14주차 과제

과 목	머신러닝을이용한재난설계
담당교수	이두호
학 번	201720970
학 과	소프트웨어·미디어·산업공학부
이 름	권대한

머신러닝을이용한재난설계 제출일 2021. 12. 06.

201720970_권대한

k = 5 로 하여 테스트 데이터 45개에 대하여 유클리드 거리, 맨하탄 거리, 마할라노비스 거리 측도로 분류예측을 수행하고, 혼동행렬을 작성하라.

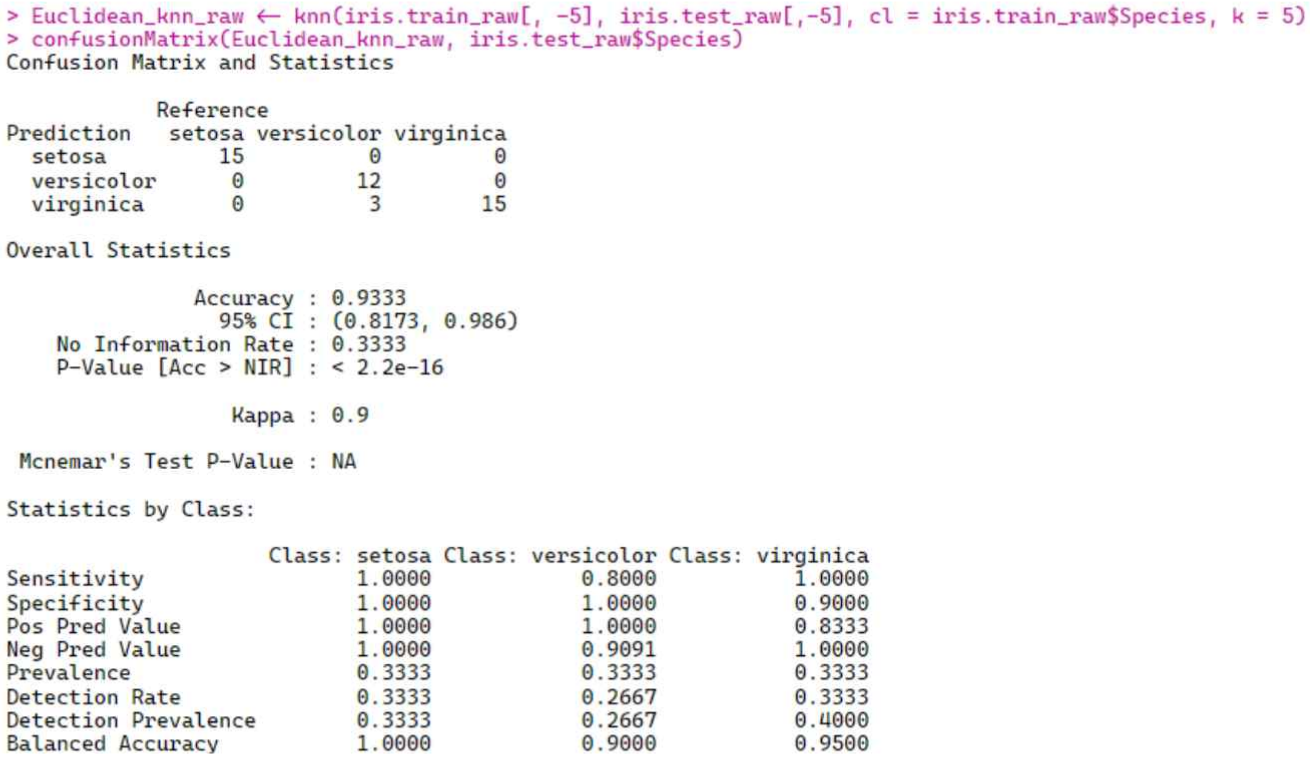
<코드>

```
library(caret)
library(class)
library(ggvis)

standardization <- function(x) {
  return ((x-min(x))/(max(x)-min(x)))
}

iris <- iris
iris %>% ggvis(~Sepal.Length, ~Sepal.Width, fill = ~Species) %>% layer_points
iris %>% ggvis(~Petal.Length, ~Petal.Width, fill = ~Species) %>% layer_points()

iris_x <- as.data.frame(lapply(iris[1:4], standardization))
set.seed(1234)
samp <- createDataPartition(iris$Species, p=0.7, list=F)
iris.train_raw <- iris[samp, ]
iris.train <- iris_x[samp,]
row.names(iris.train) <- 1:nrow(iris.train)
iris.test_raw <- iris[-samp,]
iris.test <- iris_x[-samp,]

row.names(iris.test) <- 1:nrow(iris.test)
row.names(iris.test_raw) <- 1:nrow(iris.test_raw)

Euclidean_knn_raw <- knn(iris.train_raw[, -5], iris.test_raw[,-5], cl = iris.train_raw$Species, k = 5)
confusionMatrix(Euclidean_knn_raw, iris.test_raw$Species)

Euclidean_knn <- knn(iris.train[, -5], iris.test[,-5], cl = iris.train_raw$Species, k = 5)
confusionMatrix(Euclidean_knn, iris.test_raw[, 5])

Euclidean_distance <- function(x1, x2) sqrt(((x1 - x2)^2) %>% sum())
Manhattan_distance <- function(x1, x2) sqrt(abs((x1 - x2)) %>% sum())
```

```
Mahalanobis_distance <- function(x1, x2) mahalanobis(x = x1 - x2,center = colMeans(x1 - x2),cov = cov(x1 - x2))
# sqrt(t(x1 - x2) %*% solve(cov(x1 - x2)) %*% (x1 - x2))
# mahalanobis((x1 - x2), colMeans(x1 - x2), cov(x1 - x2), tol=1e-20)
# mahalanobis(x=df[,5:6],center = colMeans(df[,5:6]),cov = cov(df[,5:6]))

knn.t <- function(train, test, cl, Exec_Func, k)
{
  test.result <- test %>% length() %>% as.numeric()

  for(i in 1:nrow(test))
    {
    train$calc = sapply(1:nrow(train), function(n) Exec_Func(train[n, 1:4], test[i, 1:4]))
    nearest = order(train$calc)[1:k]
    train.result = cl[nearest]
    result_factor = names(sort(table(train.result), decreasing=TRUE)[1])
    test.result[i]=result_factor
    }
  train$calc = NULL
  test.result %>% as.factor()
}

Euclidean_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5],Euclidean_distance, k = 5)
Euclidean_result
confusionMatrix(Euclidean_result, iris.test_raw$Species)

Manhattan_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5], Manhattan_distance, k = 5)
Manhattan_result
confusionMatrix(Manhattan_result, iris.test_raw$Species)

Mahalanobis_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5], Mahalanobis_distance, k = 5)
Mahalanobis_result
confusionMatrix(Mahalanobis_result, iris.test_raw$Species)
```

<출력 결과>

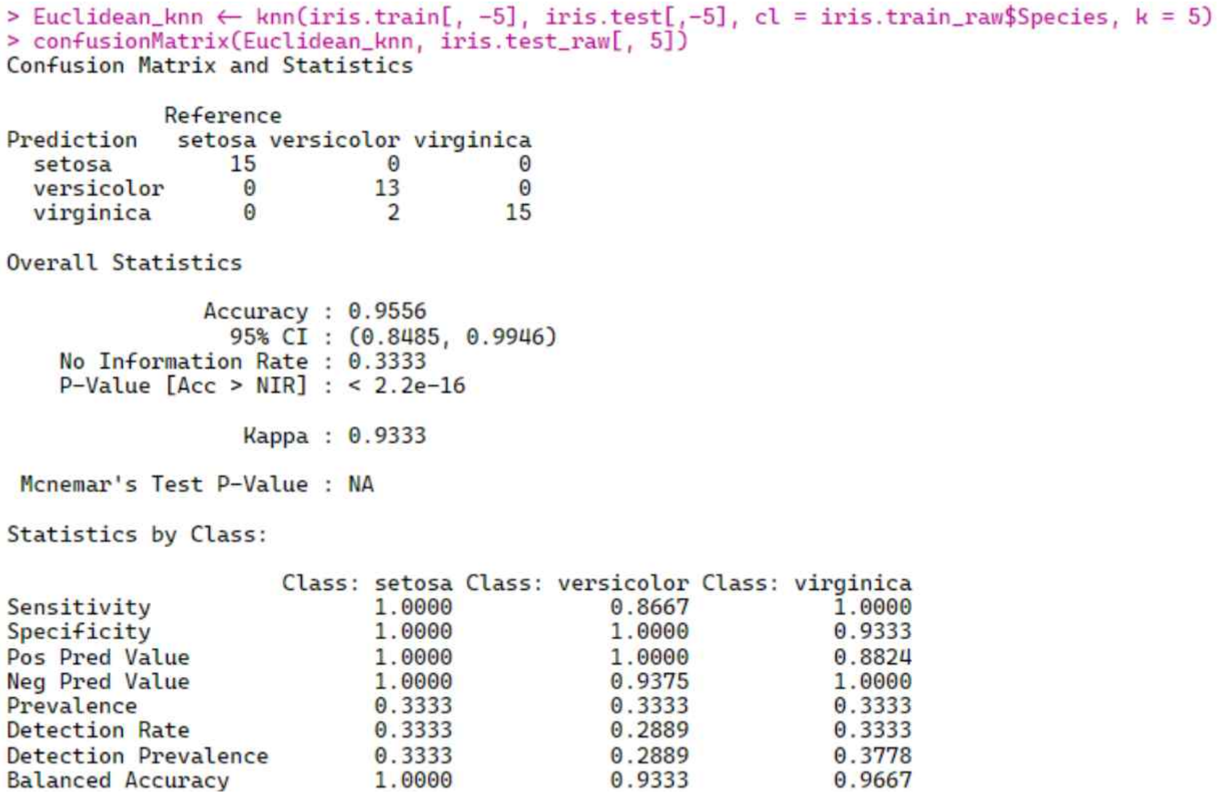
Case 1. 정규화 하지 않은 상태에서의 Euclidean_result



```
> Euclidean_knn_raw <- knn(iris.train_raw[, -5], iris.test_raw[,-5], cl = iris.train_raw$Species, k = 5)
> confusionMatrix(Euclidean_knn_raw, iris.test_raw$Species)
Confusion Matrix and Statistics

            Reference
Prediction   setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         12         0
  virginica       0          3        15

Overall Statistics
                                         
               Accuracy : 0.9333         
                 95% CI : (0.8173, 0.986)
    No Information Rate : 0.3333         
    P-Value [Acc > NIR] : < 2.2e-16      
                                         
                  Kappa : 0.9            
                                         
 Mcnemar's Test P-Value : NA             

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.8000           1.0000
Specificity                 1.0000            1.0000           0.9000
Pos Pred Value              1.0000            1.0000           0.8333
Neg Pred Value              1.0000            0.9091           1.0000
Prevalence                  0.3333            0.3333           0.3333
Detection Rate              0.3333            0.2667           0.3333
Detection Prevalence        0.3333            0.2667           0.4000
Balanced Accuracy           1.0000            0.9000           0.9500
```

Result: Accuracy 93%

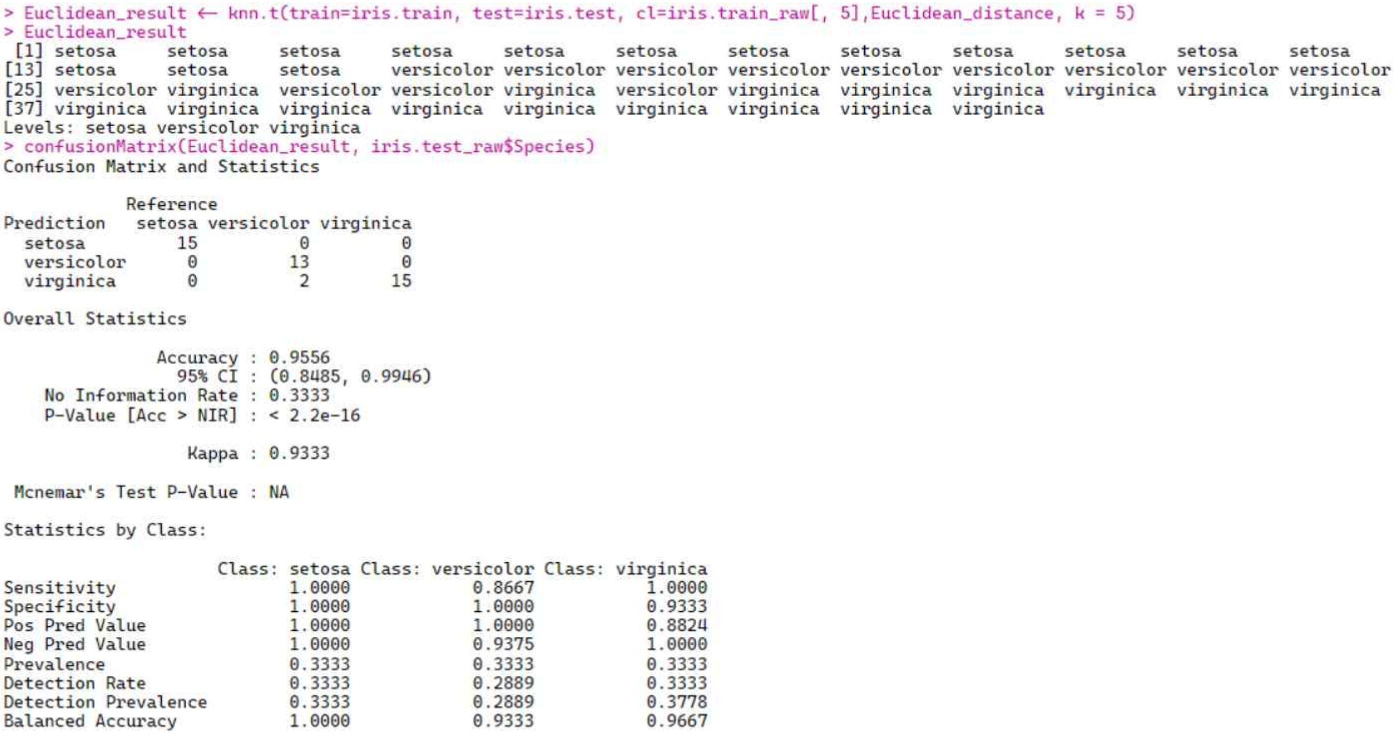
Case 2. 정규화 후 Euclidean_result



```
> Euclidean_knn <- knn(iris.train[, -5], iris.test[,-5], cl = iris.train_raw$Species, k = 5)
> confusionMatrix(Euclidean_knn, iris.test_raw[, 5])
Confusion Matrix and Statistics

            Reference
Prediction   setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         13         0
  virginica       0          2        15

Overall Statistics
                                          
               Accuracy : 0.9556          
                 95% CI : (0.8485, 0.9946)
    No Information Rate : 0.3333          
    P-Value [Acc > NIR] : < 2.2e-16       
                                          
                  Kappa : 0.9333          
                                          
 Mcnemar's Test P-Value : NA              

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.8667           1.0000
Specificity                 1.0000            1.0000           0.9333
Pos Pred Value              1.0000            1.0000           0.8824
Neg Pred Value              1.0000            0.9375           1.0000
Prevalence                  0.3333            0.3333           0.3333
Detection Rate              0.3333            0.2889           0.3333
Detection Prevalence        0.3333            0.2889           0.3778
Balanced Accuracy           1.0000            0.9333           0.9667
```

Result: Accuracy 95.56%

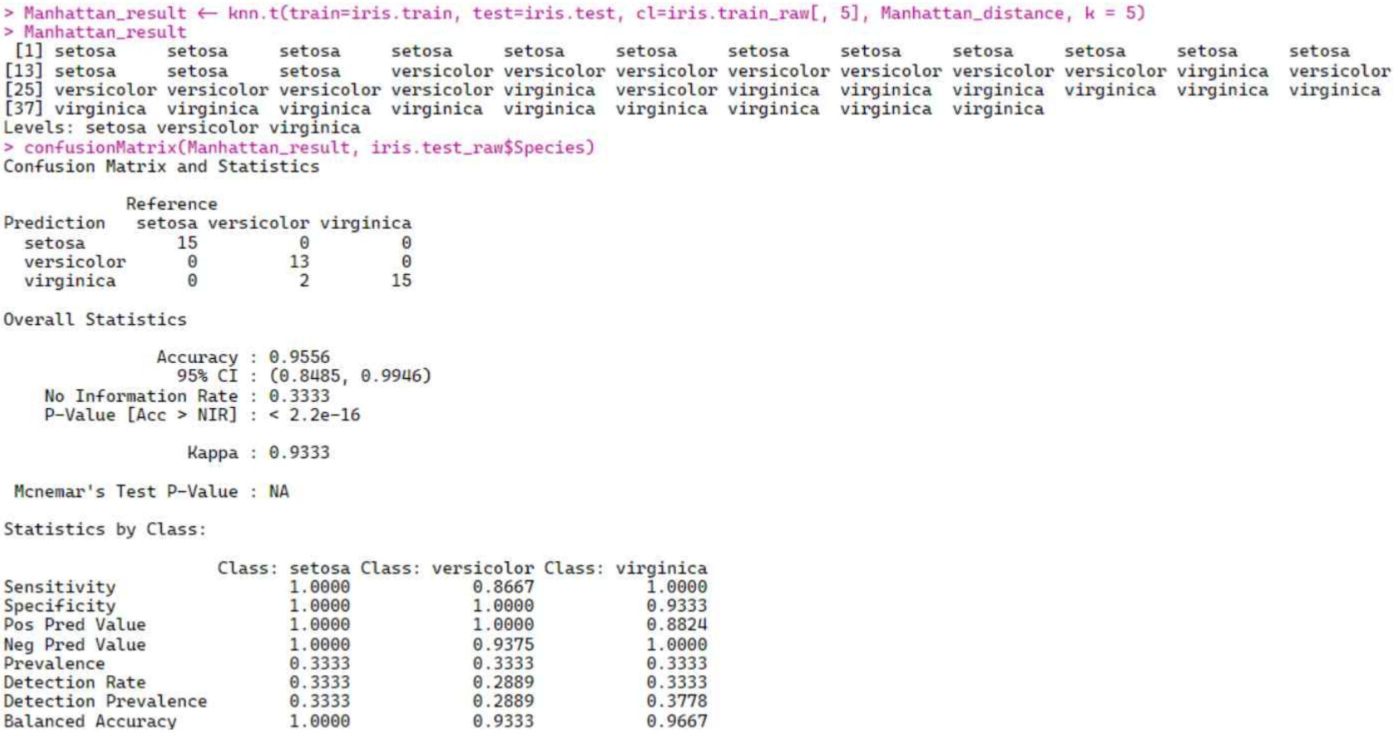
Case 3. confusionMatrix(Euclidean_result, iris.test_raw$Species)



```
> Euclidean_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5],Euclidean_distance, k = 5)
> Euclidean_result
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa    
[13] setosa     setosa     setosa     versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[25] versicolor virginica  versicolor versicolor virginica  versicolor virginica  virginica  virginica  virginica  virginica  virginica 
[37] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica 
Levels: setosa versicolor virginica
> confusionMatrix(Euclidean_result, iris.test_raw$Species)
Confusion Matrix and Statistics

            Reference
Prediction   setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         13         0
  virginica       0          2        15

Overall Statistics
                                          
               Accuracy : 0.9556          
                 95% CI : (0.8485, 0.9946)
    No Information Rate : 0.3333          
    P-Value [Acc > NIR] : < 2.2e-16       
                                          
                  Kappa : 0.9333          
                                          
 Mcnemar's Test P-Value : NA              

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.8667           1.0000
Specificity                 1.0000            1.0000           0.9333
Pos Pred Value              1.0000            1.0000           0.8824
Neg Pred Value              1.0000            0.9375           1.0000
Prevalence                  0.3333            0.3333           0.3333
Detection Rate              0.3333            0.2889           0.3333
Detection Prevalence        0.3333            0.2889           0.3778
Balanced Accuracy           1.0000            0.9333           0.9667
```

Result: Accuracy 95.56%

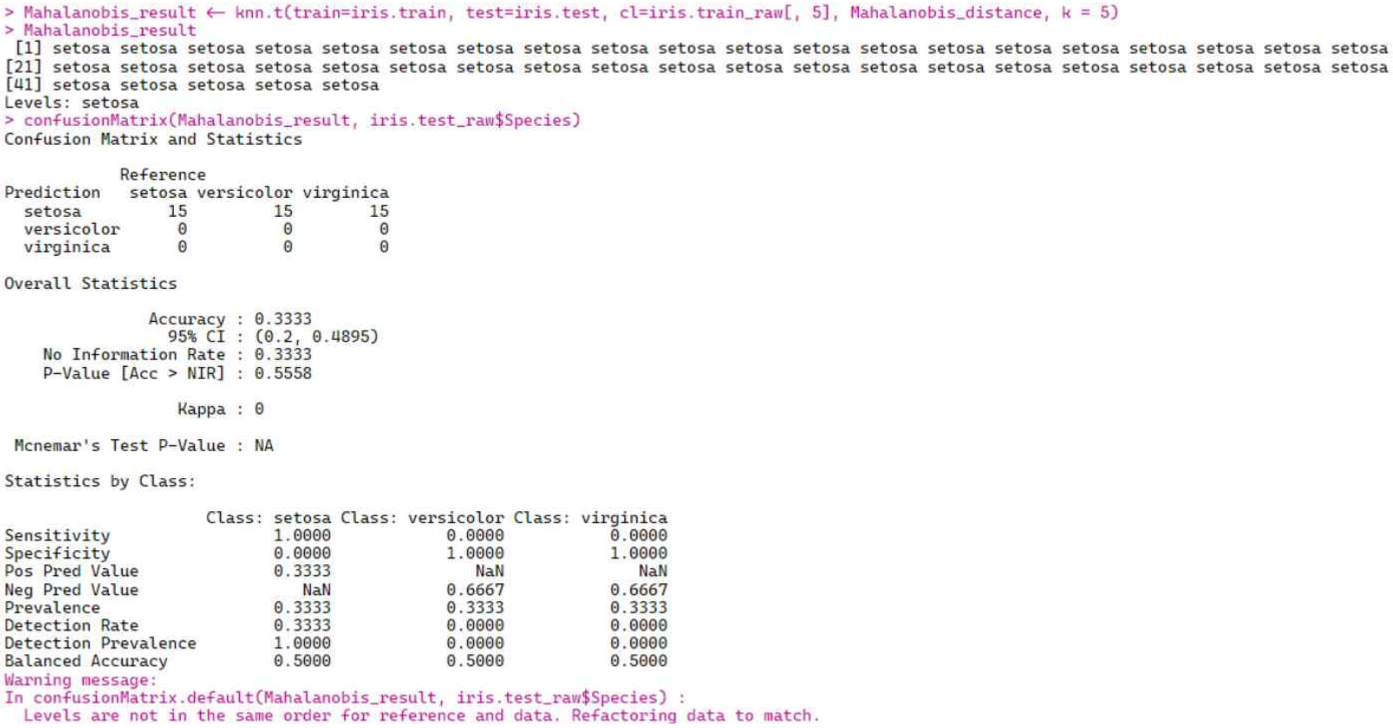
Case 4. confusionMatrix(Manhattan_result, iris.test_raw$Species)



```
> Manhattan_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5], Manhattan_distance, k = 5)
> Manhattan_result
 [1] setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa     setosa    
[13] setosa     setosa     setosa     versicolor versicolor versicolor versicolor versicolor versicolor versicolor virginica  versicolor
[25] versicolor versicolor versicolor versicolor virginica  versicolor virginica  virginica  virginica  virginica  virginica  virginica 
[37] virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica  virginica 
Levels: setosa versicolor virginica
> confusionMatrix(Manhattan_result, iris.test_raw$Species)
Confusion Matrix and Statistics

            Reference
Prediction   setosa versicolor virginica
  setosa         15          0         0
  versicolor      0         13         0
  virginica       0          2        15

Overall Statistics
                                          
               Accuracy : 0.9556          
                 95% CI : (0.8485, 0.9946)
    No Information Rate : 0.3333          
    P-Value [Acc > NIR] : < 2.2e-16       
                                          
                  Kappa : 0.9333          
                                          
 Mcnemar's Test P-Value : NA              

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.8667           1.0000
Specificity                 1.0000            1.0000           0.9333
Pos Pred Value              1.0000            1.0000           0.8824
Neg Pred Value              1.0000            0.9375           1.0000
Prevalence                  0.3333            0.3333           0.3333
Detection Rate              0.3333            0.2889           0.3333
Detection Prevalence        0.3333            0.2889           0.3778
Balanced Accuracy           1.0000            0.9333           0.9667
```

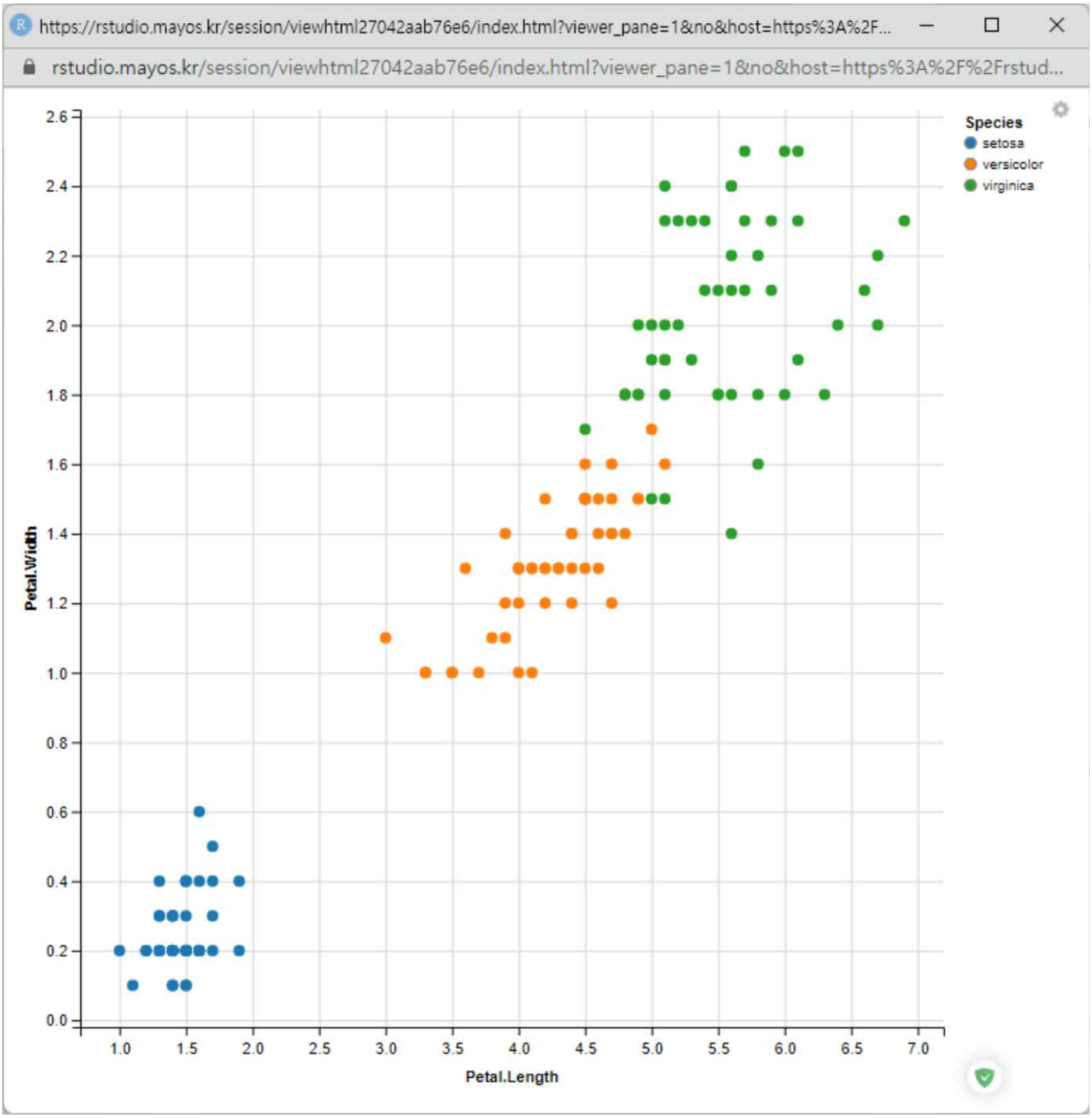
Result: Accuracy 95.56%

Case 5. confusionMatrix(Mahalanobis_result, iris.test_raw$Species)



```
> Mahalanobis_result <- knn.t(train=iris.train, test=iris.test, cl=iris.train_raw[, 5], Mahalanobis_distance, k = 5)
> Mahalanobis_result
 [1] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[21] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
[41] setosa setosa setosa setosa setosa
Levels: setosa
> confusionMatrix(Mahalanobis_result, iris.test_raw$Species)
Confusion Matrix and Statistics

            Reference
Prediction   setosa versicolor virginica
  setosa         15         15        15
  versicolor      0          0         0
  virginica       0          0         0

Overall Statistics
                                          
               Accuracy : 0.3333          
                 95% CI : (0.2, 0.4895)   
    No Information Rate : 0.3333          
    P-Value [Acc > NIR] : 0.5558          
                                          
                  Kappa : 0               
                                          
 Mcnemar's Test P-Value : NA              

Statistics by Class:

                     Class: setosa Class: versicolor Class: virginica
Sensitivity                 1.0000            0.0000           0.0000
Specificity                 0.0000            1.0000           1.0000
Pos Pred Value              0.3333               NaN              NaN
Neg Pred Value                 NaN            0.6667           0.6667
Prevalence                  0.3333            0.3333           0.3333
Detection Rate              0.3333            0.0000           0.0000
Detection Prevalence        1.0000            0.0000           0.0000
Balanced Accuracy           0.5000            0.5000           0.5000
Warning message:
In confusionMatrix.default(Mahalanobis_result, iris.test_raw$Species) :
  Levels are not in the same order for reference and data. Refactoring data to match.
```

Result: Accuracy 33.33%



Petal.Length, Petal.Width가 Result인 Species를 어떻게 군집화 하는지 그려보았으며, Mahalanobis Distance 특성상, Result가 Feature의 추세를 따른다면, 실제 나타나는 거리보다 가깝게 표현하므로, 본인의 Mahalanobis Distance KNN Model은 Setosa만 분리 가능했다.