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

Note that we have combined the original Age and Income columns into Age Group and Income Group columns, respectively, based on the specified ranges. This allows us to perform Naive Bayes classification on this updated dataset using the Age Group and Income Group columns as input

ne	Education Level	Defaulted
00	High School	No
00	Bachelor's	No
00	Master's	No
00	High School	No
00	Bachelor's	Yes
00	Master's	No
0	High School	Yes
00	Bachelor's	No
00	Bachelor's	No
00	High School	Yes

Prior

$P(buy = yes) = \frac{3}{10} = 0.3$ $P(buy = no) \frac{7}{10} = 0.7$

likelihood

$P(age = 30-39,\ income = 40,000-50,000,\ Education = Bachelor's)$

$P(age = 30-39 \mid yes) \times (income = 40,000-50,000 \mid yes) \times (Education = Bachelor's \mid yes)$

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

$= (1/3) \times (1/3) \times (1/3) = 0.037$

$= 0.037 \times 0.30 = 0.0111$

$P(age = 30-39 \mid no) \times (income = 40,000-50,000 \mid no) \times (Education = Bachelor's \mid no)$

$\frac{3}{7}$ $\frac{2}{7}$ $\frac{3}{7}$

$= (3/7) \times (1/7) \times (3/7) = 0.49$

$= 0.49 \times 0.7 = 0.34$

ตอบ no เพราะมีค่ามากกว่า yes คือเท่ากับ 0.34