191805010_cihat_çoban

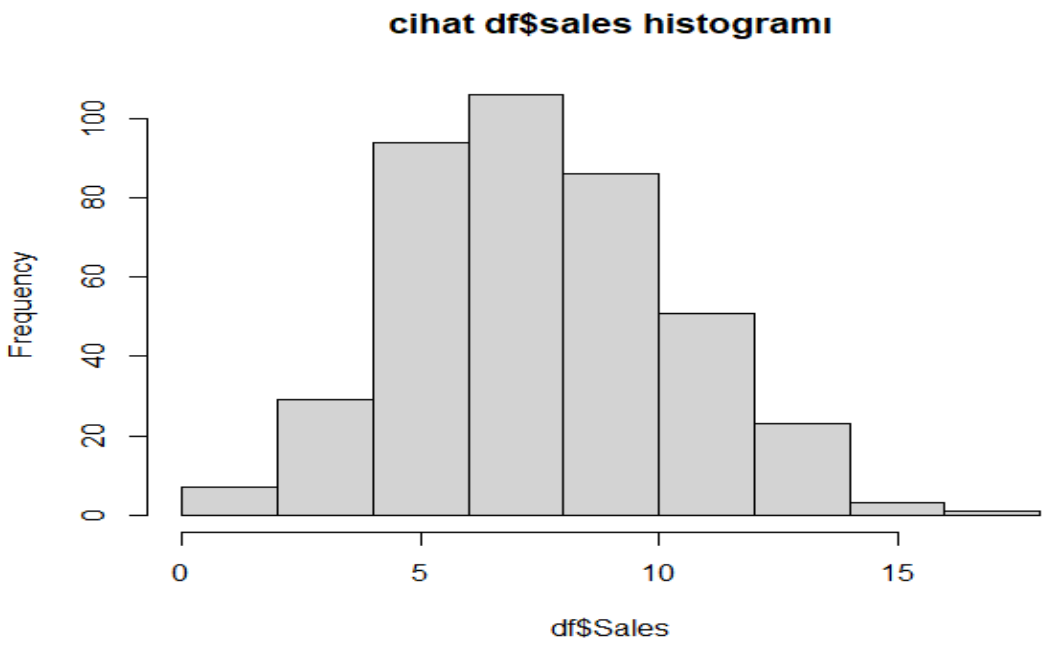
Install Librarys and Data:

```
> library(e1071)  # e1071 paketi Naive Bayes fonksiyonunu bunun içinden gelecek
> library(pROC)
> library(dplyr)
> library(ggplot2)
> library(psych)
> library(caret)
> library(rpart.plot)
> library(ISLR)
> data(Carseats)
> df <- Carseats
> str(df)
'data.frame':   400 obs. of  11 variables:
 $ Sales      : num  9.5 11.22 10.06 7.4 4.15 ...
 $ CompPrice  : num  138 111 113 117 141 124 115 136 132 132 ...
 $ Income     : num  73 48 35 100 64 113 105 81 110 113 ...
 $ Advertising: num  11 16 10 4 3 13 0 15 0 0 ...
 $ Population : num  276 260 269 466 340 501 45 425 108 131 ...
 $ Price      : num  120 83 80 97 128 72 108 120 124 124 ...
 $ ShelveLoc  : Factor w/ 3 levels "Bad","Good","Medium": 1 2 3 3 1 1 3 2 3 3 ...
 $ Age        : num  42 65 59 55 38 78 71 67 76 76 ...
 $ Education  : num  17 10 12 14 13 16 15 10 10 17 ...
 $ Urban      : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...
 $ US         : Factor w/ 2 levels "No","Yes": 2 2 2 2 1 2 1 2 1 2 ...
```

Summary:

```
> summary(df)
     Sales          CompPrice       Income        Advertising      Population      Price      
 Min.   : 0.000   Min.   : 77   Min.   : 21.00   Min.   : 0.000   Min.   : 10.0   Min.   : 24.0  
 1st Qu.: 5.390   1st Qu.:115   1st Qu.: 42.75   1st Qu.: 0.000   1st Qu.:139.0   1st Qu.:100.0  
 Median : 7.490   Median :125   Median : 69.00   Median : 5.000   Median :272.0   Median :117.0  
 Mean   : 7.496   Mean   :125   Mean   : 68.66   Mean   : 6.635   Mean   :264.8   Mean   :115.8  
 3rd Qu.: 9.320   3rd Qu.:135   3rd Qu.: 91.00   3rd Qu.:12.000   3rd Qu.:398.5   3rd Qu.:131.0  
 Max.   :16.270   Max.   :175   Max.   :120.00   Max.   :29.000   Max.   :509.0   Max.   :191.0  
  ShelveLoc        Age          Education    Urban       US     
 Bad    : 96   Min.   :25.00   Min.   :10.0   No :118   No :142  
 Good   : 85   1st Qu.:39.75   1st Qu.:12.0   Yes:282   Yes:258  
 Medium:219   Median :54.50   Median :14.0                      
              Mean   :53.32   Mean   :13.9                      
              3rd Qu.:66.00   3rd Qu.:16.0                      
              Max.   :80.00   Max.   :18.0                      
```

Plot-Histogram:



191805010_cihat_çoban

Data Partition and Train:

```
> hist(df$Sales , main = "cihat df$sales histogramı")
> df$Sales <- as.factor(ifelse(df$Sales <= 8, "Low", "High"))
> set.seed(123)
> train_indeks <- createDataPartition(df$Sales, p = 0.8, list = FALSE, times = 1)
> train <- df[train_indeks,]
> test  <- df[-train_indeks,]
> train_x <- train %>% select(-Sales)
> train_y <- train$Sales
> test_x <- test %>% select(-Sales)
> test_y <- test$Sales
> nb_model <- naiveBayes(Sales ~ ., data = train)
> summary(nb_model)
          Length Class  Mode
apriori    2     table  numeric
tables    10     -none- list
levels     2     -none- character
isnumeric 10     -none- logical
call       4     -none- call
> predicted_values <- predict(nb_model, test , type="raw")
> tb <- table(predict(nb_model, test ), test_y)
> confusionMatrix(tb, positive = "High")
Confusion Matrix and Statistics
```

Confusion Matrix and Statistics:

```
> predicted_values <- predict(nb_model, test , type="raw")
> tb <- table(predict(nb_model, test ), test_y)
> confusionMatrix(tb, positive = "High")
Confusion Matrix and Statistics

      test_y
       High Low
  High   23   5
  Low     9  42

               Accuracy : 0.8228
                 95% CI : (0.7206, 0.8996)
    No Information Rate : 0.5949
    P-Value [Acc > NIR] : 1.248e-05

                  Kappa : 0.6248

 Mcnemar's Test P-Value : 0.4227

            Sensitivity : 0.7188
            Specificity : 0.8936
         Pos Pred Value : 0.8214
         Neg Pred Value : 0.8235
             Prevalence : 0.4051
         Detection Rate : 0.2911
   Detection Prevalence : 0.3544
      Balanced Accuracy : 0.8062

       'Positive' Class : High
```

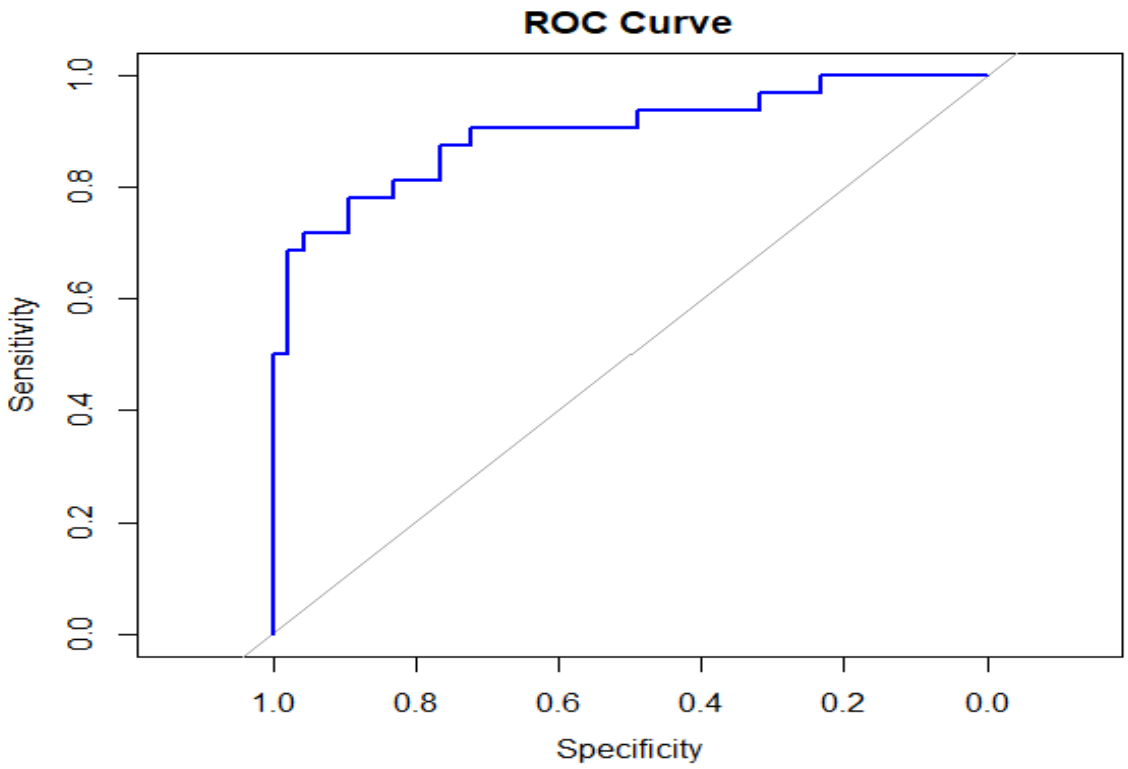
191805010_cihat_çoban

Precision – Recall – F1 Score:

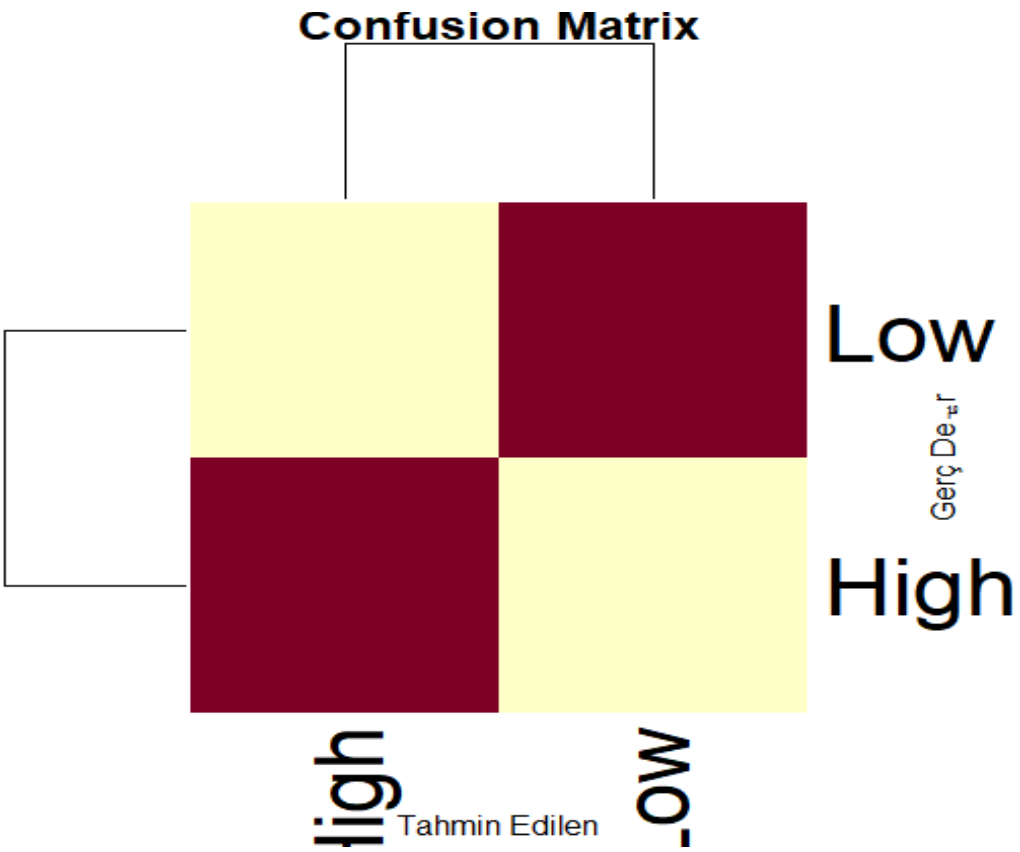
```
> roc_curve <- roc(ifelse(test$Sales == "High", 1, 0), predicted_values[, "High"])
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> plot(roc_curve, main = "ROC Curve", col = "blue", lwd = 2)
> tb <- table(predict(nb_model, test), test_y)
> conf_matrix <- as.matrix(tb)
> heatmap(conf_matrix, annot = TRUE, fmt = "d", cmap = "Blues",
+         xlab = "Tahmin Edilen", ylab = "Gerçek Değer",
+         main = "Confusion Matrix")
Warning messages:
1: In plot.window(...) : "annot" bir grafiksel parametre değil
2: In plot.window(...) : "fmt" bir grafiksel parametre değil
3: In plot.window(...) : "cmap" bir grafiksel parametre değil
4: In plot.xy(xy, type, ...) : "annot" bir grafiksel parametre değil
5: In plot.xy(xy, type, ...) : "fmt" bir grafiksel parametre değil
6: In plot.xy(xy, type, ...) : "cmap" bir grafiksel parametre değil
7: In title(...) : "annot" bir grafiksel parametre değil
8: In title(...) : "fmt" bir grafiksel parametre değil
9: In title(...) : "cmap" bir grafiksel parametre değil
> precision <- posPredValue(predict(nb_model, test), test_y, positive = "High")
There were 18 warnings (use warnings() to see them)
> recall <- sensitivity(predict(nb_model, test), test_y, positive = "High")
> f1_score <- 2 * (precision * recall) / (precision + recall)
> cat("\nPrecision:", precision, "\n")

Precision: 0.8214286
> cat("Recall (Sensitivity):", recall, "\n")
Recall (Sensitivity): 0.71875
> cat("F1 Score:", f1_score, "\n")
F1 Score: 0.7666667
>
```

ROC Curve:



Heat-Map:


Confusion Matrix
Low
High
Gerç De ır
ligh
Tahmin Edilen
_ow