A Heart Disease Prediction System using Artificial Neural Network and Naive Bayes

Ashish Bhatia[1] Mohit Matwani[2] Pawan Karira[3] Rahul Sidhwani[4] Asha Bharambe[5]
[1,2,3,4]Student [5]Professor
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]V.E.S. Institute of Technology, Chembur, Mumbai

Abstract— In the world of medical science, decision are made based on the prior available data and the experience of the doctors only. Due to this there are more chances of errors, longer time in diagnosis, increases cost as well. Hospital are unaware of the rich hidden patterns in the patient database which can be used to predict a disease using data mining techniques. In this paper we designed a system which will work on the same database and predict heart disease. It will compare artificial neural network algorithm and Naïve Bayes algorithm for prediction. Using important factors such as age, sex, cholesterol etc. it can predict the probability of person getting or prone to heart disease. The best of the above algorithms is used for implementation.

Key words: Data Mining, Artificial Neural Network, Heart Disease

I. INTRODUCTION

Today Hospitals store patient records so that when patients comes back for treatment hospital staff should know that from which disease patient was suffering previously and how their treatment should be taken forward. These stored patient records can be mined using data mining techniques for prediction of a disease. Today chance of having a heart disease is more due various factors and prediction of heart disease will help hospital staff in taking decisions in case of emergency.

Nowadays, in the world Heart disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. In 2008, 17.3 million people died due to Heart Disease. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease.

Data mining is a knowledge based technique used to analyses data and use it in more efficient form. Our paper intends to predict the heart disease based on patient dataset.

Hence Data mining can be an effective tool to assist the hospital staff in predicting the possibility of heart disease depending on patient's data.

In rural areas many hospital cannot use latest technology assistance. This assistance, with the help of hospital staff can be provided by a disease prediction system which will let the user know whether the person is prone to heart disease or not.

II. MOTIVATION

A major challenge faced by health care organization is the provision of quality services at affordable cost and on time. Poor clinical decisions can lead to disastrous consequences which are not acceptable. Irrelevant tests will increase cost of treatment. They also increase workload for hospital. They can achieve this results by adopting appropriate computer based or decision support system. These systems should provide necessary health assistance

Current database system in hospital contains huge amount of data which is rarely used. It contains wealth of hidden information which can be used for decision making. We can turn this data into a useful information that can be used to make intelligent decision making system which can predict the disease and by generating reports.

Currently available such systems will help a common man to predict whether he or she is prone to a heart disease or not. Sometimes this sensitive result can make him or her undergo stress. Hence user should not get access to such data. Hospital should be involved in such practices.

III. PROBLEM STATEMENT

Now a day hospital store database for the purpose of billing, Inventory management and generation of statics.

Many hospitals are still unaware of the rich hidden patterns in database this data can also be effectively used for prediction of a disease. This will ensure patient safety, decrease unwanted medical practices and improve output.

The clinical features like age, cholesterol etc of a particular patient should be effectively used to predict the existence of heart Diseases.

IV. DATA SOURCE

A. Creators of the used Dataset:

V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D

B. Donor:

Purushottam (puru.mit2002@gmail.com)

The Dataset which we obtained consists of 303 records. This dataset consist of 14 attributes. These 14 attributes are

1) Gender(1=Male,0=Female)
2) Age
3) Chestpain type
   - Value 1: typical angina
   - Value 2: atypical angina
   - Value 3: non-anginal pain
   - Value 4: asymptomatic
4) Maximum Heart rate thalach
5) Cholestrol
6) trestbps:
   - Resting blood pressure (in mmHg on admission to the hospital)
7) Fasting Blood sugar
8) restecg: resting electrocardiographic results
   - Value 0: normal

– Value 1:having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
– Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

9) exang: exercise induced angina
10) oldpeak = ST depression induced by exercise relative to rest
11) slope: the slope of the peak exercise ST segment
Value 1: upsloping
Value 2: flat
Value 3: downsloping
12) ca: number of major vessels (0-3) colored by flourosopy.
13) thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14) num: The num attribute specifies that whether the patient is having heart disease or not. If num=0 indicate absence of heart disease and num=1,2,3 indicate the patient is suffering from heart disease.

V. METHODOLOGY

We have developed a system using machine learning algorithms like ANN (Artificial Neural Network) and Naïve Bayes, and we have compared their efficiencies.

A. Artificial Neural Network:

In machine learning and cognitive science, artificial neural networks (ANNs) are a family of statistical learning models inspired by biological neural networks (the central nervous systems of humans brain) and are used to estimate or approximate functions that can depend on a large number of inputs. Artificial neural networks are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making training, testing adaptive to inputs and capable of learning.

The word network in the term artificial neural network refers to the interconnection between the neurons in the different layers of system.

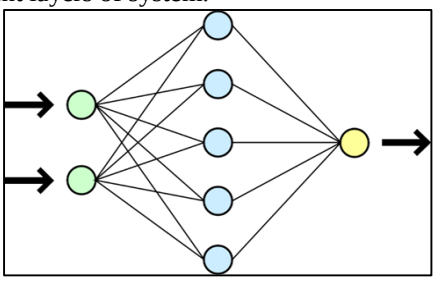

Fig. 1: Artificial Neural Network

The network has three layers namely,

1) Input layer:

Input layer consists of 13 neurons (node) which are 13 attribute of a patient.

2) Hidden layer:

This layer can contain multiple sub-layers. So the number of nodes in the hidden layer may varies. Which are used for computational purposes.

3) Output layer:

Output layer have one node which is the result of the prediction.

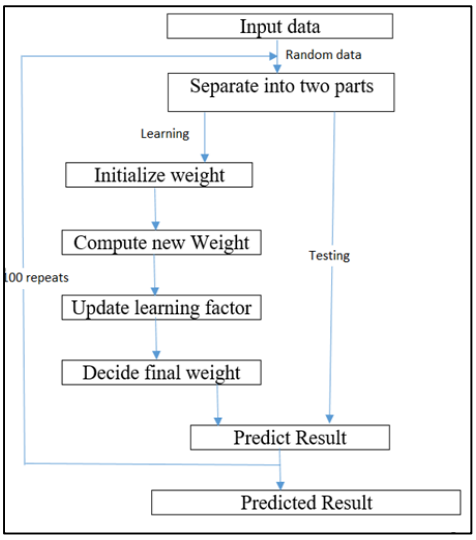

Fig. 2: Training Neural Network

B. Naive Bayes:

Naïve Bayes is a classic Bayesian theorem for classification with strong independence assumption. It gives best results when the input is multi-dimensional. It gives a probability that a particular class of a tuple can be predicted using values entered.

This method was proposed by Thomas Bayes which is given by following formulae:

$$p(h|e)=[p(e \mid h)*p(h)]/p(e)$$

The Hypothesis or eventh is calculated based on the evidences (e) that are observed.

Model proposed by naïve bayes uses conditional probability, the test case or a tuple which is to be classified is represented by a vector x(x1,x2,…xn). The different values of x represents n independent features, for k possible classes or outcomes,

$$p(ck| x1,x2,…xn).$$

If number of features n increases or n take large number of values it classification becomes difficult, therefore the above formula is remodel to,

$$P(ck|x) = P(ck)*P(x|ck)/P(x)$$

Where,
P(ck|x) is posterior probability
P(ck) is prior probability
P(x|ck) is likelihood
And P(x) is evidence.

VI. BLOCK DIAGRAM OF THE SYSTEM

The proposed System is a Web based System. All interaction are held through a web browser.

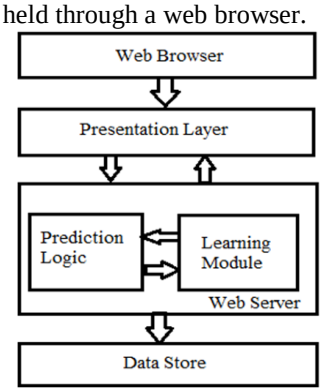

Fig. 3: Block Diagram

VII. RESULTS

A test dataset is used to calculate the efficiency of the system.

The Graph below shows the accuracy of system with different algorithms.

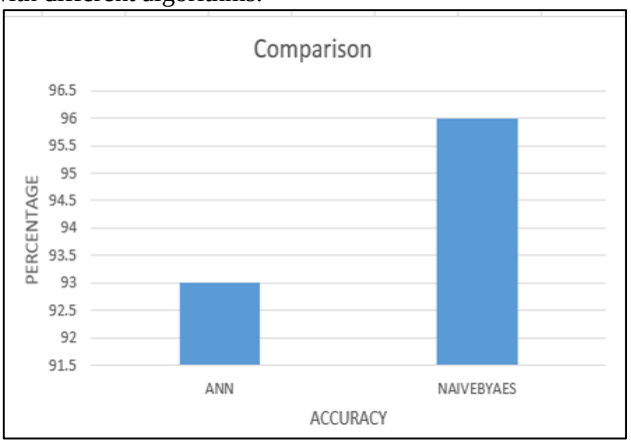

Fig. 4: Comparison of algorithms

VIII. CONCLUSION

We have developed a heart disease prediction system. Which accepts 13 clinical features of a patient and predicts the existence of heart disease using Naïve Bayes Classifier algorithm with a System accuracy of 96% for our heart disease dataset.

Our system will provide following features

1) Provide faster results
2) Access to system from any place
3) Reliability, since the system has very high accuracy on standard dataset
4) Availability, system will be available 24*7 to its users

REFERENCES

[1] HDPS: Heart disease prediction system. Chen, A. Huang, S.Y.;Hong, P.S.; Cheng, C.H.; Lin, E.J. Computing in Cardiology, 2011
[2] Genetic neural network based data mining in prediction of heart disease using risk factors Amin, S.U.; Agarwal, K.; Beg, R.
[3] Efficient heart disease prediction system using decision tree Purushottam; Saxsena, K.; Sharma, R.
[4] INTELLIGENT HEART DISEASE PREDICTION SYSTEM USING DATA MINING TECHNIQUES AWANG, R.COMPUTER SYSTEMS AND APPLICATIONS, 2008.
[5] Prediction of heart disease using learning vector quantization algorithm Sonawane, J.S.; Patil, D.R. IT in Business, Industry and Government (CSIBIG), 2014
[6] Prediction of heart disease using multilayer perceptron neural network Sonawane, J.S.; Patil, D.R. Information Communication and Embedded Systems (ICICES), 2014

