Laboratorio N° 01: Análisis Predictivo con R

Objetivos

Nombre: Jesus Antonio Huallpa Maron

- Aplicar análisis exploratorio y predictivo en un conjunto de datos utilizando R.
- Entrenar un modelo para realizar predicciones sobre los datos.

1. Carga del conjunto de datos.

```
datos <- read.csv("datos_combinados.csv", sep = ",", header = TRUE, fileEncoding = "UTF-8")

head(datos)
```

A data.frame: 6 × 18

	X.id.	X.fecha.	X.ip.	X.clase.	X.horario.	X.dia.	X.turno.	X.laboratorio.	X.total_enviado_mb.	X.total_recibido_mb.	X.tema.	X.navegador.	X.seccion.	X.docente.	X.total_mbps.	X.total_
	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<chr>	<chr>	<chr>	<chr>	<dbl>	<d
1	14	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	27.15	345.58	OSCURO	MSEdgeHTM	A	I.CHAPARRO	372.73	0
2	15	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	78.12	1463.15	OSCURO	MSEdgeHTM	A	I.CHAPARRO	1541.27	1
3	16	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	27.15	345.58	OSCURO	MSEdgeHTM	A	I.CHAPARRO	372.73	0
4	17	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	2.28	27.05	OSCURO	MSEdgeHTM	A	I.CHAPARRO	29.33	0
5	18	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	78.12	1463.15	OSCURO	MSEdgeHTM	A	I.CHAPARRO	1541.27	1
6	19	2024-09-23	172.30.107.16	PROGRAMACION I	9:40-11:20	Lunes	MAÑANA	LAB-A (P-306)	2.28	27.05	CLARO	ChromeHTML	A	I.CHAPARRO	29.33	0

2. Exploración de los datos.

```
str(datos)

summary(datos)
```

```
'data.frame':   1453 obs. of  18 variables:
 $ X.id.                 : int  14 15 16 17 18 19 21 22 25 26 ...
 $ X.fecha.              : chr  "2024-09-23" "2024-09-23" "2024-09-23" "2024-09-23" ...
 $ X.ip.                 : chr  "172.30.107.16" "172.30.107.16" "172.30.107.16" "172.30.107.16" ...
 $ X.clase.              : chr  "PROGRAMACION I" "PROGRAMACION I" "PROGRAMACION I" "PROGRAMACION I" ...
 $ X.horario.            : chr  "9:40-11:20" "9:40-11:20" "9:40-11:20" "9:40-11:20" ...
 $ X.dia.                : chr  "Lunes" "Lunes" "Lunes" "Lunes" ...
 $ X.turno.              : chr  "MAÑANA" "MAÑANA" "MAÑANA" "MAÑANA" ...
 $ X.laboratorio.        : chr  "LAB-A (P-306)" "LAB-A (P-306)" "LAB-A (P-306)" "LAB-A (P-306)" ...
 $ X.total_enviado_mb.   : num  27.15 78.12 27.15 2.28 78.12 ...
 $ X.total_recibido_mb.  : num  345.6 1463.2 345.6 27.1 1463.2 ...
 $ X.tema.               : chr  "OSCURO" "OSCURO" "OSCURO" "OSCURO" ...
 $ X.navegador.          : chr  "MSEdgeHTM" "MSEdgeHTM" "MSEdgeHTM" "MSEdgeHTM" ...
 $ X.seccion.            : chr  "A" "A" "A" "A" ...
 $ X.docente.            : chr  "I.CHAPARRO" "I.CHAPARRO" "I.CHAPARRO" "I.CHAPARRO" ...
 $ X.total_mbps.         : num  372.7 1541.3 372.7 29.3 1541.3 ...
 $ X.total_GB.           : num  0.36 1.51 0.36 0.03 1.51 0.03 0 0 0 0.23 ...
 $ X.tiempo_sesion.      : chr  "NULL" "NULL" "NULL" "NULL" ...
 $ X.consumo_energia_kwh.: chr  "NULL" "NULL" "NULL" "NULL" ...
    X.id.         X.fecha.           X.ip.             X.clase.
Min.   : 14.0   Length:1453        Length:1453        Length:1453
1st Qu.:207.0   Class :character   Class :character   Class :character
Median :373.0   Mode  :character   Mode  :character   Mode  :character
Mean   :369.2
3rd Qu.:529.0
Max.   :690.0
 X.horario.           X.dia.            X.turno.         X.laboratorio.
Length:1453        Length:1453        Length:1453        Length:1453
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character



X.total_enviado_mb. X.total_recibido_mb.   X.tema.          X.navegador.
Min.   :   0.00     Min.   :     0.00    Length:1453        Length:1453
1st Qu.:   0.60     1st Qu.:    27.04    Class :character   Class :character
Median :   6.83     Median :   202.33    Mode  :character   Mode  :character
Mean   :  72.29     Mean   :  2165.13
3rd Qu.:  23.72     3rd Qu.:   698.22
Max.   :9998.03     Max.   :233842.54
 X.seccion.         X.docente.        X.total_mbps.        X.total_GB.
Length:1453        Length:1453        Min.   :     0.00   Min.   :  0.000
Class :character   Class :character   1st Qu.:    27.51   1st Qu.:  0.030
Mode  :character   Mode  :character   Median :   209.86   Median :  0.200
                                      Mean   :  2211.30   Mean   :  2.159
                                      3rd Qu.:   818.45   3rd Qu.:  0.800
                                      Max.   :234434.87   Max.   :228.940
X.tiempo_sesion.   X.consumo_energia_kwh.
Length:1453        Length:1453
Class :character   Class :character
Mode  :character   Mode  :character
```

3. Limpieza y preparación de los datos.

```
# Verificar si la columna 'horario' tiene valores nulos
if (all(is.na(datos$horario))) {
    datos$horario <- "Sin horario"
} else {
    datos$horario[is.na(datos$horario)] <- "Sin horario"
}
```

4. Entrenamiento del modelo predictivo.

```
if (!require(caret)) install.packages("caret")
if (!require(ggplot2)) install.packages("ggplot2")
library(caret)
library(ggplot2)

datos$X.total_enviado_mb <- as.numeric(datos$X.total_enviado_mb)
datos$X.total_recibido_mb <- as.numeric(datos$X.total_recibido_mb)
datos$X.total_GB <- as.numeric(datos$X.total_GB)

set.seed(123)
modelo <- train(X.total_GB ~ X.total_enviado_mb + X.total_recibido_mb, data = datos, method = "lm")

summary(modelo)

predicciones <- predict(modelo, datos[, c("X.total_enviado_mb", "X.total_recibido_mb")])

datos$prediccion_GB <- predicciones

ggplot(datos, aes(x = X.total_GB, y = prediccion_GB)) +
  geom_point(color = "blue") +
  geom_abline(slope = 1, intercept = 0, color = "red") +
  labs(title = "Predicción vs Real", x = "Valor Real de X.total_GB", y = "Predicción de X.total_GB") +
  theme_minimal()

rmse <- sqrt(mean((predicciones - datos$X.total_GB)^2))
cat("RMSE:", rmse)
```

```
Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2407  0.0213  0.0230  0.0266  1.4131

Coefficients:
                      Estimate Std. Error  t value Pr(>|t|)
(Intercept)         -2.299e-02  8.597e-03   -2.674  0.00758 **
X.total_enviado_mb   9.741e-04  1.987e-05   49.016  < 2e-16 ***
```

5. Conclusión.

Se ha realizado un análisis exploratorio y se ha entrenado un modelo predictivo básico utilizando el conjunto de datos proporcionado.

```
F-statistic: 7.043e+05 on 2 and 1450 DF,  p-value: < 2.2e-16
RMSE: 0.3172242
```