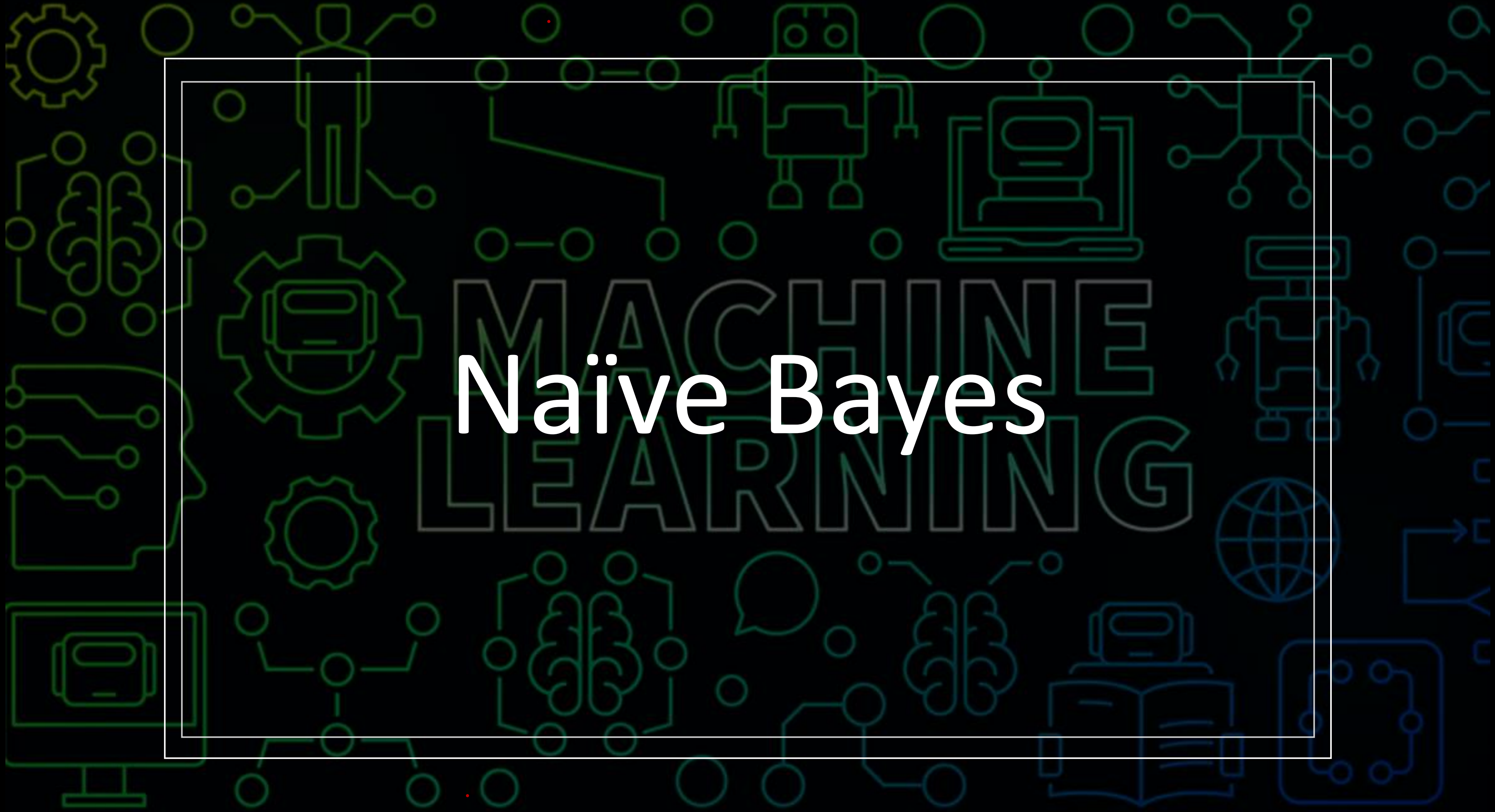
Naïve Bayes



What is Naïve Bayes ?

Naive Bayes is a simple but surprisingly powerful algorithm for predictive modeling.

NAIVE

BAYES

Classification Technique

Bayes' Theorem

Given a hypothesis H and evidence E , Bayes' theorem states that the relationship between the probability of the hypothesis before getting the evidence P(H) and the probability of the hypothesis after getting the evidence P(H|E) is

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

Conditional Probability Formula

Probability of
A and *B*

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

Probability of
A given *B*

Probability of *B*

Conditional Probability Formula

$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

PROOF OF BAYES THEOREM

The probability of two events A and B happening, $P(A \cap B)$, is the probability of A, $P(A)$, times the probability of B given that A has occurred, $P(B|A)$.

$$P(A \cap B) = P(A)P(B|A) \tag{1}$$

On the other hand, the probability of A and B is also equal to the probability of B times the probability of A given B.

$$P(A \cap B) = P(B)P(A|B) \tag{2}$$

Equating the two yields:

$$P(B)P(A|B) = P(A)P(B|A) \tag{3}$$

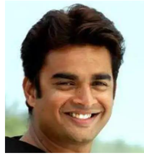
and thus

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)} \tag{4}$$

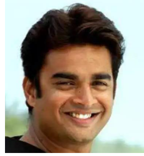
This equation, known as Bayes Theorem is the basis of statistical inference.

The probability of two events A and B happening, P(A∩B), is the probability of A, P(A), times the probability of B given that A has occurred, P(B|A).

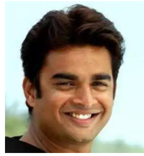
$$P(A \cap B) = P(A)P(B|A) \tag{1}$$


உண்மை

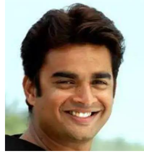

பொய்


உண்மை


பொய்


உண்மை

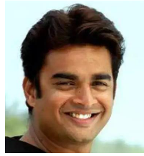

பொய்


பொய்


உண்மை

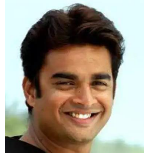
On the other hand, the probability of A and B is also equal to the probability of B times the probability of A given B.

$$P(A \cap B) = P(B)P(A|B) \tag{2}$$



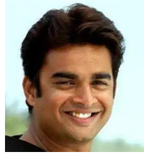
உண்மை

பொய்



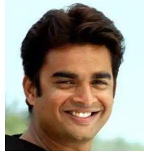
உண்மை

பொய்



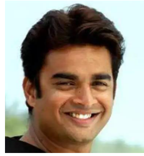
உண்மை

பொய்



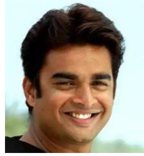
பொய்

உண்மை



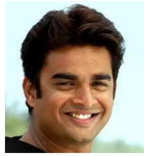
உண்மை

பொய்



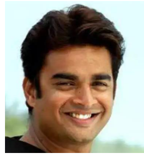
உண்மை

பொய்



உண்மை

பொய்

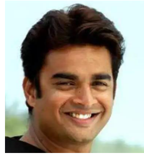


பொய்

உண்மை

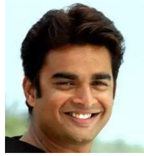
$$P(A|B) = P(A)\frac{P(B|A)}{P(B)} \tag{4}$$

This equation, known as Bayes Theorem is the basis of statistical inference.



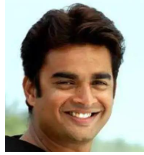
உண்மை

பொய்



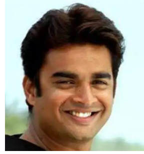
உண்மை

பொய்



உண்மை

பொய்



பொய்

உண்மை

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)} \tag{4}$$

This equation, known as Bayes Theorem is the basis of statistical inference.

?

Bayes' Theorem Proof

Likelihood
How probable is the evidence
Given that our hypothesis is true?

Prior
How probable was our hypothesis
Before observing the evidence?

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

Posterior
How probable is our Hypothesis
Given the observed evidence?
(Not directly computable)

Marginal
How probable is the new evidence
Under all possible hypothesis?

Classification Steps

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

Frequency Table		Play	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	3	2

Frequency Table		Play	
		Yes	No
Humidity	High	3	4
	Normal	6	1

Frequency Table		Play	
		Yes	No
Wind	Strong	6	2
	Weak	3	3

Classification Steps

Likelihood Table		Play: Yes	Play: No	
Outlook	Sunny	3/10	2/4	5/14
	Overcast	4/10	0/4	4/14
	Rainy	3/10	2/4	5/14
		10/14	4/14	

$P(x|c) = P(Sunny|Yes) = 3/10 = 0.3$

$P(x) = P(Sunny) = 5/14 = 0.36$

$P(c) = P(Yes) = 10/14 = 0.71$

Likelihood of 'Yes' given Sunny is

$$P(c|x) = P(Yes|Sunny) = P(Sunny|Yes)^{*}\ P(Yes) / P(Sunny) = (0.3 \times 0.71) / 0.36 = 0.591$$

Similarly Likelihood of 'No' given Sunny is

$$P(c|x) = P(No|Sunny) = P(Sunny|No)^{*}\ P(No) / P(Sunny) = (0.4 \times 0.36) / 0.36 = 0.40$$

Classification Steps

Likelihood table for Humidity

Likelihood Table		Play: Yes	Play: No	
Humidity	High	3/9	4/5	7/14
	Normal	6/9	1/5	7/14
		9/14	5/14	

Likelihood table for Wind

Likelihood Table		Play: Yes	Play: No	
Wind	Weak	6/9	2/5	8/14
	Strong	3/9	3/5	6/14
		9/14	5/14	

$P(Yes|High)$ = 0.33 x 0.6 / 0.5 = 0.42

$P(No|High)$ = 0.8 x 0.36 / 0.5 = 0.58

$P(Yes|Weak)$ = 0.67 x 0.64 / 0.57 = 0.75

$P(No|Weak)$ = 0.4 x 0.36 / 0.57 = 0.25

Classification Steps

Suppose we have a day with the following values

Outlook	=	Rain
Humidity	=	High
Wind	=	Weak
Play	=	?

Classification Steps

Suppose we have a day with the following values

Outlook	=	Rain
Humidity	=	High
Wind	=	Weak
Play	=	?

Likelihood of 'Yes' on that Day = P(Outlook = Rain|Yes)*P(Humidity= High|Yes)* P(Wind= Weak|Yes)*P(Yes)

Classification Steps

Suppose we have a day with the following values

Outlook	=	Rain
Humidity	=	High
Wind	=	Weak
Play	=	?

Likelihood of 'Yes' on that Day = P(Outlook = Rain|Yes)*P(Humidity= High|Yes)* P(Wind= Weak|Yes)*P(Yes)

= 2/9 * 3/9 * 6/9 * 9/14 = 0.0199

Likelihood of 'No' on that Day = P(Outlook = Rain|No)*P(Humidity= High|No)* P(Wind= Weak|No)*P(No)

= 2/5 * 4/5 * 2/5 * 5/14 = 0.0166

Classification Steps

P(Yes) = 0.0199 / (0.0199+ 0.0166) = 0.55

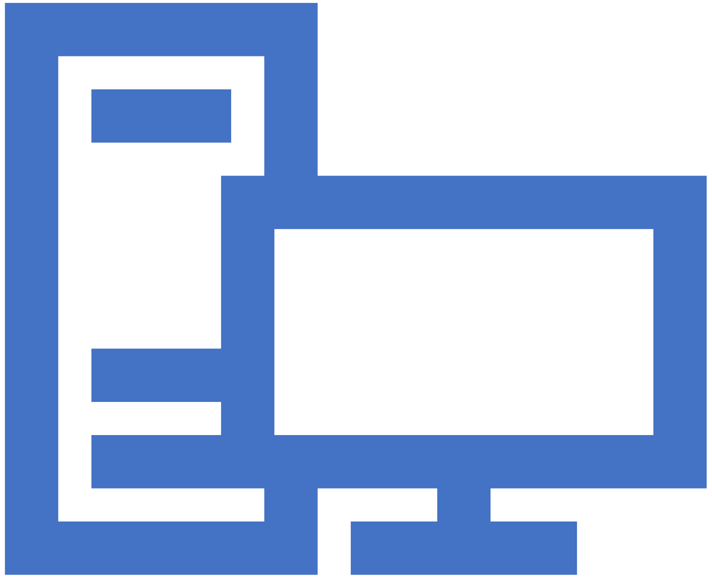
P(No) = 0.0166 / (0.0199+ 0.0166) = 0.45

Our model predicts that there is a 55% chance there will be game tomorrow

Naive Bayes (NB) is ‘naive’ because it makes the assumption that features of a measurement are independent of each other. This is naive because it is (almost) never true.

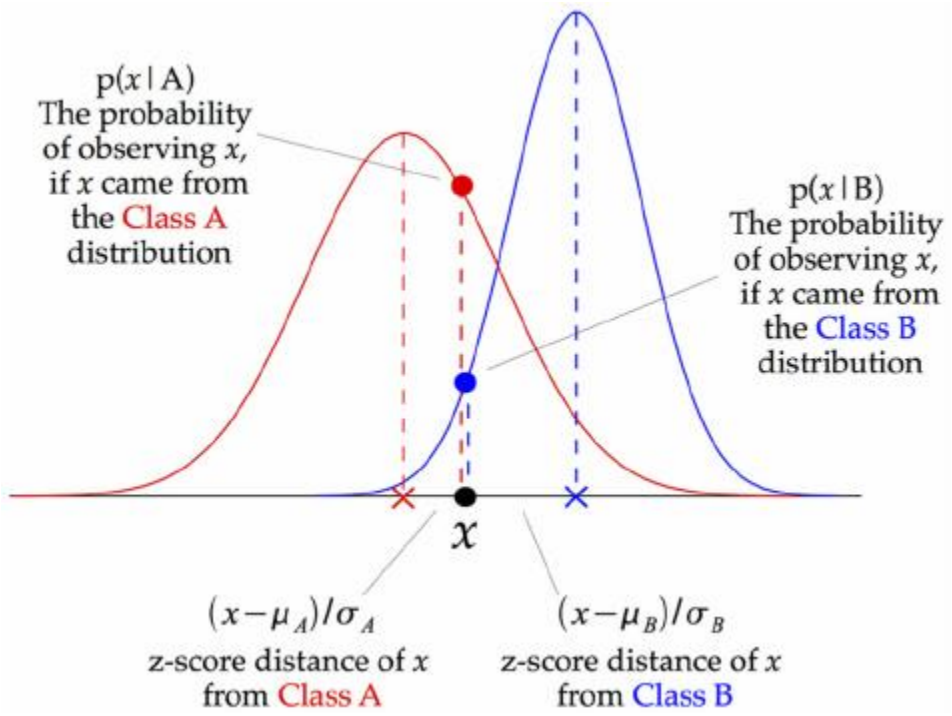
Likelihood of ‘Yes’ on that Day = P(Outlook = Rain|Yes)*P(Humidity= High|Yes)* P(Wind= Weak|Yes)*P(Yes)

What about dataset with Real Numbers?



We have a savior called **Probability Distributions**

$$p(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}}$$



Gaussian Naive Bayes

- Naive Bayes can be extended to real-valued attributes, most commonly by assuming a Gaussian distribution.
- This extension of naive Bayes is called Gaussian Naive Bayes. Other functions can be used to estimate the distribution of the data, but the Gaussian (or Normal distribution) is the easiest to work with because you only need to estimate the mean and the standard deviation from your training data.

Compute Mean and Standard Deviation

- This is as simple as calculating the mean and standard deviation values of each input variable (x) for each class value.

mean(x) = 1/n * sum(x)

- Where n is the number of instances and x are the values for an input variable in your training data.
- We can calculate the standard deviation using the following equation:

standard deviation(x) = sqrt(1/n * sum(xi-mean(x)^2))

- This is the square root of the average squared difference of each value of x from the mean value of x, where n is the number of instances, sqrt() is the square root function, sum() is the sum function, xi is a specific value of the x variable for the i’th instance and mean(x) is described above, and ^2 is the square.

Sample Data (X_i) :

10	12	14	15	17	18	18	24

n = 8 Mean = $\overline{X}$ = 16

$$S = \sqrt{\frac{(10-\overline{X})^2 + (12-\overline{X})^2 + (14-\overline{X})^2 + \cdots + (24-\overline{X})^2}{n-1}}$$

$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{7}}$$

$$= \sqrt{\frac{130}{7}} = \boxed{4.308}$$

A measure of the "average" scatter around the mean

Computing the probability

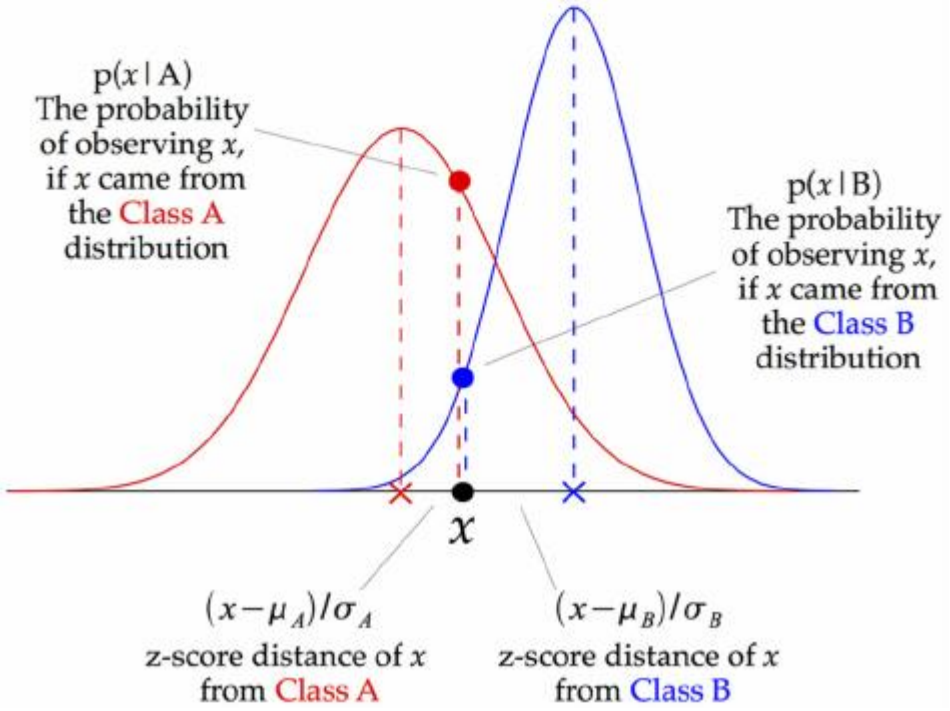
- Probabilities of new x values are calculated using the Gaussian Probability Density Function(PDF).
- When making predictions these parameters can be plugged into the Gaussian PDF with a new input for the variable, and in return the Gaussian PDF will provide an estimate of the probability of that new input value for that class.

pdf(x, mean, sd) = (1 / (sqrt(2 * PI) * sd)) * exp(-((x-mean^2)/(2*sd^2)))

- Where pdf(x) is the Gaussian PDF, sqrt() is the square root, mean and sd are the mean and standard deviation calculated above, PI is the numerical constant, exp() is the numerical constant e or Euler’s number raised to power and x is the input value for the input variable.

Gaussian Probability Distribution

$$p(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}}$$



Gaussian Naïve Bayes

- We can then plug in the probabilities into the equation above to make predictions with real-valued inputs.
- For example, adapting one of the above calculations with numerical values for weather and car:

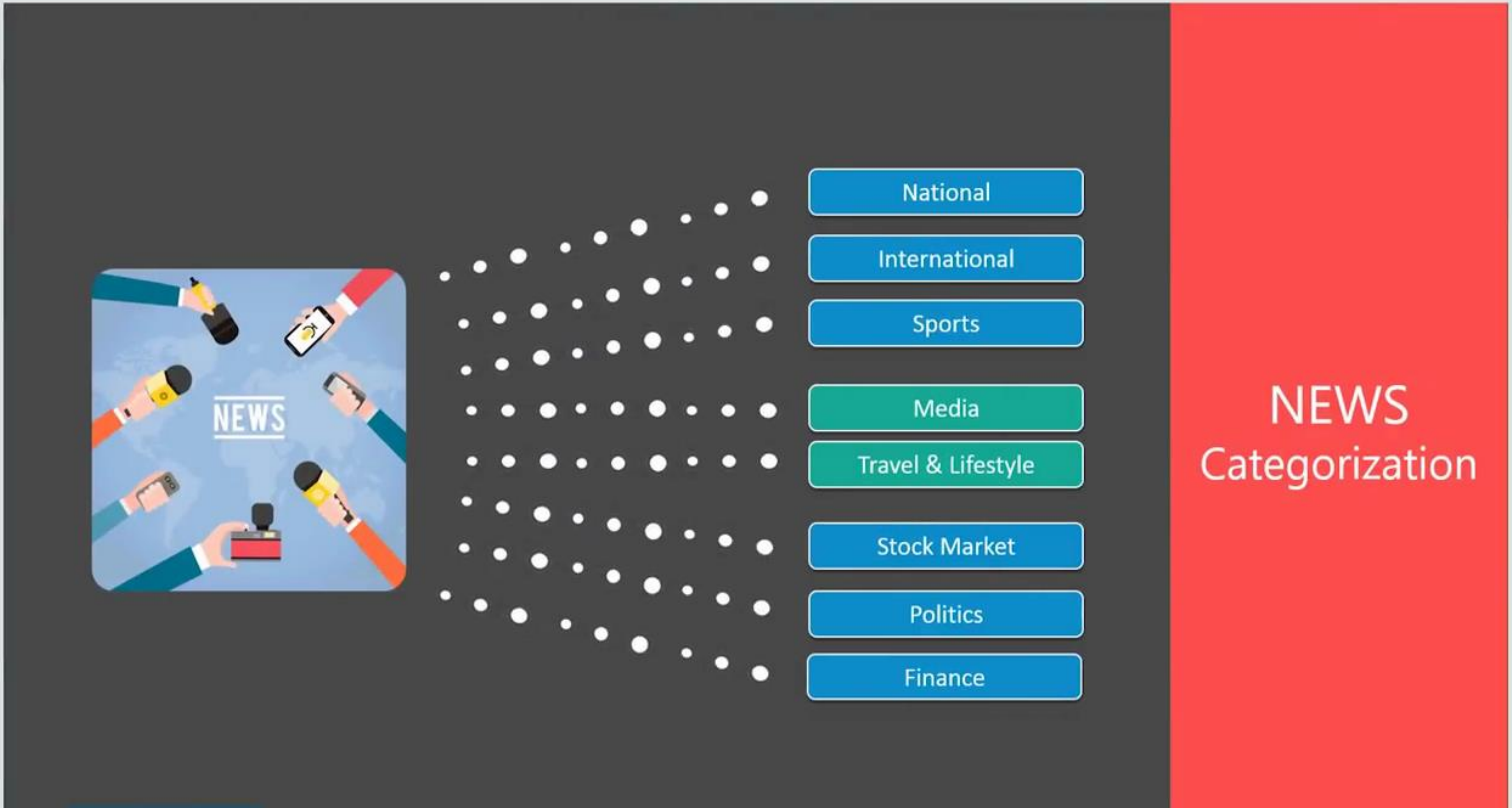
go-out = P(pdf(weather)|class=go-out) * P(pdf(car)|class=go-out) * P(class=go-out)

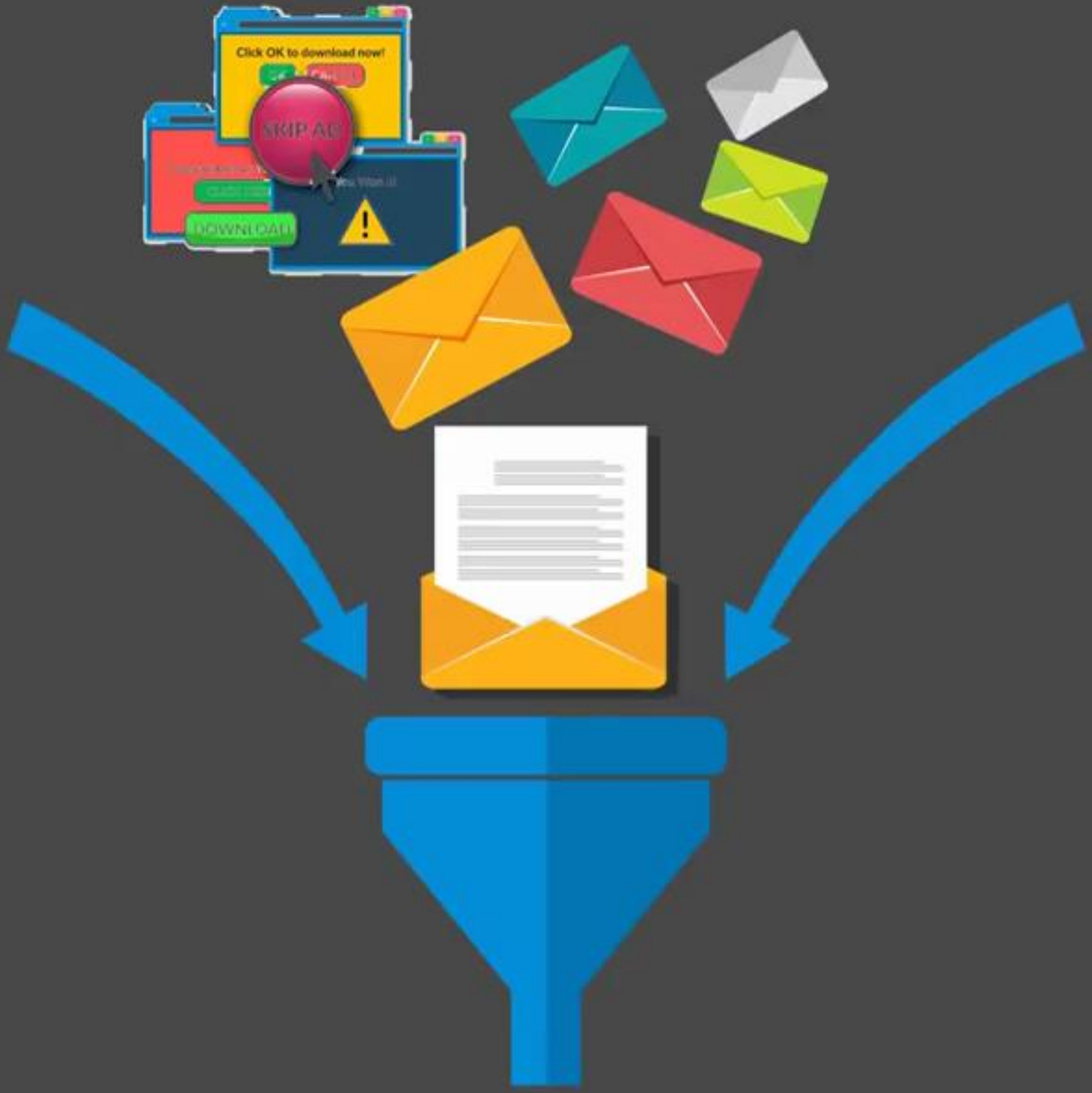
Other Naïve Bayes Types

Other popular Naive Bayes classifiers are:

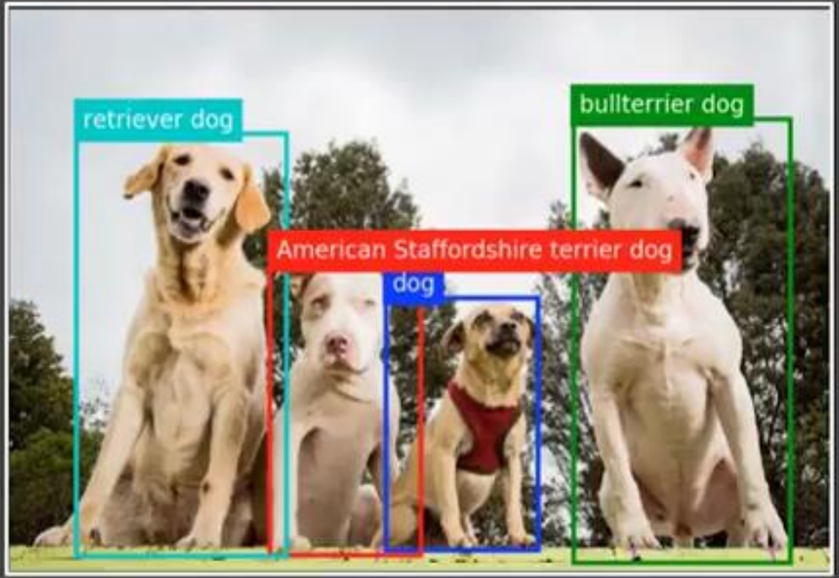
- **Multinomial Naive Bayes**: Feature vectors represent the frequencies with which certain events have been generated by a **multinomial distribution**. This is the event model typically used for document classification.
- **Bernoulli Naive Bayes**: In the multivariate Bernoulli event model, features are independent booleans (binary variables) describing inputs. Like the multinomial model, this model is popular for document classification tasks, where binary term occurrence(i.e. a word occurs in a document or not) features are used rather than term frequencies(i.e. frequency of a word in the document).

Naïve Bayes Applications

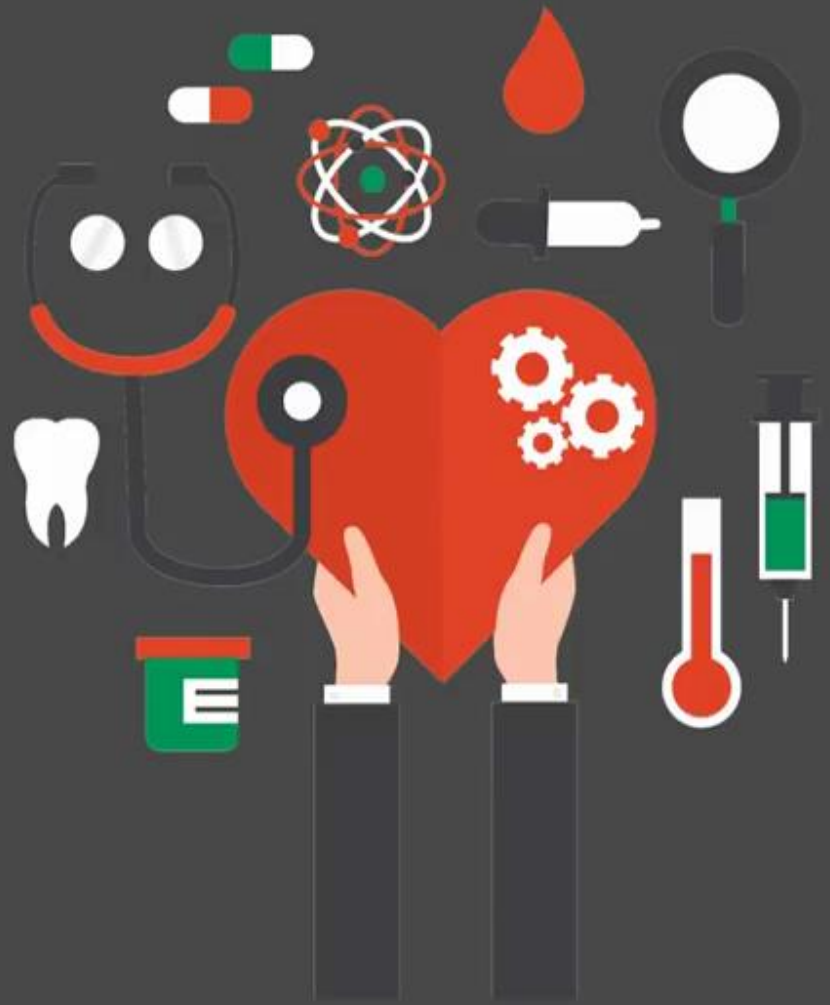

NEWS
National
International
Sports
Media
Travel & Lifestyle
Stock Market
Politics
Finance
NEWS
Categorization



SPAM Filtering



OBJECT & FACE Recognition



MEDICAL Diagnosis

WEATHER Prediction

Slide Credits:
Edureka __/__

அய்யா...
ரொ... ரொம்ப
நன்றிங்கய்யா