

# Heart Disease Prediction Using Data Mining

Ranjana Joshi<sup>1</sup>, Dr. Shilpa Sethi<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Application, YMCA University of Science & Technology, Faridabad, India

<sup>2</sup>Assistant Prof, Department of Computer Application, YMCA University of Science & Technology, Faridabad, India

**Abstract:** *The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The Healthcare industry is generally "information rich", but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. This research paper intends to use data mining Classification Modeling Technique, namely, Naïve Bayes in Heart Disease Prediction.*

**Keywords:** Naive Bayes, Technique, Data mining, Heart disease.

## 1. Introduction

Extraction of useful information is known as data mining. Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Nowadays to live luxurious life people work like machine in order to earn lot of money hence they forget to take care of their health. Because of this, there is change in the food which they consume, their lifestyle changes. They are more tensed under the pressure of earning more money. This leads to various disease like blood pressure, diabetes, cholesterol at a very young age. All these reasons lead further to negligence of their health which increases the chances of heart disease. Heart is the most essential organ of the human body and if it gets affected then it also affects the other major organs of the body. The aim of proposed system is to design an automated system which would manage complete clinical details, patient's history and their appointment details in a single database. Doctors will use this system to keep track of the patient consulting to them. The intentions of the proposed system are to reduce over-time pay and increase the number of patients that can be treated accurately.

The paper is organized as follows: Section 2 covers the literature review in the area of clinical diagnostic system. The section also identifies the existing system. Section 3 covers the proposed system.

## 2. Literature Review

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but are largely limited. They can answer simple queries like:

"What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?", "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However they cannot answer complex queries like **"Given patient records, predict the probability of patients getting a heart disease."** Clinical decisions are often made based on doctors' intuition and

experience rather than on the knowledge rich data hidden in the database. It has been observed that problems like (i). Unwanted biases, (ii). Medical errors and (iii). Large medical costs which affect the quality of service provided to patients. The proposed system aim to resolve these problems by integrating decision support system and computer based potential record systems. The system based on concepts of data mining, has the potential to generate a knowledge rich environment which helps to significantly improve the quality of clinical decisions.

Quality service for clinical system implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are obviously unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques like decision tree and naïve bayes can help remedy this situation.

### Decision Tree

Decision tree is a tree like structure. That consists of nodes and edges. Upper most nodes is called Root node and sub-nodes are called children node. Root has only one incoming edge. Node that does not contain outgoing edges are called terminal nodes. Area where it is use are:

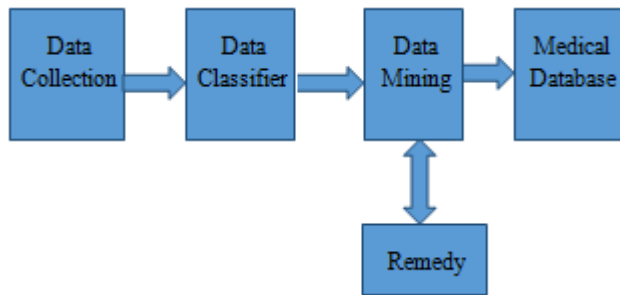
- 1) Sentiment Analyzers
- 2) Investment Solutions
- 3) Large number of Customer Satisfaction

**Naïve bayes:-** Naïve bayes is a classification technique based on the bayes theorem. Bayes theorem is used to give probabilistic value.

Area where it is used are:-

- 1) Face recognition
- 2) Sentiment analysis
- 3) Medical diagnosis
- 4) Digit recognition
- 5) Weather prediction

### 3. Architecture of proposed system



The heart disease prediction system based on data mining consists of four function components as shown above. These are as follows:

- 1) Data collection module.
- 2) Data classifier.
- 3) Data mining module.
- 4) Medical database.

The detail working of each of these components is given in following subsections.

**Data collection module:** The major task of data collection module is to gather the patient data relevant to heart disease. Various methods can be adaptable for thi purpose. Two most common are:- (i). Interviewing and (ii). Patient history. For interviewing, questions are prepared. For this purpose, the proposed system has used patient data of Cleveland clinic foundation. It is freely available in UCI repository [].

**Data classifier:** The collection data is classified into set of certain attributes. The choice of attributes is purely based upon the factors that lead to heart disease. For this purpose, 13 different attributes are identified. These attributes are listed in table1.

**Table 1**

S No.	Risk factor	Values
1	Sex	Male or Female
2	Age in Years	Age in Numeric
3	Blood Cholesterol	Below 120 mm Hg-Low 120 to 139 mm Hg-Normal Above 139 mm Hg-High
4	Blood Pressure	Below 120 mm Hg-Low 120 to 139 mm Hg-Normal Above 139 mm Hg-High
5	Hereditary	Family Member diagnosed with HD Yes Otherwise No
6	Smoking	Yes or No
7	Alcohol Intake	Yes or No
8	Physical Activity	Low , Normal or High
9	Diabetes	Yes or No
10	Diet	Poor , Normal or Good
11	Obesity	Yes or No
12	Stress	Yes or No
13	Medical Diagnosis	Heart Disease Yes or No

**Data mining module:** It is the most important component of proposed system. It is based on Naïve Bayes theorem. According to the Bayes theorem the probability of hypothesis H to be true for a given evidence E can be complete using the eq.1.

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Where:-

- 1) P(H|E) represents the probability of a hypothesis to be true for given evidence E.
- 2) P(H) (prior probability of class).
- 3) P(X/H) (likelihood).
- 4) P(X) (prior probability of predictor).

#### 3.2.4 Medical Database

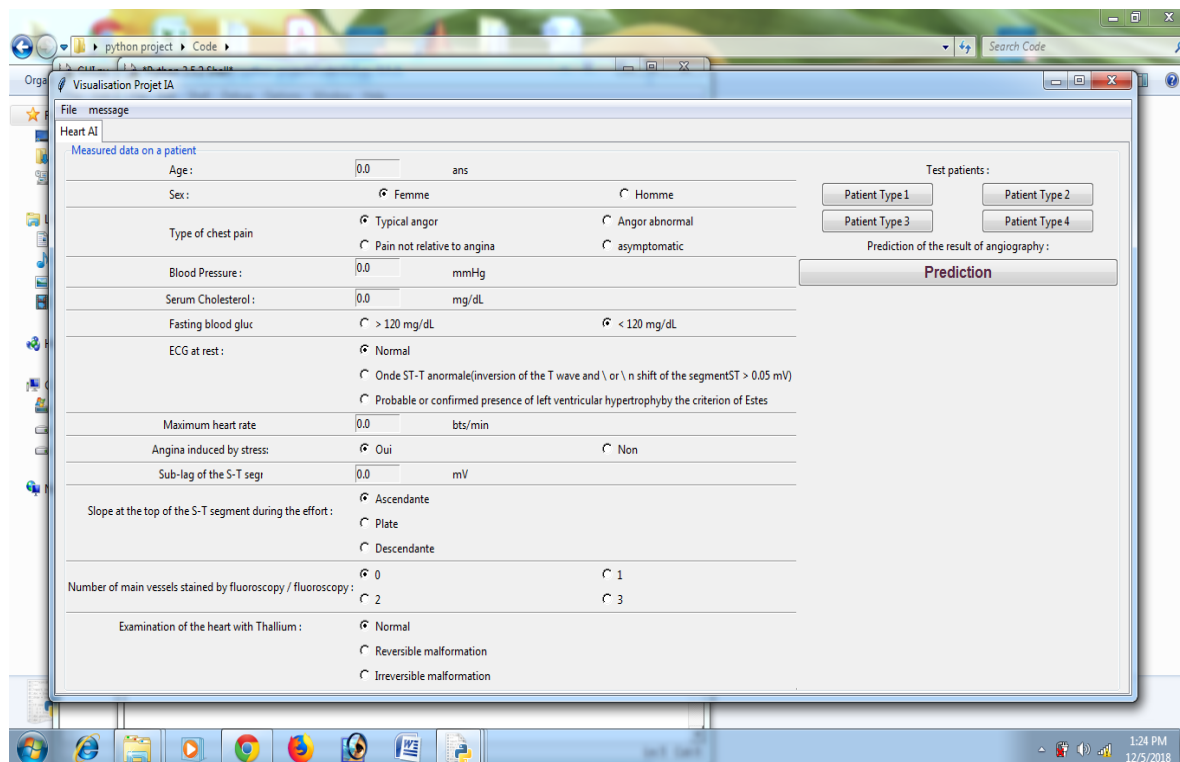
A collection of data or information. In online information retrieval, a Collection of index records in machine readable for m. This can be shown as:-

d 1	Disease	Symptoms	Remedy
-----	---------	----------	--------

Term	Definition
d 1	d 1 represents the derail number for a particular disease.
Disease	It represents infection like Cancer, Heart attack, Typhoid, Diabetes etc.
Symptoms	<ul style="list-style-type: none"> <li>• It represents Syndrome like for Cancer are:-                             <ol style="list-style-type: none"> <li>(i) Skin changes</li> <li>(ii) Breast changes</li> <li>(iii) Changes in bowel habits</li> <li>(iv) Difficult or painful urination</li> </ol> </li> <li>• For Heart attack are:-                             <ol style="list-style-type: none"> <li>(i) Chest discomfort</li> <li>(ii) Discomfort in other areas of the upper body.</li> <li>(iii) Shortness of breath.</li> <li>(iv) Other signs:-nausea or lightheadedness.</li> </ol> </li> </ul>

#### Implementation

This contains the patient's details like diseases, symptoms etc. programming language used is python.



**Figure 2**

The result shows the accuracy that if the value is greater than 50% then the person contains heart disease or otherwise not. The accuracy comes by calculating the attributes or patient database.

#### 4. Conclusion

In this paper, we have presented an intelligent and effective heart disease prediction methods using data mining. We studied an efficient approach for the extraction of significant patterns from the heart disease data warehouses for the efficient prediction of heart disease. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

#### References

[1] “Prediction System for Heart using Naive Bayes”, Shadab Adam Pattekari and AsmaParveen, International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624. Vol 3, Issue 3, 2012, pp 290-294.

[2] ‘http://www-igm.univ-mlv.fr/~lecroq/string/node14.html’, Boyer Moore, accessed on 25th April, 2014.

[3] SellappanPalaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008.

[4] Anthony M. Dymond, “Data Extraction Methods in Medium and Large Databases”, Dymond and Associates, LLC, Concord, CA.

[5] “abchomeopathy.com”

[6] NidhiBhatla, KiranJyoti”An Analysis of Heart Disease Prediction using Different Data Mining Techniques”.International Journal of Engineering Research & Technology

[8] JyotiSoniUjma Ansari Dipesh Sharma, SunitaSoni. “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”.

[9] Chaitrali S. DangareSulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques” International Journal of Computer Applications (0975 – 888)