

Imagine you work for a bank and you want to predict whether a loan applicant will default on their loan or not based on some demographic and financial data. Here is a sample dataset containing 10 loan applicants and whether they defaulted on their loan or not:

Applicant ID	Age	Income	Education Level	Defaulted
1	20-29	<20000	High School	No
2	30-39	20001-39999	Bachelor's	No
3	20-29	20001-39999	Master's Bachelor's	No
4	30-39	60000-80000	High School Master's	No
5	40-49	<20000	Bachelor's High School	Yes
6	30-39	40000-59999	Master's Bachelor's	No
7	20-29	40000-59999	High School Master's	Yes
8	40-49	60000-80000	Bachelor's Bachelor's	No
9	20-29	<20000	Bachelor's High School	No
10	30-39	20001-39999	High School Master's	Yes

11 30-39 40000-59999 Bachelor's

$P(\text{Defaulted} | \text{age } 30-39, \text{Income} = 40,000-59,999, \text{Education level} = \text{Bachelor's})$

$$\text{Prior} \rightarrow P(\text{Defaulted} = \text{yes}) = \frac{3}{10}$$

$$P(\text{Defaulted} = \text{No}) = \frac{7}{10}$$

$P(\text{age } 30-39, \text{Income} = 40,000-59,999, \text{Education level} = \text{Bachelor's})$

$$\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}$$

$$P(X|C_i) : P(X | \text{Defaulted} = \text{"yes"}) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = 0.037$$

$$= 0.037 \times 0.30 = 0.011$$

$$P(\text{age} = 30-39 | \text{no}) \times (\text{income} = 40,000-59,999 | \text{no}) \times (\text{Education} = \text{Bachelor's} | \text{no})$$

$$\frac{3}{7}$$

$$\frac{1}{3}$$

$$\frac{3}{7}$$

$$= \frac{3}{7} \times \frac{1}{3} \times \frac{3}{7} = 0.49$$

$$= 0.49 \times 0.7 = 0.34$$

more no than defaulting yes