

“**Experiment 2.3”**

**Student Name: SUMIT KUMAR UID: 20BCS8226**

**Branch: CSE Section/Group: DM-720 A**

**Subject Name: Data Mining Lab Subject Code: 20CSP-376**

**Aim:**

To perform the cluster analysis by k-means method using R.

**Theory:**

[K Means Clustering](https://www.geeksforgeeks.org/k-means-clustering-introduction/) in [R Programming](https://www.geeksforgeeks.org/introduction-to-r-programming-language/) is an Unsupervised Non-linear algorithm that cluster data based on similarity or similar groups. It seeks to partition the observations into a pre-specified number of clusters. Segmentation of data takes place to assign each training example to a segment called a cluster. In the unsupervised algorithm, high reliance on raw data is given with large expenditure on manual review for review of relevance is given. It is used in a variety of fields like Banking, healthcare, retail, Media, etc.

K-Means clustering groups the data on similar groups. The algorithm is as follows:

* Choose the number **K** clusters.
* Select at random K points, the centroids(Not necessarily from the given data).
* Assign each data point to closest centroid that forms K clusters.
* Compute and place the new centroid of each centroid.
* After final reassignment, name the cluster as Final cluster.

**Code:**

# Loading data

data(iris)



# Structure

str(iris)

|  |
| --- |
| # Installing Packages  install.packages("ClusterR")  install.packages("cluster")    # Loading package  library(ClusterR)  library(cluster)    # Removing initial label of  # Species from original dataset  iris\_1 <- iris[, -5]    # Fitting K-Means clustering Model  # to training dataset  set.seed(240) # Setting seed  kmeans.re <- kmeans(iris\_1, centers = 3, nstart = 20)  kmeans.re |

Cluster identification for

# each observation

kmeans.re$cluster

# Confusion Matrix

cm <- table(iris$Species, kmeans.re$cluster)

cm

# Model Evaluation and visualization

plot(iris\_1[c("Sepal.Length", "Sepal.Width")])

plot(iris\_1[c("Sepal.Length", "Sepal.Width")],

     col = kmeans.re$cluster)

plot(iris\_1[c("Sepal.Length", "Sepal.Width")],

     col = kmeans.re$cluster,

     main = "K-means with 3 clusters")

## Plotiing cluster centers

kmeans.re$centers

kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")]

# cex is font size, pch is symbol

points(kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")],

       col = 1:3, pch = 8, cex = 3)

## Visualizing clusters

y\_kmeans <- kmeans.re$cluster



clusplot(iris\_1[, c("Sepal.Length", "Sepal.Width")],

         y\_kmeans,

         lines = 0,

         shade = TRUE,

         color = TRUE,

         labels = 2,

         plotchar = FALSE,

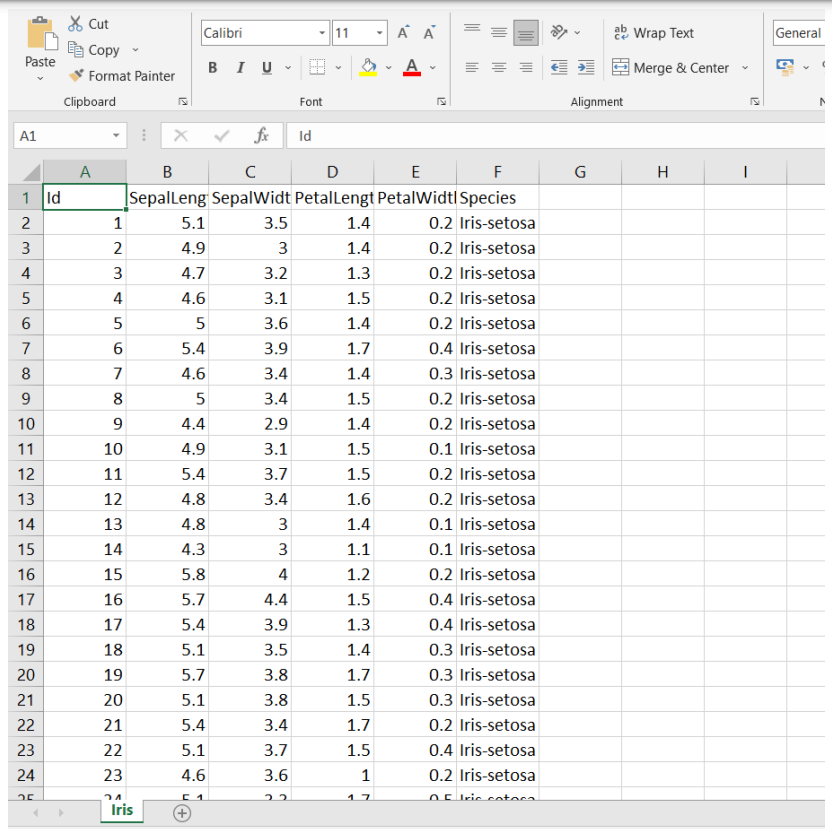
         span = TRUE,

         main = paste("Cluster iris"),

         xlab = 'Sepal.Length',

        ylab = 'Sepal.Width')

**DATASET USED:**





**OUTPUT:**

