr. Praveen Kumar Reddy is working in the area of energy aware applications for Internet of Things (IoT) and high-performance Computing.



**Dr T Sunil Kumar Reddy** received his PhD in Computer Science and Engineering from JNTU Anantapur in 2015. He did his Masters in Information Technology (Networking) from VIT, Vellore in 2007. He is Professor in the Department of Computer Science Engineering, Sri Venkateswara College of Engineering and Technology, Chittoor. His area of interest includes Cloud Computing, Networks, Data Analytics, Internet of Things (IoT), Information Security, Grid Computing. He published more than 20 journal articles in highly reputed journals and conferences.



**Dr S. Balakrishnan** is working as Professor in Computer Science and Business Systems, Sri Krishna college of Engineering and Technology



**Dr. Syed Muzamil Basha**, completed his Ph.D. degree in School of computer science and Engineering from the VIT (Deemed by university), Vellore, India. In 2011. He has 8 years of teaching and research experience. His current research interests include BigData Analytics, Block Chain Management, Internet of Things. He has 30+ publications, which includes 18- research papers (Scopus, SCI) and conferences (International) -7, Book Chapters-6, Letters - 2. He is one of the editor for the text books - 4, Conducted Guest Lecture in various Engineering Colleges in India. He received Award for this contribution towards number of publications in the year 2018 from VIT.



**Ravi Kumar Poluru** received M.Tech. from JNTU Anantapur, Anantapuram in 2014. Currently, he is a Research Scholar in the School of Computer Science & Engineering, Vellore Institute of Technology, Vellore and pursuing his Ph.D. work in the field of Internet of Things. His main areas of research include the Internet of Things, Machine Learning, Wireless Sensor Networks and Nature Inspired Optimization Techniques.

