Load necessary libraries and load dataset:

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.1.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.1.2

## Loading required package: lattice
```

```
library(class)
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.1.2
```

```
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.1.2
```

```
UniversalBank <- read.csv('C:/Users/lmszr/Documents/School/Fundamentals of Machine Learning/UniversalBa
UniversalBank$Personal.Loan = as.factor(UniversalBank$Personal.Loan)
```

A: Divide data into 60% and 40% Create a pivot table for the training data with Online as a column variable, CC as a row variable, and Loan as a secondary row variable. The values inside the table should convey the count.

```
Train_Index = createDataPartition(UniversalBank$Personal.Loan, p=0.6, list=FALSE)
Train.df=UniversalBank[Train_Index,]
Validation.df=UniversalBank[-Train_Index,]

mytable <- xtabs(~ CreditCard++Online+Personal.Loan, data=Train.df)
ftable(mytable)
```

```
##                   Personal.Loan    0    1
## CreditCard Online
## 0          0                     778   74
##            1                    1144  124
## 1          0                     301   40
##            1                     489   50
```

B: Consider the task of classifying a customer who owns a bank credit card and is actively using online banking services. Looking at the pivot table, what is the probability that this customer will accept the loan offer? [This is the probability of loan acceptance (Loan = 1) conditional on having a bank credit card (CC = 1) and being an active user of online banking services (Online = 1)].

Looking at the pivot table, a customer who owns a credit card and uses online banking has a probability of 49/514 = 0.09533 or rounded 0.10 of having a personal loan.

C: Create two separate pivot tables for the training data. One will have Loan (rows) as a function of Online (columns) and the other will have Loan (rows) as a function of CC.

```r
table(PersonalLoan=Train.df$Personal.Loan, Online=Train.df$Online)
```

```
##              Online
## PersonalLoan    0    1
##             0 1079 1633
##             1  114  174
```

```r
table(PersonalLoan=Train.df$Personal.Loan, CreditCard=Train.df$CreditCard)
```

```
##              CreditCard
## PersonalLoan    0    1
##             0 1922  790
##             1  198   90
```

D: Compute the following quantities [P(A | B) means "the probability of A given B"]:

i. P(CC = 1 | Loan = 1) (the proportion of credit card holders among the loan acceptors): 85/288 = 0.30
ii. P(Online = 1 | Loan = 1): 169/288 = 0.59
iii. P(Loan = 1) (the proportion of loan acceptors): 288/3000 = 0.10
iv. P(CC = 1 | Loan = 0): 779/2712 = 0.29
v. P(Online = 1 | Loan = 0): 1622/2712 = 0.60
vi. P(Loan = 0) 2712/3000 = 0.90

E: Use the quantities computed above to compute the naive Bayes probability P(Loan = 1 | CC = 1, Online = 1).

P(Loan = 1 | CC = 1, Online = 1) = P(CC = 1 | Loan = 1)* P (Online = 1 | Loan = 1) * P(Loan = 1)/P(CC = 1, Online = 1) P(Loan = 1 | CC = 1, Online = 1) = (0.30 * 0.59 * 0.10)/P(CC = 1, Online = 1) P(Loan = 1 | CC = 1, Online = 1) = 0.0177/P(CC = 1, Online = 1)

P(Loan = 0 | CC = 1, Online = 1) = P(CC = 1 | Loan = 0)* P (Online = 1 | Loan = 0) * P(Loan = 0)/P(CC = 1, Online = 1) P(Loan = 0 | CC = 1, Online = 1) = (0.29 * 0.60 * 0.90)/P(CC = 1, Online = 1) P(Loan = 0 | CC = 1, Online = 1) = (0.29 * 0.60 * 0.90)/P(CC = 1, Online = 1) P(Loan = 0 | CC = 1, Online = 1) = 0.1566/P(CC = 1, Online = 1)

P(Loan = 1 | CC = 1, Online = 1) + P(Loan = 0 | CC = 1, Online = 1) = 1

P(Loan = 1 | CC = 1, Online = 1) = 0.0177 / (0.0177 + 0.1566) = 0.10155

F: Compare this value with the one obtained from the pivot table in (B). Which is a more accurate estimate?

Both methods give the same rounded outcome of 0.10 in question B and 0.10 in question E. Question B's method may be more accurate as it uses the actual data to calculate the probability. The Naive Bayes method also assumes that each variable is independent when that may not be the case in reality. However, both outcomes are almost the same.

G: Which of the entries in this table are needed for computing P(Loan = 1 | CC = 1, Online = 1)?

Personal.Loan, CreditCard, Online

Run naive Bayes on the data. Examine the model output on training data, and find the entry that corresponds to P(Loan = 1 | CC = 1, Online = 1). Compare this to the number you obtained in (E).

```r
nb.model<-naiveBayes(Personal.Loan~CreditCard+Online, data=Train.df)
To_Predict=data.frame(CreditCard= 1, Online=1)
predict(nb.model,To_Predict,type='raw')
```

```
##              0         1
## [1,] 0.8948275 0.1051725
```

The probability found of having a loan using this method can be rounded to 0.10, the same as was found in questions B and E.