

Quiz

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	25 20-29	20,000 $\leq 20,000$	High School	No
2	35 30-39	40,000 - 59,999 50,000	Bachelor's	No
3	45 40-49	60,000 - 80,000 80,000	Master's	No
4	28 20-29	20,001 - 39,999 22,000	High School	No
5	32 30-39	40,000 - 59,999 45,000	Bachelor's	Yes
6	46 40-49	60,000 - 80,000 70,000	Master's	No
7	24 20-29	18,000 $\leq 20,000$	High School	Yes
8	38 30-39	60,000 - 80,000 60,000	Bachelor's	No
9	32 30-39	40,000 - 59,999 48,000	Bachelor's	No
10	28 20-29	20,001 - 39,999 25,000	High School	Yes

Applicant ID	Age	Income	Education Level	Defaulted
11	31 30-39	40,000 - 59,999 55,000	Bachelor's	?

In this example, we have a new applicant who is 31 years old, has an annual income of \$55,000, and has a Bachelor's degree. The question mark in the Defaulted column indicates that we do not know whether this applicant will default on their loan or not. We can use our Naive Bayes classifier to predict the value of the Defaulted column for this new applicant based on the values of the other columns.

Question $P(\text{Defaulted} = \text{Yes} \mid \text{Age} = 30-39, \text{Income} = 40,000-59,999, \text{Education} = \text{Bachelor's})$

Likelihood $\rightarrow P(\text{Age} = 30-39, \text{Income} = 40,000-59,999, \text{Education} = \text{Bachelor's} \mid \text{Defaulted} = \text{Yes})$

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

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

$$= \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27} = 0.037$$

Likelihood $\rightarrow P(\text{Age} = 30-39, \text{Income} = 40,000-59,999, \text{Education} = \text{Bachelor's} \mid \text{Defaulted} = \text{No})$

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

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

$$= \frac{3}{7} \times \frac{2}{7} \times \frac{3}{7} = 0.052$$

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

$P(X | \text{Defaulted} = \text{Yes}) \times P(\text{Defaulted} = \text{Yes})$

$$= 0.031 \times 0.3 = 0.0093$$

$P(X | \text{Defaulted} = \text{No}) \times P(\text{Defaulted} = \text{No})$

$$= 0.052 \times 0.7 = 0.0364$$

สรุป ค่าที่ 11 สอดคล้องกับค่าที่ Defaulted = No