Capstone Project

Assignment 2

Course code: CSA1643

Course: Data Mining and Data Warehousing

S.No: 18

Name: G.Naga Sai Vishnu Sanjeev

Reg. No: 192210101

Slot: C

Title: Educational Data Mining for adaptive learning systems

Assignment Release Date :

Assignment Preliminary Stage (Assignment 2) submission Date :

Mentor Name : Dr Murphy M

Mentor Phone number and Department :80562 84236 , Electrochemistry

CODE:

```r
# Load necessary libraries
library(dplyr) # For data manipulation
library(ggplot2) # For data visualization

# Load the dataset
data <- read.csv("dataset.csv")

# Explore the structure of the dataset
str(data)

# Summary statistics of the dataset
summary(data)

# Data preprocessing (cleaning, transformation, feature engineering, etc.)
# Example:
# Remove missing values
data <- na.omit(data)

# Perform exploratory data analysis (EDA)
# Example:
# Plot distribution of student grades
ggplot(data, aes(x = grade)) +
  geom_histogram(fill = "skyblue", color = "black", bins = 20) +
  labs(title = "Distribution of Student Grades")

# Implement machine learning algorithms for educational data mining
# Example:
# Predict student performance using linear regression
model <- lm(grade ~ hours_studied + interaction_time, data = data)

# Evaluate model performance
summary(model)

# Make predictions
predictions <- predict(model, newdata = data)

# Assess model accuracy
accuracy <- mean((data$grade - predictions)^2)
```

```r
print(paste("Mean Squared Error:", accuracy))
# Conduct further analyses based on the objectives of educational data mining
# Example:
# Analyze patterns in student interactions with learning materials
interaction_analysis <- data %>%
  group_by(student_id) %>%
  summarize(total_interactions = sum(interaction_time))

# Visualize interaction patterns
ggplot(interaction_analysis, aes(x = total_interactions)) +
  geom_histogram(fill = "lightgreen", color = "black", bins = 20) +
  labs(title = "Distribution of Total Interactions with Learning Materials")

# Implement adaptive learning algorithms based on EDM insights
# Example:
# Adjust learning content based on student performance and interaction patterns
# This would typically involve more complex algorithms and integration with a learning
platform.

# Continuous improvement and iteration based on insights gained from EDM
# Example:
# Collect feedback on the effectiveness of adaptive learning interventions and make adjustments
accordingly
```

Output:

R ▾ Global Environment ▾		
Data		
▶ data	10 obs. of 4 variables	▦
▶ interaction_analy…	10 obs. of 2 variables	▦
▶ model	List of 12	🔍
▶ mtcars	32 obs. of 11 variables	▦
▶ plot	List of 11	🔍
▶ test_data	200 obs. of 4 variables	▦
▶ train_data	800 obs. of 4 variables	▦
train_index	int [1:800, 1] 1 3 4 5 6 7 8 9 10 11 ...	▦
Values		
accuracy	1.4485628742515	
levels	chr [1:2] "Yes" "No"	
num_students	1000	
predictions	Named num [1:10] 82.7 72.2 93.2 77.9 64.3 ...	

GRAPH:

Distribution of Total Interactions with Learning Materials