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	20,000	High School	No
2	35	50,000	Bachelor's	No
3	45	80,000	Master's	No
4	28	22,000	High School	No
5	32	45,000	Bachelor's	Yes
6	46	70,000	Master's	No
7	24	18,000	High School	Yes
8	38	60,000	Bachelor's	No
9	32	48,000	Bachelor's	No
10	29	25,000	High School	Yes

Applicant ID	Age	Income	Education Level	Defaulted
11	31	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.

≥20000 : >10
20,001 - 39,999 : 10-19
40,000 - 59999 : 20-29
60,000 - 79999 : 30-39
80,000 - 99999 : 40-49

Applicant ID	Age	Income	Education Level	Defaulted
1	20-29	20,001 - 39,999	High School	No
2	30-39	40,000 - 59,999	Bachelor's	No
3	40-49	80,000 - 99,999	Master's	No
4	20-29	20,001 - 39,999	High School	No
5	30-39	40,000 - 59,999	Bachelor's	Yes
6	40-49	60,000 - 79,999	Master's	No
7	20-29	≥20,000	High school	Yes
8	30-39	60,000 - 79,999	Bachelor's	No.
9	30-39	40,000 - 59,999	Bachelor's	No
10	20-29	20,001 - 39,999	High school	Yes.

X = (age = 30-39, income = 40,000 - 59,999, Education Leivel = Bachelor's)

P(Yes | age = 30-39, income = 40,000 - 59,999, Education Leivel = Bachelor's)

Likelihood

$P(X|C_i)$: $P(\text{age}=30-39, \text{income}=40{,}000-59{,}999, \text{Education Level}=\text{Bachelor's}|\text{Yes})$

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

$P(X|C_i)$: $P(\text{age}=30-39, \text{income}=40{,}000-59{,}999, \text{Education Level}=\text{Bachelor's}|\text{No})$

$= \frac{3}{7} \times \frac{2}{7} \times \frac{3}{7} = 0.0524$

Prior

$P(C_i)$: $P(\text{Defaulted}=\text{'Yes'}) = \frac{3}{10} = 0.3$

$P(C_i)$: $P(\text{Defaulted}=\text{'No'}) = \frac{7}{10} = 0.7$

∴ $P(X|C_i) \times P(C_i)$: $P(X|\text{Defaulted}=\text{yes}) \times P(\text{Defaulted}=\text{'Yes'}) = 0.037 * 0.3 = 0.0111$

∴ $P(X|\text{Defaulted}=\text{No}) \times P(\text{Defaulted}=\text{'No'}) = 0.0524 \times 0.7 = 0.03668$

ดังนั้น, สรุปว่า ลูกค้าที่มีรายได้ 50,000 บาท ถึง 59,999 บาท และมีวุฒิการศึกษาระดับ Bachelor's มีแนวโน้มที่จะ ผิดนัดชำระหนี้ #