BAYESIAN CLASSIFICATION

CSE 634 DATA MINING | PROF. ANITA WASILEWSKA
References

- Bayes Theorem: https://www.investopedia.com/terms/b/bayes-theorem.asp
- Bayes Classification: https://www.tutorialspoint.com/data_mining/dm_bayesian_classification.html
- http://users.sussex.ac.uk/~christ/crs/ml/lec02b.html
- Example of Bayes Classification: https://mlcorner.wordpress.com/2013/04/28/bayesian-classifier/
- Data Mining Concepts and Techniques 2nd Edition by Jiawei Han and Micheline Kamber
- Classify mail as spam or ham using Bayes: https://github.com/varunon9/naive-bayes-classifier
- Applications of Bayes Classification: https://www.quora.com/In-what-real-world-applications-is-Naive-Bayes-classifier-used
- Sentiment Analysis using Bayes: http://suruchifialoke.com/2017-06-10-sentiment-analysis-movie/
- Classify mail using Bayes: https://medium.com/swlh/classify-emails-into-ham-and-spam-using-naive-bayes-classifier-ffddd7faa1ef
Topics

1) Introduction and Bayes Theorem
2) Naive Bayes Classification
3) Bayesian Belief Networks
4) Applications of Naive Bayes
5) Research Paper - Comparing Bayes
Introduction

- Bayesian classifiers are the statistical classifiers based on Bayes' Theorem.

- Bayesian classifiers can predict class membership probabilities i.e. the probability that a given tuple belongs to a particular class.

- It uses the given values to train a model and then it uses this model to classify new data.

Source: https://www.tutorialspoint.com/data_mining/dm_bayesian_classification.htm
Where is it used?
Trying to find the answer

There are only two possible events possible for the given question:

A: It is going to rain tomorrow
B: It will not rain tomorrow.

If you think intuitively

- It's either going to be raining today or it is NOT going to be raining today
- So technically there is 50% CHANCE OF RAIN tomorrow. Correct?
That's too *Naive* even for Bayes!

Bayesian theorem argues that the probability of an event taking place changes if there is information available about a related event.

- This means that if you recall the previous weather conditions for the last week, and you remember that it has actually rained every single day, your answer will no longer be 50%.
- The Bayesian approach provides a way of explaining how you should change your existing beliefs in the light of new evidence.
- Bayesian rule’s emphasis on prior probability makes it better suited to be applied in a wide range of scenarios.

Source: https://mlcorner.wordpress.com/2013/04/28/bayesian-classifier/
What is Bayes Theorem?

- Bayes' theorem, named after 18th-century British mathematician Thomas Bayes, is a mathematical formula for determining conditional probability.

- The theorem provides a way to revise existing predictions or theories given new or additional evidence.

- In finance, Bayes' theorem can be used to rate the risk of lending money to potential borrowers.

Source: https://www.investopedia.com/terms/b/bayes-theorem.asp
Bayes Formula

- The formula for the Bayes theorem is given as follows:

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B|A)}{P(B)}
\]

- Bayes' theorem is also called Bayes' Rule or Bayes' Law.

Source: https://www.investopedia.com/terms/b/bayes-theorem.asp
Small Example

This is Bill.

Bill is 35 years old.

Bill earns $40000/yr

Bill has a very fair credit rating.

Will Bill buy a computer?
Bayes theorem to the rescue!

\[ P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \]

**H**: Hypothesis that Bill will buy the computer

**X**: Bill is 35 years old with fair credit rating and income of 40000$/year

\( P(H|X) \): The probability that Bill will buy the computer **GIVEN** that we know his age, income and credit rating **[Posterior]**

\( P(H) \): Probability that Bill will buy computer **( REGARDLESS** of knowing age, income and credit rating ) **[Prior]**

\( P(X|H) \): Probability that someone is 35 years old, has fair credit rating, earns 40000$/yr AND has **BOUGHT** the computer. **[Likelihood]**

\( P(X) \): Probability that Bill is 35 years old, has fair credit rating, earns 40000$/yr **[Evidence]**
Big Example

Bill now wants to play football!
(Because he is tired of using his computer)
The Naive Bayes **nerd** is here!

I have noted down all the days it was good/bad to play football and the combination of weather metrics on that particular day.

That is perfect. We will be using Naive Bayes algorithm to predict if you should play on a particular day or not.

Source: http://qr.ae/TUTR3L
Let's identify all the factors!

All possible weather combinations:

- Summer
- Monsoon
- Winter
- Sunny
- No Sun
- Windy
- No Wind

Source: http://qr.ae/TUTR3L
Draw frequency tables for each factor

For each of the frequency tables, we will find the likelihoods for each of the cases.

Here, $c = \text{Play}$ and $x = \text{Variables like Season, Sunny & Windy}$.

**Likelihood Table**

<table>
<thead>
<tr>
<th>Season</th>
<th>Play</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>Yes</td>
<td>3/9</td>
<td>2/5</td>
</tr>
<tr>
<td>Monsoon</td>
<td>Yes</td>
<td>4/9</td>
<td>0/5</td>
</tr>
<tr>
<td>Winter</td>
<td>Yes</td>
<td>2/9</td>
<td>3/5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/14</td>
<td>5/14</td>
</tr>
</tbody>
</table>

**Likelihood of 'Yes' given Summer is:**

$$P(c | x) = P(Yes | Summer) = P(Summer | Yes) \times P(Yes) / P(Summer) = (0.33 \times 0.64) / 0.36 = 0.60$$

Source: http://qr.ae/TUTR3L
Find the probability

Let us use the likelihood table to predict whether to play football on
(Season = Winter, Sunny = No, Windy = Yes)

Since the probability is greater than 0.5, we should play football on that day.

\[
P(c \mid x) = P(\text{Play} = \text{Yes} \mid \text{Winter, Sunny = No, Windy = Yes})
\]

\[
= \frac{P(\text{Winter} \mid \text{Yes}) \cdot P(\text{Sunny} = \text{No} \mid \text{Yes}) \cdot P(\text{Windy} = \text{Yes} \mid \text{Yes}) \cdot P(\text{Yes})}{P(\text{Winter}) \cdot P(\text{Sunny} = \text{No}) \cdot P(\text{Windy} = \text{Yes})}
\]

\[
= \frac{(2/9) \cdot (6/9) \cdot (6/9) \cdot (9/14)}{(5/14) \cdot (7/14) \cdot (8/14)} = 0.6223
\]

Source: http://qr.ae/TUTR3L
How to know if results are correct?

The Accuracy of Classification can be found out using a Confusion Matrix.
Confusion Matrix

- **True Positives (TP):** number of positive examples, labeled as such.

- **False Positives (FP):** number of negative examples, labeled as positive.

- **True Negatives (TN):** number of negative examples, labeled as such.

- **False Negatives (FN):** number of positive examples, labeled as negative.

Source: https://rasbt.github.io/mlxtend/user_guide/evaluate/confusion_matrix/
Finding accuracy of classification

**Accuracy** = \((\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})\)

Accuracy gives us the result of total correct predictions out of all the predictions.

**Precision**: \(\text{TP}/(\text{TP} + \text{FP})\)

Precision answers the following question: Out of all the examples the classifier labeled as positive, what fraction were correct?

**Recall**: \(\text{TP}/(\text{TP} + \text{FN})\)

Recall answers: out of all the positive examples there were, what fraction did the classifier pick up?

• The **Homo apriorius** establishes the probability of an hypothesis, no matter what data tell.
• The **Homo pragamiticus** establishes that it is interested by the data only.
• The **Homo frequentistus** measures the probability of the data given the hypothesis.
• The **Homo sapients** measures the probability of the data and of the hypothesis.
• The **Homo bayesianis** measures the probability of the hypothesis, given the data.

Source: http://www.brera.mi.astro.it/~andreon/inference/Inference.html
Data Mining Process

TESTING AND EVALUATION

LEARNING

Preprocessing

CLEANING

SELECTED

Rules or Descriptions

knowledge

Transformed data

Processed Data

Target data

Data

http://www3.cs.stonybrook.edu/~cse634/L3ch6classification.pdf
What are Bayesian Classifiers?

- Statistical classifiers.
- Predict class membership probabilities, such as the probability that a given tuple belongs to a particular class.
- Based on Bayes’ theorem
- Exhibits high accuracy and speed when applied to large databases.
Naive Bayes Classification

Before explaining the mathematical representations, let us see the basic principle of Bayesian classification:

**Predict the most probable class for each instance. How?**

Find out the probability of the previously unseen instance belonging to each class, and then select the most probable class.

A Naive Bayes Classifier is a program which predicts a class value given a set of attributes.

For each known class value,

- Calculate probabilities for each attribute, conditional on the class value.
- Use the product rule to obtain a joint conditional probability for the attributes.
- Use Bayes rule to derive conditional probabilities for the class variable.

Once this has been done for all class values, output the class with the highest probability.
For the Bayes classifier, we need to “learn” two functions, the likelihood and the prior.

Prior: $P(Y = y)$
Likelihood: $P(X = x | Y = y)$

Model Parameters

- **Instance Attributes**: Instances are represented as a vector of attributes.

\[ X = (X_1, \ldots, X_k) \]

- Let there are ‘m’ classes: \( C_1, C_2, \ldots, C_m \).

- Classification is to derive the maximum posteriori, i.e., maximal \( P(C_i|X) \).

- The likelihood now becomes

\[
P(X_1 = x_1, \ldots, X_k = x_k | Y = y)
\]

This affects the number of model parameters.
Model Parameters

The problem with explicitly modeling $P(X_1,\ldots,X_n|Y)$ is that there are usually way too many parameters:

- We’ll run out of space
- We’ll run out of time
- And we’ll need tons of training data (which is usually not available)
- It is computationally expensive to evaluate $P(X|C_i)$
A Naive Bayes Classifier assumes that attributes are conditionally independent given the class. Each feature is conditionally independent of every other feature for a particular class label. This reduces the number of model parameters.

\[ P(X_1, \ldots, X_k|Y) = \prod_{i=1}^{k} P(X_i|Y) \]
Bayes Classification

Naive Bayes works equally well for multi valued attributes also.

- **Model parameters:**
  
  \[ P(Y = y) \quad \text{for all classes } y \]
  
  \[ P(X_i = x \mid Y = y) \quad \text{for all attributes } X_i, \text{values } x \text{ and classes } y \]

- **Decision rule:**

  \[
  f(d) = \arg\max_y P(Y = y) \prod_{i=1}^{k} P(X_i(d) \mid Y = y)
  \]
“Zero” Problem

What if there is a class, $C_i$ and $X$ has an attribute $X_k$ such that none of the samples in $C_i$ has that attribute value?

In that case $P(x_k|C_i) = 0$, which results in $P(X|C_i) = 0$
even though $P(x_k|C_i)$ for all the other attributes in $X$ may be large.
“Zero” Problem - Remedy

- The class conditional probability can be re-estimated with the ‘m-estimate’: m is the number of virtual samples ~ upto 1% of training example.

- Using the Laplacian correction to avoid computing probability values of zero. Here we have 1 more tuple for each attribute-class pair. The “corrected” probability estimates are close to their “uncorrected” counterparts, yet the zero probability value is avoided.
Numeric Underflow Problem

- What’s nice about Naïve Bayes is that it returns probabilities. These probabilities can tell us how confident the algorithm is.

- Since we are multiplying these probabilities, it could lead to a floating-point underflow.

- So it is better to sum logs of probabilities rather than multiplying probabilities.
Bayesian Belief Networks

- Naive Bayesian classifier assumes class conditional independence
- This assumption simplifies computation
- When this assumption is true, Naive Bayesian classifier is the most accurate in comparison with all other classifiers
- However, dependencies can exist between variables
- Bayesian Belief Networks makes no class conditional independence assumption – improvement over Naive Bayesian classifier
Bayesian Belief Networks

- Specifies joint conditional probability distributions
- Allows class conditional independencies to be defined between subsets of variables
- Provides graphical model of causal relationships
- Trained Bayesian Belief Networks can be used for classification
- Also known as Belief Networks, Bayesian Networks and Probabilistic Networks
Bayesian Belief Networks

- Defined by two components
  1. A Directed Acyclic Graph
  2. Set of conditional probability tables
- Each node in the DAG represents a random variable (Discrete or Continuous valued)
- Each node may correspond to actual attributes given in the data or to “hidden variables” believed to form a relationship
- Each edge represents a probabilistic dependence
Bayesian Belief Networks

- If there is an edge from node Y to a node Z, then Y is a parent or immediate predecessor of Z, and Z is a descendant of Y

- Each variable is conditionally independent of its non-descendants in the graph, given its parents
Bayesian Belief Networks

A simple Bayesian Belief Network

Reference: Data Mining Concepts and Techniques 2nd Edition by Jiawei Han and Micheline Kamber, Page 316
Bayesian Belief Networks

- Has Conditional Probability Table (CPT) for each variable.
- CPT for a variable Y specifies the conditional distribution $P(Y | \text{Parents}(Y))$, where Parents(Y) are the parents of Y.
- From previous example:

  \[
  P(\text{LungCancer} = \text{yes} | \text{FamilyHistory} = \text{yes}, \text{Smoker} = \text{yes}) = 0.8
  \]
  \[
  P(\text{LungCancer} = \text{no} | \text{FamilyHistory} = \text{no}, \text{Smoker} = \text{no}) = 0.9
  \]
Bayesian Belief Networks

- Let $X = (x_1, \ldots, x_n)$ be a data tuple described by the variables or attributes $Y_1, \ldots, Y_n$ respectively

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(Y_i))$$

- $P(x_1, \ldots, x_n)$ is the probability of a particular combination of values of $X$, and the values for $P(x_i | Parents(Y_i))$ correspond to the entries in the CPT for $Y$
Bayesian Belief Networks

- A node within the network can be selected as an output node, representing a class label attribute
- There may be more than one output node
- Various algorithms for learning can be applied to the network
- Rather than returning a single class label, the classification process can return a probability distribution that gives the probability of each class
Bayesian Belief Networks - Training

Case 1

- Network topology is given in advance
- Human experts have knowledge of the conditional dependencies which helps in designing the network
- Experts specifies conditional probabilities for the nodes that participate in direct dependencies
- There are various methods for training the belief network
- Example: Gradient Descent
Bayesian Belief Networks - Training

Case 2

- Network topology is inferred from data
- There are several algorithms for learning the topology from the training data given observable variables
- The problem is one of Discrete Optimization
Naive Bayes in Real Life
Text classification:

- Naive Bayes Classifier application.

Why Text Classification?
- Classify web pages by topic
- Learning which articles are of interest
- Information extraction
- Internet filters.

https://www.slideshare.net/ashrafmath/naive-bayes-15644818
Examples of Text classification:

- **CLASSES=BINARY**
  - “spam” / “not spam”

- **CLASSES =TOPICS**
  - “finance” / “sports” / “politics”

- **CLASSES =OPINION**
  - “like” / “hate” / “neutral”

- **CLASSES =TOPICS**
  - “AI” / “Theory” / “Graphics”

- **CLASSES =AUTHOR**
  - “Shakespeare” / “Marlowe” / “Ben Jonson”

https://www.slideshare.net/ashrafmath/naive-bayes-15644818
Naive Bayes Approach

- Build the vocabulary as the list of all distinct words that appear in all the documents in the training set.
- Remove stop words and markings
- Words in the vocabulary becomes the attributes, classification is independent of position of words
- Train the classifier based on the training data set
- Evaluate the results on Test data.
Simple text classifier.

```
MariaDB [naiveBayes]> select * from trainingSet;

+---------+---------------------------------------------+------+
<table>
<thead>
<tr>
<th>S_NO</th>
<th>document</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>Have a pleasurable stay! Get up to 30% off + Flat 20% Cashback on Oyo Room bookings done via Paytm</td>
<td>spam</td>
</tr>
<tr>
<td>78</td>
<td>Lets Talk Fashion! Get flat 40% Cashback on Backpacks, Watches, Perfumes, Sunglasses &amp; more</td>
<td>spam</td>
</tr>
<tr>
<td>79</td>
<td>Opportunity with Product firm for Fullstack</td>
<td>ham</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backend</td>
</tr>
<tr>
<td>80</td>
<td>Javascript Developer, Fullstack Developer in</td>
<td>ham</td>
</tr>
<tr>
<td></td>
<td>Bangalore- Urgent Requirement</td>
<td></td>
</tr>
</tbody>
</table>

ssh root@dhcp230.fsl.cs.sunysb.edu
```
```
<table>
<thead>
<tr>
<th>S_NO</th>
<th>word</th>
<th>count</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>have</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>98</td>
<td>pleasurable</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>99</td>
<td>stay</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>100</td>
<td>get</td>
<td>2</td>
<td>spam</td>
</tr>
<tr>
<td>101</td>
<td>off</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>102</td>
<td>flat</td>
<td>2</td>
<td>spam</td>
</tr>
<tr>
<td>103</td>
<td>cashback</td>
<td>2</td>
<td>spam</td>
</tr>
<tr>
<td>104</td>
<td>oyo</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>105</td>
<td>room</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>106</td>
<td>bookings</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>107</td>
<td>done</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>108</td>
<td>via</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>109</td>
<td>paytm</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>110</td>
<td>lets</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>111</td>
<td>talk</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>112</td>
<td>fashion</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>113</td>
<td>backpacks</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>114</td>
<td>watches</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>115</td>
<td>perfumes</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>116</td>
<td>sunglasses</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>117</td>
<td>more</td>
<td>1</td>
<td>spam</td>
</tr>
<tr>
<td>118</td>
<td>opportunity</td>
<td>1</td>
<td>ham</td>
</tr>
<tr>
<td>119</td>
<td>product</td>
<td>1</td>
<td>ham</td>
</tr>
<tr>
<td>120</td>
<td>firm</td>
<td>1</td>
<td>ham</td>
</tr>
</tbody>
</table>
```

```
[root@dhcp230 naive-bayes-classifier]# php main.php "100% cashback offer" spam
[root@dhcp230 naive-bayes-classifier]# php main.php "javascript developer" ham
[root@dhcp230 naive-bayes-classifier]#
```
Advantages:

- Requires a small amount of training data to estimate the parameters.
- Good results obtained in most cases
- Easy to implement

Disadvantages:

- Practically, dependencies exist among variables.
  
  eg: hospitals : patients: Profile: age, family history etc.
  Symptoms: fever, cough etc., Disease: lung cancer, diabetes, etc.

- Dependencies among these cannot be modelled by Naive bayesian classifier.
Weather prediction:

TODAY
62°F | 37°C
morning fog, partly cloudy

TOMORROW
58°F | 41°C
rain showers, cloudy

Recommendation System:

🌟🌟🌟🌟🌟
Awesome movie!

https://www.quora.com/In-what-real-world-applications-is-Naive-Bayes-classifier-used
http://suruchifialoke.com/2017-06-10-sentiment-analysis-movie
Medical Diagnosis

Digit Recognition

https://www.quora.com/In-what-real-world-applications-is-Naive-Bayes-classifier-used
Title: Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques

Authors: Chaitrali S. Dangare & Sulabha S. Apte, PhD


Publishing Period: June 2012
Bayes Network has been present since time immemorial. For years, it proved to be a simple yet powerful classifier that could be used for prediction.

The computation power required to run a bayesian classifier is considerately simpler when compared to most of the modern day classification algorithms.

This paper debates the use of bayesian classifier along with IDT and NN and their usage in a Heart Disease Prediction System in the medical field.
We have used a cutting edge classifier to look into your medical report and analyse you. It is our new innovation.

So, What Does it Say?
That You are Pregnant!

Are You Serious!

Classifiers are Important!!!

Source: bigstock-healthcare-technology-and-med-83203256.jpg
Data Set Used

- The publicly available heart disease database is used.
- The Cleveland Heart Disease database consists of 303 records & Statlog Heart Disease database consists of 270 records.
- The data set consists of 3 types of attributes: Input, Key & Predictable attribute.
- The analysis was performed on 13 attributes initially followed by 2 more attributes separately.
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>age</td>
<td>1 = male 0 = female</td>
</tr>
<tr>
<td>sex</td>
<td>male or female</td>
<td>1 = typical type 1 2 = typical type agina 3 = non-agina pain 4 = asymptomatic</td>
</tr>
<tr>
<td>cp</td>
<td>chest pain</td>
<td>Continuous value in mm hg</td>
</tr>
<tr>
<td>thetspb</td>
<td>resting blood pressure</td>
<td>Continuous value in mm/dl</td>
</tr>
<tr>
<td>chol</td>
<td>serum cholestrol</td>
<td>Continuous value in mm/dl</td>
</tr>
<tr>
<td>restecg</td>
<td>rest ecg results</td>
<td>0 = normal 1 = having_ST_T wave abnormal 2 = left ventricular hypertrophy</td>
</tr>
<tr>
<td>fbs</td>
<td>fasting blood sugar</td>
<td>1 ≥ 120 mg/dl 0 ≤ 120 mg/dl</td>
</tr>
<tr>
<td>thealach</td>
<td>max heart rate</td>
<td>Continuous value</td>
</tr>
<tr>
<td>exang</td>
<td>exercise induced agina</td>
<td>0= no 1 = yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>oldpeak</td>
<td>ST depression induced by exercise relative to rest</td>
<td>Continuous value</td>
</tr>
<tr>
<td>slope</td>
<td>Slope of the peak exercise</td>
<td>1 = unsloping 2 = flat 3 = ownsloping</td>
</tr>
<tr>
<td>ca</td>
<td>Number of major vessels colored by flourish</td>
<td>0-3 value</td>
</tr>
<tr>
<td>thal</td>
<td>Defect type</td>
<td>3 = normal 6 = fixed 7 = reversible defect</td>
</tr>
</tbody>
</table>

### Table 2.1 Primary Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>obes</td>
<td>obesity</td>
<td>1 = yes 0 = no</td>
</tr>
<tr>
<td>smoke</td>
<td>smoking</td>
<td>1= past 2 = current 3 = never</td>
</tr>
</tbody>
</table>

### Table 2.1 Additional Attributes
Performing Naive Bayes:

- Naive Bayes classifier is based on Bayes theorem.
- This classifier algorithm uses **conditional independence**.

Let $X=\{x_1, x_2, \ldots, x_n\}$ be a set of $n$ attributes.
- In Bayesian, $X$ is considered as evidence and $H$ be some hypothesis means, the data of $X$ belongs to specific class $C$.
- We have to determine $P(H|X)$, the probability that the hypothesis $H$ holds given evidence i.e. data sample $X$.
- According to Bayes theorem the $P(H|X)$ is expressed as

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$
# Performance

<table>
<thead>
<tr>
<th>Actual Output/Prediction</th>
<th>a ( has heart disease )</th>
<th>b ( no heart disease )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a ( has heart disease )</td>
<td>TP</td>
<td>TN</td>
</tr>
<tr>
<td>b ( no heart disease )</td>
<td>FP</td>
<td>FN</td>
</tr>
</tbody>
</table>

## Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>110</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>10</td>
<td>145</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>123</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>5</td>
<td>138</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td>b</td>
<td>18</td>
<td>145</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>85</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>185</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>117</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>106</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>164</td>
</tr>
</tbody>
</table>

## Source

[Source](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.258.8158&rep=rep1&type=pdf) | Figure 3-6
Comparison with IDT & NN

ID3 and NN classifiers are also implemented.

<table>
<thead>
<tr>
<th>Classification Techniques</th>
<th>13 attributes</th>
<th>15 attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>94.44</td>
<td>90.74</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>96.66</td>
<td>99.62</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>99.25</td>
<td>100</td>
</tr>
</tbody>
</table>

Prediction Accuracy

![Chart showing prediction accuracy for Naïve Bayes, Decision Trees, and Neural Networks with 13 and 15 attributes.](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.258.8158&rep=rep1&type=pdf) | Figure 2
Conclusion

- The overall objective of the work is to predict more accurately the presence of heart disease.
- It has been seen that Neural Networks provides accurate results as compared to Decision trees & Naive Bayes.
- Naive Bayes has a serious drawback where the events are considered mutually independent of each other.
- In Real life, it is very much difficult for events to be exclusively unrelated and naive bayes fails to make use of the correlation.
- However given the compute power required ,it is a reasonably efficient classifier.
FREE COMPUTE POWER !!!

Source: https://i.ytimg.com/vi/E7myDAKBgRs/maxresdefault.jpg
Want Medical Records Accessed?

YES

NO

Thank You

Questions?

Source: bigstock-healthcare-technology-and-med-83203256.jpg