

## # KNN Model

```
library(tidyverse)
```

```
data("iris")
```

```
str(iris)
```

```
head(iris)
```

## # Loading package

```
library(e1071)
```

```
library(caTools)
```

```
library(class)
```

## # Splitting data into train and test data

```
split <- sample.split(iris, SplitRatio = 0.7)
```

```
train_cl <- subset(iris, split == "TRUE")
```

```
test_cl <- subset(iris, split == "FALSE")
```

## # Feature Scaling

```
train_scale <- scale(train_cl[, 1:4])
```

```
test_scale <- scale(test_cl[, 1:4])
```

```
head(train_scale)
```

```
head(test_scale)
```

## # Fitting KNN Model to training dataset

```
classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 1)
```

```
classifier_knn
```

## # Confusion Matrix

```
cm <- table(test_cl$Species, classifier_knn)
```

```
cm
```

## # Model Evaluation - Choosing K

### # Calculate out of Sample error

```
misClassError <- mean(classifier_knn != test_cl$Species)
```

```
print(paste('Accuracy =', 1-misClassError))
```

## # K = 3

```
classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 3)
```

```
misClassError <- mean(classifier_knn != test_cl$Species)
```

```

print(paste('Accuracy =', 1-misClassError))

# K = 5

classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 5)
misClassError <- mean(classifier_knn != test_cl$Species)
print(paste('Accuracy =', 1-misClassError))

# K = 7

classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 7)
misClassError <- mean(classifier_knn != test_cl$Species)
print(paste('Accuracy =', 1-misClassError))

# K = 15

classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 15)
misClassError <- mean(classifier_knn != test_cl$Species)
print(paste('Accuracy =', 1-misClassError))

# K = 19

classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = 19)
misClassError <- mean(classifier_knn != test_cl$Species)
print(paste('Accuracy =', 1-misClassError))

library(ggplot2)

# Data preparation

k_values <- c(1, 3, 5, 7, 15, 19)

# Calculate accuracy for each k value

accuracy_values <- sapply(k_values, function(k) {
  classifier_knn <- knn(train = train_scale, test = test_scale, cl = train_cl$Species, k = k)
  1 - mean(classifier_knn != test_cl$Species) })

# Create a data frame for plotting

accuracy_data <- data.frame(K = k_values, Accuracy = accuracy_values)

# Plotting

ggplot(accuracy_data, aes(x = K, y = Accuracy)) +
  geom_line(color = "lightblue", size = 1) + geom_point(color = "lightgreen", size = 3) +
  labs(title = "Model Accuracy for Different K Values",
       x = "Number of Neighbors (K)", y = "Accuracy") + theme_minimal()

```