Naive Bayesian Classification

- Probabilistic learning → har mumkin hypothesis ka clear probability nikalti hy.
- Incremental training → training example se probability update hoti rehti hy (increase/decrease). Model gets better with time.
- Probabilistic Prediction → To handle uncertainty, there can be multiple hypothesis phr un ko probabilities kay mutabiq weight diye gye.

- We know whether each of events S, M, $\bar{E}$ occured.
- We want to find out whether event P is true

$Pr(P \mid S, M, \bar{E})$ ⇒ Means a is banday ka computer lanay ka chance kitna hy.

By Baye's Theorum:

$$Pr(A \mid B) = \frac{Pr(A) \times Pr(B \mid A)}{Pr(B)}$$

$$Pr(P \mid S, M, \bar{E}) = \frac{Pr(P)}{Pr(S, M, \bar{E})} \times Pr(S, M, \bar{E} \mid P)$$

$$= \frac{Pr(P)}{Pr(S, M, \bar{E})} \times Pr(S \mid P) \times P(M \mid P) \times P(\bar{E} \mid P)$$

- Naive Bayes assumes S, M, $\bar{E}$ are independent from each other. when P is fixed.

- The class (buyer or non-buyer) ka product karke nikalay ge, final prediction is that.

Naive Bayes sometimes called Idiot Bayes

- Naive Bayesian Classification.
  - Test set represents a young student with fair credit rating.

$$Pr(P \mid S, \neg M, \neg E)$$

$$\frac{Pr(P)}{Pr(S, \neg M, \neg E)} \times Pr(S, \neg M, \neg E \mid P)$$

  - statistical independence is assumed

$$\frac{Pr(P)}{Pr(S, \neg M, \neg E)} \times Pr(S \mid P) \cdot Pr(\neg M \mid P) \cdot Pr(\neg E \mid P)$$

- Example:-

Positives (P) → Buy Computers (9/14)

Negetives (n) → do not buy computer (5/14)

- Now we calculate probabilities for each column.

1) Age

<=30	youth	P(youth | P)	2/9	P(youth | n)	3/5
31-40	mid	P(mid | P)	4/9	P(youm | n)	0/5
>40	senior	P(senior | P)	3/9	P(youm | n)	2/5

2) Income.

P(low | P)	3/9	P(low | n)	1/5
P(medium | P)	4/9	P(medium | n)	2/5
P(high | P)	2/9	P(high | n)	2/5.

3) Student

P (yes | P) 6/9 P (yes | n) 1/5

P (No | P) 3/9 P (No | n) 4/5

4) Credit

P (fair | P) 6/9 P (fair | n) 2/5

P (excellent | P) 3/9 P (excellent | n) 3/5

- One new record. X

X = age = youth, student = yes.

income = low, credit rating = fair.

	Positives	negatives.
youth = age	2/9	3/5
Yes = student	6/9	1/5
fair = credit rating	6/9	2/5
low = income	3/9	1/5

Positives = P (youth | P) x P (yes | P) x P (fair | P) x P (low | P) x P(P)

= 2/9 x 6/9 x 6/9 x 3/9 = 0.0211

Negitives = 3/5 x 1/5 x 1/5 x 2/5 x 5/14.

= 0.0034

P > n

So he will buy.