**DATA SCIENCE FINAL PROJECT**

**Student Mental Health**

Dataset Description:

The 'Student Mental Health' dataset is a comprehensive collection aimed at exploring factors related to students' mental health. It includes key variables such as age, gender, CGPA (Cumulative Grade Point Average), course enrolment, and indicators for depression, anxiety, panic attacks, and seeking help. This dataset provides insights into demographic distributions, academic performance correlations with mental health, the prevalence of mental health conditions, and patterns of help-seeking behaviour. Potential use cases involve identifying risk factors, developing predictive models, and informing targeted interventions. Ethical considerations emphasize responsible data handling due to the sensitivity of mental health information. Overall, the dataset is a valuable resource for researchers, educators, and policymakers interested in addressing mental health challenges in the student population.

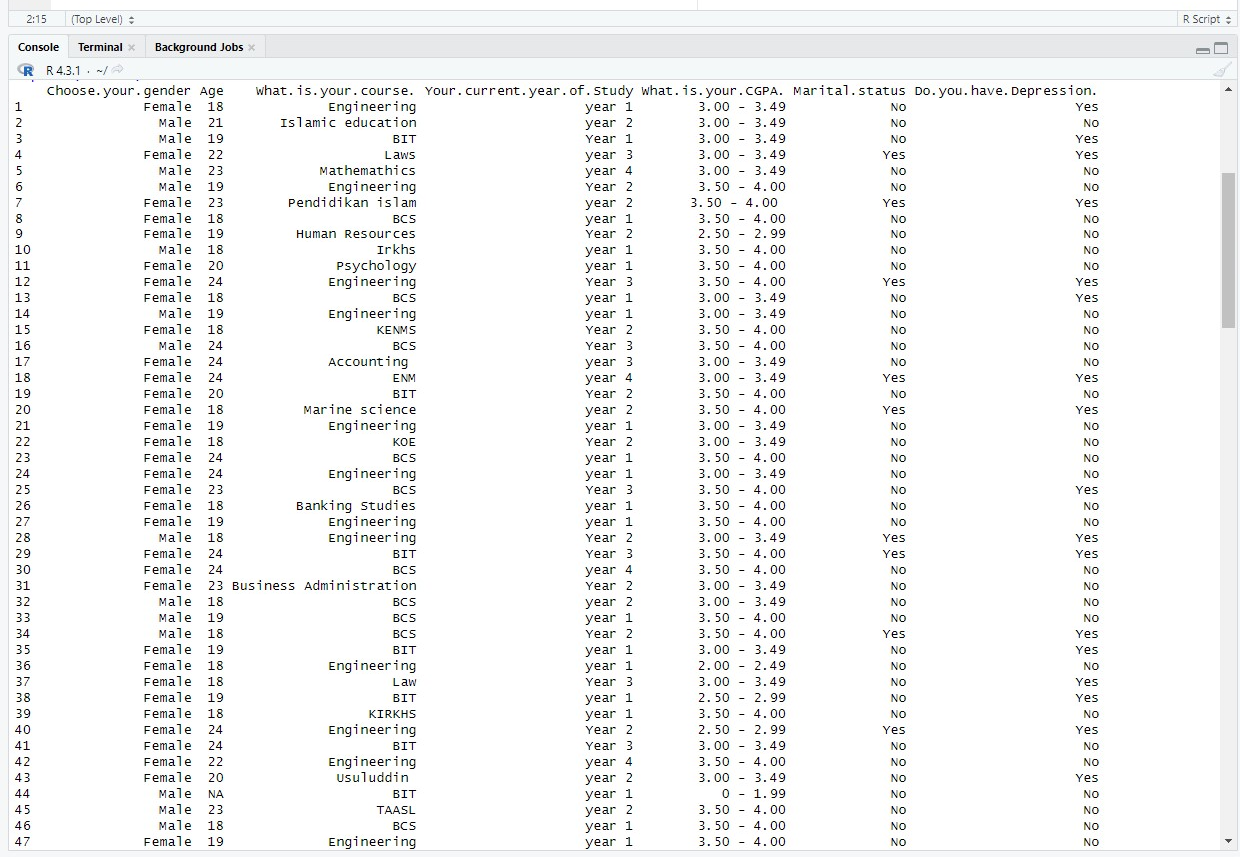
CODES – CONSOLE – DETAILS:

1. Import CSV file.

Code:

dataset <- read.csv("D:/FALL2023/IntroToDataScience/Student\_Mental\_health.csv", header = TRUE, sep = ",")

print(dataset)



*This code is used to import external Excel files (in CSV format) into R.*

1. Find missing values.

Code:

missing\_values <- is.na(dataset)

print(missing\_values)

A screenshot of a computer

Description automatically generated

*This code is used to identify the missing values in a dataset. The one that has TRUE written in it is a missing value.*

1. Discard missing values.

Code:

dataset <- na.omit(dataset)

print(dataset)A screenshot of a computer

Description automatically generated

*This code is used to remove the instance that had a missing value.*

1. Converting Numeric Age values to Categorical Age values.

Code:

catage <- "Age"

breaks <- c(0, 19, 25, Inf)

labels <- c("Teenager", "Young Adult", "Adult")

dataset$CatAge <- cut(dataset[[catage]], breaks = breaks, labels = labels, include.lowest = TRUE)

print(dataset)A screenshot of a computer

Description automatically generated

*This code is used to convert the categorize the age range to teenager, young adult, and adult. It is saved as a new column.*

1. Pearson’s Chi-squared Test

Code:

dataset <- data.frame(

"what.is.your.cgpa" = sample(c("0-1.99", "2-2.49", "2.5-2.99", "3-3.49", "3.5-4.00"), 100, replace = TRUE),

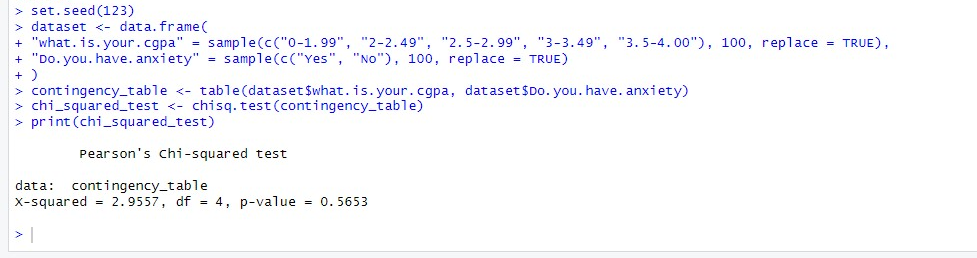
"Do.you.have.anxiety" = sample(c("Yes", "No"), 100, replace = TRUE)

)

contingency\_table <- table(dataset$What.is.your.cgpa., dataset$Do.you.have.anxiety.)

chi\_squared\_test <- chisq.test(contingency\_table)

print(chi\_squared\_test)



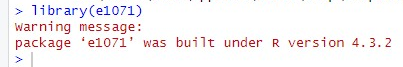
*This code is used to find the significant attributes using Pearson’s Chi-Squared Test.*

1. Install Package
2. e1071

Code:

install.packages("e1071")

library(e1071)



*This code is used to install the necessary package for the ‘Naïve Bayes’ function.*

1. caret

Code:

install.packages("caret")

library(caret)

A screen shot of a computer code

Description automatically generated

*This code is used to install the necessary package for the ‘Naïve Bayes’ classification.*

1. Naïve Bayes

Code:

nb\_model <- naiveBayes(What.is.your.course. ~ ., data = dataset)

print(nb\_model)

A screenshot of a computer

Description automatically generated

*This code is used to find the Naïve Bayes function to predict, based on the course of the student.*

1. Confusion Matrix

Code:

pred <- predict(nb\_model, dataset)

confusion\_matrix <- table(dataset$What.is.your.course., pred)

print(confusion\_matrix)

A screenshot of a computer

Description automatically generated

*This code has been used to find the confusion matrix on the student’s courses.*

1. Accuracy

Code:

accuracy <- sum(diag(confusion\_matrix)) / sum(confusion\_matrix)

print(paste("Accuracy:", accuracy))



*This code is used to find the predictive accuracy of the Naïve Bayes classification.*

1. 10-Fold Cross Validation

Code:

1. Train Set and Test Set

Code:

train\_indices <- sample(1:nrow(dataset), 0.7 \* nrow(dataset))

train\_data <- dataset[train\_indices, ]

test\_data <- dataset[-train\_indices, ]



*This code is used to generate the train set and test set according to this dataset.*

1. Recall, Precision and F-measure value

Code: