Experiment-2.3

Student Name: Nabha Varshney **UID: 20BCS4995**

Branch: CSE **Section/Group: 20BCS-DM-704 (A)**

Semester: 6th **Date of Performance: 21thApr 2023**

Subject Name: Data Mining **Subject Code: 20CSP- 351**

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

Objective-

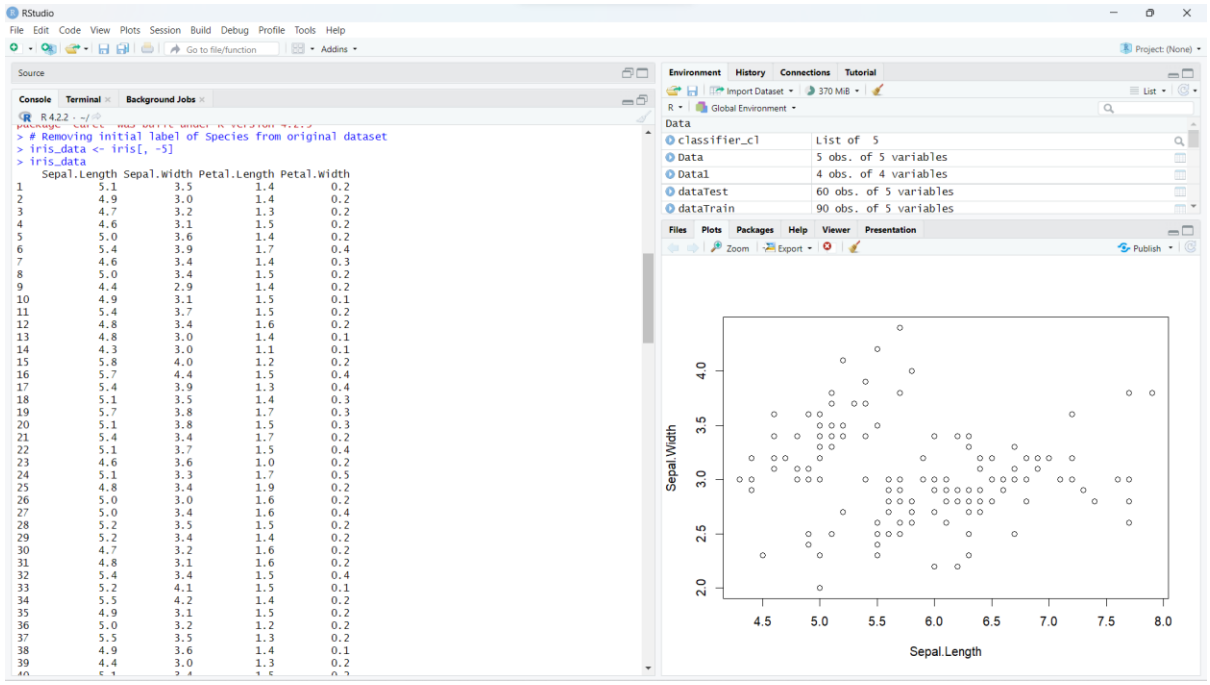
- Represent the reading of file using R studio
- Displaying the graph using clusterR, caret and cluster.
- Demonstration of Cluster analysis by K-means method.

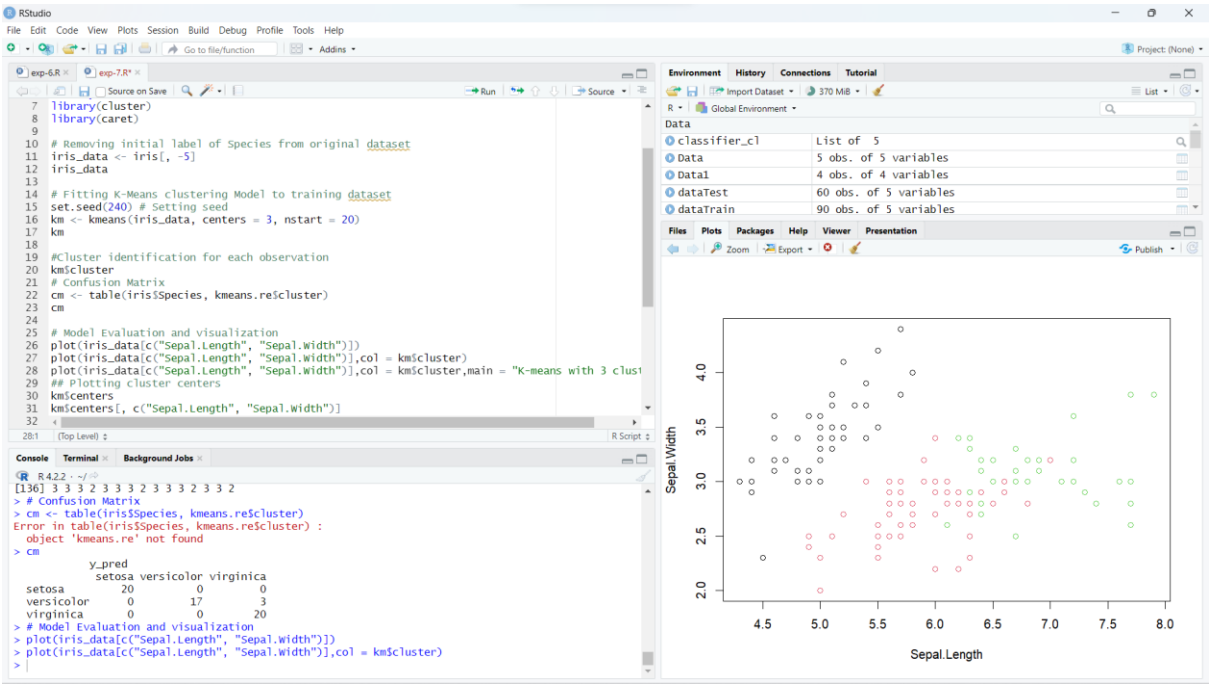
Script and Output-

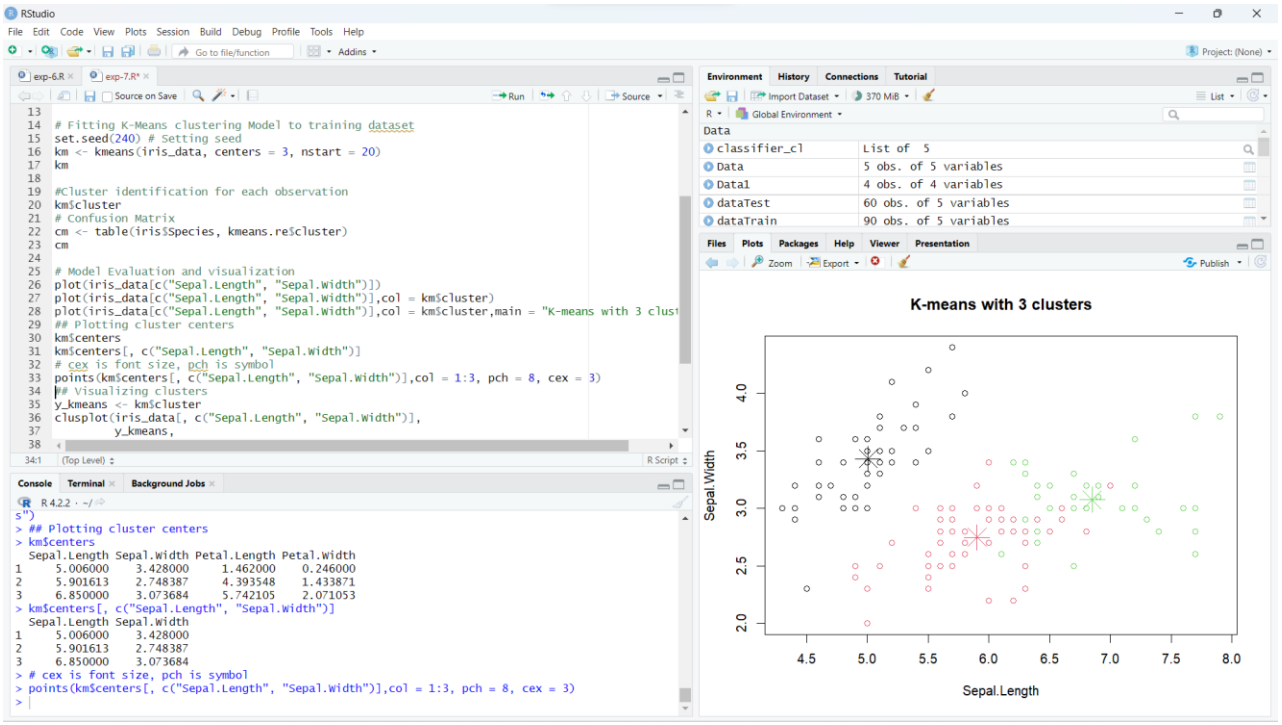
```
str(iris)
# Installing Packages
install.packages("ClusterR")
install.packages("cluster")
# Loading package
library(ClusterR)
library(cluster)
library(caret)

# Removing initial label of Species from original dataset
iris_data <- iris[, -5]
iris_data

# Fitting K-Means clustering Model to training dataset
set.seed(240) # Setting seed
```

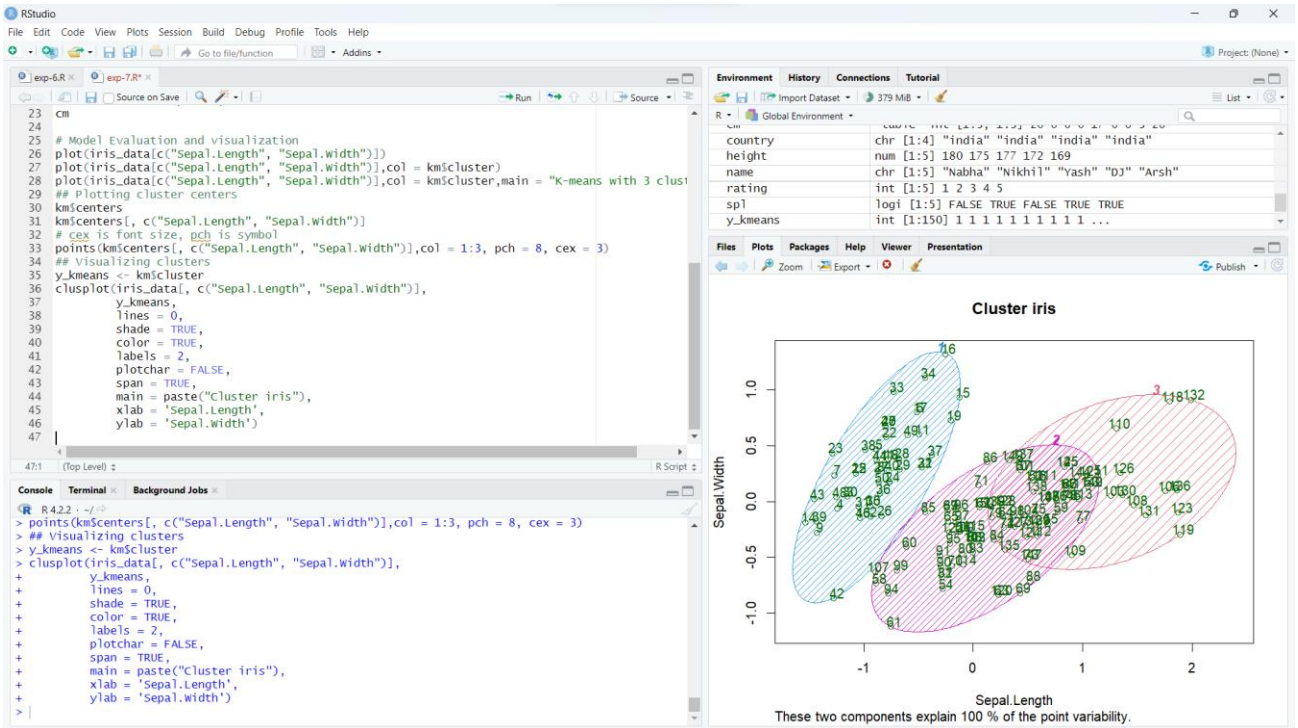
```
km <- kmeans(iris_data, centers = 3, nstart = 20)
km

#Cluster identification for each observation
km$cluster
# Confusion Matrix
cm <- table(iris$Species, kmeans.re$cluster)
cm

# Model Evaluation and visualization
plot(iris_data[c("Sepal.Length", "Sepal.Width")])
plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)
plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster,main = "K-
means with 3 clusters")
## Plotting cluster centers
km$centers
km$centers[, c("Sepal.Length", "Sepal.Width")]
# cex is font size, pch is symbol
points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex =
3)
## Visualizing clusters
y_kmeans <- km$cluster
clusplot(iris_data[, 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')
```

Output-


RStudio
> # Removing initial label of Species from original dataset
> iris_data <- iris[, -5]
> iris_data
Sepal.Length Sepal.Width Petal.Length Petal.Width
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Sepal.Width
Sepal.Length


RStudio
library(cluster)
library(caret)
# Removing initial label of Species from original dataset
iris_data <- iris[, -5]
iris_data
Fitting K-Means clustering Model to training dataset
set.seed(240) # Setting seed
km <- kmeans(iris_data, centers = 3, nstart = 20)
km
#Cluster identification for each observation
km$cluster
# Confusion Matrix
cm <- table(iris$Species, kmeans.re$cluster)
cm
# Model Evaluation and visualization
plot(iris_data[c("Sepal.Length", "Sepal.Width")])
plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)
plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster,main = "K-means with 3 clust
Plotting cluster centers
km$centers
km$centers[, c("Sepal.Length", "Sepal.Width")]
[136] 3 3 3 2 3 3 3 2 3 3 3 2 3 3 2
> # Confusion Matrix
> cm <- table(iris$Species, kmeans.re$cluster)
Error in table(iris$Species, kmeans.re$cluster) :
object 'kmeans.re' not found
> cm
y_pred
setosa versicolor virginica
setosa 20 0 0
versicolor 0 17 3
virginica 0 0 20
> # Model Evaluation and visualization
> plot(iris_data[c("Sepal.Length", "Sepal.Width")])
> plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)
Sepal.Width
Sepal.Length


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
exp-7.R*
Source on Save
Run
Source
13
14 # Fitting K-Means clustering Model to training dataset
15 set.seed(240) # Setting seed
16 km <- kmeans(iris_data, centers = 3, nstart = 20)
17 km
18
19 #Cluster identification for each observation
20 km$cluster
21 # Confusion Matrix
22 cm <- table(iris$Species, kmeans.re$cluster)
23 cm
24
25 # Model Evaluation and visualization
26 plot(iris_data[c("Sepal.Length", "Sepal.Width")])
27 plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)
28 plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster,main = "K-means with 3 clust
29 ## Plotting cluster centers
30 km$centers
31 km$centers[, c("Sepal.Length", "Sepal.Width")]
32 # cex is font size, pch is symbol
33 points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex = 3)
34 ## Visualizing clusters
35 y_kmeans <- km$cluster
36 clusplot(iris_data[, c("Sepal.Length", "Sepal.Width")],
37 y_kmeans,
38
34:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.2.2 · ~/
s")
> ## Plotting cluster centers
> km$centers
Sepal.Length Sepal.Width Petal.Length Petal.Width
1 5.006000 3.428000 1.462000 0.246000
2 5.901613 2.748387 4.393548 1.433871
3 6.850000 3.073684 5.742105 2.071053
> km$centers[, c("Sepal.Length", "Sepal.Width")]
Sepal.Length Sepal.Width
1 5.006000 3.428000
2 5.901613 2.748387
3 6.850000 3.073684
> # cex is font size, pch is symbol
> points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex = 3)
>
Environment History Connections Tutorial
Import Dataset 370 MiB
List
R Global Environment
Data
classifier_cl List of 5
Data 5 obs. of 5 variables
Data1 4 obs. of 4 variables
dataTest 60 obs. of 5 variables
dataTrain 90 obs. of 5 variables
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
K-means with 3 clusters
Sepal.Width
Sepal.Length
2.0 2.5 3.0 3.5 4.0
4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
exp-6.R
exp-7.R*
Source on Save
Run
Source
23 cm
24
25 # Model Evaluation and visualization
26 plot(iris_data[c("Sepal.Length", "Sepal.Width")])
27 plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster)
28 plot(iris_data[c("Sepal.Length", "Sepal.Width")],col = km$cluster,main = "K-means with 3 clust
29 ## Plotting cluster centers
30 km$centers
31 km$centers[, c("Sepal.Length", "Sepal.Width")]
32 # cex is font size, pch is symbol
33 points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex = 3)
34 ## Visualizing clusters
35 y_kmeans <- km$cluster
36 clusplot(iris_data[, c("Sepal.Length", "Sepal.Width")],
37 y_kmeans,
38 lines = 0,
39 shade = TRUE,
40 color = TRUE,
41 labels = 2,
42 plotchar = FALSE,
43 span = TRUE,
44 main = paste("Cluster iris"),
45 xlab = 'Sepal.Length',
46 ylab = 'Sepal.Width')
47
47:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.2.2 · ~/
> points(km$centers[, c("Sepal.Length", "Sepal.Width")],col = 1:3, pch = 8, cex = 3)
> ## Visualizing clusters
> y_kmeans <- km$cluster
> clusplot(iris_data[, 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')
>
Environment History Connections Tutorial
Import Dataset 379 MiB
List
R Global Environment
country chr [1:4] "india" "india" "india" "india"
height num [1:5] 180 175 177 172 169
name chr [1:5] "Nabha" "Nikhil" "Yash" "DJ" "Arsh"
rating int [1:5] 1 2 3 4 5
spl logi [1:5] FALSE TRUE FALSE TRUE TRUE
y_kmeans int [1:150] 1 1 1 1 1 1 1 1 1 1 ...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Cluster iris
Sepal.Width
Sepal.Length
-1.0 -0.5 0.0 0.5 1.0
-1 0 1 2
These two components explain 100 % of the point variability.