# Ensure the 'new\_students.csv' file is in the current working directory. Check with getwd() and change with setwd() as needed.

# Loading the student data using base R read.csv function

new\_students <- read.csv("new\_students.csv")

# Display the dataset

new\_students

# Data preprocessing: Transform 'result' into a binary variable and 'disability\_status' into a factor

new\_students$is\_passed <- as.factor(ifelse(new\_students$result == "Pass", 1, 0))

new\_students$studied\_credits <- as.factor(new\_students$credits\_completed)

# Convert 'economic\_status' to a numeric scale based on given categories

economic\_scale <- c("Low", "Medium", "High")

new\_students$economic\_numeric <- as.numeric(factor(new\_students$economic\_status, levels = economic\_scale))

# Creating train and test sets manually

set.seed(12345) # Setting seed for reproducibility

sample\_count <- floor(0.75 \* nrow(new\_students)) # Splitting the data into 75% as train data and 25% as test data

training <- sample(seq\_len(nrow(new\_students)), size = sample\_count)

train\_data <- new\_students[training, ]

test\_data <- new\_students[-training, ]

# Building a logistic regression model with glm (Generalized Linear Model) in base R

logit\_model <- glm(is\_passed ~ studied\_credits + economic\_numeric, family = binomial(link = "logit"), data = train\_data)

# Model summary display

summary(logit\_model)

# Tidying the model

library(broom)

tidy(logit\_model)

# Plotting the model

plot(logit\_model)

# Prediction on test data

test\_predictions <- predict(logit\_model, test\_data, type = "response")

predicted\_outcome <- ifelse(test\_predictions > 0.5, 1, 0)

# Accuracy calculation

true\_outcomes <- as.numeric(test\_data$is\_passed) - 1 # Adjusting factor levels from 1 to 0 and 1 for comparison

model\_accuracy <- mean(predicted\_outcome == true\_outcomes)

print(paste("Model Accuracy:", model\_accuracy))

licence()