



Formato de Informe

1. Portada

- **Título del informe:** Aplicación de herramientas de aplicaciones de minería de datos
- **Asignatura:** Desarrollo web Avanzado
- **Nombre del autor o autores:** Ariel Reyes, Anthony Villareal
- **Departamento:** Departamento de Ciencias de la Computación
- **Universidad:** Universidad de las Fuerzas Armadas ESPE
- **Fecha de entrega:** 03/03/2025

2. Índice

Formato de Informe	1
1. Portada	1
2. Índice	1
3. Resumen	1
4. Introducción.....	1
5. Objetivos	2
6. Desarrollo o Cuerpo del Informe	2
7. Resultados	6
8. Conclusiones.....	8
9. Recomendaciones.....	8
10. Bibliografía o Referencias.....	8

3. Resumen

Este informe presenta un análisis detallado de minería de datos aplicado al dataset mtcars utilizando R y la librería Shiny para crear una interfaz interactiva. Se implementaron diversas técnicas de minería de datos, como clustering (K-Means y Jerárquico), Análisis de Componentes Principales (PCA) y Regresión Lineal. Los resultados obtenidos proporcionan insights clave sobre la estructura de los datos y permiten realizar predicciones para mejorar la toma de decisiones

4. Introducción



El objetivo de este informe es desarrollar y analizar una aplicación de minería de datos en R utilizando Shiny, aplicando técnicas de clustering, reducción de dimensionalidad y modelos predictivos sobre el dataset mtcars.

La minería de datos es una disciplina clave en el análisis de grandes volúmenes de información. En este informe se emplea el lenguaje de programación R y la plataforma Shiny para diseñar una herramienta interactiva que facilite el análisis exploratorio y la modelización de datos.

El dataset mtcars contiene información sobre el rendimiento y características de diferentes modelos de automóviles, lo que lo convierte en un excelente recurso para aplicar técnicas de minería de datos. A través del análisis, se pretende extraer patrones significativos que permitan comprender mejor la relación entre las variables del conjunto de datos.

El alcance del informe incluye el desarrollo de un dashboard en Shiny para realizar análisis visuales y predictivos sobre mtcars. Se explorarán diversas técnicas de minería de datos y se evaluará su efectividad en la segmentación y predicción de datos automotrices.

Sin embargo, este informe no abarca la optimización de hiperparámetros en modelos de machine learning ni la integración con bases de datos externas. Además, los resultados obtenidos estarán limitados a las características del dataset utilizado y no se realizarán pruebas en conjuntos de datos adicionales.

5. Objetivos

- General: Desarrollar una aplicación en R con Shiny que implemente técnicas de minería de datos sobre el dataset mtcars.
- Específicos:
 - Implementar análisis exploratorio de datos mediante visualizaciones interactivas.
 - Aplicar técnicas de clustering K-Means y Jerárquico para segmentación de datos.
 - Realizar Análisis de Componentes Principales (PCA) para reducción de dimensionalidad.
 - Construir un modelo de Regresión Lineal para predicción de variables.

6. Desarrollo o Cuerpo del Informe

- **Conceptos base:**

La minería de datos implica la extracción de información relevante de grandes volúmenes de datos mediante técnicas estadísticas y de machine learning. En este informe se aplican técnicas de clustering, PCA y regresión lineal para explorar patrones en los datos y mejorar la toma de decisiones basada en evidencia.

El clustering es una técnica utilizada para agrupar objetos similares dentro de un conjunto de datos. Por otro lado, el PCA es una herramienta útil para reducir la dimensionalidad de los datos, permitiendo representar la información en menos variables sin perder

características esenciales. Finalmente, la regresión lineal se emplea para modelar relaciones entre variables y realizar predicciones basadas en los datos existentes.

- **Herramientas de desarrollo:**

Para la implementación de este análisis, se utilizaron las siguientes herramientas:

- R y RStudio: Lenguaje de programación estadístico y entorno de desarrollo para análisis de datos.
- Shiny: Framework para crear aplicaciones web interactivas en R, facilitando la visualización y manipulación de datos.
- ggplot2: Librería para visualización de datos que permite generar gráficos de alta calidad.
- factoextra y cluster: Paquetes especializados en clustering y visualización de resultados de segmentación.

- **Exploración de Datos:**

El primer paso en el análisis consistió en la exploración del dataset mtcars. Se realizaron gráficos de dispersión y matrices de correlación para identificar patrones y relaciones entre variables clave, como el consumo de combustible (mpg), la potencia del motor (hp) y el peso del vehículo (wt).

```
# Cargar el dataset mtcars
dataset <- mtcars
dataset_numeric <- dataset[, c("mpg", "hp", "wt", "disp", "drat")]

# Normalizar los datos
normalize <- function(x) {
  return((x - min(x)) / (max(x) - min(x)))
}
dataset_scaled <- as.data.frame(lapply(dataset_numeric, normalize))
```

- **Clustering K-Means y Jerárquico:**

Para segmentar los datos, se aplicaron los métodos de clustering K-Means y Jerárquico. El clustering permitió agrupar los automóviles en categorías basadas en similitudes de



características, facilitando la identificación de grupos con comportamientos similares.

```
# Definir la lógica del servidor
server <- function(input, output) {

  # 🚩 Gráfico de dispersión para explorar relaciones
  output$scatterPlot <- renderPlot({
    ggplot(dataset, aes_string(x = input$var_x, y = input$var_y)) +
      geom_point(color = "blue", size = 3) +
      theme_minimal() +
      labs(title = "Gráfico de Dispersión", x = input$var_x, y = input$var_y)
  })

  # 🚩 Matriz de correlaciones
  output$corPlot <- renderPlot({
    corr_matrix <- cor(dataset_numeric)
    corrpplot(corr_matrix, method = "color", type = "upper", tl.cex = 0.8)
  })

  # 🚩 Clustering K-Means
  kmeans_result <- eventReactive(input$run_kmeans, {
    set.seed(123)
    kmeans(dataset_scaled, centers = input$num_clusters, nstart = 25)
  })

  output$kmeansPlot <- renderPlot({
    req(kmeans_result())
    fviz_cluster(kmeans_result(), data = dataset_scaled, geom = "point",
      ellipse.type = "convex", ggtheme = theme_minimal())
  })

  output$kmeansTable <- renderTable({
    req(kmeans_result())
    data.frame(Car_Model = rownames(dataset), Cluster = kmeans_result()$cluster)
  })
}
```



```
# 📌 Clustering Jerárquico
hclust_result <- eventReactive(input$run_hclust, {
  hclust(dist(dataset_scaled), method = "ward.D2")
})

output$hclustPlot <- renderPlot({
  req(hclust_result())
  dend <- as.dendrogram(hclust_result())
  plot(dend, main = "Dendrograma Jerárquico", cex = 0.8)
})

# 📌 Análisis de Componentes Principales (PCA)
pca_result <- eventReactive(input$run_pca, {
  prcomp(dataset_scaled, center = TRUE, scale. = TRUE)
})

output$pcaPlot <- renderPlot({
  req(pca_result())
  fviz_pca_biplot(pca_result(), repel = TRUE, col.var = "blue", col.ind = "red")
})

# 📌 Regresión Lineal
lm_result <- eventReactive(input$run_lm, {
  formula <- as.formula(paste(input$pred_var, "~ ."))
  train(formula, data = dataset_numeric, method = "lm")
})

output$lmPlot <- renderPlot({
  req(lm_result())
  plot(lm_result()$finalModel$fitted.values, dataset_numeric[[input$pred_var]],
    col = "blue", pch = 16, xlab = "valores Predichos", ylab = "valores Reales",
    main = "Regresión Lineal - valores Reales vs. Predichos")
  abline(0, 1, col = "red", lwd = 2)
})

output$lmsummary <- renderPrint({
  req(lm_result())
  summary(lm_result()$finalModel)
})
}
```

- **Definir Interfaz:**

Se define como se va a comportar y que elementos va a tener el interfaz.

```
# Definir la interfaz de usuario (UI)
ui <- fluidPage(
  titlePanel("Minería de Datos en mtcars"),

  sidebarLayout(
    sidebarPanel(
      h4("Exploración de Datos"),
      selectInput("var_x", "Variable X:", choices = colnames(dataset_numeric), selected = "mpg"),
      selectInput("var_y", "Variable Y:", choices = colnames(dataset_numeric), selected = "hp"),

      h4("Agrupamiento"),
      sliderInput("num_clusters", "Número de clusters (K):", min = 2, max = 6, value = 3),
      actionButton("run_kmeans", "Ejecutar K-Means"),
      actionButton("run_hclust", "Ejecutar Clustering Jerárquico"),

      h4("Análisis de Componentes Principales"),
      actionButton("run_pca", "Ejecutar PCA"),

      h4("Predicción con Regresión Lineal"),
      selectInput("pred_var", "Variable Objetivo:", choices = colnames(dataset_numeric), selected = "mpg"),
      actionButton("run_lm", "Ejecutar Regresión Lineal")
    ),

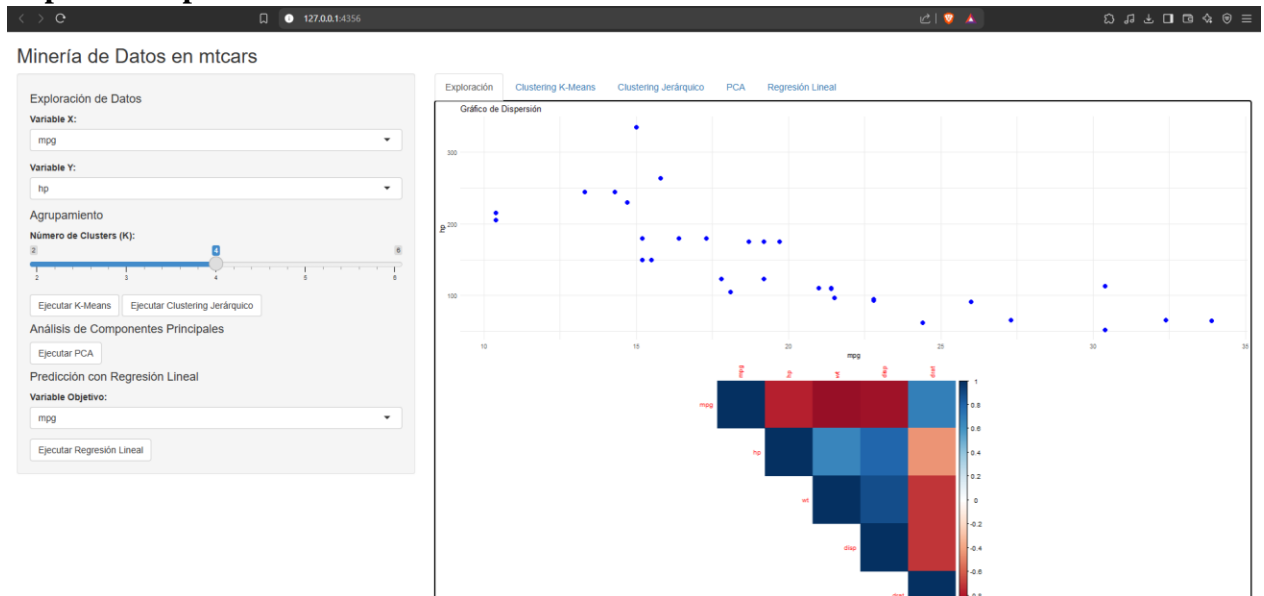
    mainPanel(
      tabsetPanel(
        tabPanel("Exploración", plotOutput("scatterPlot"), plotOutput("corPlot")),
        tabPanel("Clustering K-Means", plotOutput("kmeansPlot"), tableOutput("kmeansTable")),
        tabPanel("Clustering Jerárquico", plotOutput("hclustPlot")),
        tabPanel("PCA", plotOutput("pcaPlot")),
        tabPanel("Regresión Lineal", plotOutput("lmPlot"), verbatimTextOutput("lmSummary"))
      )
    )
  )
)
```

- Ejecutar la aplicación

```
# Ejecutar la aplicación shiny
shinyApp(ui = ui, server = server)
```

7. Resultados

- Exploración preliminar





- Clustering K-Means

Minería de Datos en mtcars

Exploración de Datos

Variable X: mpg

Variable Y: hp

Agrupamiento

Número de Clusters (K): 4

Ejecutar K-Means Ejecutar Clustering Jerárquico

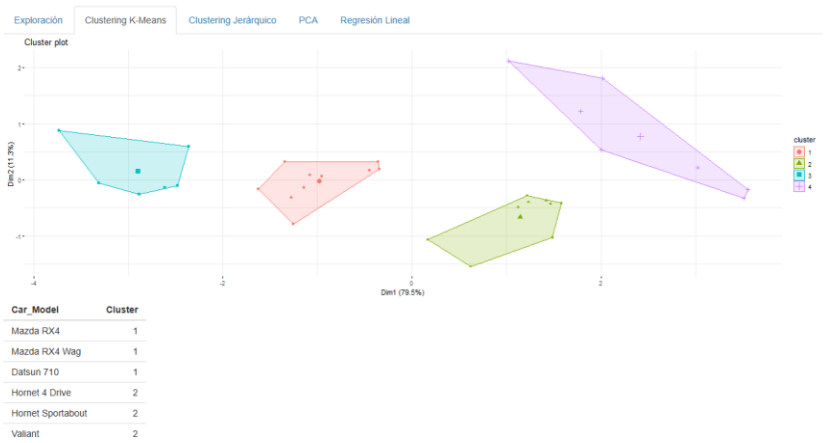
Análisis de Componentes Principales

Ejecutar PCA

Predicción con Regresión Lineal

Variable Objetivo: mpg

Ejecutar Regresión Lineal



- Clustering Jerárquico

Exploración de Datos

Variable X: mpg

Variable Y: hp

Agrupamiento

Número de Clusters (K): 4

Ejecutar K-Means Ejecutar Clustering Jerárquico

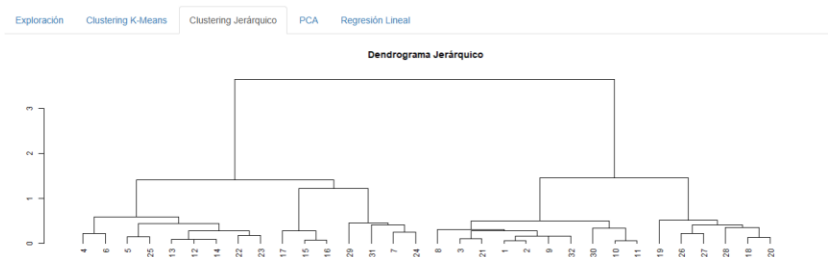
Análisis de Componentes Principales

Ejecutar PCA

Predicción con Regresión Lineal

Variable Objetivo: mpg

Ejecutar Regresión Lineal



- PCA

Minería de Datos en mtcars

Exploración de Datos

Variable X: mpg

Variable Y: hp

Agrupamiento

Número de Clusters (K): 4

Ejecutar K-Means Ejecutar Clustering Jerárquico

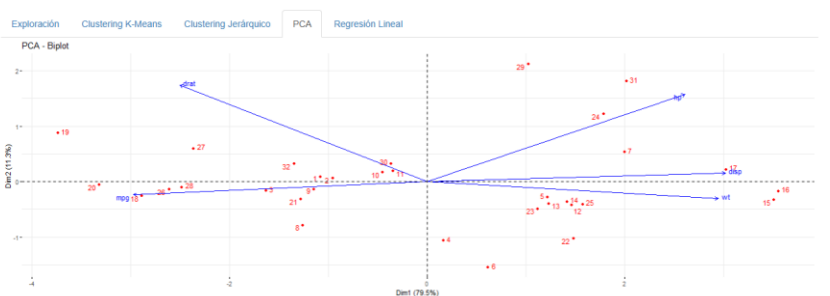
Análisis de Componentes Principales

Ejecutar PCA

Predicción con Regresión Lineal

Variable Objetivo: mpg

Ejecutar Regresión Lineal



- **Regresión lineal**

Exploración de Datos

Variable X: mpg

Variable Y: hp

Agrupamiento

Número de Clusters (K): 4

Ejecutar K-Means Ejecutar Clustering Jerárquico

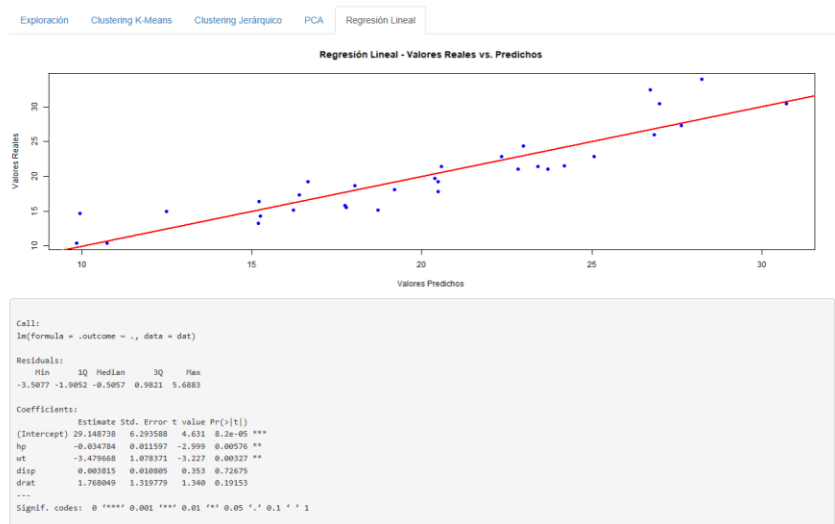
Análisis de Componentes Principales

Ejecutar PCA

Predicción con Regresión Lineal

Variable Objetivo: mpg

Ejecutar Regresión Lineal



GITHUB:

<https://github.com/ANTHONYNESTORVILLARREALMACIAS/Desarrollo-Web-Avanzado-AnthonyV-ArielR.git>

8. Conclusiones

Este análisis ha demostrado cómo las técnicas de minería de datos pueden proporcionar información valiosa sobre conjuntos de datos complejos. Se ha logrado segmentar, reducir dimensionalidad y predecir variables con éxito, validando la efectividad de estas metodologías. Además, los hallazgos obtenidos pueden servir como base para futuras investigaciones en el ámbito del análisis de datos y la optimización del rendimiento de vehículos.

9. Recomendaciones

Es recomendable incorporar técnicas avanzadas de aprendizaje automático para mejorar la precisión de los modelos empleados en este estudio. Además, explorar métodos de optimización de hiperparámetros permitirá obtener un mejor desempeño en los modelos de clustering y regresión, proporcionando segmentaciones y predicciones más precisas.

Por otro lado, ampliar el análisis incluyendo nuevos conjuntos de datos ayudaría a evaluar la generalización de los modelos, asegurando que sus resultados sean aplicables a otros escenarios. Finalmente, mejorar la interfaz de Shiny con visualizaciones más interactivas y personalizables aumentaría la usabilidad y accesibilidad de la aplicación para diversos usuarios.

10. Bibliografía o Referencias

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.



- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.