Baye's Classification Method :-

Grouping itemsets based on decision boundary.

↳ created by historical data (trained data)

Advantage: Events are independent of each other.

→ Features are independent of Each Other.

→ Probability Calculation or Assumption B Faster.

Baye's Theorem → Based on Conditional Probability

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

Posterior Probability

Likelihood Probability

Marginal probability

Prior probability

Ex:- Bag 1 → 2R + 3B $P(R|1) = \frac{2}{5}$

2 → 3R + 4B $P(R|2) = \frac{3}{7}$

A : Red

B : 1

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

$\frac{2/5 \times 5/12}{1/2}$

$= \frac{\frac{2}{5} \times \frac{5}{12\,6}}{\frac{1}{2}}$ $\frac{2}{1}$ $\frac{1}{6\,3}$

Steps of Naive Baye's Th →

1) Labelled data

2) Prior classes

3) Likelihood (P {feature/class})

4) Apply Naive Bayes Formula

5)

Example 2)

Outlook	Temp	Humidity	Windy
Sunny/P = 1/3	Hot/P = 1/2	High/P = 3/7	Weak/P = 6/8
Sunny/N = 2/3	Hot/N = 1/2	High/N = 4/7	Weak/N = 2/8
Overcast/P = 1	Mid/P = 4/6 = 2/3	Normal/P = 6/7	Strong/P = 1/2
Overcast/N = 0	Mid/N = 2/6 = 1/3	Normal/N = 1/7	Strong/N = 1/2
Rain/P = 1/2	Cool/P = 3/4		
Rain/N = 1/2	Cool/N = 1/4		

x

Step 1)

$$P(\text{Class } P) = \frac{\text{Count of } P}{\text{Total Rows}}$$

$$N(\text{Class } N) = \frac{\text{Count of } N}{\text{Total Rows}}$$

$$P(P) = \frac{9}{14} = 0.6429$$

$$P(N) = \frac{5}{14} = 0.3571$$

Step 2) Likelihood Probabilities. Probabilities →

Feature: Outlook →

Sunny →	$P(P) = 2/9 = 0.222$	$P(N) = 3/5 = 0.6$
Overcast →	$P(P) = 4/9 = 0.444$	$P(N) = 0/5 = 0$
Rain →	$P(P) = 3/9 = 0.333$	$P(N) = 2/5 = 0.4$

Feature: Temparature →

	P (Class = P)	P (Class = N)
Hot	$P(P) = 2/9 = 0.222$	$P(N) = 2/5 = 0.4$
Mild	$4/9 = 0.444$	$2/5 = 0.4$
Cool	$3/9 = 0.333$	$1/5 = 0.2$

Feature: Humidity

	P (class P)	P (class N)
High	$3/9 = 0.333$	$4/5 = 0.8$
Normal	$6/9 = 0.667$	$1/5 = 0.2$

Feature: Windy

	P (class P)	P (class N)
Strong	$3/9 = 0.333$	$3/5 = 0.6$
Weak	$6/9 = 0.667$	$2/5 = 0.4$

① Naive Bayes →

Outlook	Temperature	Humidity	windy	Class
Sunny	hot	high	weak	N
sunny	hot	high	Strong	N
overcast	hot	high	weak	P
rain	mild	high	weak	P
rain	cool	normal	weak	P
rain	cool	normal	Strong	N
overcast	cool	normal	Strong	P
Sunny	mild	high	weak	N
Sunny	cold	normal	weak	P
rain	mild	normal	weak	P
sunny	mild	normal	strong	P
overcast	mild	high	Strong	P
overcast	hot	normal	weak	P
rain	mild	high	Strong	N

Steps:—

1. Extract the dataset (convert to tabular format)
2. Preprocess the data (convert categorical variables into frequencies)
3. Calculate Prior probabilitees for class labels (P and N).
4. Calculate Likelihood Probability for each feature given a class
5. Apply Baye's theorem to classify a new instances.
6. Explain each step with calculations.

Step 3) Classify a New Instance :-

Outlook = Sunny.
Temparature = Cool
Humidity = High
Windy = Strong

Using Naive Bayes →

$P(P/\text{feature}) = P(P) \times P(\text{Sunny}/P) \times P(\text{Cool}/P) \times P(\text{High}/P) \times P(\text{Strong}/P)$

$= 0.6427 \times 0.222 \times 0.333 \times 0.333$

$= \boxed{0.0053}$

$P(N/\text{Feature}) = P(N) \times P(\text{Sunny}/N) \times P(\text{Cool}/N) \times P(\text{High}/P) \times P(\text{Strong}/P)$

$= 0.3571 \times 0.6 \times 0.2 \times 0.8 \times 0.6$

$= \boxed{0.0206}$

$\therefore$ Since, $P(N) > P(P)$, the instance is classified as N.

Q.2) Steps for Decision Tree →

1. Extracting the Dataset
2. Computing Entropy (Buys_Computer)
3. Computing Information Gain (Attribute)
4. Choosing the best Attribute (Entropy ↓ I.G ↑)
5. Constructing the decision tree
6. Final Answer for Output

→ Entropy of Whole dataset →

$$E(S) = -P_{yes}\log_2 P_{yes} - P_{no}\log_2 P_{no}$$

$$= -\left(\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14}\right)$$

$$= -(0.643 \times -0.639 + 0.357 \times -1.485)$$

$$= \boxed{0.940}$$

→ Computing Information Gain for Each Attribute →

Attribute : AGE →

Youth → $E = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

Middle-Aged → $E = -\frac{4}{4}\log_2\frac{4}{4} - 0 = 0.000$

Senior → $E = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$

$$E(Age) = \frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971)$$

$$= 0.693$$

$$IG(Age) = E(S) - E(Age) = 0.940 - 0.693$$

$$\boxed{IG(Age) = 0.247}$$

Attribute : Income →

High → $E = 0.971$

Medium → $E = 0.981$

Low → $E = 0.00$

$$E(Income) = \frac{5}{14}(0.971) + \frac{6}{14}(0.981) + \frac{3}{14}(0.00)$$

$$= 0.911$$

$$\boxed{I.G(Income) = 0.029}$$

Attribute: Student

No → $E = 0.985$
Yes → $E = .591$

$E(\text{Student}) = \frac{7}{14}(0.985) + \frac{7}{14}(0.591) =$

$= 0.775$

$\boxed{IG(\text{Student}) = 0.165}$

Attribute: Credit_Rating →

Fair → $E = 0.811$
Excellent → $E = 1$

$E(\text{Credit}) = \frac{8}{14}(0.811) + \frac{6}{14}(1)$

$= 0.892$

$\boxed{I.G(\text{Credit}) = 0.048}$

→ Comparing I.G for each Attribute →

✓ Age → 0.247 (Highest I.G)
Income → 0.029
Student → 0.165
Credit → 0.048

So, Decision Tree →