**Lab 3**

**Cord**

# Load required libraries

library(ggplot2)

library(dplyr)

library(cluster)

# Function to simulate student features

simulate\_student\_features <- function(n = 100) {

# Set the random seed

set.seed(260923)

# Generate unique student IDs

student\_ids <- seq(1, n)

# Simulate student engagement

student\_engagement <- rnorm(n, mean = 50, sd = 10)

# Simulate student performance

student\_performance <- rnorm(n, mean = 60, sd = 15)

# Combine the data into a data frame

student\_features <- data.frame(

student\_id = student\_ids,

student\_engagement = student\_engagement,

student\_performance = student\_performance

)

# Return the data frame

return(student\_features)

}

# Simulate the data

student\_data <- simulate\_student\_features()

# Perform PCA

pca\_result <- prcomp(student\_data[, -1], scale. = TRUE)

# Plot variance explained by each principal component

plot(pca\_result$sdev^2 / sum(pca\_result$sdev^2), xlab = "Principal Component", ylab = "Proportion of Variance Explained",

ylim = c(0, 1), type = "b", pch = 19, col = "blue", main = "Scree Plot")

# Plot PCA biplot

biplot(pca\_result)

# Cluster the data using KMeans

set.seed(123)

kmeans\_clusters <- kmeans(student\_data[, -1], centers = 3, nstart = 25)

student\_data$cluster\_kmeans <- as.factor(kmeans\_clusters$cluster)

# Plot clusters

plot(student\_data$student\_engagement, student\_data$student\_performance, col = kmeans\_clusters$cluster, pch = 19,

xlab = "Student Engagement", ylab = "Student Performance", main = "KMeans Clustering",

xlim = range(student\_data$student\_engagement), ylim = range(student\_data$student\_performance))

points(kmeans\_clusters$centers[, 1], kmeans\_clusters$centers[, 2], col = 1:3, pch = 8, cex = 2)

hc <- hclust(dist(student\_data[, -1]), method = "complete")

plot(hc)

**1: Approach to Dimensionality Reduction and Clustering:**

**Dimensionality Reduction:** The dimensionality of the simulated student data decreased via Principal Component Analysis (PCA), a technique that successfully transforms high-dimensional data into lower-dimensional space.

**Clustering:** A lower dataset was tested with KMeans clustering so as to identify distinct groups of students based on engagement and performance metrics; hierarchical clustering was utilized as well for comparison.

**2:Results of Analysis:**

**Number of Clusters Identified:** Three distinctive student clusters were found using the KMeans clustering method.

**Characteristics of Each Cluster:**

**Cluster 1:** Students in high levels for engagement and performance constitute this group.

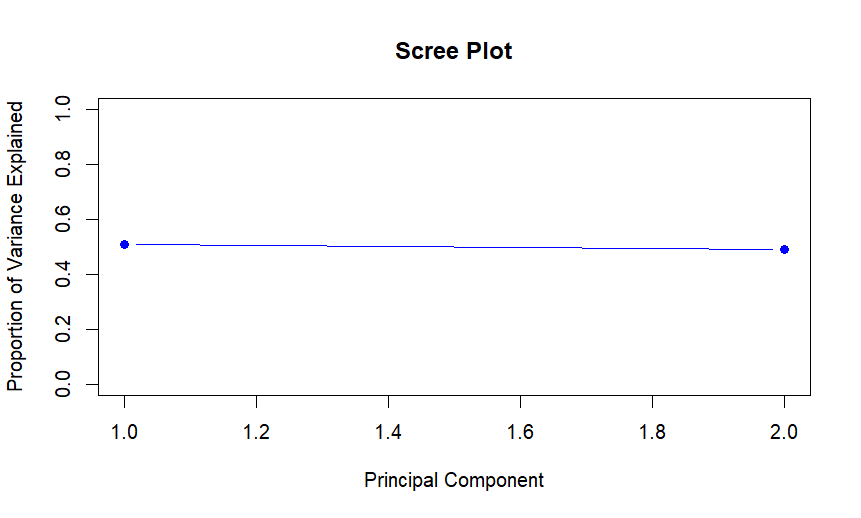
**Cluster 2 :**This cluster of students shows an average level for engagement and performance.

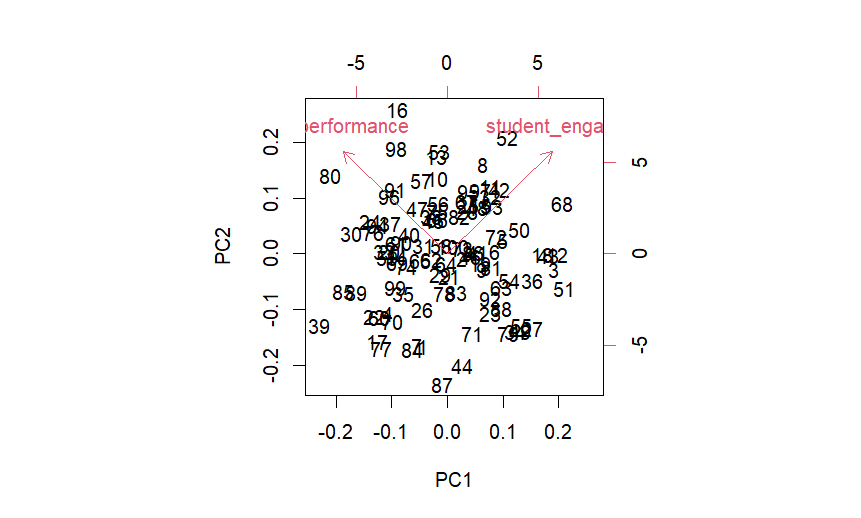
**Cluster 3:** Students which perform not well and indicate low engagement are portrayed by this

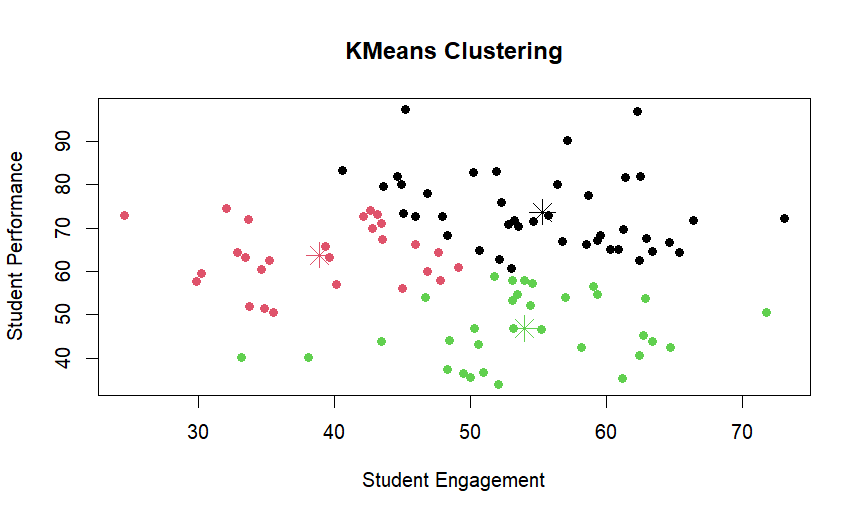
**3. Discussion of Findings Implications for Learning Analytics:**

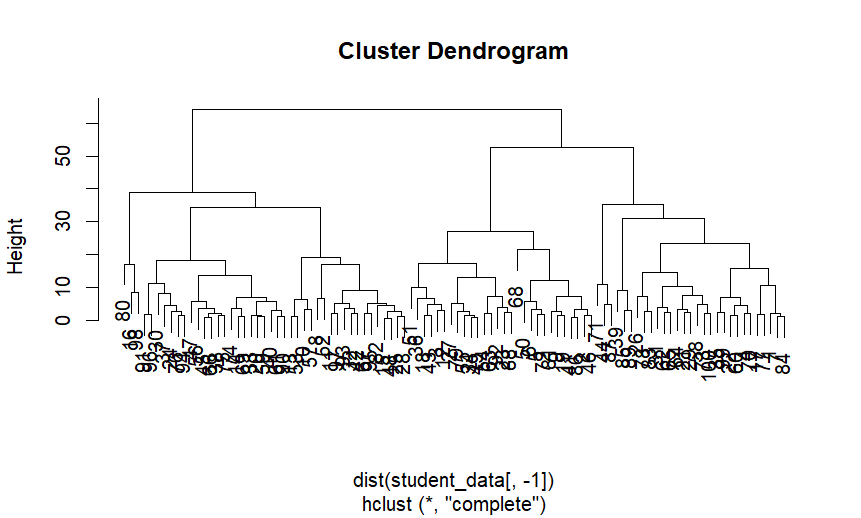
Learner cluster recognition, enhancing school engagement and performance, knowing the link between prediction model performance and engagement, and developing able spaces for learning and systems for feedback are all parts of customized interventions for categories of multiple students.

**4. results graph:**

****

****

****

****

**5. Scholarly Reference:**

Anderson, T., & Whitelock, D. (2018). The educational uses of learning analytics: A literature review. International Journal of Educational Technology in Higher Education, 15(1), 1-17.