SIM Code and their output with Screen-short

---

title: "R Notebook"

output: html\_notebook

---

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the \*Run\* button within the chunk or by placing your cursor inside it and pressing \*Ctrl+Shift+Enter\*.

```{r}

if (!require(caret)) install.packages("caret", dependencies=TRUE)

if (!require(corrplot)) install.packages("corrplot", dependencies=TRUE)

library(ggplot2)

library(dplyr)

library(caret)

library(randomForest)

library(corrplot)

library(reshape2)

plot(cars)

```

A graph with numbers and dots

AI-generated content may be incorrect.

# Load the dataset

stroke\_data <- read.csv("healthcare-dataset-stroke-data.csv")

# Remove 'id' column as it is not useful for prediction

stroke\_data <- stroke\_data[, !names(stroke\_data) %in% "id"]

# Check for missing values

sum(is.na(stroke\_data))



# Replace non-numeric values in 'bmi' with NA and impute missing values with mean

stroke\_data$bmi <- as.numeric(gsub("[^0-9.]", "", stroke\_data$bmi))

stroke\_data$bmi[is.na(stroke\_data$bmi)] <- mean(stroke\_data$bmi, na.rm = TRUE)

#Imputing missing values in 'bmi' with mean

stroke\_data$bmi[is.na(stroke\_data$bmi)] <- mean(stroke\_data$bmi, na.rm = TRUE)

# Convert categorical variables to factors

cat\_cols <- c("gender", "hypertension", "heart\_disease", "ever\_married",

"work\_type", "Residence\_type", "smoking\_status", "stroke")

stroke\_data[cat\_cols] <- lapply(stroke\_data[cat\_cols], as.factor)

# Distribution of Numerical Features

numeric\_features <- stroke\_data[, sapply(stroke\_data, is.numeric)]

par(mfrow = c(2,2))

lapply(names(numeric\_features), function(col) hist(numeric\_features[[col]],

main = paste("Histogram of", col), xlab = col, col = "blue"))

par(mfrow = c(1,1))

A screenshot of a computer

AI-generated content may be incorrect.

A white background with black numbers

AI-generated content may be incorrect.

A graph of different age groups

AI-generated content may be incorrect.

# Convert numeric features to long format explicitly

numeric\_features\_long <- melt(numeric\_features, id.vars = NULL, measure.vars = names(numeric\_features))

# Boxplot for Outlier Detection

ggplot(numeric\_features\_long, aes(x = variable, y = value)) +

geom\_boxplot(aes(fill = variable)) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +

ggtitle("Boxplot of Numerical Features")

A screenshot of a graph

AI-generated content may be incorrect.

# Correlation matrix

cor\_matrix <- cor(numeric\_features, use = "complete.obs")

corrplot(cor\_matrix, method = "color", tl.cex = 0.8)

A blue and white squares with red text

AI-generated content may be incorrect.

# Standardization (Scaling Numeric Features)

stroke\_data[names(numeric\_features)] <- scale(stroke\_data[names(numeric\_features)])

# Splitting data into training and testing sets

set.seed(123)

train\_index <- createDataPartition(stroke\_data$stroke, p = 0.8, list = FALSE)

train\_data <- stroke\_data[train\_index, ]

test\_data <- stroke\_data[-train\_index, ]

# Logistic Regression Model

stroke\_model <- glm(stroke ~ ., data = train\_data, family = binomial)

pred\_probs <- predict(stroke\_model, test\_data, type = "response")

pred\_labels <- ifelse(pred\_probs > 0.5, 1, 0)

conf\_matrix <- confusionMatrix(factor(pred\_labels, levels = levels(test\_data$stroke)), test\_data$stroke)

print(conf\_matrix)

A screenshot of a computer

AI-generated content may be incorrect.

# Decision Tree Model

tree\_model <- train(stroke ~ ., data = train\_data, method = "rpart")

tree\_pred <- predict(tree\_model, test\_data)

tree\_conf\_matrix <- confusionMatrix(tree\_pred, test\_data$stroke)

print(tree\_conf\_matrix)

A screenshot of a computer

AI-generated content may be incorrect.

# Random Forest Model

rf\_model <- randomForest(stroke ~ ., data = train\_data, ntree = 100)

rf\_pred <- predict(rf\_model, test\_data)

rf\_conf\_matrix <- confusionMatrix(rf\_pred, test\_data$stroke)

print(rf\_conf\_matrix)

A screenshot of a computer

AI-generated content may be incorrect.