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.

Age : 10-19, 20-29, 30-39, 40-49

Income : < 20,000 , 20,001-39,999 , 40,000-59,999 , 60,000-80,000

Class :

C1 : Defaulted = "yes" = 3/10
C2 : Defaulted = " no" = 7/10

compute P(x|Ci) for each class

P(Age = " 10-19 " | C1) = 0
P(Age = " 20-29 " | C1) = 2/3
P(Age = " 30-39 " | C1) = 1/3
P(Age = " 40-49 " | C1) = 0

P(Age = " 10-19 " | C2) = 0
P(Age = " 20-29 " | C2) = 2/7
P(Age = " 30-39 " | C2) = 3/7
P(Age = " 40-49 " | C2) = 2/7

P(Income = "< 20,000 " | C1) = 1/3
P(Income = " 20,001-39,999 " | C1 = 1/3
P(Income = " 40,000-59,999 " | C1 = 1/3
P(Income = " 60,000-80,000 " | C1 = 0

P(Income = "< 20,000 " | C2) = 0
P(Income = " 20,001-39,999 " | C2 = 2/7
P(Income = " 40,000-59,999 " | C2 = 2/7
P(Income = " 60,000-80,000 " | C2 = 3/7

P(Education Level = " High school " | C1) = 2/3
P(Education Level = " Bachelor's " | C1) = 1/3
P(Education Level = " Master's " | C1) = 0

P(Education Level = " High school " | C2) = 2/7
P(Education Level = " Bachelor's " | C2) = 3/7
P(Education Level = " Master's " | C2) = 2/7

D11 = (Age = "30-39 , Income = "40,000-59,999 " , Education Level = " Bachelor's ")

P(x|Ci) : (P|C1) = 1/3 x 1/3 x 1/3 = 1/27 = 0.037 x 0.3 = 0.0111
: (P|C2) = 3/7 x 2/7 x 3/7 = 18/343 = 0.052 x 0.7 = 0.0364

∴ Therefore , ID11 belong to class (" Defaulted " = " no ")

น.ส. วรรณนภา สั่งเสริม