PAM CLUSTERING

CODE

# Load the required library

library(cluster)

# Read the dataset

data <- read.csv("C:/Users/sriva/OneDrive/Desktop/ObesityDataSet.csv")

# Select only the numeric columns for clustering

numeric\_data <- data[, c("Age", "Height", "Weight", "FCVC", "NCP", "CH2O", "FAF", "TUE")]

# Scale the numeric data

scaled\_data <- scale(numeric\_data)

# Determine the number of clusters using silhouette width

max\_clusters <- 10

sil\_width <- numeric(max\_clusters)

for (i in 2:max\_clusters) {

pam\_fit <- pam(scaled\_data, k = i, diss = TRUE)

sil\_width[i] <- pam\_fit$silinfo$avg.width

}

# Plot silhouette width to determine the optimal number of clusters

plot(1:max\_clusters, sil\_width, type = "b", xlab = "Number of Clusters",

ylab = "Silhouette Width")

# Determine the optimal number of clusters based on silhouette width

optimal\_clusters <- which.max(sil\_width)

# Perform PAM clustering with the optimal number of clusters

pam\_fit <- pam(scaled\_data, k = optimal\_clusters, diss = TRUE)

# Print cluster results

print(pam\_fit)

# Visualize clustering results

plot(scaled\_data, col = pam\_fit$clustering, pch = 20,

main = paste("PAM Clustering with", optimal\_clusters, "Clusters"))

points(pam\_fit$medoids, col = 1:optimal\_clusters, pch = 8, cex = 2)

legend("topleft", legend = 1:optimal\_clusters, col = 1:optimal\_clusters, pch = 8, title = "Cluster")