Dataset Recap

Age	Sex	Chest Pain Type	Target (Heart Disease)
45	Male	Typical Angina	Yes
54	Female	Atypical Angina	No
60	Male	Non-Anginal Pain	Yes
50	Male	Atypical Angina	No

New Person to Predict

Age: 50, **Sex**: Male, **Chest Pain Type**: Typical Angina.
We want to calculate whether this person is more likely to have heart disease (**Yes**) or not (**No**).

Step 1: Bayes' Theorem

$$P(\text{Yes} \mid \text{Features}) = P(\text{Age} \mid \text{Yes}) \cdot P(\text{Sex} \mid \text{Yes}) \cdot P(\text{Chest Pain} \mid \text{Yes}) \cdot P(\text{Yes})$$

$$P(\text{No} \mid \text{Features}) = P(\text{Age} \mid \text{No}) \cdot P(\text{Sex} \mid \text{No}) \cdot P(\text{Chest Pain} \mid \text{No}) \cdot P(\text{No})$$

We compute these step by step.

Step 2: Prior Probabilities ($P(Yes)$ and $P(No)$)

From the dataset:

- Total rows = 4
- $P(\text{Yes}) = \frac{2}{4} = 0.5$
- $P(\text{No}) = \frac{2}{4} = 0.5$

Step 3: Likelihoods

1. **For Age** (Continuous Feature)

We use the **Gaussian distribution formula**:

$$P(\text{Age} \mid \text{Target}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **For Heart Disease = Yes**:
  - Ages = [45, 60] → Mean (μ) = 52.5, Variance (σ^2) = 56.25.
  - $P(\text{Age} = 50 \mid \text{Yes}) = \frac{1}{\sqrt{2\pi(56.25)}} e^{-\frac{(50-52.5)^2}{2(56.25)}}$
- **For Heart Disease = No**:
  - Ages = [54, 50] → Mean (μ) = 52, Variance (σ^2) = 4.
  - $P(\text{Age} = 50 \mid \text{No}) = \frac{1}{\sqrt{2\pi(4)}} e^{-\frac{(50-52)^2}{2(4)}}$

2. **For Sex** (Categorical Feature)

- $P(\text{Male} \mid \text{Yes}) = \frac{1}{2} = 0.5$
- $P(\text{Male} \mid \text{No}) = \frac{1}{2} = 0.5$

3. **For Chest Pain Type** (Categorical Feature)

- **Heart Disease = Yes**:
  - Typical Angina = 1 occurrence → $P(\text{Typical Angina} \mid \text{Yes}) = \frac{1}{2} = 0.5$
- **Heart Disease = No**:
  - Typical Angina = 0 occurrences → $P(\text{Typical Angina} \mid \text{No}) = 0$

Step 4: Calculate Posterior Probabilities

1. **For Heart Disease = Yes:**

$$P(\text{Yes} \mid \text{Features}) = P(\text{Age} = 50 \mid \text{Yes}) \cdot P(\text{Male} \mid \text{Yes}) \cdot P(\text{Typical Angina} \mid \text{Yes}) \cdot P(\text{Yes})$$

Substitute values:

$$P(\text{Yes} \mid \text{Features}) = (\text{Age calculation}) \cdot 0.5 \cdot 0.5 \cdot 0.5$$

2. **For Heart Disease = No**:

$$P(\text{No} \mid \text{Features}) = P(\text{Age} = 50 \mid \text{No}) \cdot P(\text{Male} \mid \text{No}) \cdot P(\text{Typical Angina} \mid \text{No}) \cdot P(\text{No})$$

Substitute values:

$$P(\text{No} \mid \text{Features}) = (\text{Age calculation}) \cdot 0.5 \cdot 0 \cdot 0.5 = 0$$

Step 5: Decision

Compare $P(\text{Yes} \mid \text{Features})$ and $P(\text{No} \mid \text{Features})$:

- If $P(\text{Yes} \mid \text{Features}) > P(\text{No} \mid \text{Features})$, predict **Heart Disease = Yes**.
- Since $P(\text{No} \mid \text{Features}) = 0$, the prediction is **Yes**.

Summary

Naive Bayes uses the probability of each feature given the class (Yes or No) and multiplies them with the prior probabilities to make a prediction. In this case, the model predicts the person **has heart disease (Yes)**.