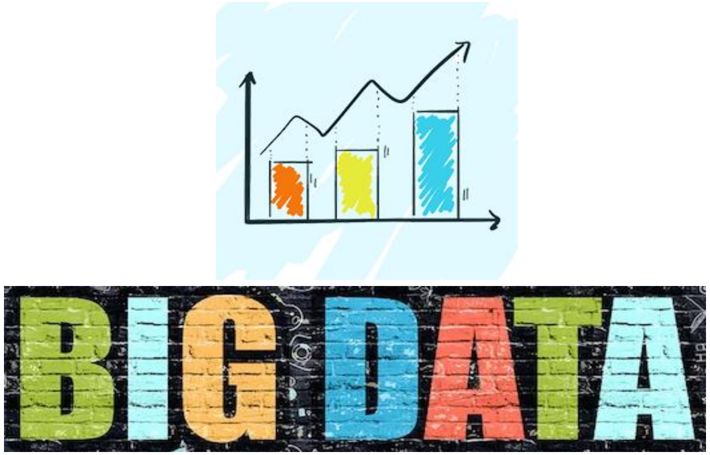
sohailimran@yahoo.com



Sohail IMRAN سهيل عمران

Classification

Ref: Web

naive bayes

Naive Bayes is among one of the simplest, but most powerful algorithms for **classification** based on Bayes' Theorem with an assumption of independence among predictors.
The Naive Bayes model is easy to build and particularly useful for very large data sets.
There are two parts to this algorithm:

- Naive
- Bayes

The Naive Bayes classifier assumes that the presence of a feature in a class is unrelated to any other feature.
Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that a particular fruit is an apple or an orange or a banana, and that is why it is known as **"Naive."**

bayes' theorem

In statistics and probability theory, Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. It serves as a way to figure out conditional probability.

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 :

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

For this reason,

P(H) is called the prior probability, while

P(H|E) is called the posterior probability.

The factor that relates the two, P(H|E)/P(E), is called the likelihood ratio.
Using these terms, Bayes' theorem can be rephrased as:

"**The posterior probability equals the prior probability times the likelihood ratio**."

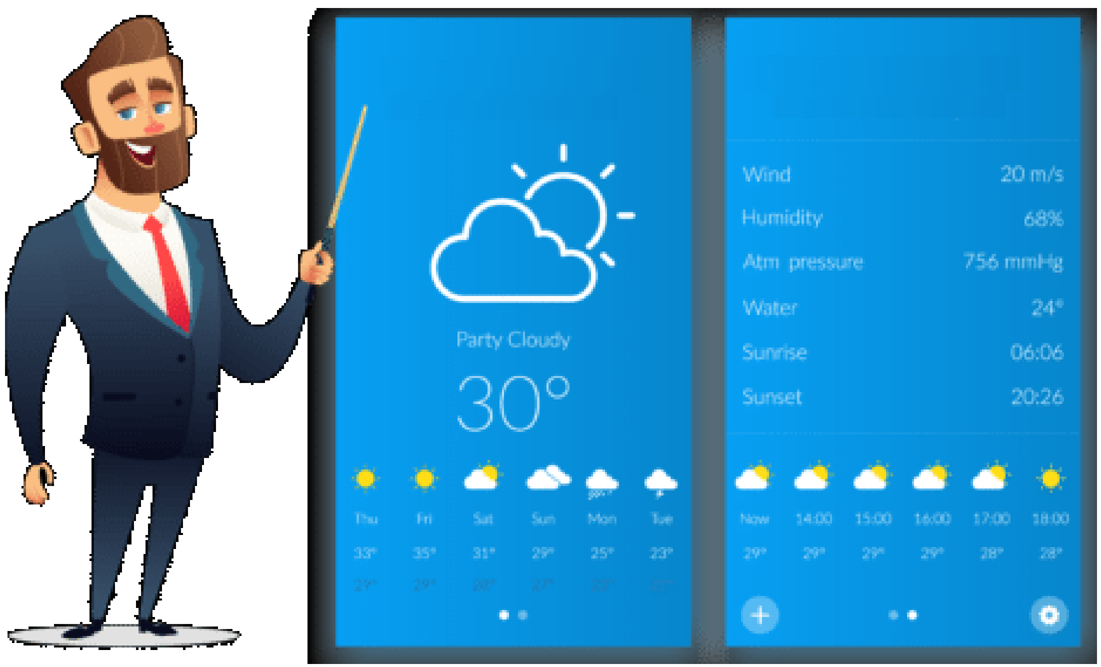
steps

01	CALCULATE PRIOR PROBABILITY FOR GIVEN CLASS LABELS
02	CALCULATE CONDITIONAL PROBABILITY WITH EACH ATTRIBUTE FOR EACH CLASS
03	MULTIPLY SAME CLASS CONDITIONAL PROBABILITY.
04	MULTIPLY PRIOR PROBABILITY WITH STEP 3 PROBABILITY.
05	SEE WHICH CLASS HAS HIGHER PROBABILITY, HIGHER PROBABILITY CLASS BELONGS TO GIVEN INPUT SET STEP.

example

Let's continue our Naive Bayes tutorial and predict the future with some weather data.

Here we have our data, which comprises the day, Weather and Temperature. The final column is 'Play,' i.e., can we play outside, which we have to predict.



Whether	Temperature	Play
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No
Overcast	Cool	Yes
Sunny	Mild	No
Sunny	Cool	Yes
Rainy	Mild	Yes
Sunny	Mild	Yes
Overcast	Mild	Yes
Overcast	Hot	Yes
Rainy	Mild	No

frequency table

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table

Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9

likelihood

Prior Probabilities

Likelihood Table 1

Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

Posterior Probabilities

Likelihood Table 2

Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

Likelihood of 'Yes' given 'Overcast' is:

P(c|x) = P(Yes|Overcast)

= P(Overcast|Yes)* P(Yes) / P(Overcast)

= (0.44. x 0.64) /0.29 = 0.98

The likelihood of 'No' given 'Overcast' is:

P(c|x) = P(No|Overcast)

= P(Overcast|No)* P(No) / P(Overcast)

= (0 x 0.36) /0.29 = 0

prediction

Suppose we have a **Day** with the following values :

- Outlook = Overcast
- Temperature = Mild
- Play = ?

P(Play= Yes | Weather=Overcast, Temp=Mild) = P(Weather=Overcast, Temp=Mild | Play= Yes) P(Play=Yes) ....(1)
P(Weather=Overcast, Temp=Mild | Play= Yes)= P(Overcast |Yes) P(Mild |Yes) ………..(2)

1.Calculate Prior Probabilities: P(Yes)= 9/14 = 0.64
2.Calculate Posterior Probabilities: P(Overcast |Yes) = 4/9 = 0.44 P(Mild |Yes) = 4/9 = 0.44
3.Put Posterior probabilities in equation (2)
P(Weather=Overcast, Temp=Mild | Play= Yes) = 0.44 * 0.44 = 0.1936
1.Put Prior and Posterior probabilities in equation (1)
P(Play= Yes | Weather=Overcast, Temp=Mild)
= 0.1936*0.64 = 0.124

Similarly
1.Calculate Prior Probabilities: P(No)= 5/14 = 0.36
2.Calculate Posterior Probabilities: P(Weather=Overcast |Play=No) = 0/9 = 0 P(Temp=Mild | Play=No)=2/5=0.4
3.Put posterior probabilities in equation
P(Weather=Overcast, Temp=Mild | Play= No) = 0 * 0.4= 0
1.Put prior and posterior probabilities in equation)
P(Play= No | Weather=Overcast, Temp=Mild) = 0*0.36=0

Our model predicts that there is a 12**%** chance there will be a game tomorrow.