Assignment 4

2025-10-13

```
library(dplyr)
library(class)
library(caret)
library(e1071)
library(readr)
library(ggplot2)
library(e1071)
```

```
Heart_disease <- read.csv("/Users/dylandoyle-rowan/Downloads/Heart_disease.csv")
```

Dummy Variables

```
Heart_disease$Target <- ifelse(Heart_disease$MAX_HeartRate > 170, "Yes", "No")
Heart_disease$BP_New <- ifelse(Heart_disease$Blood_Pressure > 120, "Yes", "No")
```

Factors

```
Heart_disease$Target <- as.factor(Heart_disease$Target)
Heart_disease$BP_New <- as.factor(Heart_disease$BP_New)
```

Data Structure

```
str(Heart_disease[, c("Target", "BP_New", "chest_pain_type", "MAX_HeartRate", "Blood_Pre
ssure")])
```

```
## 'data.frame':    303 obs. of  5 variables:
##  $ Target         : Factor w/ 2 levels "No","Yes": 1 2 2 2 1 1 1 2 1 2 ...
##  $ BP_New         : Factor w/ 2 levels "No","Yes": 2 2 2 1 1 2 2 1 2 2 ...
##  $ chest_pain_type: int  0 1 1 1 0 0 1 1 1 1 ...
##  $ MAX_HeartRate  : int  150 187 172 178 163 148 153 173 162 174 ...
##  $ Blood_Pressure : int  145 130 130 120 120 140 140 120 172 150 ...
```

Question #1

```
Target_table <- table(Heart_disease$Target)
print(Target_table)
```

```
##
##  No Yes
## 245  58
```

```
Target_prop <- prop.table(Target_table)
print(Target_prop)
```

```
##
##        No       Yes
## 0.8085809 0.1914191
```

If someone shows up with chest pain and there is no other information, we should predict that They will not develop a heart disease because a heavy majority of people in our data don't have heart disease and its much more common to not have it.

#Question 2

```
Heart_disease30 <- Heart_disease[1:30, c("Target", "BP_New", "chest_pain_type")]

head(Heart_disease30)
```

```
##   Target BP_New chest_pain_type
## 1     No    Yes               0
## 2    Yes    Yes               1
## 3    Yes    Yes               1
## 4    Yes     No               1
## 5     No     No               0
## 6     No    Yes               0
```

#Pivot Table

```
Piv_table <- ftable(Heart_disease30)
print("Pivot Table:")
```

```
## [1] "Pivot Table:"
```

```
print(Piv_table)
```

```
##               chest_pain_type 0 1
## Target BP_New
## No     No                     2 2
##        Yes                    7 8
## Yes    No                     0 3
##        Yes                    3 5
```

Contingency Table

```
contingency_table <- table(Heart_disease30$Target, Heart_disease30$BP_New, Heart_disease
30$chest_pain_type)
print("Contingency Table:")
```

```
## [1] "Contingency Table:"
```

```
print(contingency_table)
```

```
## , ,  = 0
##
##
##       No Yes
##   No   2   7
##   Yes  0   3
##
## , ,  = 1
##
##
##       No Yes
##   No   2   8
##   Yes  3   5
```

Bayes conditional probabilities

When BP yes and Chest type 1

```
count_yes_no_0 <- sum(Heart_disease30$Target == "Yes" &
                        Heart_disease30$BP_New == "No" &
                        Heart_disease30$chest_pain_type == 0)
count_total_no_0 <- sum(Heart_disease30$BP_New == "No" &
                          Heart_disease30$chest_pain_type == 0)
P1 <- ifelse(count_total_no_0 > 0, count_yes_no_0 / count_total_no_0, 0)
```

When BP no and Chest type 0

```
count_yes_no_1 <- sum(Heart_disease30$Target == "Yes" &
                        Heart_disease30$BP_New == "No" &
                        Heart_disease30$chest_pain_type == 1)
count_total_no_1 <- sum(Heart_disease30$BP_New == "No" &
                          Heart_disease30$chest_pain_type == 1)
P2 <- ifelse(count_total_no_1 > 0, count_yes_no_1 / count_total_no_1, 0)
```

When BP No and Chest type 1

```
count_yes_yes_0 <- sum(Heart_disease30$Target == "Yes" &
                         Heart_disease30$BP_New == "Yes" &
                         Heart_disease30$chest_pain_type == 0)
count_total_yes_0 <- sum(Heart_disease30$BP_New == "Yes" &
                           Heart_disease30$chest_pain_type == 0)
P3 <- ifelse(count_total_yes_0 > 0, count_yes_yes_0 / count_total_yes_0, 0)
```

When BP Yes and Chest type 0

```
count_yes_yes_1 <- sum(Heart_disease30$Target == "Yes" &
                         Heart_disease30$BP_New == "Yes" &
                         Heart_disease30$chest_pain_type == 1)
count_total_yes_1 <- sum(Heart_disease30$BP_New == "Yes" &
                           Heart_disease30$chest_pain_type == 1)
P4 <- ifelse(count_total_yes_1 > 0, count_yes_yes_1 / count_total_yes_1, 0)

cat("Bayes Conditional Probabilities P(Target=Yes | predictors):\n")
```

```
## Bayes Conditional Probabilities P(Target=Yes | predictors):
```

```
cat("P(Target=Yes | BP_New=No, chest_pain_type=0) =", round(P1, 3), "\n")
```

```
## P(Target=Yes | BP_New=No, chest_pain_type=0) = 0
```

```
cat("P(Target=Yes | BP_New=No, chest_pain_type=1) =", round(P2, 3), "\n")
```

```
## P(Target=Yes | BP_New=No, chest_pain_type=1) = 0.6
```

```
cat("P(Target=Yes | BP_New=Yes, chest_pain_type=0) =", round(P3, 3), "\n")
```

```
## P(Target=Yes | BP_New=Yes, chest_pain_type=0) = 0.3
```

```
cat("P(Target=Yes | BP_New=Yes, chest_pain_type=1) =", round(P4, 3), "\n")
```

```
## P(Target=Yes | BP_New=Yes, chest_pain_type=1) = 0.385
```

#b.)

```
Probability_Target <- rep(0, 30)
for (i in 1:30) {
  bp_val <- Heart_disease30$BP_New[i]
  cp_val <- Heart_disease30$chest_pain_type[i]

  if (bp_val == "No" & cp_val == 0) {
    Probability_Target[i] <- P1
  } else if (bp_val == "No" & cp_val == 1) {
    Probability_Target[i] <- P2
  } else if (bp_val == "Yes" & cp_val == 0) {
    Probability_Target[i] <- P3
  } else if (bp_val == "Yes" & cp_val == 1) {
    Probability_Target[i] <- P4
  }
}
```

Adding probabilities to dataset

```
Heart_disease30$Probability_Target <- Probability_Target
Heart_disease30$Pred_Probability <- ifelse(Heart_disease30$Probability_Target > 0.5, "Ye
s", "No")

print("Results for First 30 Records:")
```

```
## [1] "Results for First 30 Records:"
```

```
print(Heart_disease30)
```

```
##    Target BP_New chest_pain_type Probability_Target Pred_Probability
## 1      No    Yes               0          0.3000000               No
## 2     Yes    Yes               1          0.3846154               No
## 3     Yes    Yes               1          0.3846154               No
## 4     Yes     No               1          0.6000000              Yes
## 5      No     No               0          0.0000000               No
## 6      No    Yes               0          0.3000000               No
## 7      No    Yes               1          0.3846154               No
## 8     Yes     No               1          0.6000000              Yes
## 9      No    Yes               1          0.3846154               No
## 10    Yes    Yes               1          0.3846154               No
## 11     No    Yes               0          0.3000000               No
## 12     No    Yes               1          0.3846154               No
## 13    Yes    Yes               1          0.3846154               No
## 14     No     No               0          0.0000000               No
## 15     No    Yes               0          0.3000000               No
## 16     No     No               1          0.6000000              Yes
## 17    Yes     No               1          0.6000000              Yes
## 18     No    Yes               0          0.3000000               No
## 19    Yes    Yes               0          0.3000000               No
## 20     No    Yes               0          0.3000000               No
## 21     No    Yes               0          0.3000000               No
## 22    Yes    Yes               1          0.3846154               No
## 23    Yes    Yes               0          0.3000000               No
## 24     No    Yes               1          0.3846154               No
## 25    Yes    Yes               0          0.3000000               No
## 26     No    Yes               1          0.3846154               No
## 27     No    Yes               1          0.3846154               No
## 28     No     No               1          0.6000000              Yes
## 29     No    Yes               1          0.3846154               No
## 30     No    Yes               1          0.3846154               No
```

c.)

```
P_Target_Yes <- sum(Heart_disease30$Target == "Yes") / nrow(Heart_disease30)


P_BP_Yes_given_Target_Yes <- sum(Heart_disease30$BP_New == "Yes" & Heart_disease30$Targe
t == "Yes") / sum(Heart_disease30$Target == "Yes")
P_CP1_given_Target_Yes <- sum(Heart_disease30$chest_pain_type == 1 & Heart_disease30$Tar
get == "Yes") / sum(Heart_disease30$Target == "Yes")
```

Evidence

```
P_BP_Yes <- sum(Heart_disease30$BP_New == "Yes") / nrow(Heart_disease30)
P_CP1 <- sum(Heart_disease30$chest_pain_type == 1) / nrow(Heart_disease30)
```

Naive Bayes probability

```
naive_bayes_prob <- (P_Target_Yes * P_BP_Yes_given_Target_Yes * P_CP1_given_Target_Yes)
/ (P_BP_Yes * P_CP1)
cat("Manual Naive Bayes Calculation:\n")
```

```
## Manual Naive Bayes Calculation:
```

```
cat("P(Target=Yes) =", round(P_Target_Yes, 3), "\n")
```

```
## P(Target=Yes) = 0.367
```

```
cat("P(BP_New=Yes | Target=Yes) =", round(P_BP_Yes_given_Target_Yes, 3), "\n")
```

```
## P(BP_New=Yes | Target=Yes) = 0.727
```

```
cat("P(chest_pain_type=1 | Target=Yes) =", round(P_CP1_given_Target_Yes, 3), "\n")
```

```
## P(chest_pain_type=1 | Target=Yes) = 0.727
```

```
cat("P(BP_New=Yes) =", round(P_BP_Yes, 3), "\n")
```

```
## P(BP_New=Yes) = 0.767
```

```
cat("P(chest_pain_type=1) =", round(P_CP1, 3), "\n")
```

```
## P(chest_pain_type=1) = 0.6
```

```
cat("\nNaive Bayes Probability P(Target=Yes | BP_New=Yes, chest_pain_type=1) =", round(n
aive_bayes_prob, 3), "\n")
```

```
##
## Naive Bayes Probability P(Target=Yes | BP_New=Yes, chest_pain_type=1) = 0.422
```

Question 3

```
set.seed(123)
```

Training Data Split

```
train.index <- sample(row.names(Heart_disease), 0.6 * nrow(Heart_disease))
valid.index <- setdiff(row.names(Heart_disease), train.index)

train.df <- Heart_disease[train.index, ]
valid.df <- Heart_disease[valid.index, ]
```

Naive Bayes Model

```
nb_model <- naiveBayes(Target ~ chest_pain_type + BP_New, data = train.df)
```

#Predictions

```
train_pred <- predict(nb_model, train.df)
valid_pred <- predict(nb_model, valid.df)
```

Confusion Matrix

```
conf_matrix_train <- confusionMatrix(train_pred, train.df$Target)
print("Confusion Matrix - Training Set:")
```

```
## [1] "Confusion Matrix - Training Set:"
```

```
print(conf_matrix_train)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No Yes
##        No  149  32
##        Yes   0   0
##
##                Accuracy : 0.8232
##                  95% CI : (0.7597, 0.8758)
##     No Information Rate : 0.8232
##     P-Value [Acc > NIR] : 0.5471
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : 4.251e-08
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.8232
##          Neg Pred Value :    NaN
##              Prevalence : 0.8232
##          Detection Rate : 0.8232
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : No
##
```

```
conf_matrix_valid <- confusionMatrix(valid_pred, valid.df$Target)
print("Confusion Matrix - Validation Set:")
```

```
## [1] "Confusion Matrix - Validation Set:"
```

```
print(conf_matrix_valid)
```

```
## Confusion Matrix and Statistics
##
##          Reference
## Prediction No Yes
##        No  96  26
##        Yes  0   0
##
##               Accuracy : 0.7869
##                 95% CI : (0.7035, 0.8558)
##    No Information Rate : 0.7869
##    P-Value [Acc > NIR] : 0.5523
##
##                  Kappa : 0
##
##  Mcnemar's Test P-Value : 9.443e-07
##
##            Sensitivity : 1.0000
##            Specificity : 0.0000
##         Pos Pred Value : 0.7869
##         Neg Pred Value :    NaN
##             Prevalence : 0.7869
##         Detection Rate : 0.7869
##   Detection Prevalence : 1.0000
##      Balanced Accuracy : 0.5000
##
##       'Positive' Class : No
##
```

Model Summary

```
print("Model Summary:")
```

```
## [1] "Model Summary:"
```

```
print(nb_model)
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##        No       Yes
## 0.8232044 0.1767956
##
## Conditional probabilities:
##      chest_pain_type
## Y          [,1]      [,2]
##   No  0.3892617 0.4892273
##   Yes 0.7187500 0.4568034
##
##      BP_New
## Y            No       Yes
##   No  0.3288591 0.6711409
##   Yes 0.3750000 0.6250000
```