ISSN 2047-3338

Naive Bayesian Classification Approach in Healthcare Applications

R. Bhuvaneswari and K. Kalaiselvi
Department of CSE, Saveetha Engineering College Chennai, India
bhuvaneswari.cit@gmail.com, kalaime@gmail.com

***Abstract*– In data mining, classification is a form of data analysis that can be used to extract models describing important data classes. Two of the well known algorithms used in data mining classification are Backpropagation Neural Network (BNN) and Naïve Bayesian (NB). Bayesian approaches are a fundamentally important DM technique. Given the probability distribution, Bayes classifier can provably achieve the optimal result. Bayesian method is based on the probability theory. Bayes Rule is applied here to calculate the posterior from the prior and the likelihood, because the later two is generally easier to be calculated from a probability model. Statistics provide a strong fundamental background for quantification and evaluation of results. However, algorithms based on statistics need to be modified and scaled before they are applied to data mining.**

***Index Terms*– Decision Support Systems, Naive Bayesian Classification (NBC) and Health Care**

I. INTRODUCTION

DATA mining techniques provide people with new power to research and to manipulate the existing large volume of data. Data mining process discovers interesting information from the hidden data, which can either be used for future prediction and/or intelligent summarization of the data details. There are many achievements of application from data mining techniques to various areas such as engineering, marketing, medical, financial, and car manufacturing. The design and manufacturing domain is a natural candidate for data-mining applications because it contains extensive data. Besides enhancing innovation, data-mining methods can reduce the risks associated with conducting business and improve decision-making [1].

Especially in profiling practices such as surveillance and fraud detection, before data mining algorithms can be used, a target dataset must be assembled. As data mining can only uncover patterns already present in the data, the target dataset must be large enough to contain huge number of patterns while at the same time, remain to be concise enough to be mined in an acceptable timeframe. A common source for data is a data mart or data warehouse. Because data mart and data warehouse are significant repository, preprocessing is essential to perform analysis on the multivariate datasets before any clustering or data mining task is performed [2].

Data mining tasks like clustering, association rule mining, sequence pattern mining, and classification are used in many applications. Some of the widely used data mining algorithms in classification include Bayesian algorithms and neural networks.

II. NAIVE BAYESIAN CLASSIFIER

The objects can be classified as either GREEN or RED. Our task is to classify new cases as they arrive (i.e., decide to which class label they belong, based on the currently exiting objects).

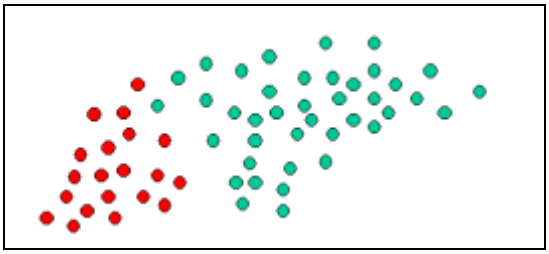

Fig. 1. Objects are classified to GREEN or RED

We can then calculate the priors (i.e. the probability of the object among all objects) based on the previous experience. Thus,

$$\text{Prior probability for GREEN} \propto \frac{\text{Number of GREEN objects}}{\text{Total number of objects}}$$

$$\text{Prior probability for RED} \propto \frac{\text{Number of RED objects}}{\text{Total number of objects}}$$

Since there is a total of 60 objects, 40 of which are GREEN and 20 RED, our prior probabilities for class membership are:

$$\text{Prior probability for GREEN} \propto \frac{40}{60}$$

$$\text{Prior probability for RED} \propto \frac{20}{60}$$

Having formulated our prior probability, we are now ready to classify a new object (WHITE circle in Figure 2). Since the objects are well clustered, it is reasonable to assume that the more GREEN (or RED) objects in the vicinity of X, the more likely that the new cases belong to that particular color. To

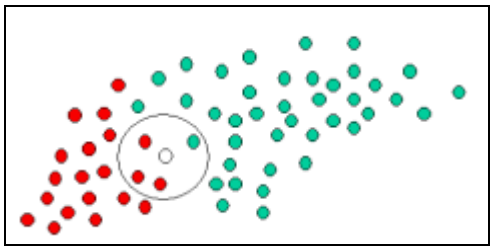

Fig. 2. Classify the WHITE circle

measure this likelihood, we draw a circle around X which encompasses a number (to be chosen a priori) of points irrespective of their class labels. Then we calculate the number of points in the circle belonging to each class label.

$$Likelihood\ of\ X\ given\ GREEN \propto \frac{Number\ of\ GREEN\ in\ the\ vicinity\ of\ X}{Total\ number\ of\ GREEN\ cases}$$

$$Likelihood\ of\ X\ given\ RED \propto \frac{Number\ of\ RED\ in\ the\ vicinity\ of\ X}{Total\ number\ of\ RED\ cases}$$

In Fig. 2, it is clear that Likelihood of X given RED is larger than Likelihood of X given GREEN, since the circle encompasses 1 GREEN object and 3 RED ones. Thus:

$$Probability\ of\ X\ given\ GREEN \propto \frac{1}{40}$$

$$Probability\ of\ X\ given\ RED \propto \frac{3}{40}$$

Although the prior probabilities indicate that X may belong to GREEN (given that there are twice as many GREEN compared to RED) the likelihood indicates otherwise; that the class membership of X is RED (given that there are more RED objects in the vicinity of X than GREEN). In the Bayesian analysis, the final classification is produced by combining both sources of information (i.e. the prior and the likelihood) to form a posterior probability using Bayes Rule.

$$Posterior\ probability\ of\ X\ being\ GREEN \propto$$

$$Prior\ probability\ of\ GREEN \times Likelihood\ of\ X\ given\ GREEN$$

$$= \frac{4}{6} \times \frac{1}{40} = \frac{1}{60}$$

$$Posterior\ probability\ of\ X\ being\ RED \propto$$

$$Prior\ probability\ of\ RED \times Likelihood\ of\ X\ given\ RED$$

$$= \frac{2}{6} \times \frac{3}{40} = \frac{1}{40}$$

Finally, we classify X as RED since its class membership achieves the largest posterior probability.

III. GAUSSIAN BAYESIAN CLASSIFIERS

The problem with the Naïve Bayes Classifier is that it assumes all attributes are independent of each other which in general cannot be applied. Gaussian PDF can be plug-in here to estimate the attribute probability density function (PDF) [4]. Because the well developed Gaussian PDF theories, we can classify the new object easier through the same Bayes Classifier Model but with certain degree recognition of the covariance. Normally, this gives more accurate classification result.

A. How to apply the Gaussian to the Bayes Classifier

The application here is very intuitive. We assume the Density Estimation follows a Gaussian distribution. Then the prior and the likelihood can be calculated through the Gaussian PDF [3]. The critical thing here is to identify the Gaussian distribution (i.e. find the mean and variance of the Gaussian). The following 5 steps are a general model to initialize the Gaussian distribution to fit our input dataset.

i) Choose a probability estimator form (Gaussian)

ii) Choose an initial set of parameters for the estimator (Gaussian mean and variance)

iii) Given parameters, compute posterior estimates for hidden variable

iv) Given posterior estimates, find distributional parameters that maximize expectation (mean) of joint density for data and hidden variable (Guarantee to also maximize improvement of likelihood)

v) Assess goodness of fit (i.e. log likelihood) If not stopping criterion, return to (3).

From research perspective, Gaussian may not be the only PDF to be applied to the Bayes Classifier, although it has very strong theoretical support and nice properties. The general model of applying those PDF's should be the same. The estimation results highly depend on whether or how close a PDF can simulate the given dataset.

$$p(x_k \mid C_j) = \begin{Bmatrix} \frac{1}{\sigma_{ij}\sqrt{2\pi}} \exp\left(\frac{-(x-\mu_{ij})^2}{2\sigma_{ij}}\right), \quad -\infty < x < \infty, -\infty < \mu_{ij} <, \sigma_{ij} > 0 & \text{Normal} \\ \mu_{ij}: \text{mean}, \sigma_{ij}: \text{standard deviation} & \\ \frac{1}{x\sigma_{ij}(2\pi)^{1/2}} \exp\left\{\frac{-[\log(x/m_{ij})]^2}{2\sigma_{ij}^2}\right\}, \quad 0 < x < \infty, m_{ij} > 0, \sigma_{ij} > 0 & \text{Lognormal} \\ m_{ij}: \text{scale parameter}, \sigma_{ij}: \text{shape parameter} & \\ \frac{\left(\frac{x}{b_{ij}}\right)^{c_{ij}-1}}{b_{ij}\Gamma(c_{ij})} \exp\left(\frac{-x}{b_{ij}}\right), \quad 0 \le x < \infty, b_{ij} > 0, c_{ij} > 0 & \text{Gamma} \\ b_{ij}: \text{scale parameter}, c_{ij}: \text{shape parameter} & \\ \frac{\lambda_{ij}\exp(-\lambda_{ij})}{x!}, \quad 0 \le x < \infty, \lambda_{ij} > 0, x = 0,1,2,\ldots & \text{Poisson} \\ \lambda_{ij}: \text{mean} & \end{Bmatrix}$$

IV. BAYESIAN NETWORKS

First of all, there is a nice introduction to Bayesian Networks and their Contemporary Applications by Daryle Niedermayer .It is generally hard for me to come up with a better one here. So the following may just a simpler version .We notice that the Gaussian model helps to integrate some correlation which improves the classification performance against the Naïve model assuming the independence. However, using Gaussian model with Bayes Classifier still has its limitation of generating the correlations [10]. So it is where the Bayesian Networks (Bayes Nets) get involved.

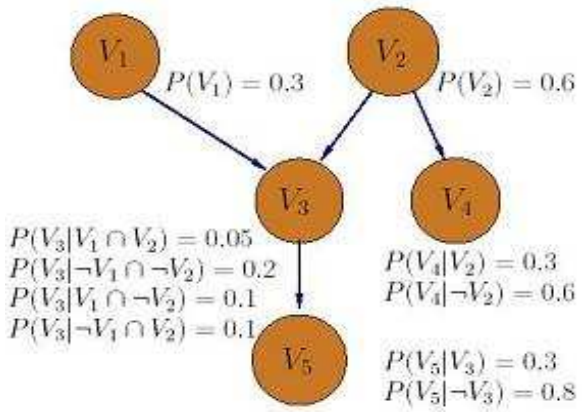

Fig. 3. A Bayes net for 5 attributes

Bayes Net is a model of utilizing the conditional probabilities among different variables. It is generally impossible to generate all conditional probabilities from a given dataset. Our task is to pick important ones and use them in the classification process. So essentially, a Bayes net is a set of "Generalized Probability Density Function".

Informally, we can define a Bayes net as an augmented directed acyclic graph, represented by the vertex set *V* and directed edge set *E*. Each vertex from *V* represents an attribute, and each edge from *E* represents a correlation between two attributes. One example of five attributes Bayes net is shown in Fig. 3.

Some important observations here:

- There is no loop in the graph representation since it is acyclic.
- Two variable *vi* and *vj* may still correlated even if they are not connected.
- Each variable *vi* is conditionally independent of all non-descendants, given its parents.

Consider a domain U of *n* variables, $x_1,...x_n$. Each variable may be discrete having a finite or countable number of states, or continuous. Given a subset X of variables x_i where $x_i \in$U, if one can observe the state of every variable in X, then this observation is called an instance of X and is denoted as X=$p(x_i|x_1,\dots x_{i-1},\xi) = p(x_i|\Pi_i,\xi)$ for the observations $x_i = k_i,\ x_i \in X$. The "joint space" of U is the set of all instances of U. $p\left(X = \vec{k}_X \mid Y = \vec{k}_Y,\ \xi\right)$ denotes the "generalized probability density" that X=$p(x_i|x_1,\dots x_{i-1},\xi) = p(x_i|\Pi_i,\xi)$ given Y=$\vec{k}_Y$ for a person with current state information ξ. p(X|Y, ξ) then denotes the *gpdf* for X, given all possible observations of Y. The *joint gpdf* over U is the *gpdf* for U.

A Bayesian network for domain U represents a *joint gpdf* over U. This representation consists of a set of local *conditional gpdfs* combined with a set of conditional independence assertions that allow the construction of a global *gpdf* from the local *gpdfs*. Then these values can be ascertained as:

$$p(x_1,\dots x_R|\xi) = \prod_{i=1}^{R} p(x_i|x_1,\dots x_{i-1},\xi) \quad (1)$$

A Bayesian Network Structure then encodes the assertions of conditional independence in Equation 1 above. Essentially then, a Bayesian Network Structure B_s is a directed acyclic graph such that (1) each variable in U corresponds to a node in B_s, and (2) the parents of the node [9].

A. Corresponding to x_i are the nodes corresponding to the variables in $[Pi]_i$

How to build a Bayes Net? A general model can be followed below:

- Choose a set of relevant variables.
- Choose an ordering for the variables.
- Assume the variables are *X1, X2, ..., Xn* (where *X1* is the first, and *Xi* is the *ith*).
- for *i = 1* to *n*:
- Add the *Xi* vertex to the network
- Set *Parent(Xi)* to be a minimal subset of *X1, ..., Xi-1*, such that we have conditional independence of *Xi* and all other members of *X1, ..., Xi-1* given *Parents(Xi)*.
- Define the probability table of *P(Xi=k | Assignments of Parent(Xi))*.

There are many choices of how to select relevant variables, as well as how to estimate the conditional probabilities. If we imagine the network as a connection of Bayes classifiers, then the probability estimation can be done applying some PDF like Gaussian [5]. In some cases, the design of the network can be rather complicated. There are some efficient ways of getting relevant variables from the dataset attributes. Assume the coming signal to be stochastic will give a nice way of extracting the signal attributes. And normally, the likelihood weighting is another way to getting attributes.

V. APPLYING BAYESIAN APPROACH ON DATASETS

A. Microarray Data Cleaning

- We remove the "Gene Description" column first, and change the "Gene Accession Number" to "ID".
- We normalize the data such that for each value set the minimum field value to 20 and the maximum to 16,000. (i.e., the expression values less than 20 or over 16,000 were considered by biologists as unreliable for this experiment)
- We transpose the data making each column representing an attribute and each row representing a record. (i.e., the data transpose is needed to get the CSV file compatible with Weka. The transpose is done by MatLab)
- We further add a "Class" attributes to indicate the kind of leukemia. (i.e., Class {ALL, AML})

B. Attribute Selection/Feature Reduction

There are in total 7070 genes (attributes) in the dataset. It is critical to do the attribute selection. We take the standard Signal to Noise (S2N) ratio and T-value to get significant attribute set. Let Avg1, Avg2 be the average expression

values and Stdev1, Stdev2 be the sample standard deviations [8].

S2N = (Avg1 - Avg2) / (Stdev1 + Stdev2)

*T-value = (Avg1 - Avg2) / sqrt (Stdev1*Stdev1 / N1 + Stdev2 * Stdev2 / N2)*

Where N1 is the number of ALL observations, and N2 is the number of AML observations.

T-value: T-value is the observed value of the T-statistic that is used to test the hypothesis that two attributes are correlated. The T-value can range between -infinity and +infinity. A T-value near 0 is evidence for the null hypothesis that there is no correlation between the attributes. A T-value far from 0 (either positive or negative) is evidence for the alternative hypothesis that there is correlation between the attributes.

C. Top 20 Genes with the highest S2N Ratio

Table I: Genes with Highest S2N Ratio

Rank	Gene Name	S2N	T-value
1	M55150_at	1.467641	8.091951
2	X95735_at	1.444531	5.727643
3	U50136_rna1_at	1.421708	6.435952
4	U22376_cds2_s_at	1.339308	7.9043
5	M81933_at	1.204042	6.164965
6	M16038_at	1.203218	4.930437
7	M84526_at	1.20142	4.057042
8	M23197_at	1.195974	4.737778
9	U82759_at	1.192556	6.24302
10	Y12670_at	1.184737	5.324928
11	D49950_at	1.143704	5.588063
12	M27891_at	1.133427	3.986204
13	X59417_at	1.124637	6.803106
14	X52142_at	1.122589	5.833144
15	M28170_at	1.116756	6.253971
16	X17042_at	1.105975	5.388623
17	U05259_rna1_at	1.103966	6.175126
18	Y00787_s_at	1.081995	4.701085
19	M96326_rna1_at	1.07719	3.869518
20	U12471_cds1_at	1.069731	6.146299

D. Common Genes from the above two Sets

Thirty one attributes in the common set: {D26156_s_at, D49950_at, L13278_at, L47738_at, L49229_f_at, M11147_at, M21551_rna1_at, M28170_at, M31211_s_at, M31523_at, M55150_at, M62762_at, M81933_at, M91432_at, M92287_at, S50223_at, U05259_rna1_at, U09087_s_at, U12471_cds1_at, U22376_cds2_s_at, U32944_at, U50136_rna1_at, U82759_at, X15949_at, X17042_at, X52142_at, X59417_at, X74262_at, X95735_at, Y12670_at, Z15115_at}

From statistic aspect, the common attribute set contains those attributes that have high discriminating ability to distinguish ALL and AML leukemia.

E. Preliminary result with NaiveBayes and BayesNet

Table II: NaiveBayes Vs BayesNet Datas

ID	*NB(S2N)*	*BN(S2N)*	*NB(T)*	*BN(T)*	*NB(C)*	*BN(C)*
39	ALL	ALL	ALL	ALL	ALL	ALL
40	ALL	ALL	ALL	ALL	ALL	ALL
42	ALL	ALL	ALL	ALL	ALL	ALL
47	ALL	ALL	ALL	ALL	ALL	ALL
48	ALL	ALL	ALL	ALL	ALL	ALL
49	ALL	ALL	ALL	ALL	ALL	ALL
41	ALL	ALL	ALL	ALL	ALL	ALL
43	ALL	ALL	ALL	ALL	ALL	ALL
44	ALL	ALL	ALL	ALL	ALL	ALL
45	ALL	ALL	ALL	ALL	ALL	ALL
46	ALL	ALL	ALL	ALL	ALL	ALL
70	ALL	ALL	ALL	ALL	ALL	ALL
71	ALL	ALL	ALL	ALL	ALL	ALL
72	ALL	ALL	ALL	ALL	ALL	ALL
68	ALL	ALL	ALL	ALL	ALL	ALL
69	ALL	ALL	ALL	ALL	ALL	ALL
67	AML	AML	ALL	ALL	AML	AML
55	ALL	ALL	ALL	ALL	ALL	ALL
56	ALL	ALL	ALL	ALL	ALL	ALL
59	ALL	ALL	ALL	ALL	ALL	ALL
52	AML	AML	AML	AML	AML	AML
53	AML	AML	AML	AML	AML	AML
51	AML	AML	ALL	AML	ALL	AML
50	AML	AML	AML	AML	AML	AML
54	AML	ALL	ALL	ALL	ALL	ALL
57	AML	AML	AML	AML	AML	AML
58	AML	AML	AML	AML	AML	AML
60	AML	ALL	ALL	ALL	ALL	ALL
61	AML	ALL	AML	ALL	AML	ALL
65	AML	AML	AML	AML	AML	AML
66	ALL	ALL	ALL	ALL	ALL	ALL
63	AML	AML	ALL	AML	AML	AML
64	AML	AML	ALL	AML	AML	AML
62	AML	AML	ALL	ALL	ALL	ALL
Total ALL	20	23	27	25	24	24
Total AML	14	11	7	9	10	10

NB: NaiveBayes; BN: BayesNet. S2N uses the top 50 S2N ratio gene set; T uses the top 50 T-value gene set; C uses the common 31 gene set from previous two.

ID field uses original sample sequence number. Red columns show the samples classified to different kind through different attribute set or different technique. Since the limitation of Bayes methods on this dataset, we do not try to push the mining further on to expect some "better" result. Some major problems with the Bayes methods here are listed below:

- Validation (either training set or testing set) is hard to tell.
- Comparison between the methods is generally not applicable.
- Detection for new type of leukemia is impossible with the given training set.

F. Statistic and Syntactic (preamble)

We propose to use rule based systems to do the mining of this dataset besides some Bayes methods. The interesting point to examine here is how these two techniques work and compared. Clearly, Bayes approach is based on the statistic model built through the dataset, and rule based system is a syntactic approach in some sense more like our thinking process [6]. Theoretically, the statistical techniques have a well-founded mathematical theory support, and thus, usually computationally inexpensive to be applied. On the other hand, syntactical techniques give nice structural descriptions/rules, and thus, simple to be understand and validated.

G. Data Preprocessing

We add a top row for descriptions of the columns. The last columns are normalized in two ways to make a two-class list and a three-class list.

Two-class: "8" -> "1" (i.e. class 1) and "4" and "7" -> "2" (i.e., class 2).

Three-class: "8" -> "1" (i.e. class1) ,"7" -> "2" (i.e., class 2), and "4" -> "3" (i.e. class3).

H. Preliminary Analysis

The two-class case seems very optimistic if we take a look at the graph of attributes plotted with respect to the "class" Fig. 5:

Attribute 7, 8, 9, and 10 separate the two classes completely, which means any of the attribute can be used to give a promised mining outcome. It is thus of less interest here to be further discussed. The three-class case is unfortunately and fortunately not as simple as the two-class case. The plots of attributes demonstrate the difficulties.

The main problem is the class "2" and "3" (i.e., red and light blue in the graph) as hardly an attribute can clearly separate them. And above all, attribute 10 is probably the most useful one in three-class case.

I. Results

The two-class case is trivial; both techniques can guarantee the 100% correctness from mining.

The three-class case can be directly mined by the two techniques (need to discritize attributes when using PRISM). The confusion matrices are listed in the Table III:

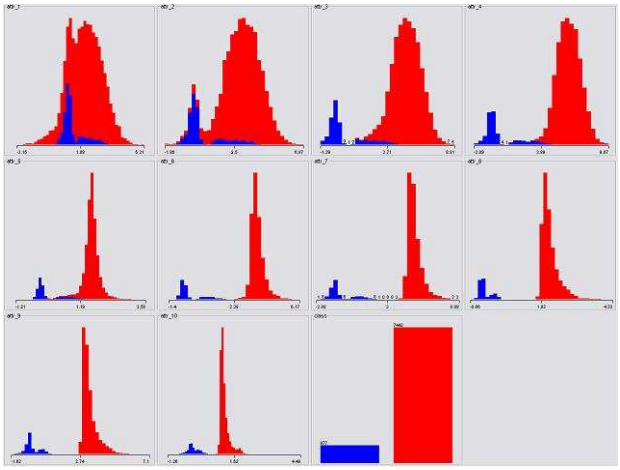

Fig. 4. Analysis for Two class case

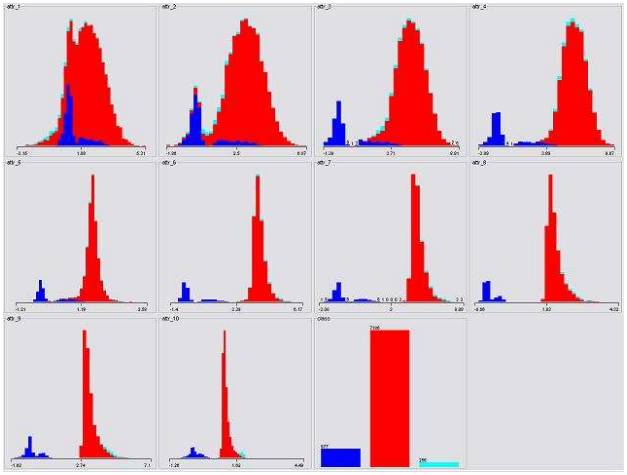

Fig. 5. Analysis for Three class case

Table III: Calculation for three classes

	BayesNet - 95.2497%			**DecisionTable - 98.9521%**			**PRISM - 98.7426%**		
Class =>	a	b	c	a	b	c	a	b	c
a = 1	330	0	0	330	0	0	330	0	0
b = 2	0	2328	126	0	2438	16	0	2437	11
c = 3	0	10	69	0	14	65	0	17	60

DecisionTable uses an attribute set of *{1,4,7,10}*. *PRISM* uses all attributes but based on the information gain to rank the importance.

The information gain ranking is: 10, 9, 7, 8, 6, 4, 5, 3, 2, 1 (based on attribute sequence).

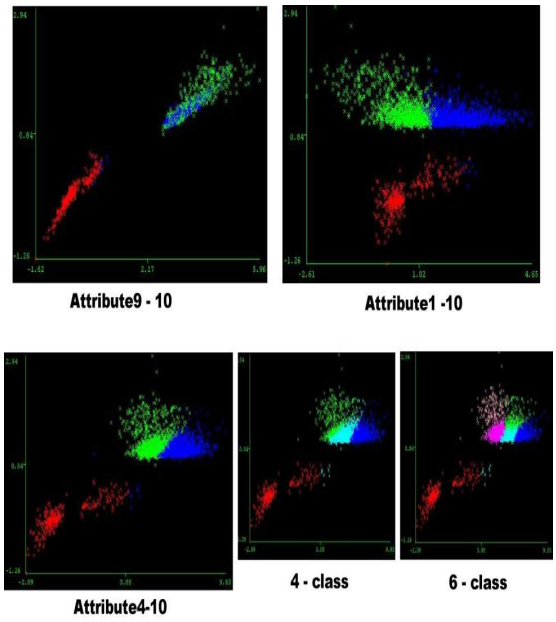

Fig. 6. Clustering result using K-mean algorithm

VI. VISUALIZATION THROUGH CLUSTERING

Some interesting information can be gained with clustering the dataset without predefined knowledge (i.e. class). The following plots show different clustering results from applying the k-mean algorithm using Weka.

Class 1 is generally clustered in red which is a very pleasant case. Class 2 and 3 are confusing from the cluster plots.

A. Dataset Recovery

The most intuitive method to deal with the missing values is just to ignore them. The plots of attributes ignoring the missing values show very likely distributions from the previous dataset.

We also try to recover the missing values with the means [7]. Following table lists the means of the 10 attributes subjected to classes (i.e., "4, 7, or 8") omitting the missing cells.

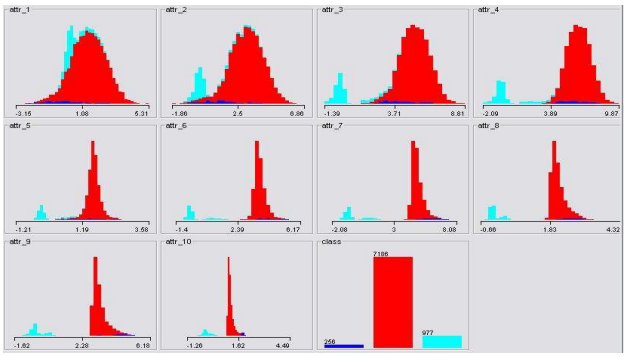

Fig. 7. Analysis omitting missing values

Table IV: Recovery of missing values

Class/Mean	*attr_1*	*attr_2*	*attr_3*	*attr_4*	*attr_5*	*attr_6*	*attr_7*	*attr_8*	*attr_9*	*attr_10*	*Total Sample*
4	-0.098	1.317	4.051	5.958	1.475	4.147	6.047	2.657	4.542	1.888	256
7	1.498	3.081	5.198	6.370	1.594	3.701	4.871	2.108	3.277	1.170	7186
8	0.559	0.478	0.325	0.150	-0.060	-0.200	-0.367	-0.145	-0.314	-0.169	976

The process is then straight forward replacing missing values with the mean values in respect to the class and attributes number. The plots of attributes after recovery are shown in Fig. 8:

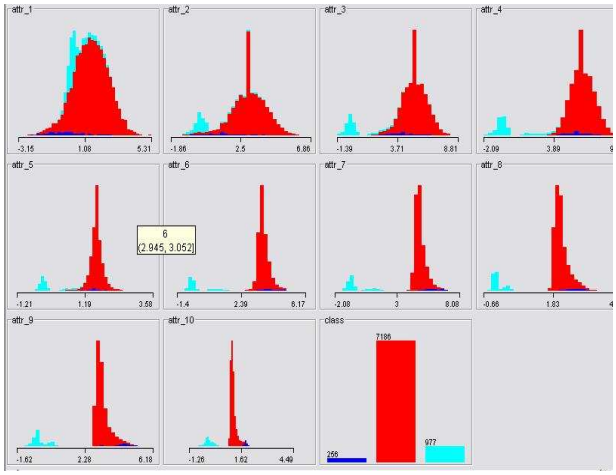

Fig. 8. Analysis after recovery

B. General Result

The recovered dataset with the means in general helps the Bayes methods to classify the dwarfs, but kind of misleading for the rule base systems.

The standard derivation table for each attribute after two preprocessing methods somehow states the point.

Table V: Preprocessing Method

Method	*Class*	*attr_1*	*attr_2*	*attr_3*	*attr_4*	*attr_5*	*attr_6*	*attr_7*	*attr_8*	*attr_9*	*attr10*
Ignore	4	1.216	1.250	1.164	1.068	0.773	0.707	0.717	0.446	0.568	0.374
Replace	4	1.216	1.188	1.105	0.999	0.734	0.692	0.699	0.439	0.553	0.365
Ignore	7	1.082	1.120	1.048	1.020	0.278	0.333	0.431	0.267	0.398	0.158
Replace	7	1.081	1.056	0.994	0.975	0.267	0.321	0.418	0.260	0.388	0.154
Ignore	8	0.699	1.151	1.312	1.493	0.457	0.657	0.850	0.201	0.408	0.240
Replace	8	0.699	1.114	1.294	1.475	0.453	0.652	0.845	0.200	0.405	0.238

VII. CONCLUSION

This paper predicts the use of Naive Baye's classifier in medical applications. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Decision Support in Heart Disease Prediction System is also developed using Naive Bayesian Classification technique. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care. Naïve Bayes classification can be used as a best decision support system.

REFERENCES

[1] Bressan, M. and J. Vitria, *On the selection and classification of independent features.* Pattern Analysis and Machine Intelligence,IEEE Transactions on, 2003. 25(10): p. 1312-1317

[2] 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.

[3] Giorgio Corani, Marco Zaffalon, JNCC2: The Java Implementation of Naive Credal Classifier 2, Journal of Machine Learning Research9 (2008) 2695-2698 ISSN [4] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.

[4] Ho, T. J.: "Data Mining and Data Warehousing", Prentice Hall, 2005.

[5] Harleen Kaur and Siri Krishan Wasan, Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science 2 (2): 194-200, 2006 ISSN 1549-3636 © 2006 Science Publications.

[6] Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.

[7] Neapolitan, R., *Learning Bayesian Networks*. 2004, London: Pearson Printice Hall

[8] Shantakumar B.Patil, Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656 © EuroJournals Publishing, Inc. 2009.

[9] Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

[10] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Techniques, 978-1-4244-1968- 5/08/$25.00 ©2008 IEEE.

R. Bhuvaneswari, received her B.E and M.Tech in Information Technology. Currently, she is an Assistant Professor in the Department of Computer Science Engineering Saveetha Engineering College, Chennai. Her area of interests includes, Data Mining and Data Warehousing, Database Management Systems, Operating Systems, Web Based Services.

K. Kalaiselvi, received her B.E and M.E in Computer Science Engineering. Currently, she is an Assistant Professor in the Department of Computer Science Engineering, Saveetha Engineering College, Chennai. Her area of interests includes Data Mining and Data Warehousing, Database Management Systems, Data structures, Operating System concepts, Multimedia.