Prepare a classification model using Naive Bayes for salary data.

**Ans :**

**R Code :**

## Naive Bayes

########## Salary Data Set #########

SalaryData\_Train <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Naive Bayes\\SalaryData\_Train.csv')

SalaryData\_Test <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Naive Bayes\\SalaryData\_Test.csv')

## Training a model on the data ----

library(e1071)

salary\_classifier <- naiveBayes(SalaryData\_Train, SalaryData\_Train$Salary)

## Step 4: Evaluating model performance ----

salary\_test\_pred <- predict(salary\_classifier, SalaryData\_Test)

library(gmodels)

CrossTable(salary\_test\_pred, SalaryData\_Test$Salary,

prop.chisq = FALSE, prop.t = FALSE, prop.r = FALSE,

dnn = c('predicted', 'actual'))

## Step 5: Improving model performance ----

salary\_classifier2 <- naiveBayes(SalaryData\_Train, SalaryData\_Train$Salary, laplace = 1)

salary\_test\_pred2 <- predict(salary\_classifier2, SalaryData\_Test)

CrossTable(salary\_test\_pred2, SalaryData\_Test$Salary,

prop.chisq = FALSE, prop.t = FALSE, prop.r = FALSE,

dnn = c('predicted', 'actual'))

**Results :**

> CrossTable(salary\_test\_pred, SalaryData\_Test$Salary,

+ prop.chisq = FALSE, prop.t = FALSE, prop.r = FALSE,

+ dnn = c('predicted', 'actual'))

Cell Contents

|-------------------------|

| N |

| N / Col Total |

|-------------------------|

Total Observations in Table: 15060

| actual

predicted | <=50K | >50K | Row Total |

-------------|-----------|-----------|-----------|

<=50K | 11113 | 202 | 11315 |

| 0.978 | 0.055 | |

-------------|-----------|-----------|-----------|

>50K | 247 | 3498 | 3745 |

| 0.022 | 0.945 | |

-------------|-----------|-----------|-----------|

Column Total | 11360 | 3700 | 15060 |

| 0.754 | 0.246 | |

-------------|-----------|-----------|-----------|

> ## Step 5: Improving model performance ----

> salary\_classifier2 <- naiveBayes(SalaryData\_Train, SalaryData\_Train$Salary, laplace = 1)

> salary\_test\_pred2 <- predict(salary\_classifier2, SalaryData\_Test)

> CrossTable(salary\_test\_pred2, SalaryData\_Test$Salary,

+ prop.chisq = FALSE, prop.t = FALSE, prop.r = FALSE,

+ dnn = c('predicted', 'actual'))

Cell Contents

|-------------------------|

| N |

| N / Col Total |

|-------------------------|

Total Observations in Table: 15060

| actual

predicted | <=50K | >50K | Row Total |

-------------|-----------|-----------|-----------|

<=50K | 11186 | 97 | 11283 |

| 0.985 | 0.026 | |

-------------|-----------|-----------|-----------|

>50K | 174 | 3603 | 3777 |

| 0.015 | 0.974 | |

-------------|-----------|-----------|-----------|

Column Total | 11360 | 3700 | 15060 |

| 0.754 | 0.246 | |

-------------|-----------|-----------|-----------|

**Inference :**

Getting little better results by adding laplace term in algorithm.