Naive Bayes

→ It is a supervised machine learning algorithm which is used to solve classification problem, by using baye's theorem.

Note: It is called Naive because it assumes that all of the independent features are independent of each other.

* **Independent Events :**

Those events whose occurrence is not dependent on any other event.

Eg: → Rolling a die $\in \{1, 2, 3, 4, 5, 6\}$

$P(1) = 1/6, \quad P(2) = 1/6, \quad P(6) = 1/6$

→ Tossing a coin $\in \{H, T\}$

$P(H) = 1/2, \quad P(T) = 1/2$

* **Dependent Events :**

Those events whose occurrence is dependent on previous event.

Eg: Box → | 3 Red, 2 Green | $P(R) = 3/5$

↓

| 2 Red, 2 Green | $P(G) = 2/4$

$\therefore P(R \text{ and } G) = P(R) * P(G|R)$

↳ conditional probability

* **Baye's Theorem :**

It determines the probability of an event based on prior knowledge of other related events which have already happened.

$$P(A \text{ and } B) = P(B \text{ and } A)$$

$$\Rightarrow P(A) * P(B|A) = P(B) * P(A|B)$$

$$\Rightarrow \boxed{P(B|A) = \frac{P(B) * P(A|B)}{P(A)}} \rightarrow \text{Baye's Theorem}$$

Dependent Event

Independent Event

→ Generalized Form for ML :

$$P(Y|x_1, x_2, \ldots x_n) = \frac{P(Y) * P(x_1|Y) * P(x_2|Y) \ldots P(x_n|Y)}{P(x_1) * P(x_2) \ldots P(x_n)}$$

$$P(N|x_1, x_2, \ldots x_n) = \frac{P(N) * P(x_1|N) * P(x_2|N) \ldots P(x_n|N)}{P(x_1) * P(x_2) * \ldots P(x_n)}$$

Note : As we can see the determinant is constant or never changing, therefore we can ignore it.

$$P(Y|x_1, x_2 \ldots x_n) = P(Y) * P(x_1) * P(x_2) * \ldots P(x_n)$$

→ How theorem predict target variable !

outlook → Independent Feature

	Yess	NO	P(y)	P(n)
Sunny	2	3	2/9	3/5
overcast	4	0	4/9	0
Rain	3	2	3/9	2/5
	9	5		

Temperature → Independent Feature

	Yes	NO	P(y)	P(n)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
cold	3	1	3/9	1/5
	9	5		

Play

Yes	9	P(Y) =	9/14
NO	5	P(N) =	5/14
	14		

• Predicting probabibrlity (Yes/NO) for Sunny & Hot.

$$P(Y|\text{Sunny}, \text{Hot}) = P(Y) * P(\text{Sunny}|Y) * P(\text{Hot}|Y)$$

$$= \frac{\cancel{9}}{\cancel{14}_7} \times \frac{\cancel{2}}{\cancel{9}} \times \frac{2}{9}$$

$$= \frac{2}{63} = 0.031$$

$$P(N|\text{Sunny}, \text{Hot}) = P(N) * P(\text{Sunny}|N) * P(\text{Hot}|N)$$

$$= \frac{\cancel{8}}{\cancel{14}_{7}} \times \frac{3}{\cancel{8}} \times \frac{\cancel{2}^{1}}{5}$$

$$= \frac{3}{35} = 0.085$$

$$\therefore \quad P(Y|\text{Sunny}, \text{Hot}) = \frac{0.031}{0.031 + 0.085} = 0.267 = 26.7\%$$

$$P(N|\text{Sunny}, \text{Hot}) = \frac{0.085}{0.031 + 0.085} = 0.732 = 73.2\%$$