Exp.No: 9

IMPLEMENT CLUSTERING TECHNIQUES – HIERARCHICAL AND KMEANS

AIM:

To write an R code to implement hierarchical and k-means clustering techniques.

PROCEDURE:

- 1. Load the iris dataset and use only the numeric columns for clustering by excluding the Species column.
- 2. Standardize the data to ensure all variables have equal weight in the clustering process.
- 3. Compute the distance matrix using the Euclidean method and perform hierarchical clustering using the "complete" linkage method, plot the dendrogram, and cut the tree to form 3 clusters.
- 4. Perform K-means clustering by setting the number of clusters, run the clustering algorithm, and add cluster assignments to the original dataset.
- 5. Display the first few rows of the updated dataset and plot the clusters using ggplot2 for visualization.

PROGRAM CODE:

a) HIERARCHIAL CLUSTERING

```
# Load the iris dataset data(iris)

# Use only the numeric columns for clustering (exclude the Species column) iris_data
<- iris[, -5]

# Standardize the data iris_scaled
<- scale(iris_data)

# Compute the distance matrix distance_matrix <- dist(iris_scaled,
method = "euclidean")

# Perform hierarchical clustering using the "complete" linkage method hc_complete
<- hclust(distance_matrix, method = "complete")

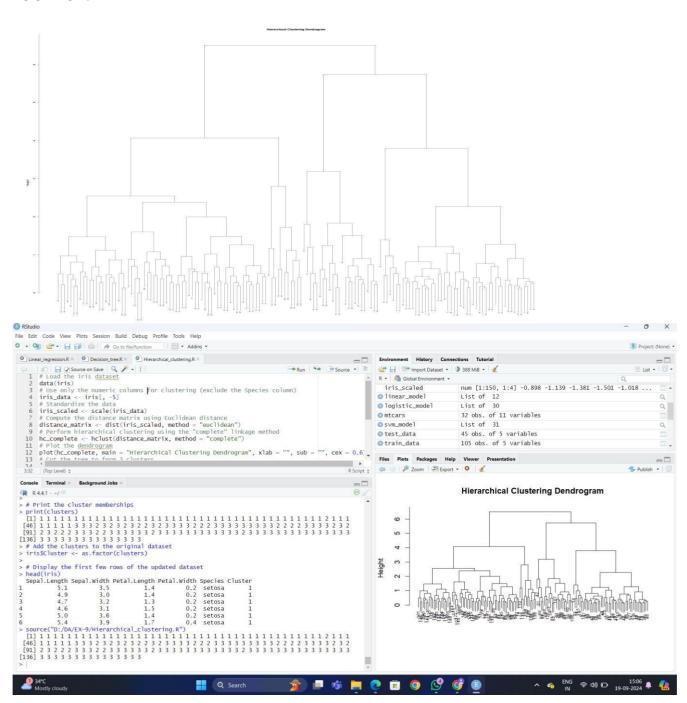
# Plot the dendrogram plot(hc_complete, main = "Hierarchical Clustering Dendrogram", xlab
= "", sub = "", cex =

0.6)
```

Cut the tree to form 3 clusters clusters <- cutree(hc_complete, k = 3) # Print the cluster memberships print(clusters)

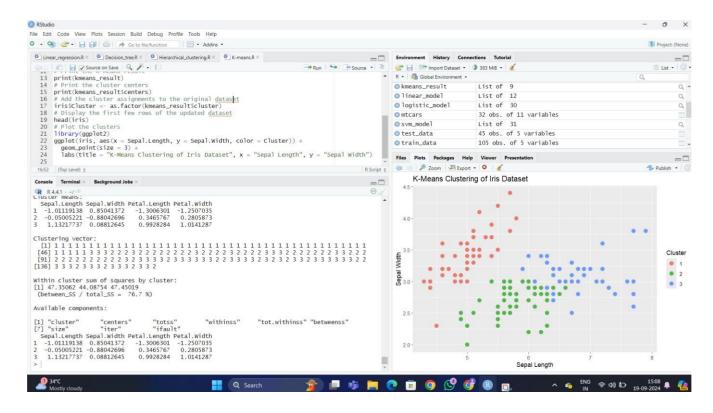
- # Add the clusters to the original dataset iris\$Cluster
- <- as.factor(clusters)
- # Display the first few rows of the updated dataset head(iris)

OUTPUT:



b) K-MEANS CLUSTERING

```
# Load the iris
dataset data(iris)
 # Use only the numeric columns for clustering (exclude the Species column) iris data
 <- iris[, -5]
 # Standardize the data iris scaled
 <- scale(iris data)
 # Set the number of clusters set.seed(123)
 # For reproducibility k <- 3
 # Number of clusters
# Perform K-Means clustering
 kmeans result <- kmeans(iris scaled, centers = k, nstart = 25)
# Print the K-Means result print(kmeans_result)
 # Print the cluster centers
 print(kmeans result$centers)
 # Add the cluster assignments to the original dataset iris$Cluster
 <- as.factor(kmeans result$cluster)
# Display the first few rows of the updated dataset head(iris)
 # Plot the clusters library(ggplot2)
 ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Cluster)) +
 geom point(size = 3) +
  labs(title = "K-Means Clustering of Iris Dataset", x = "Sepal Length", y = "Sepal Width")
                                                                                 OUTPUT:
```



RESULT:

Thus the R program to implement hierarchical and k-means clustering techniques has been executed and verified successfully.