International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

Heart disease prediction using Naïve Bayes

Garima Singh[1], Kiran Bagwe[2], Shivani Shanbhag[3], Shraddha Singh[4], Sulochana Devi[5]

[1,2,3,4]student, IT, Xavier institute of engineering, Maharashtra, India
[5]Professor, IT, Xavier institute of engineering, Maharashtra, India

---

Abstract - *It might have happened such a lot of times that you just or somebody yours would like doctors facilitate right away, however they're not obtainable thanks to some reason. The heart disease Prediction application is a user support and on-line consultation project. Here, we tend to propose a web application that enables users to induce instant steerage on their cardiopathy through an intelligent system online. The application is fed with numerous details and also the heart disease related to those details. The application permits user to share their heart connected problems. It then processes user specific details to see for numerous health problem that might be related to it. Here we tend to use some intelligent data processing techniques to guess the foremost correct illness that might be related to patient's details. Supported result, the will contact doctor consequently for any treatment. The system permits user to look at doctor's details too. The system may be used without charge heart disease consulting on-line.*

Key Words: Heart, Disease, Prediction, Application, Illness, Intelligent, Data, Processing, System, Technique, Naïve Bayes" Algorithm.

1.INTRODUCTION

A major challenge facing health care organizations (hospitals, medical centers) is that the provision of quality services at reasonable prices. Quality service implies diagnosing patients properly and administering treatments that are effective. Poor clinical selections will result in fateful consequences that are so unacceptable. Hospitals should additionally reduce the price of clinical tests. They'll come through these results by using applicable computer-based info and/or call support systems. Most hospitals today use some form of hospital information systems to manage their health care or patient knowledge. These systems usually generate huge amounts of knowledge that take the shape of numbers, text, charts and pictures. Sadly, these knowledges are rarely accustomed support clinical decision-making. There's a wealth of hidden information in these knowledges that's mostly untapped. This raises a crucial question: "How will we tend to flip knowledge into helpful info that may change health care practitioners to create intelligent clinical decisions?"

Although data mining has been around for over 20 years, its potential is barely being accomplished now. Data mining combines applied mathematics analysis, machine learning and information technology to extract hidden patterns and relationships from giant databases.

Naive Bayes or Bayes' Rule is that the basis for several machine-learning and data processing ways. The rule (algorithm) is employed to make models with predictive capabilities. It provides new ways that of exploring and understanding knowledge. It learns from the "evidence" by calculating the correlation between the target (i.e., dependent) and different (i.e., independent) variables

1.1 Objective

The main aim of this analysis is to develop a prototype Health Care Prediction System using, Naive Bayes. The System will discover and extract hidden data related to diseases (heart attack, cancer and diabetes) from a historical heart disease database. It will answer complicated queries for diagnosing sickness and so assist care practitioners to form intelligent clinical selections which ancient call support systems cannot. By providing effective treatments, it conjointly helps to reduce treatment prices. To reinforce visualization and easy interpretation, it displays the results in tabular and PDF forms.

1.2 Scope

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome [1].

The application is fed with varied details and therefore the cardiovascular disease related to those details. The application permits user to share their heart connected problems.

It then processes user specific details to ascertain for varied illness that might be related to it. Here we tend to use some intelligent data mining techniques to guess the foremost correct illness that might be related to patient's details.

Based on result, system automatically shows the result specific doctors for more treatment. The system permits user to look at doctor's details. The system can be use in case of emergency.

2. DATA SOURCE

Questionnaires have advantages over some other types of medical symptoms that they are cheap, do not require as much effort from the questioner as verbal or telephone surveys, and often have standardized answers that make it

simple to compile data [2]. With the help of the dataset, the patterns important to the heart attack prediction are extracted. The records were split equally into 2 datasets: training dataset and testing dataset. To avoid bias, the records for every set were selected indiscriminately

Table -1: Input Attributes

srno	attribtes	Values
1	Age	Patient's age
2	Gender	Male, female
3	CP	(chest pain)
4	Trest bps	resting blood pressure
5	Cholesterol	
6	FBS	fasting blood sugar>120? yes=1, no = 0
7	Rest ECG	resting electrocardiographic results 0,1,2
8	Thalach	maximum heart rate achieved
9	Exang : exercise induced angina	1= yes; 0= no
10	Old peak	ST depression induced by exercise relative to rest
11	Slope	the slope of the peak exercise ST segment
12	CA	:no. of major vessels (0 to 3) colored fluoroscopy
13	Thal	3 =normal ,6=fixed defect ,7= reversible defect
14	Diagnosis of heart disease	output

3. IMPLEMENTATION OF NAÏVE BAYES

In probability theory, Bayes' theorem (often called Bayes' law after Thomas Bayes) relates the conditional and marginal probabilities of two random events. It is often used to compute posterior probabilities given observations [3]. For example, a patient could also be ascertained to possess certain symptoms. Bayes' theorem is often used to compute the likelihood that a projected diagnosing is correct, as long as observation.

3.1 Implementation of Naïve Bayes

A Naive Bayes' classifier may be a term addressing a simple probabilistic classification supported applying Bayes' theorem. In easy terms, a Naïve Bayes classifier assumes that the presence (or absence) of a specific feature of a category is unrelated to the presence (or absence) of the other feature. As an example, a fruit could also be thought of to be an apple if it's red, round, and regarding 4" in diameter. Even supposing these options rely on the existence of the opposite options, a Naive Bayes' classifier considers all of those properties to independently contribute to the likelihood that this fruit is an apple.

Naive Bayes algorithm is based on Bayesian Theorem.

3.2 Bayesian theorem

Given training data X, posterior probability of a hypothesis H, P(H|X), follows the Bayes theorem

P(H|X)=P(X|H)P(H)/P(X) (1.1)

Algorithm

The Naive Bayes algorithm is based on Bayesian theorem as given by equation (1.1)

Steps in algorithm are as follows:

1. Each data sample is represented by an n dimensional feature vector, X = (x1, x2….. xn), depicting n measurements made on the sample from n attributes, respectively A1, A2, An.

2. Suppose that there are m classes, C1, C2……Cm. Given an unknown data sample, X (i.e., having no class label), the classifier will predict that X belongs to the class having the highest posterior probability, conditioned if and only if:

P(Ci/X)>P(Cj/X) for all 1< = j< = m and j!= i

Thus we maximize P(Ci|X). The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes theorem,

3. As P(X) is constant for all classes, only P(X|Ci)P(Ci) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e. P(C1) = P(C2) = …..= P(Cm), and we would therefore maximize P(X|Ci). Otherwise, we maximize P(X|Ci)P(Ci).

Note that the class prior probabilities may be estimated by P(Ci) = si/s , where Si is the number of training samples of class Ci, and s is the total number of training samples. on X. That is, the naive probability assigns an unknown sample X to the class Ci [4]

4. SYSTEM ARCHITECTURE

Considering the anomalies in the existing system computerization of the whole activity is being suggested after initial analysis. It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason.

Here, we propose a web application that allows users to get instant guidance on their heart disease through an intelligent system online. The application is fed with various details and the heart disease associated with those details. The application allows user to share their heart related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details.

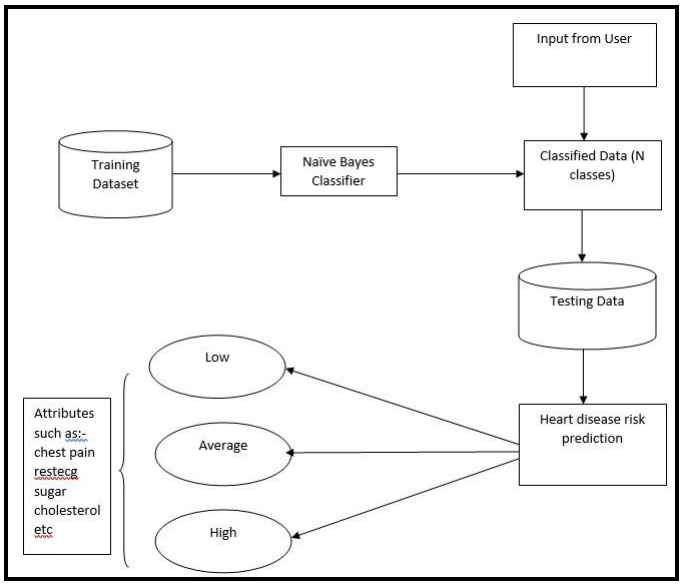

Fig -1: System architecture

Based on result, system automatically shows the result specific doctors for further treatment. The system allows user to view doctor's details. The system can be use in case of emergency. The main goal of this system is to predict heart disease using data mining technique such as Naive Bayesian Algorithm. Raw hospital data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as Naïve Bayes algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and the user is given the result based on the prediction whether the risk of heart disease is low, average or high.

5. CONCLUSIONS

The Heart Disease Prediction, historically viewed as a necessary burden in medical offices, healthcare facilities and wellness centers, can be completely automated through an inefficient online software program. The benefits of implementing this technology touch everyone involved in the scheduling process, as administrators and users can conduct their tasks more efficiently and accurately. The system extracts hidden knowledge from a historical heart disease database. This system can be further enhanced and expanded for many more disease prediction.

Heart disease prediction can be additional increased and expanded. For instance, it can incorporate different medical attributes besides the 14 listed It may also incorporate different data processing techniques, e.g., time series, clustering and association Rules. Continuous information may also be used rather than simply categorical information. Another space is to use Text Mining to mine the vast quantity of unstructured information offered in healthcare databases. Another challenge would be to integrate data mining and text mining

REFERENCES

[1] Obenshain, M.K," Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[2] Sellappan Palaniappan, Rafiah Awang ,Intelligent Heart Disease Prediction System Using

Data Mining ""978-1-4244-1968-5/08/ ©2008 IEEE

[3] Chapman, P., Clinton, J., Kerber, R. Khabeza, T.,

Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0:

Step by step data mining guide", SPSS, 1-78, 2000.

[4] N. Aaditya Sunder, P. PushpaLatha, "Performance analysis of classification data mining techniques over heart disease

database" Inernational Journal Of Engineering Science and Advance Technology"-vol-2 issue-3,470-478,May-June 2012.

[5] Shadab Adam Pattekari and Asma Parveen" PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES" International Journal of Advanced Computer and Mathematical Sciences

ISSN 2230-9624. Vol 3, Issue 3, 2012.