International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Heart Disease Prediction and Recommendation

Maunish Shah[1], Akshit Modi[2], Jay Jethwa[3], Harsh Bhor[4]

[1,2,3]Student, Dept. of IT Engineering, K. J. Somaiya Institute of Engineering And Information Technology, Sion, Mumbai
[4]Professor, Dept. of IT Engineering, K. J. Somaiya Institute of Engineering and Information Technology, Sion, Mumbai

---

Abstract - *Over the last few decades, heart-related diseases are the main reason for a huge amount of death in the world and has emerged as the most life-threatening disease, not only in India but in the whole world. There's a need for a reliable and accurate system to diagnose and treat such diseases in time for proper treatment. To automate the process of diagnosis and treatment Machine Learning algorithms and techniques is applied to the medical dataset. As technology is advancing, researchers have been using Machine Learning techniques in the diagnosis of heart-related diseases to help the healthcare industry.*

Key Words: cardiovascular, decision tree, Cleveland, Gini, recommendation, Scikit-learn

1. INTRODUCTION

One of the biggest cause of death nowadays is heart disease. Abnormal variations in blood pressure, cholesterol, pulse rate etc. are the major reasons for heart disease. The heart is one of the important functional part of the body and if it gets affected, the functions of the whole body get disturbed. In the medical field, heart disease is one of the major challenges; because a lot of parameters and technicality is involved for accurately predicting this disease. According to the latest survey conducted by WHO, the medical professionals were able to correctly predict only 67% of heart disease, so there is a vast scope of research in the area of predicting heart disease in humans[3]. A huge amount of clinical data is available, where vital information is hidden is seldom visited and remains untapped, researchers use data mining techniques to help health care professionals in the diagnosis of heart disease[7]. In order to predict the risk level of heart disease, the Decision Tree algorithm is being used. The tree takes in 14 attributes as it's input such as age, sex, chest pain type, resting blood pressure, cholesterol in mg/dl, fasting blood sugar, etc. Once the prediction of the risk level is performed, the precautionary suggestions would be provided to the user, which would help them to control their risk level temporarily. "How can we turn data into useful information that can enable users to make effective clinical decisions"[8] This is the main objective of this paper. This paper is dedicated to a wide scope in the field of heart-related disease. Later part of the scope discusses how the user can go about understanding the various parameters related to the prediction and also the various ways through which they can control those parameters.

2. LITERATURE SURVEY

The existing papers that are surveyed have their focus on the prediction of heart disease. The processed dataset that is collected from various sources is used to train the algorithm and are then used for prediction.

T. Princy, J. Thomas[1] intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analyzed on medical data sets using algorithms. While Priyanga and Naveen[2] mainly focus on to create a decision support system using Naive Bayes algorithm for predicting heart disease. A web application is created to get user input and the application can retrieve hidden knowledge related to heart disease from a historical database (Cleveland dataset). M. Gandhi, S. Singh[4] focusesmon classification methods of data mining used in data discovery. Different classification techniques of data mining have merits and demerits for data classification and knowledge extraction. Purushottam, K. Saxena and Richa Sharma[8] depicted the extraction of risk level from the heart disease database. The input database contains the screening of clinical data of heart patients. Initially, the database is pre-processed to make the mining process more efficient. Decision tree is used for the prediction. The papers majorly focus on finding the efficiency of various algorithms as the predictions in the domain of medical has to be accurate.

3. PROPOSED SYSTEM

The heart report when generated has terms that are only understood by the doctors and according to that, they provide the necessary medications. The aim is to provide an Android application to the users, wherein they can provide the application the above-mentioned attributes which would then be used to predict the risk level of having a heart disease. The users are then provided with an in-depth explanation of all the terms that are responsible for the risk level generated in the prediction and are recommended with various ways to reduce their risk level.

3.1 ALGORITHM USED

Decision Tree - Decision Trees being a non-parametric supervised learning method they are used for classification. The goal is to create a model that predicts the value of a target variable (risk level) by learning simple decision rules inferred from the data features (Dataset). The major challenge in the decision tree is to identify the attribute for the root node in each level. This process is known as attribute selection. One of the main advantages of using decision tree was that it is able to handle both numerical and categorical data. Other techniques are usually specialized in analyzing datasets that have only one type of variable.

The decision tree learning algorithm learns recursively as follows:

Steps:

1.Compute the Gini index for data-set

2.For every attribute/feature:

1.Calculate Gini index for all categorical values

2.Take average information entropy for the current attribute

3.Calculate the Gini gain

3. Pick the best Gini gain attribute.

4. Repeat until we get the desired tree.

Machine learning library Scikit-learn is used for prediction. Scikit-learn has a range of supervised and unsupervised learning algorithms provided through a consistent interface in Python. The sklearn.tree module includes decision tree-based models for classification.

class sklearn.tree.DecisionTreeClassifier(criterion='gini',splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None)

The method uses gini as the splitting criterion for attribute selection. Gini provides a statistical measure of the degree of variation or inequality represented in the dataset. Gini index - Given a training set S, the target attribute takes on k different values (i.e. classes), the Gini index of S is defined as

$$G(S) = \sum_{i=1}^{n} P(i) * (1 - P(i))$$

Where P(i) is the probability of a certain classification i, per the training data set.

3.2 TECHNOLOGY STACK

A. Android application

An Android app is a software application which runs on Android platform. Because the Android platform is built for mobile devices, a typical Android app is designed for a smartphone or a tablet PC running on the Android OS.

The main advantages of building an android application that it is easier to access and also the android phones are much cheaper than other alternatives. The target audience is also comparatively very high and hence it makes a suitable platform for the idea being suggested here.

B. Python

Python being an object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, makes it a very attractive language for machine learning applications.

Many libraries are available to perform data analysis and mining on large sets of medical datasets:

NumPy is important to perform scientific computing with Python. It is used to operate on multi-dimensional arrays and matrices.

SciPy works in association with NumPy arrays and offers effective routines for numerical integration and up-gradation.

Pandas, also developed on top of NumPy, is used to manage and manipulate the CSV (Comma Separated Value) files generated from the android application.

Developed on NumPy, SciPy. Scikit-learn acts as a machine learning library that leads to the generation of the decision tree used in predictions.

C. Flask

Flask is used for connecting the backend server with the frontend android application.

Flask is a microweb framework written in Python that does not require particular tools or libraries. Flask can be used for RESTful request dispatching in Python. A RESTful architecture uses HTTP coding for much of its functionality. A REST API defines a set of functions which developers can perform like requests and receive responses via HTTP protocol such as GET and POST. The data can be passed to the server with the help of URL, with which the server would fetch the data and the server can send a response back for the request.

4. SYSTEM DESIGN

The system design includes the overall structure of the system including its components, flow of data and also it explains how the input is accepted from and the output is presented to the user.

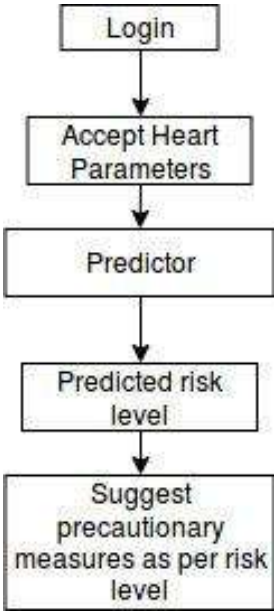

Fig -1: System flow-diagram

As shown in the above figure, the user would first login into the system, verifying which the system would redirect the user to the home page through which the he/she could navigate the app. The user will enter the various parameters which would then be sent to the server. According to the input data entered by the user, the algorithm will predict the risk level of heart disease. The details entered by the user would then serve as an input to the 'Predictor'. The Predictor would then predict the risk level for the user. The risk levels would range from 0 to 1. Along with the prediction, the application will also recommend various control/precautionary measures based on the level of risk the user is present in. Recommendation means the measures provided by a system or any individual entity to provide the best course of action in order to improve any deteriorated condition.

Recommendation will provide user the necessary control measures or recommendations based on the risk level of heart disease. The control measures provided by our system will act as a temporary solution. These measures will temporarily help the user to bring down their heart parameters which surpassed the threshold level to normal or at least reduce the risk of facing the heart disease.

5. DATASET DESCRIPTION

Cleveland dataset: Robert Detrano, M.D., Ph.D., collected this data at V.A. Medical Centre. All published experiments related to using a subset of 14 of the 76 attributes present in the processed Cleveland heart disease database.

14 attributes are used:

Table -1: Attribute Description

No.	Clinical Features	Description
1	Age	Instance age in years
2	Sex	Instance gender
3	Cp	Chest pain type
4	Trestbp (mmHg)	Resting blood pressure
5	Chol (mg/dl)	Serum cholesterol
6	Fbs	Fasting blood sugar
7	Restecg	Resting electrocardiographic results
8	Thalach	Maximum heart rate achieved
9	Exang	Exercised induced angina
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	The slope of the peak exercise ST segment
12	Ca	Number of major vessels (0-3) coloured by fluoroscopy
13	Thal	3= normal; 6= fixed defect; 7= reversible defect
14	Num	Diagnosis of heart disease

Table -2: Result of Prediction

Diagnosis	Value
No risk of Heart Disease	0
Risk of heart disease	1

Table 2 depicts the result of the system i.e. whether the user has a risk of heart disease or not i.e. 0 represents no risk and 1 represents the presence of risk. This diagnosis (0 or 1) is what the decision tree predicts based upon the values of input parameters as shown in Table 1.

6. CONCLUSION

To make effective clinical decisions in the medical field plays a vital role as there is a need for a reliable and accurate system. The paper studies how the machine learning and decision tree algorithm works can be used for predicting the heart disease by taking into consideration certain parameters such as thal, cholesterol, heart rate, age, gender. Using these parameters, the algorithm predicts the risk level of heart disease between the range 0 - 1 where 0 depicts no risk and 1 depicts the presence of risk. Based on the risk level predicted, the user is suggested some precautionary measures in order to lower the risk and stay fit.

ACKNOWLEDGEMENT

We are grateful to K J Somaiya Institute of Engineering and IT and express our thanks to Principal, Dr. Suresh Ukarande for extending his support. We are highly indebted to Prof. Harsh Bhor for his guidance and constant supervision as well as for providing necessary information.

REFERENCES

[1] T. Princy, J. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques," International Conference on Circuit, Power and Computing Technologies [ICCPCT], 5090-1277, 2016.

[2] Priyanga, Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naive Bayes Weighted Approach (NBwa)," Asia Modelling Symposium (AMS), 2376-1172, 2017.

[3] H. Sharma, M. Rizvi, "Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms", 2319-7242, International Journal Of Engineering And Computer Science [ISSN], pp. 21623-21631, June 2017.

[4] M. Gandhi, S. Singh, "Predictions in Heart Disease using Techniques of Data Mining," 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management, 4799-8433, pp. 520-525, June 2015.

[5] M. Jabbar, S. Samreen, "Heart disease prediction system based on hidden naïve bayes classifier," International Conference on Circuits, Controls, Communications and Computing (I4C), 1721-4328, October 2016.

[6] H. Sharma, M. Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey," International Journal on Recent and Innovation Trends in Computing and Communication [IJRITCC], 2321-8169, August 2017.

[7] M. Kirmani, S. Ansarullah, "Prediction of Heart Disease using Decision Tree a Data Mining Technique", 2277-5420, International Journal of Computer Science and Network [IJCSN], December 2016.

[8] Purushottam, K. Saxena, R. Sharma, "Efficient Heart Disease Prediction System using Decision Tree", 4799-8890, International Conference on Computing, Communication and Automation (ICCCA), pp. 72-77, July 2015.

[9] V.V. Ramalingam, A. Dandapath, M. Raja, "Heart disease prediction using machine learning techniques: a survey," International Journal of Engineering & Technology [IJET], pp. 684-687, 2018.

[10] A. Pandey, P. Pandey, K. Jaiswal, A. Sen, "A Heart Disease Prediction Model using Decision Tree," IOSR Journal of Computer Engineering [IOSR-JCE], e-ISSN: 2278-0661, p-ISSN: 2278-8727, pp.83-86, July-August 2013.

[11] A. Mahmood1, M. Kuppa, "Early detection of clinical parameters in heart disease by improved decision tree algorithm," Second Vaagdevi International Conference on Information Technology for Real World Problems, 2010.

[12] J. Yang, J. Kim, Un-Gu Kang, Y. Lee, "Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS–LDA)," Pers Ubiquit Comput (2014), pp. 1351–1362, 2013.

[13] T. Tang, G. Zheng, Y. Huang, G. Shu, P. Wang, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis," IEMS, Vol. 4, No. 1, pp. 102-108, June 2005.

[14] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," IEEE/ACS Int. Conf. Comput. Syst. Appl., pp. 108–115, 2008.

[15] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," Neural Comput. Appl., vol. 23, no. 7–8, pp. 2387–2403, Dec. 2013.

[16] C. Dangare, S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," Volume 47– No.10, June 2012.