Heart disease prediction
Report

First 3-4 pages will be the college first page, logo, acknowledgement etc.

ABSTRACT
The main objective of this research is to develop an Intelligent System using data mining modelling technique, namely, Naive Bayes. It retrieves hidden data from stored databases and compares the user values with a trained data set. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs.
Heart is the most essential or crucial portion of our body. Heart is used to maintain and conjugate blood in our body. There are a lot of cases in the world related to heart diseases. People are dying due to heart disease. Various symptoms are mentioned. The health care industries found a large amount of data. This paper gives the idea of predicting heart disease using machine learning algorithms. Here, we will use machine learning algorithms **knn and naïve bayes**. The algorithms are used on the basis of features and for predicting heart disease.

Keyword: Data mining Naive bayes, knn, heart disease, prediction

Introduction

A major challenge facing healthcare organisations (hospitals, medical centres) is the provision of quality services at reasonable prices. Quality service implies diagnosing patients properly and administering treatments that are effective. Poor clinical selections will result in fateful consequences that are so unacceptable. Hospitals should additionally reduce the price of clinical tests. They'll come through these results by using applicable computer-based info and/or call support systems. Most hospitals today use some form of hospital information systems to manage their health care or patient knowledge. These systems usually generate huge amounts of knowledge that take the shape of numbers, text, charts and pictures. Sadly, these knowledge are rarely accustomed to supporting clinical decision-making. There's a wealth of hidden information in this knowledge that's mostly untapped. This raises a crucial question: “How will we tend to flip knowledge into helpful info that may change health care practitioners to create intelligent clinical decisions?” Although data mining has been around for over 20 years, its potential is barely being accomplished now. Data mining combines applied mathematics analysis, machine learning and information technology to extract hidden patterns and relationships from giant databases. Naive Bayes or Bayes' Rule is the basis for several machine-learning and data processing ways. The rule (algorithm) is employed to make models with predictive capabilities. It provides new ways of exploring and understanding knowledge. It learns from the "evidence" by calculating the correlation between the target (i.e., dependent) and different (i.e., independent) variables.

Data Mining is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data with the wide use of databases and the explosive growth in their sizes. Data mining refers to extracting or "mining" knowledge from large amounts of data. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden among large amounts of data. The essential process of Knowledge Discovery is the conversion of data into knowledge in order to aid in decision making, referred to as data mining. The Knowledge Discovery process consists of an iterative sequence of data cleaning, data integration, data selection, data mining pattern recognition and knowledge presentation. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden among large amounts of data. 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 have resulted in hospital stays longer than 10 days?", "Identify 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. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The proposed system that integrates clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modelling and analysis tools, e.gData mining, have the potential to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions.

Research objectives

Most hospitals today employ sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data. There is a wealth of hidden information in these data that is largely untapped. How data is turned into useful information that can enable healthcare practitioners to make intelligent clinical decisions. The main objective of this research is to develop a Decision Support in Heart Disease Prediction System (DHDPS) using one data mining modelling technique, namely, Naïve Bayes. DSHDPS is implemented as web based questionnaire application. Based on user answers, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. We provide the report of the patient in two ways using chart and pdf which indicates whether that particular patient having the heart disease or not. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge rich environment which can help to significantly improve the quality of clinical decisions. The diagnosis of diseases is a significant and tedious task in medicine. The detection of heart disease from various factors or symptoms is a multi-layered issue which is not free from false presumptions often accompanied by unpredictable effects. Thus the effort to utilize knowledge and experience of numerous specialists and clinical screening data of patients collected in databases to facilitate the diagnosis process is considered a valuable option. Providing precious services at affordable costs is a major constraint encountered by the healthcare organizations (hospitals, medical centers). Valuable quality service denotes the accurate diagnosis of patients and providing efficient treatment. Poor clinical decisions may lead to disasters and hence are seldom entertained. Besides, it is essential that the hospitals decrease the cost of clinical test. Appropriate computer-based information and/or decision support systems can aid in achieving clinical tests at a reduced cost. Naive Bayes or Bayes' Rule is the basis for many machine-learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data. It learns from the "evidence" by calculating the correlation between the target (i.e., dependent) and other (i.e., independent) variables.

Scope

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. The application is fed with varied details and therefore the cardiovascular disease related to those details. The application permits user to share their heart connected problems. It then processes user specific details to ascertain for varied illness that might be related to it. Here we tend to use some intelligent data mining techniques to guess the foremost correct illness that might be related to patient's details. Based on result, system automatically shows the result specific doctors for more treatment. The system permits user to look at doctor's details. The system can be use in case of emergency.

Implementation

In probability theory, Bayes' theorem (often called Bayes' law after Thomas Bayes) relates the conditional and marginal probabilities of two random events. It is often used to compute posterior probabilities given observations . For example, a patient could also be ascertained to possess certain symptoms. Bayes' theorem is often used to compute the likelihood that a projected diagnosis is correct, as long as observation.

A Naive Bayes' classifier may be a term addressing a simple probabilistic classification supported by applying Bayes' theorem. In easy terms, a Naïve Bayes classifier assumes that the presence (or absence) of a specific feature of a category is unrelated to the presence (or absence) of the other feature. As an example, a fruit could also be thought of to be an apple if it's red, round, and regarding 4" in diameter. Evensupposing these options rely on the existence of the opposite options, a Naive Bayes' classifier considers all of those properties to independently contribute to the likelihood that this fruit is an apple. Naive Bayes algorithm is based on Bayesian Theorem.

Bayesian Statistics continues to remain incomprehensible in the ignited minds of many analysts. Being amazed by the incredible power of machine learning, a lot of us have become unfaithful to statistics. Our focus has narrowed down to exploring machine learning. Isn't it true?

We fail to understand that machine learning is not the only way to solve real world problems. In several situations, it does not help us solve business problems, even though there is data involved in these problems. To say the least, knowledge of statistics will allow you to work on complex analytical problems, irrespective of the size of data.

In 1770s, Thomas Bayes introduced 'Bayes Theorem'. Even after centuries later, the importance of 'Bayesian Statistics' hasn't faded away. In fact, today this topic is being taught in great depths in some of the world's leading universities.

1. Frequentist Statistics

The debate between *frequentist* and *Bayesians* have haunted beginners for centuries. Therefore, it is important to understand the difference between the two and how does there exists a thin line of demarcation!

It is the most widely used inferential technique in the statistical world. Infact, generally it is the first school of thought that a person entering into the statistics world comes across.

Frequentist Statistics tests whether an event (hypothesis) occurs or not. It calculates the probability of an event in the long run of the experiment (i.e the experiment is repeated under the same conditions to obtain the outcome).

Here, the sampling distributions of fixed size are taken. Then, the experiment is theoretically repeated infinite number of times but practically done with a stopping intention. For example, I perform an experiment with a stopping intention in mind that I will stop the experiment when it is repeated 1000 times or I see minimum 300 heads in a coin toss.

Let's go deeper now.

Now, we'll understand *frequentist statistics* using an example of coin toss. The objective is to estimate the fairness of the coin. Below is a table representing the frequency of heads:

no. of tosses	no. of heads	difference
10	4	-1
50	25	0
100	44	-6
500	255	5
1000	502	2
5000	2533	33
10000	5067	67

We know that probability of getting a head on tossing a fair coin is 0.5. `No. of heads` represents the actual number of heads obtained. `Difference` is the difference between `0.5*(No. of tosses) - no. of heads`.

An important thing is to note that, though the difference between the actual number of heads and expected number of heads(50% of number of tosses) increases as the number of tosses are increased, the proportion of number of heads to total number of tosses approaches 0.5 (for a fair coin).

This experiment presents us with a very common flaw found in frequentist approach i.e. *Dependence of the result of an experiment on the number of times the experiment is repeated.*

To know more about frequentist statistical methods, you can head to this excellent course on inferential statistics.

2. The Inherent Flaws in Frequentist Statistics

Till here, we've seen just one flaw in *frequentist statistics*. Well, it's just the beginning.

20th century saw a massive upsurge in the *frequentist statistics* being applied to numerical models to check whether one sample is different from the other, a parameter is important enough to be kept in the model and variousother manifestations of hypothesis testing. But *frequentist statistics* suffered some great flaws in its design and interpretation which posed a serious concern in all real life problems. For example:

1. `p-values` measured against a sample (fixed size) statistic with some stopping intention changes with change in intention and sample size. i.e If two persons work on the same data and have different stopping intention,

they may get two different `p- values` for the same data, which is undesirable.

For example: Person A may choose to stop tossing a coin when the total count reaches 100 while B stops at 1000. For different sample sizes, we get different t-scores and different p-values. Similarly, intention to stop may change from fixed number of flips to total duration of flipping. In this case too, we are bound to get different *p-values*.

2- Confidence Interval (C.I) like `p-value` depends heavily on the sample size. This makes the stopping potential absolutely absurd since no matter how many persons perform the tests on the same data, the results should be consistent.

3- Confidence Intervals (C.I) are not probability distributions therefore they do not provide the most probable value for a parameter and the most probable values.

These three reasons are enough to get you going into thinking about the drawbacks of the *frequentist approach* and why there is a need for *a Bayesian approach*. Let's find it out.

From here, we'll first understand the basics of Bayesian Statistics.

3. Bayesian Statistics

“Bayesian statistics is a mathematical procedure that applies probabilities to statistical problems. It provides people the tools to update their beliefs in the evidence of new data.”

You got that? Let me explain it with an example:

Suppose, out of all the 4 championship races (F1) between Niki Lauda and James hunt, Niki won 3 times while James managed only 1.

So, if you were to bet on the winner of the next race, who would he be ?

I bet you would say Niki Lauda.

Here’s the twist. What if you are told that it rained once when James won and once when Niki won and it is definite that it will rain on the next date. So, who would you bet your money on now ?

By intuition, it is easy to see that chances of winning for James have increased drastically. But the question is: how much ?

To understand the problem at hand, we need to become familiar with some concepts, first of which is conditional probability (explained below).

In addition, there are certain prerequisites:

<u>Pre-Requisites:</u>

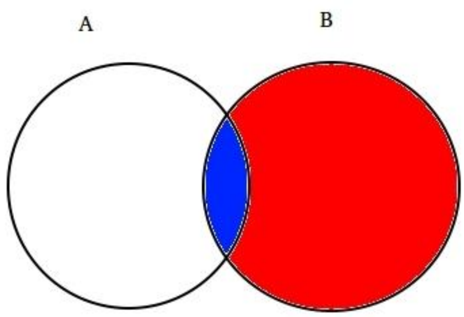
1. Linear Algebra
2. Probability and Basic Statistics

3.1 Conditional Probability

It is defined as the: Probability of an event A given B equals the probability of B and A happening together divided by the probability of B."

For example: Assume two partially intersecting sets A and B as shown below.

Set A represents one set of events and Set B represents another. We wish to calculate the probability of A given B has already happened. Lets represent the happening of event B by shading it with red.



Now since B has happened, the part which now matters for A is the part shaded in blue which is interestingly $A \cap B$. So, the probability of A given B turns out to be:

$$\frac{BlueArea}{RedArea + BlueArea}$$

Therefore, we can write the formula for event B given A has already occurred by:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

or

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Now, the second equation can be rewritten as :

$$P(A|B) = \frac{P(B|A) X P(A)}{P(B)}$$

This is known as Conditional Probability.

Let’s try to answer a betting problem with this technique.

Suppose, B be the *event of winning of James Hunt*. A be the *event of raining*. Therefore,

1. P(A) =1/2, since it rained twice out of four days.
2. P(B) is 1/4, since James won only one race out of four.
3. P(A|B)=1, since it rained every time when James won.

Substituting the values in the conditional probability formula, we get the probability to be around 50%, which is almost the double of 25% when rain was not taken into account (Solve it at your end).

This further strengthened our belief of James winning in the light of new *evidence* i.e rain. You must be wondering that this formula bears close resemblance to something you might have heard a lot about. Think!

Probably, you guessed it right. It looks like Bayes Theorem.

Bayes theorem is built on top of conditional probability and lies in the heart of Bayesian Inference. Let’s understand it in detail now.

3.2 Bayes Theorem

Bayes Theorem comes into effect when multiple events A_i form an exhaustive set with another event B. This could be understood with the help of the below diagram.

A1	**B**	
A2		
A3		

Now, B can be written as

$$B = \sum_{i=1}^{n} B \cap A_i$$

So, probability of B can be written as,

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i)$$

But $P(B \cap A_i) = P(B|A_i) \times P(A_i)$

So, replacing P(B) in the equation of conditional probability we get

$$P(A_i|B) = (P(B|A_i) \times P(A_i))/(\sum_{i=1}^{n}(P(B|A_i) \times P(A_i)))$$

This is the equation of Bayes Theorem.

4. Bayesian Inference

There is no point in diving into the theoretical aspect of it. So, we'll learn how it works! Let's take an example of coin tossing to understand the idea behind *Bayesian inference*.

An important part of *Bayesian inference* is the establishment of *parameters* and *models.*

Models are the mathematical formulation of the observed events. Parameters are the factors in the models affecting the observed data. For example, in tossing a coin, fairness of coin may be defined as the parameter of coin denoted by θ. The outcome of the events may be denoted by D.

Answer this now. What is the probability of 4 heads out of 9 tosses(D) given the fairness of the coin (θ). i.e `P(D|θ)`

Wait, did I ask the right question? No.

We should be more interested in knowing : Given an outcome (D) what is the probability of coin being fair (θ=0.5)

Lets represent it using Bayes Theorem:

$$P(\theta|D)=(P(D|\theta) \times P(\theta))/P(D)$$

Here, $P(\theta)$ is the *prior* i.e the strength of our belief in the fairness of coin before the toss. It is perfectly okay to believe that coin can have any degree of fairness between 0 and 1.

$P(D|\theta)$ is the likelihood of observing our result given our distribution for θ. If we knew that coin was fair, this gives the probability of observing the number of heads in a particular number of flips.

$P(D)$ is the evidence. This is the probability of data as determined by summing (or integrating) across all possible values of θ, weighted by how strongly we believe in those particular values of θ.

If we had multiple views of what the fairness of the coin is (but didn't know for sure), then this tells us the probability of seeing a certain sequence of flips for all possibilities of our belief in the coin's fairness.

$P(\theta|D)$ is the posterior belief of our parameters after observing the evidence i.e the number of heads .

From here, we'll dive deeper into the mathematical implications of this concept. Don't worry. Once you understand them, getting to its *mathematics* is pretty easy.

To define our model correctly , we need two mathematical models beforehand. One to represent the *likelihood function P(D|θ)* and the other for representing the distribution of *prior beliefs .* The product of these two gives the *posterior belief P(θ|D)* distribution.

Since prior and posterior are both beliefs about the distribution of fairness of coin, intuition tells us that both should have the same mathematical form. Keep this in mind. We will come back to it again.

So, there are several functions which support the existence of bayes theorem. Knowing them is important, hence I have explained them in detail.

4.1. Bernoulli likelihood function

Lets recap what we learned about the likelihood function. So, we learned that:

It is the probability of observing a particular number of heads in a particular number of flips for a given fairness of coin. This means our probability of observing heads/tails depends upon the fairness of coin (θ).

`P(y=1|θ)=` θ^y [If coin is fair θ=0.5, probability of observing heads (y=1) is 0.5]

`P(y=0|θ)=` $(1-\theta)^{1-y}$ [If coin is fair θ=0.5, probability of observing tails(y=0) is 0.5]

It is worth noticing that representing 1 as heads and 0 as tails is just a mathematical notation to formulate a model. We can combine the above mathematical definitions into a single definition to represent the probability of both the outcomes.

P(y|θ)= $\theta^y.(1-\theta)^{1-y}$

This is called the Bernoulli Likelihood Function and the task of coin flipping is called Bernoulli's trials.

`y={0,1},θ=(0,1)`

And, when we want to see a series of heads or flips, its probability is given by:

$$P(y_1, y_2, ...y_n|\theta) = \prod_1^n P(y_i|\theta)$$

$$P(y_1, y_2, ..., y_n|\theta) = \prod_1^n \theta^{y_i}.(1-\theta)^{1-y_i}$$

Furthermore, if we are interested in the probability of number of heads *z* turning up in *N* number of flips then the probability is given by:

$$P(z, N|\theta) = \theta^z.(1-\theta)^{N-z}$$

4.2. Prior Belief Distribution

This distribution is used to represent our strengths on beliefs about the parameters based on the previous experience.

But, what if one has no previous experience?

Don't worry. Mathematicians have devised methods to mitigate this problem too. It is known as `uninformative priors`. I would like to inform you beforehand that it is just a misnomer. Every uninformative prior always provides some information event the constant distribution prior.

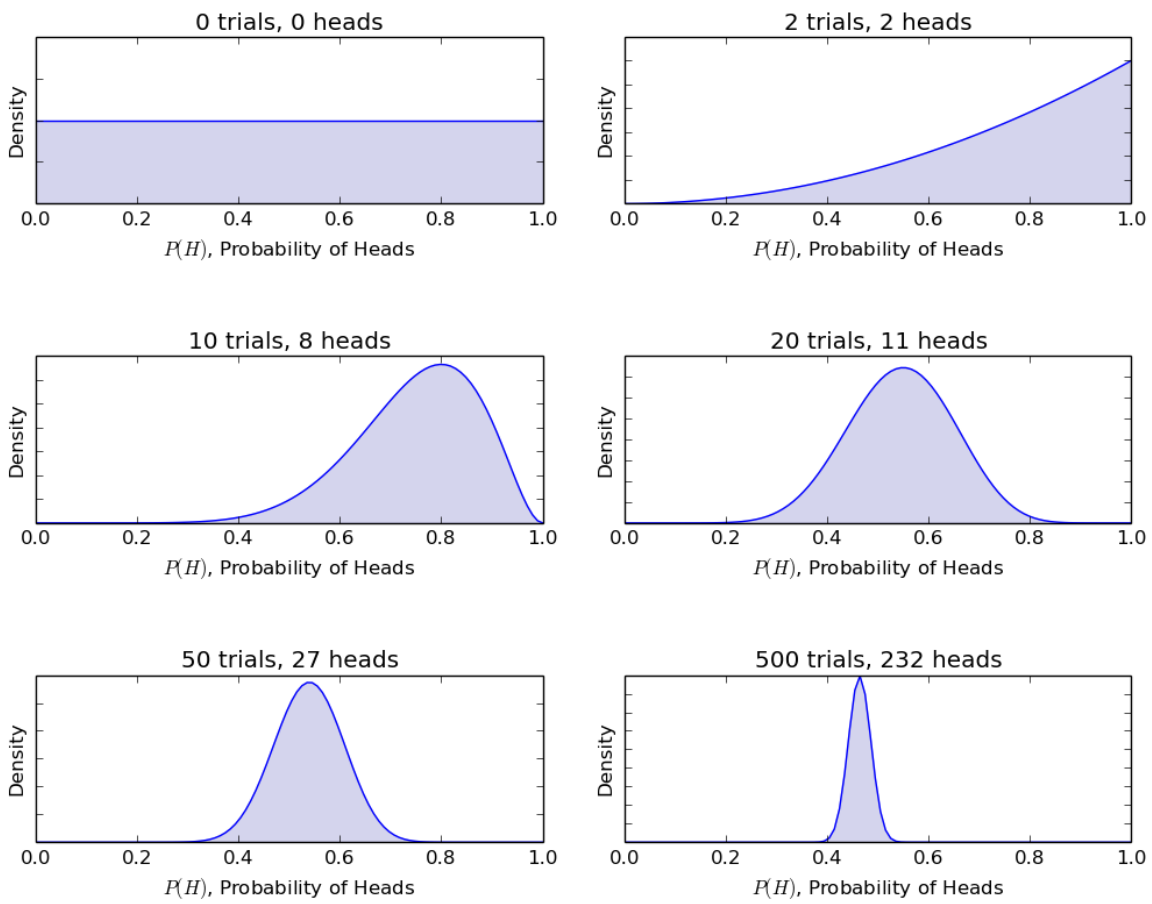
Well, the mathematical function used to represent the prior beliefs is known as *`beta distribution`*. It has some very nice mathematical properties which enable us to model our beliefs about a binomial distribution.

Probability density function of beta distribution is of the form :

$$x^{\alpha-1}.(1-x)^{\beta-1}/B(\alpha,\beta)$$

where, our focus stays on numerator. The denominator is there just to ensure that the total probability density function upon integration evaluates to 1.

α and β are called the shape deciding parameters of the density function. Here α is analogous to number of heads in the trials and β corresponds to the number of tails. The diagrams below will help you visualize the beta distributions for different values of α and β



You too can draw the beta distribution for yourself using the following code in R:

```
> library(stats)

> par(mfrow=c(3,2))

> x=seq(0,1,by=o.1)

> alpha=c(0,2,10,20,50,500)

> beta=c(0,2,8,11,27,232)

> for(i in 1:length(alpha)){
```

```
        y<-dbeta(x,shape1=alpha[i],shape2=beta[i])

        plot(x,y,type="l")

}
```

Note: α and β are intuitive to understand since they can be calculated by knowing the mean (μ) and standard deviation (σ) of the distribution. In fact, they are related as :

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$$

If mean and standard deviation of a distribution are known , then there shape parameters can be easily calculated.

Inference drawn from graphs above:

1. When there was no toss we believed that every fairness of coin is possible as depicted by the flat line.
2. When there were more number of heads than the tails, the graph showed a peak shifted towards the right side, indicating higher probability of heads and that coin is not fair.
3. As more tosses are done, and heads continue to come in larger proportion the peak narrows increasing our confidence in the fairness of the coin value.

4.3. Posterior Belief Distribution

The reason that we chose prior belief is to obtain a beta distribution. This is because when we multiply it with a likelihood function, posterior distribution yields a form similar to the prior distribution which is much easier to relate to and understand. If this much information whets your appetite, I'm sure you are ready to walk an extra mile.

Let's calculate posterior belief using bayes theorem.

Calculating posterior belief using Bayes Theorem

$$P(\theta|z, N) = P(z, N|\theta)P(\theta)/P(z, N)$$

$$= \theta^z(1-\theta)^{N-z}.\theta^{\alpha-1}(1-\theta)^{\beta-1}/[B(\alpha,\beta)P(z,N)]$$

$$= \theta^{z+\alpha-1}(1-\theta)^{N-z+\beta-1}/[B(z+\alpha, N-z+\beta)]$$

Now, our posterior belief becomes,

$$P(\theta|z+\alpha, N-z+\beta)$$

This is interesting. Just knowing the mean and standard distribution of our belief about the parameter θ and by observing the number of heads in N flips, we can update our belief about the model parameter(θ).

Lets understand this with the help of a simple example:

Suppose, you think that a coin is biased. It has a mean (μ) bias of around 0.6 with standard deviation of 0.1.

Then ,

`α= 13.8 , β=9.2`

i.e our distribution will be biased on the right side. Suppose, you observed 80 heads (`z=80`) in 100 flips(`N=100`). Let's see how our prior and posterior beliefs are going to look:

`prior = P(θ|α,β)=P(θ|13.8,9.2)`

`Posterior = P(θ|z+α,N-z+β)=P(θ|93.8,29.2)`

Lets visualize both the beliefs on a graph:

The R code for the above graph is as:

```
> library(stats)

> x=seq(0,1,by=0.1)

> alpha=c(13.8,93.8)

> beta=c(9.2,29.2)

> for(i in 1:length(alpha)){

      y<-dbeta(x,shape1=alpha[i],shape2=beta[i])

      plot(x,y,type="l",xlab = "theta",ylab = "density")
```

```
}
```

As more and more flips are made and new data is observed, our beliefs get updated. This is the real power of Bayesian Inference.

The **Naïve Bayes Classifier technique** is mainly applicable when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. The Naïve Bayes model recognizes the characteristics of patients with heart disease. It shows the probability of each input attribute for the predictable state. Naive Bayes algorithm is preferred in the following cases. When the dimensionality of data is high. When the attributes are independent of each other. Otherwise, attributes are assumed to be independent in order to simplify the computations involved and, in this sense, is considered “naïve”. When we expect more efficient output, as compared to other methods output. Exhibits high accuracy and speed when applied to large databases.

KNN (K- Nearest Neighbor)

The K nearest neighbour algorithm is an example-based learning algorithm that is widely used in real-life scenarios. The K Nearest Neighbors algorithm can be used to solve both classification and regression problems. The K Nearest Neighbour algorithm is another name for lazy learning. The K Nearest Neighbour algorithm involves preprocessing the dataset, training the model, and testing the model. Cleaning and removing erroneous and outlier values from a dataset is usually part of the preprocessing phase. In the algorithmic process, this is the most important step. It's also crucial to check the dataset's accuracy before running algorithmic tests on it. The K Nearest Neighbour uses the curve to plot the latest test results. The K element is the number of neighbours that the classification takes into account. In most cases, the K value should be a single digit number. After the test data point

SYSTEM ARCHITECTURE

Considering the anomalies in the existing system computerization of the whole activity is being suggested after initial analysis. It might have happened so many times that you or someone of yours needs doctors help immediately, but they are not available due to some reason. Here, we propose a web application that allows users to get instant guidance on their heart disease through an intelligent system online. The application is fed with various details and the heart disease associated with those details. The application allows users to share their heart related issues. It then processes user specific details to check for various illnesses that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with a patient's details.

Based on the result, the system automatically shows the result to specific doctors for further treatment. The system allows users to view doctor’s details. The system can be used in case of emergency. The main goal of this system is to predict heart disease using data mining techniques such as Naive Bayesian Algorithm. Raw hospital data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as Naïve Bayes algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and the user is given the result based on the prediction whether the risk of heart disease is low, average or high.

CONCLUSIONS

The Heart Disease Prediction, historically viewed as a necessary burden in medical offices, healthcare facilities and wellness centres, can be completely automated through an inefficient online software program. The benefits of implementing this technology touch everyone involved in the scheduling process, as administrators and users can conduct their tasks more efficiently and accurately. The system extracts hidden knowledge from a historical heart disease database. This system can be further enhanced and expanded for many more disease predictions. Heart disease prediction can be increased and expanded. For instance, it can incorporate different medical attributes besides the listed. It may also incorporate different data processing techniques, e.g., time series, clustering and association Rules. Continuous information may also be used rather than simply categorical information.

Machine learning based solutions are widely used in the healthcare sector for analysing patients' data, predicting diseases and suggesting possible treatments. With a number of machine learning techniques available today, it is important to identify the most efficient and accurate technique especially in critical domains like healthcare. A comparative analysis of the various Machine learning algorithms used in the heart disease prediction is presented. KNN and Naive Bayes are discussed and compared to identify the best suited classifier for heart disease prediction. Many previous researches and studies related to heart disease prediction were identified and analysed. findings shows that in most cases machine learning based approaches have shown significant potential to transform the healthcare sector and improve the entire process of disease predictions and suggesting treatments.

References

[1] Blake, C.L., Mertz, C.J.: “UCI Machine Learning Databases”, http://mlearn.ics.uci.edu/databases/heartdisease/, 2004.

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

[4] Han, J., Kamber, M.: “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2006.

[5] Ho, T. J.: “Data Mining and Data Warehousing”, Prentice Hall, 2005.

[6] 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

[7] Tang, Z. H., MacLennan, J.: “Data Mining with SQL Server 2005”, Indianapolis: Wiley, 2005.

[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] Wu, R., Peters, W., Morgan, M.W.: “The Next Generation Clinical Decision Support: Linking Evidence to Best Practice”, Journal Healthcare Information Management. 16(4), 50-55, 2002.

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