

# Heart Disease Prediction Using Modified K Means and Using Naive Baiyes

Sairabi Mujawar<sup>1\*</sup> and Prakash Devale<sup>2</sup>

<sup>1,2</sup> Dept. of Information Technology, BVDUCOE, Pune, Maharashtra, India

[www.ijcseonline.org](http://www.ijcseonline.org)

Received: Sep/17/2015

Revised: Oct/01/2015

Accepted: Oct /16/2015

Published: Oct /31/ 2015

**Abstract-** The health care industry is generally rich in information which is not feasible to handle manually. These large amounts of data are very important in the field of Data Mining to extract useful information and generate relationship amongst the attributes. In the health care industry, for predicting the diseases from the datasets data mining is used. Heart disease prediction is treated as most complicated task in the field of medical sciences. This paper investigates a number of techniques in the detection of heart disease. This paper includes a blueprint of application of data mining in heart disease prediction.

**Keywords—** Naive Baiye, Decision Tree, Data mining, Classification, Clustering.

## I. INTRODUCTION

Heart disease is one of the most common factors of death in India.

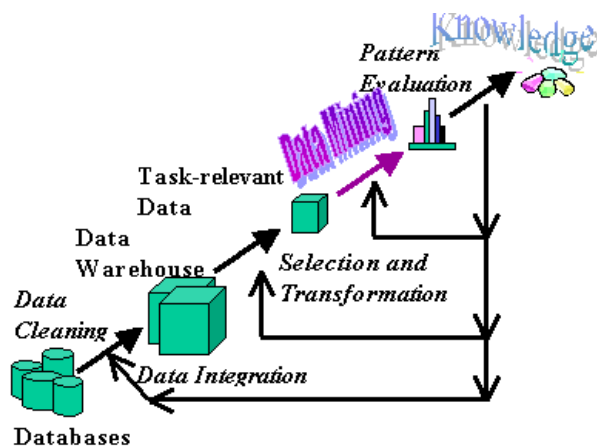


Fig. 1: Data Mining Process

Medical data consists of a number of tests which is important diagnose a particular disease. Data mining plays an important role as an intelligent diagnostic tool.

## II. HEART DISEASE

Table 1  
Papers Comparisons

Paper Topic	Author Name	Year of Publics	Algo.
Computational intelligence for heart disease diagnosis	Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping	2012 Elsevier Ltd.	Naive Bayes, SMO, IBK, AdaBoostM1, J48, PART

	Phoebe Chen		
An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease	Sivagowry .S, Dr. Durairaj. M, Persia.A	2013 IEEE.	Naive Bayes,
Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm	M.Akhil jabbar, B.L Deekshatulu, Priti Chandra	CIMTA Elsevier Ltd, 2013.	K- Nearest Neighbor and Genetic Algorithm
Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification	Ankita Dewan, Meghna Sharma	2015 IEEE.	Genetic Algorithm

### a. Papers Comparisons

## III. DATA SETS

Through Data mining repository data set from University of California, Irvine (UCI) is taken. Data set from Cleveland, Hungary, Switzerland, Long beach and Stat log data set are used. They contain 76 attributes, in which only 14 attributes are used. Among all those Cleveland data set and Stat log data set are the most widely used data set. Rest all contains missing Values. The Figure 2 shows datasets inherited from UCI repository.

Age	sex	cp	trestbps	chol	fbg	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
44	1	2	120	263	0	0	173	0	0	1	0	7	0

Fig. 2: Sample Data Set

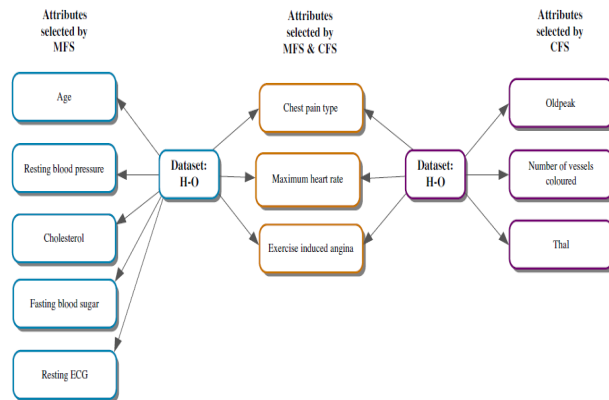


Fig. 3: MFS and CFS selection for datasets attributes.

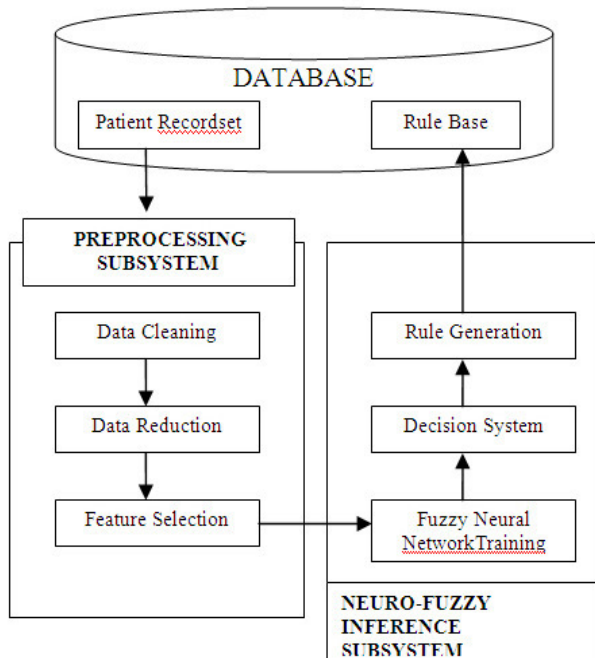


Fig. 4: Logical view of Rule Generation Process

IV. BASIC CONCEPTS

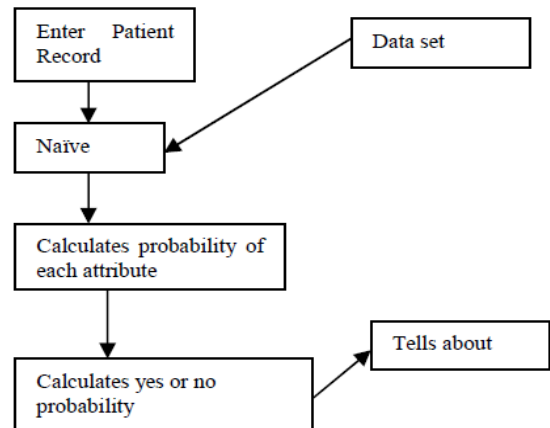


Fig. 5: Naive Bayes Algorithm on the patient data

In Naive Bayesian Classification Approach in Health Care Application. Naive Bayes is found to be the best in heart disease prediction. Naïve Bayes with predictive capability is used to create models. Understanding data and exploring is the important function of Naive Bayes. Naive Bayes is used to get efficient output.

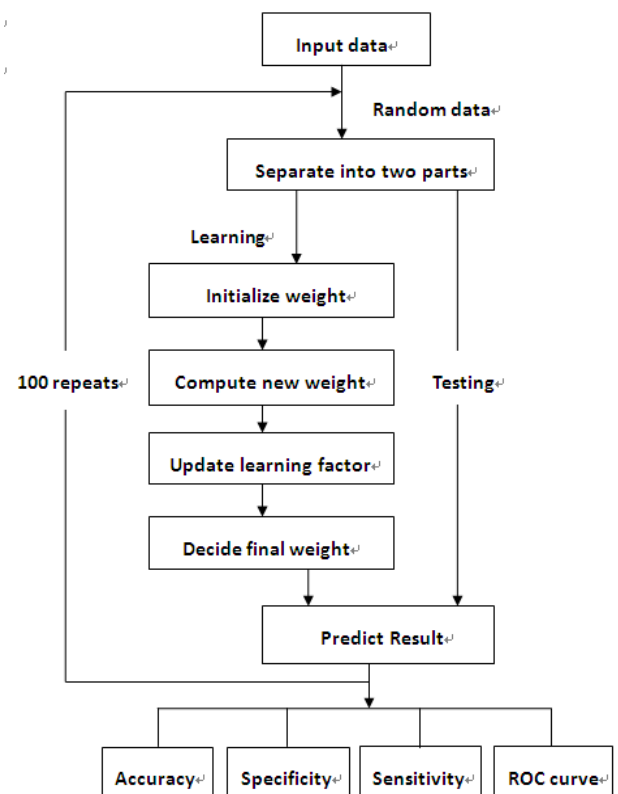


Fig. 6: The Computational Steps of LAV Model

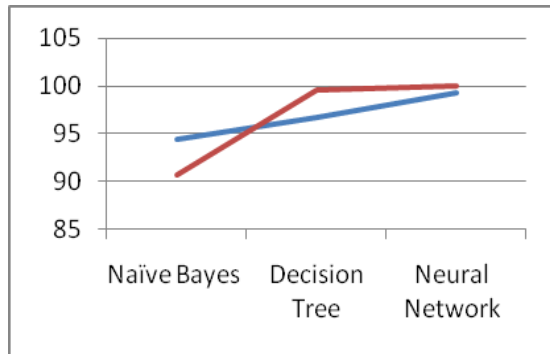


Fig. 7: Graphical Representation of Accuracy for each method

A Fuzzy logic and data mining techniques are good approach with less number of attributes which reduces number of tests for patients. Fuzzy logic combined with Decision Tree and Naive Bayes with other Data Mining techniques plays an important role in predicting Heart disease.

performance. Each technique has its own advantages and disadvantages.

#### REFERENCES

- [1] Jesmin Nahar, Tasadduq Imam, and Kevin S.Tickle, "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach", Elsevier Ltd, **2012**.
- [2] Sivagowry .S, Dr. Durairaj. M, Persia.A, "An Empirical Study on applying Data Mining Techniques for the Analysis and Prediction of Heart Disease", IEEE, **2013**.
- [3] M.Akhil jabbar, B.L Deekshatulua, Priti Chandra, "Classification of Heart Disease Using K-Neighbor and Genetic Algorithm, CIMTA", Elsevier Ltd, **2013**
- [4] Ankita Dewan, Meghna Sharma, "Prediction of Heart Disease using Hybrid Technique in Data Mining Classification", IEEE, **2015**.
- [5] Jyothi Soni, Uzma ansari and Dipesh Ansari "Intelligent and Effective Heart Disease Prediction System using Weighted Associate Classifier", IJCSE, Vol 3(6), pp 2385-2392, June **2011**.

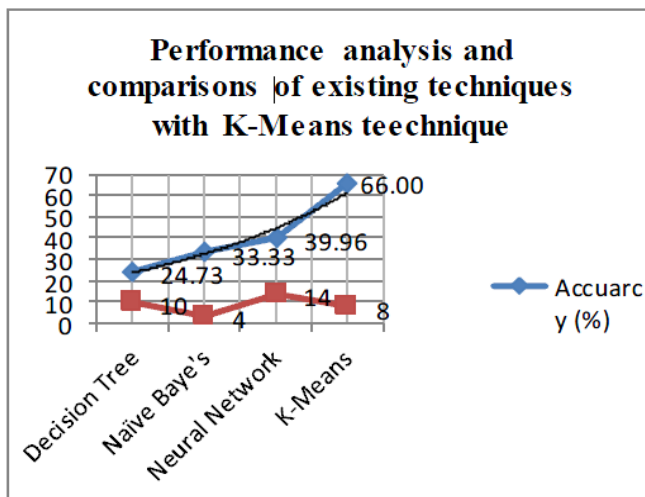


Fig. 8: Comparisons of k-means with other techniques.

K-means Clustering Algorithm is used in Bala Sundar for heart disease prediction. 'n' observations to 'k' clusters is the aim to partition. To minimize the sum of square of distance between data, Euclidean distance formula is used. The Figure 8 explains the Comparisons of k-means with other techniques.

#### ACKNOWLEDGMENT

Heart disease prediction is a major challenge in the health care industry. With less number of attributes, heart disease predictions is a challenging task in Data Mining. In some cases, Decision Tree outperforms in Classifications, whereas Neural Network and Naive Bayes has best