Predicting cardiovascular disease

ASSIGN BUSTER

Author: Bhuiyan, Md Mahbub Islam

Proper Diagnosis Prediction of Cardiovascular Diseases Using Naive Bayesian Algorithm

Abstract

Data can play a very important role to predict different types of diseases. Health care industries collects huge amount of data from different resources, but the unfortunate thing is that these data are not mined properly to find out hidden information for effective decision making or prediction. Mining data from different variables of attributes by their patterns can be used as the inputs of any decision technique. Here in this paper the prediction of cardiovascular disease will be done by Naive Bayes method where patient's medical history like age, sex, cholesterol, angina blood pressure and blood sugar will be used as input.

Keywords: heart, predict, naive bayes , data mining.

Introduction

Cardiovascular disease generally refers to protuberance that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's brawn, valves or rhythm, also are reified forms of heart disease. A manifestation of heart disease varies depending on the particular type of heart diseases. World Health Organization in the year 2012 reported that 11. 8% of total global deaths are due to Cardio Vascular Disease (CVD). There is sometimes silent massive heart attack which caused premature death as a

result proper cardiovascular disease prediction is so much necessary so that proper diagnosis can be done immediately.

Different researchers use different techniques to predict Cardio Vascular Diseases where they used Genetic Algorithmcite (Shruti, M. Akhil, Latha), Fuzzy Logiccite (Shruti), UCI repositorycite (Mary), Naive Bayescite (Mary, Shamsher, Subbalakshmi, Sellappan), Decision Treecite (Mary, Shamsher, ANBARASI, Sellappan), Linear regressioncite (Mary), Association rulecite (Mary), Neural networkscite (Mary, Sellappan), Z-Statisticscite (M. Akhil), Apriori Algorithmcite (Shruti), Data miningcite (Mary), Clustering data miningcite (Shamsher), Canfiscite (Latha), J48 Algorithmcite (Hlaudi), Reptree Algorithmcite (Hlaudi), Cart Algorithmcite (Hlaudi) etc. (Subbalakshmi) proposed Naive Bayes which was less efficient to predict accurately and was improved in (Mary) where they used UCI repository, Naive Bayes and Neural Networks to predict more accurate and consistence result.

CVD is expected to be the leading cause for deaths in developing countries likes Bangladesh due to change in life style, work culture and food habits. Hence, more measured and streamlined methods of cardiac diseases and periodic examination are of high importance. My proposal is the improvement of the prediction consistency by Naive Bayescite (Mary, Shamsher, Subbalakshmi, Sellappan) to obtain proper prediction of the diagnosis of cardiovascular diseases.

Research Objectives

The main objective of this research is to develop a Cardiovascular Disease Prediction System using Naive Bayes to increase prediction accuracy which can find out and acquire hidden patterns and relationships allied with heart disease from a historical heart disease database. By providing actual prediction, it also helps to reduce treatment costs which were unnecessarily spent by the patients for several tests. This system will transform data into useful information which will be a huge asset for the healthcare practitioners to make sensible clinical decisions.

Literature Review

Different types of techniques and methods are used by different researchers to predict cardiovascular diseases properly. Large number of work is carried out in finding out efficient methods of medical diagnosis for various diseases like Cancer, Diabetes, and Heart Disease (Mary). Some methods worked well and some methods were failed to fulfill the desired result of prediction. Few methods and techniques were efficient and few were less efficient. Efficient association classification for heart disease prediction using Gini index to produce a compact rule set and filter rules further by applying Z-Statistics and Genetic Algorithm was proposed by M. Akhil. Shruti proposed An Intelligent Heart Disease Decision Support System based on Apriori Algorithm, Genetic Algorithm and Fuzzy Logic. Several data mining techniques such as Decision Trees, Naive Bayes, Neural Network, Association Rule, and Linear Regression were proposed by Mary. A prototype Intelligent Heart Disease Foretelling System with Canfis and genetic algorithm using diachronic heart disease databases to make intelligent clinical decisions was proposed by (Latha). Different classification algorithms like J48 Algorithm,

Reptree Algorithm were proposed by Hlaudi. Comparison between the function techniques in data mining of Naive Bayes, Decision Tree and Classification by Clustering was proposed by Shamsher. DSHDPS can be served as a standardization tool to train nurses and medical students to diagnose patients with heart disease (Subbalakshmi). Naive Bayes performed with good promulgation probability of 96. 6% (ANBARASI).

Proposed Work

Proper Cardiac disease diagnosis is a fundamental and a tedious work to perform. There are various ways to diagnosis a disease. In the proposed system cardiac disease prediction is done by extracting the data from different data repository and mining it. Standard of measurement elegance is obtained after mining the datasets . Mining of the dataset is done using Naive Bayes algorithm. The results obtained from the mining is combined together to obtain the optimal result.

Data Analysis

In health care industries datasets contain huge amount of information about the patients and as well as their medical history.

Equations

Now if we take each of the attributes as cause and cardiac diseases as effect, then according to Naive Bayes Theorem we can do the prediction by the following formula and the proposed algorithm given below:

Formula,

begin (equation)

P (Cause| Effect) =

frac (substack ((P (Effect| Cause)*P (Cause))))

(substack (P (Effect)))

end (equation)

Naive Bayes classification algorithm works as Follows:

Let consider, Cause= C, Effect= E, Attribute= A, N-dimensional attribute vector = V.

1. Let M be a training set of tuples and their associated class labels. Each tuple is represented by

begin (equation*)

V= (v_1, v_2, v_3, ……., v_n)

end (equation*)

Depicting n-measurements made on the tuple from n-attributes, respectively

A1 , A2 , A3 , ……, An.

2. In order to calculate P (V/Ci) , the Naive Prediction of class contingent independence is made. There are no mutual relationships amongst the attributes.

Thus,

begin (equation*)

P (V/Ci) = P(v1/Ci)*P(v2/Ci)*…..*P(vm/Ci)

end (equation*)

3. In order to predict the class level of T, P (V/Ci)P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of tuple V is the class Ci iff P(V/Ci)P(Ci) > P(V/Ci)P(Ci)

for 1 <= j <= m , j!= i. The predicted class label of tuple V is the class Ci for which P(V/Ci)P(Ci) is the

Maximum.

Cardiac Disease Dataset

The term Cardiac disease attach to diverse diseases that cramp the heart. Cardiac disease is denoted as the leading cause of several health issues around world. Sets of medical attribute are gathered and with the help of it correlate significant to the cardiac attack prediction are acquired.

Different types of stored records n database are analyzed to find out similar pattern for the prediction.

Ten attributes were concerned in predicting the cardiac disease. The core attributes for predicting the cardiac disease are described below.

Algorithm Explanation

Begin

1. Read Patient empirical data from Data Base.

2. Calculate naive bayes output for each individual attribute.

3. Calculate probability of having disease.

4. If prediction result is satisfied Calculate Risk.

5. If prediction result is not satisfied go to step 2.

6. Repeat the process until proper prediction achieved.

7. If Output best then stop and display prediction result.

Pattern Materialization

This supplement results to the formation of a pattern based on the implementation of the algorithm. The algorithm will create a pattern based upon the different data sets. Pattern formation is the process of obtaining certain empirical data values based upon the implementation of the algorithm.

Conclusion and Future Work

In this Paper Decision Support System for the prediction of cardiovascular diseases is developed using NaÃ¨ve Bayesian Classification technique. The containment acquires hidden knowledge from a historical heart disease database. The model is very competent to predict patients with cardiac disease. This model works consistently before and after alleviation of attributes with the same model construction time. This model can be made more advanced by using other classification techniques like neural network

and genetic algorithm by which the attributes could be decreased to make the prediction. Incessant data can be used rather than just categorical data.