UI.R

dashboardPage(

  skin = "black",

  dashboardHeader(title = "Stroke Analysis"),

  dashboardSidebar(

    sidebarMenu(

      menuItem("About", tabName = "group", icon = icon("house")),

      menuItem("Dataset",

               tabName = "data",

               icon = icon("database")),

      menuItem(

        "Categorical Feature",

        tabName = "categorial",

        icon = icon("chart-simple")

      ),

      menuItem(

        "Exploratory Data Analysis",

        tabName = "eda",

        icon = icon("chart-pie")

      ),

      menuItem(

        "Logistic Regression",

        tabName = "lr",

        icon = icon("medal")

      )

    )

  ),

  dashboardBody(

    tags$head(

      tags$link(rel = "stylesheet", type = "text/css", href = "style.css")

    ),

    tags$script(HTML("$('body').addClass('fixed');")),

    tabItems(

      #Group tab content

      tabItem(

        tabName = "group",

        tags$div(

          class = "about",

          id = "about",

          tags$canvas(id = "canvas"),

          tags$div(class = "logo",

                   tags$img(src = "vku.png")),

          tags$div(

            class = "title",

            tags$h2("Data Analysis with R"),

            tags$h3("Topic: Stroke Analysis"),

            tags$div("Group 11")

          ),

          fluidRow(

            infoBox(

              "19IT1",

              "Nguyen Van An",

              "19IT001",

              icon = icon("pagelines"),

              color = "olive",

            ),

            infoBox("19IT1",

                    "Tran Quang Dat",

                    "19IT006",

                    icon = icon("snowflake")),

            infoBox(

              "19IT1",

              "Ngo Thi Huong Giang",

              "19IT008",

              icon = icon("wand-magic-sparkles"),

              color = "purple"

            ),

          )

        ),

      ),

      # Dataset tab content

      tabItem(

        tabName = "data",

        tabBox(

          id = "t1",

          width = 12,

          tabPanel("About", icon = icon("address-card"),

                   fluidRow(

                     column(

                       width = 8,

                       tags$img(

                         src = "stroke1.jpg",

                         width = '100%',

                         height = 'auto',

                         object = 'cover'

                       ),

                     ),

                     column(

                       width = 4,

                       tags$br() ,

                       tags$h4(

                         "According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths."

                       ),

                       tags$h4(

                         "This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relavant information about the patient."

                       )

                     )

                   )),

          tabPanel("Data", dataTableOutput("dataT"), icon = icon("table")),

          tabPanel(

            "Structure",

            fluidRow(

              column(

                width = 12,

                verbatimTextOutput("structure"),

                tags$p("Attribute Information"),

                tags$ul(

                  tags$li("gender: Male, Female"),

                  tags$li("age: age of the patient"),

                  tags$li("hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension"),

                  tags$li("heart\_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease"),

                  tags$li("ever\_married: No or Yes"),

                  tags$li("work\_type: children, Govt\_jov, Never\_worked, Private, Self-employed"),

                  tags$li("Residence\_type: Rural or Urban"),

                  tags$li("avg\_glucose\_level: average glucose level in blood"),

                  tags$li("bmi: body mass index"),

                  tags$li("smoking\_status: formerly smoked, never smoked, smokes or Unknown"),

                  tags$li("stroke: 1 if the patient had a stroke or 0 if not")

                )

              )

            ),

            icon = icon("uncharted")

          ),

          tabPanel("Summary",

                   verbatimTextOutput("summary"),

                   icon = icon("pen-to-square"))

        )

      ),

      # Second tab content

      tabItem(

        tabName = "categorial",

        fluidRow(

          box(

            width = 3,

            plotOutput("g1\_gender", height = 300),

            tags$p(

              "There are more Female patients than Male. The one entry that was stated as Other was added to the Female section since majority are female patients."

            )

          ),

          box(

            width = 3,

            plotOutput("g1\_married", height = 300),

            tags$p(

              "Roughly double the amount of patients have been married before than those who have not."

            )

          ),

          box(

            width = 3,

            plotOutput("g1\_residence", height = 300),

            tags$p(

              "The patients are nearly evenly distributed between rural and urban residences"

            )

          ),

          box(

            width = 3,

            plotOutput("g1\_heart", height = 300),

            tags$p(

              "The gap between patients with and without heart disease more closely resembles the gap between those with and without strokes."

            )

          ),

          box(

            width = 3,

            plotOutput("g1\_hyper", height = 300),

            tags$p(

              "The number of patients without hypertension is vastly greater than the number of patients with hypertension, but the gap is slightly less than the gap seen for stroke victims."

            )

          ),

          box(

            width = 3,

            plotOutput("g1\_stroke", height = 300),

            tags$p(

              "The number of patients who have not had strokes is vastly greater than the number of patients who have."

            ),

          ),

          box(width = 6,

              plotOutput("g1\_age"),

              tags$p("")),

          box(width = 6,

              plotOutput("g1\_bmi"),

              tags$p("")),

          box(width = 6,

              plotOutput("g1\_glu"),

              tags$p("")),

          box(

            width = 6,

            plotOutput("g1\_work"),

            tags$p(

              "There are approximately even amounts of patients that are working government jobs, are self-employed, and are children. The majority of patients work for private companies, and a small number have never worked."

            )

          ),

          box(

            width = 6,

            plotOutput("g1\_smoking"),

            tags$p(

              "The unknown data was randomly added to the three categories above based of the probability. Most patients have either never smoked. The data for formerly and currently smokers are similar."

            )

          ),

        )

      ),

      tabItem(tabName = "eda",

              fluidRow(

                box(

                  width = 6,

                  plotOutput("g2\_gender"),

                  tags$p(

                    "We can see from the plots that the gender is not a feature that descriminate a person having a stroke or not."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_married"),

                  tags$p(

                    "The graphs show that married people have a higher rate of stroke than those who are not..."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_residence"),

                  tags$p(

                    "There are not many differences in these two graphs for residence type. Perhaps, this variable can be insignificant."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_hyper"),

                  tags$p(

                    "Blood pressure that is higher than normal is called hypertension. The graph shows that people with high blood pressure have a higher rate of stroke than people without the disease."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_heart"),

                  tags$p(

                    "The graphs show that people with heart disease have a higher rate of stroke than people without the disease."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_smoking"),

                  tags$p(

                    "The 'formerly smoked' and 'smokes' percentages are slightly higher in the stroke = 1 data."

                  )

                ),

                box(

                  width = 6,

                  plotOutput("g2\_work"),

                  tags$p("")

                ),

                box(

                  width = 6,

                  plotOutput("g2\_age"),

                  tags$p("The older you get the higher the chance of getting stroke.")

                ),

                box(

                  width = 6,

                  plotOutput("g2\_glu"),

                  tags$p("Observations with stroke tend to have higher glucose level")

                ),

              )),

      tabItem(

        tabName = "lr",

        h2("Logistic Regression"),

        fluidRow(

          box(

            width = 12,

            title = "Data preprocessing",

            verbatimTextOutput("t3\_str")

          ),

          box(width = 12,

              dataTableOutput("t3\_table")),

          box(

            width = 12,

            title = "Model Summary",

            verbatimTextOutput("modal\_summary"),

            uiOutput("equationLogistic"),

          ),

          box(

            width = 12,

            title = "Model Evaluation",

            column(width = 6,

                   verbatimTextOutput("t3\_test")),

            column(width = 6,

                   plotOutput("t3\_rocr")),

          )

        )

      )

    ),

    tags$script(src = "app.js")

  )

)

Server.R

function(input, output) {

  # Rendering table dataset

  output$dataT <- DT::renderDataTable({

    datatable(df\_clean, options = list(scrollX = TRUE))

  })

  # For Structure output

  output$structure  <- renderPrint({

    str(df\_clean)

  })

  # For Summary Output

  output$summary <- renderPrint({

    summary(df\_clean)

  })

  # Menu Item 3 (Categorical Feature)

  ## Gender Distribution

  output$g1\_gender <- renderPlot({

    ggplot(data = df\_copy, aes(x = gender, fill = gender)) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Gender", title = "Gender Distribution")

  })

  ## Residence Type Distribution

  output$g1\_residence <- renderPlot({

    ggplot(df\_copy, aes(x = Residence\_type, fill = Residence\_type)) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Residence Type", title = "Residence Type Distribution")

  })

  ## Ever Married Distribution

  output$g1\_married <- renderPlot({

    ggplot(data = df\_copy, aes(x = ever\_married, fill = ever\_married)) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Ever Married", title = "Ever Married Distribution")

  })

  ## Hypertension Distribution

  output$g1\_hyper <- renderPlot({

    ggplot(df\_copy, aes(as.factor(hypertension), fill = as.factor(hypertension))) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Hypertension", title = "Hypertension Distribution")

  })

  ## Heart Disease Distribution

  output$g1\_heart <- renderPlot({

    ggplot(df\_copy, aes(as.factor(heart\_disease), fill = as.factor(heart\_disease))) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Heart Dsiease", title = "Heart Disease Distribution")

  })

  ## Smoking Status Distribution

  output$g1\_smoking <- renderPlot({

    ggplot(df\_copy, aes(x = smoking\_status, fill = smoking\_status)) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[1],

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3],

        wes\_palette("GrandBudapest2")[4]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Smoking Status", title = "Smoking Status Distribution")

  })

  ## Work Type Distribution

  output$g1\_work <- renderPlot({

    ggplot(data = df\_copy, aes(x = work\_type, fill = work\_type)) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(

        values = c(

          wes\_palette("GrandBudapest2")[1],

          wes\_palette("GrandBudapest2")[2],

          wes\_palette("GrandBudapest2")[3],

          wes\_palette("GrandBudapest2")[4],

          wes\_palette("GrandBudapest2")[5]

        )

      ) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Work Type", title = "Work Type Distribution")

  })

  ## Stroke Distribution

  output$g1\_stroke <- renderPlot({

    ggplot(df\_copy, aes(as.factor(stroke), fill = as.factor(stroke))) +

      geom\_bar() +

      geom\_text(stat = 'count', aes(label = ..count..), vjust = 1) +

      scale\_fill\_manual(values = c(

        wes\_palette("GrandBudapest2")[2],

        wes\_palette("GrandBudapest2")[3]

      )) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold"),

            legend.position = "None") +

      labs(y = "Count", x = "Stroke", title = "Stroke Distribution")

  })

  ## Age Distribution

  output$g1\_age <- renderPlot({

    ggplot(df\_copy) +

      geom\_histogram(

        data = df\_copy,

        aes(x = age),

        fill = wes\_palette("IsleofDogs1")[1],

        color = "gray"

      ) +

      theme\_minimal() +

      theme(plot.title = element\_text(size = 15, face = "bold")) +

      labs(y = "Count", x = "Age (years)", title = "Age Distribution")

  })

  ## BMI Distribution

  output$g1\_bmi <- renderPlot({

    ggplot(df\_copy) +

      geom\_histogram(

        data = df\_copy,

        aes(x = bmi),

        fill = wes\_palette("IsleofDogs1")[1],

        color = "gray"

      ) +

      theme\_minimal() +

      theme(plot.title = element\_text(size = 15, face = "bold")) +

      labs(y = "Count", x = "BMI", title = "BMI Distribution")

  })

  ## Average Glucose Level Distribution

  output$g1\_glu <- renderPlot({

    ggplot(df\_copy) +

      geom\_histogram(

        data = df\_copy,

        aes(x = avg\_glucose\_level),

        fill = wes\_palette("IsleofDogs1")[1],

        color = "gray"

      ) +

      theme\_minimal() +

      theme(plot.title = element\_text(size = 15, face = "bold")) +

      labs(y = "Count", x = "Average Glucose Level", title = "Average Glucose Level Distribution")

  })

  # Menu Item 4(Exploratory Data Analysis)

  ##  Gender Distribution by Stroke Status

  output$g2\_gender <- renderPlot({

    tbg2\_gender <-

      df\_copy %>% group\_by(gender) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(tbg2\_gender,

           aes(

             x = gender,

             y = pct,

             fill = as.factor(stroke),

             label = scales::percent(pct)

           )) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Gender",

        title = "Gender Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Hypertension Distribution by Stroke Status

  output$g2\_hyper <- renderPlot({

    tbg2\_hyper <-

      df\_copy %>% group\_by(hypertension) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(tbg2\_hyper,

           aes(

             x = as.factor(hypertension),

             y = pct,

             fill = as.factor(stroke),

             label = scales::percent(pct)

           )) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Hypertension",

        title = "Hypertension Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Heart Disease Distribution by Stroke Status

  output$g2\_heart <- renderPlot({

    tbg2\_heart <-

      df\_copy %>% group\_by(heart\_disease) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(tbg2\_heart,

           aes(

             x = as.factor(heart\_disease),

             y = pct,

             fill = as.factor(stroke),

             label = scales::percent(pct)

           )) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Heart Disease",

        title = "Heart Disease Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Ever Married Distribution by Stroke Status

  output$g2\_married <- renderPlot({

    tbg2\_married <-

      df\_copy %>% group\_by(ever\_married) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(

      tbg2\_married,

      aes(

        x = ever\_married,

        y = pct,

        fill = as.factor(stroke),

        label = scales::percent(pct)

      )

    ) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Ever Married",

        title = "Ever Married Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Smoking Status Distribution by Stroke Status

  output$g2\_smoking <- renderPlot({

    tbg2\_smoking <-

      df\_copy %>% group\_by(smoking\_status) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(

      tbg2\_smoking,

      aes(

        x = smoking\_status,

        y = pct,

        fill = as.factor(stroke),

        label = scales::percent(pct)

      )

    ) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Smoking Status",

        title = "Smoking Status Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Work Type Distribution by Stroke Status

  output$g2\_work <- renderPlot({

    tbg2\_work <-

      df\_copy %>% group\_by(work\_type) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(tbg2\_work,

           aes(

             x = work\_type,

             y = pct,

             fill = as.factor(stroke),

             label = scales::percent(pct)

           )) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Work Type",

        title = "Work Type Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Residence Type Distribution by Stroke Status

  output$g2\_residence <- renderPlot({

    tbg2\_residence <-

      df\_copy %>% group\_by(Residence\_type) %>% count(stroke) %>% mutate(pct = prop.table(n))

    ggplot(

      tbg2\_residence,

      aes(

        x = Residence\_type,

        y = pct,

        fill = as.factor(stroke),

        label = scales::percent(pct)

      )

    ) +

      geom\_col(position = 'dodge') +

      geom\_text(

        position = position\_dodge(width = .9),

        vjust = -0.5,

        size = 3

      ) +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      theme\_minimal() +

      theme(plot.title = element\_text(face = "bold")) +

      labs(

        y = "Proportion",

        x = "Residence Type",

        title = "Residence Type Distribution by Stroke Status",

        fill = "Stroke"

      ) +

      scale\_y\_continuous(labels = scales::percent)

  })

  ## Age Distribution by Stroke Status

  output$g2\_age <- renderPlot({

    ggplot(df\_copy) +

      geom\_boxplot(aes(

        x = as.factor(stroke),

        y = age,

        fill = as.factor(stroke)

      )) +

      coord\_flip() +

      theme\_minimal() +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      labs(x = "Stroke", y = "Age (years)", title = "")

  })

  ## BMI Distribution by Stroke Status

  output$g2\_bmi <- renderPlot({

    ggplot(df\_copy) +

      geom\_boxplot(aes(

        x = as.factor(stroke),

        y = bmi,

        fill = as.factor(stroke)

      )) +

      coord\_flip() +

      theme\_minimal() +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      labs(x = "Stroke", y = "BMI", title = "")

  })

  ## Glu Distribution by Stroke Status

  output$g2\_glu <- renderPlot({

    ggplot(df\_copy) +

      geom\_boxplot(aes(

        x = as.factor(stroke),

        y = avg\_glucose\_level,

        fill = as.factor(stroke)

      )) +

      coord\_flip() +

      theme\_minimal() +

      scale\_fill\_manual(values = c(wes\_palette("Royal2")[2], wes\_palette("Royal2")[5])) +

      labs(x = "Stroke", y = "Average Glucose Level", title = "")

  })

  output$g\_corrplot <- renderPlot({

    res <- cor(df\_num)

    corrplot(

      res,

      type = "upper",

      order = "hclust",

      tl.col = "black",

      tl.srt = 45

    )

  })

  output$t3\_table <- DT::renderDataTable({

    datatable(df\_num, options = list(scrollX = TRUE))

  })

  output$t3\_str <- renderPrint({

    str(df\_num)

  })

  output$modal\_summary <- renderPrint({

    summary(model)

  })

  output$equationLogistic <- renderUI({

    text = extract\_eq(

      model,

      use\_coefs = TRUE,

      # display coefficients

      wrap = TRUE,

      # multiple lines

      terms\_per\_line = 5

    )

    withMathJax(tags$p(text))

  })

  output$t3\_test <- renderPrint({

    confusionMatrix(pred\_glm, as.factor(test$stroke))

  })

  output$t3\_rocr <- renderPlot({

    pred <- prediction(pred\_test, test$stroke)

    perf <- performance(pred, "tpr", "fpr")

    plot(perf, colorize = TRUE)

  })

}