Informe Proyecto 2 entrega 2

Juan Luis Solórzano (carnet: 201598) Micaela Yataz (carnet: 18960)

2025-01-20

git: https://github.com/JusSolo/Mineria_Proyecto2.git

1 Se usaran los mismos conjuntos de entrenamiento y prueba que usó para los modelos de regresión lineal en la entrega anterior.

Pero antes se agregará la varible nueva CategoriaPrecios, que agrupe los precios de las casas en 3 categorías: Económicas, Intermedias o Caras.

```
##
##  Económicas Intermedias       Caras
##         487         490         483
```

```
y<- datosC$SalePrice
set.seed(123)
trainI<- createDataPartition(y, p=0.8, list=FALSE)
train<-datosC[trainI, ]
test<-datosC[-trainI, ]
```

```
## Conjunto de entrenamiento (cantidad de muestras: 1169 )
##
##   SalePrice LotFrontage LotArea OverallQual OverallCond YearBuilt YearRemodAdd
## 1    208500          65    8450           7           5      2003         2003
## 2    181500          80    9600           6           8      1976         1976
## 3    223500          68   11250           7           5      2001         2002
## 4    140000          60    9550           7           5      1915         1970
## 5    250000          84   14260           8           5      2000         2000
## 6    143000          85   14115           5           5      1993         1995
##   MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF X1stFlrSF X2ndFlrSF
## 1        196        706          0       150         856       856       854
## 2          0        978          0       284        1262      1262         0
## 3        162        486          0       434         920       920       866
## 4          0        216          0       540         756       961       756
## 5        350        655          0       490        1145      1145      1053
## 6          0        732          0        64         796       796       566
##   LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
## 1            0      1710            1            0        2        1
## 2            0      1262            0            1        2        0
## 3            0      1786            1            0        2        1
## 4            0      1717            1            0        1        0
## 5            0      2198            1            0        2        1
## 6            0      1362            1            0        1        1
##   BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars
```

```
## 1            3            1            8          0        2003          2
## 2            3            1            6          1        1976          2
## 3            3            1            6          1        2001          2
## 4            3            1            7          1        1998          3
## 5            4            1            9          1        2000          3
## 6            1            1            5          0        1993          2
##   GarageArea WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch
## 1        548          0          61             0          0           0
## 2        460        298           0             0          0           0
## 3        608          0          42             0          0           0
## 4        642          0          35           272          0           0
## 5        836        192          84             0          0           0
## 6        480         40          30             0        320           0
##   PoolArea MiscVal MoSold YrSold CategoriaPrecio
## 1        0       0      2   2008           Caras
## 2        0       0      5   2007     Intermedias
## 3        0       0      9   2008           Caras
## 4        0       0      2   2006     Intermedias
## 5        0       0     12   2008           Caras
## 6        0     700     10   2009     Intermedias
```

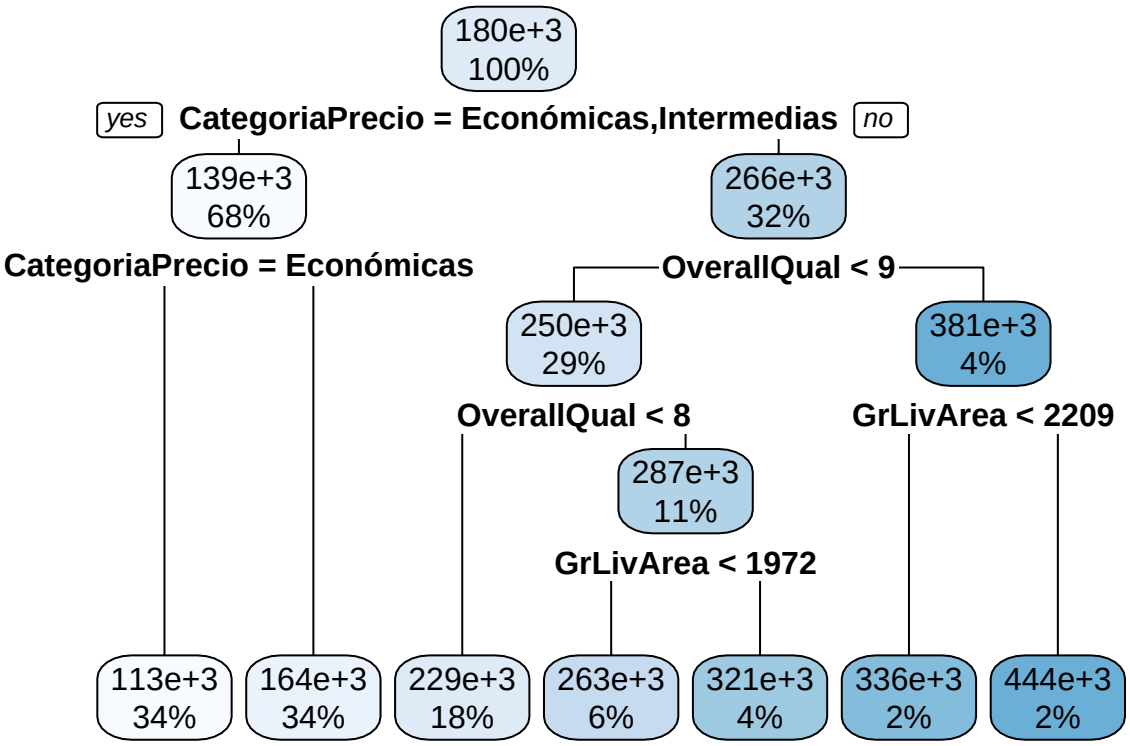
```
## Conjunto de prueba (cantidad de muestras: 291 )
##
```

```
##    SalePrice LotFrontage LotArea OverallQual OverallCond YearBuilt YearRemodAdd
## 8     200000           0   10382           7           6      1973         1973
## 11    129500          70   11200           5           5      1965         1965
## 13    144000           0   12968           5           6      1962         1962
## 30     68500          60    6324           4           6      1927         1950
## 37    145000         112   10859           5           5      1994         1995
## 39    109000          68    7922           5           7      1953         2007
##    MasVnrArea BsmtFinSF1 BsmtFinSF2 BsmtUnfSF TotalBsmtSF X1stFlrSF X2ndFlrSF
## 8         240        859         32       216        1107      1107       983
## 11          0        906          0       134        1040      1040         0
## 13          0        737          0       175         912       912         0
## 30          0          0          0       520         520       520         0
## 37          0          0          0      1097        1097      1097         0
## 39          0        731          0       326        1057      1057         0
##    LowQualFinSF GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath
## 8             0      2090            1            0        2        1
## 11            0      1040            1            0        1        0
## 13            0       912            1            0        1        0
## 30            0       520            0            0        1        0
## 37            0      1097            0            0        1        1
## 39            0      1057            1            0        1        0
##    BedroomAbvGr KitchenAbvGr TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars
## 8             3            1            7          2        1973          2
## 11            3            1            5          0        1965          1
## 13            2            1            4          0        1962          1
## 30            1            1            4          0        1920          1
## 37            3            1            6          0        1995          2
## 39            3            1            5          0        1953          1
##    GarageArea WoodDeckSF OpenPorchSF EnclosedPorch X3SsnPorch ScreenPorch
## 8         484        235         204           228          0           0
## 11        384          0           0             0          0           0
```

```
## 13        352        140           0             0          0         176
## 30        240         49           0            87          0           0
## 37        672        392          64             0          0           0
## 39        246          0          52             0          0           0
##    PoolArea MiscVal MoSold YrSold CategoriaPrecio
## 8         0     350     11   2009           Caras
## 11        0       0      2   2008      Económicas
## 13        0       0      9   2008     Intermedias
## 30        0       0      5   2008      Económicas
## 37        0       0      6   2009     Intermedias
## 39        0       0      1   2010      Económicas
```

2. Arbol de regresión para predecir el precio de las casas usando todas las variables.

```
arbol1 <- rpart(SalePrice~.,data = train)
rpart.plot(arbol1)
```



3. Úselo para predecir y analice el resultado. ¿Qué tal lo hizo?

```
# Calcular las predicciones
predicciones <- predict(arbol1, newdata = test)

# Calcular MSE (Error Cuadrático Medio)
mse <- mean((train$SalePrice - predicciones)^2)
```

```
## Warning in train$SalePrice - predicciones: longer object length is not a
## multiple of shorter object length
```

```r
# Calcular MAE (Error Absoluto Medio)
mae <- mean(abs(train$SalePrice - predicciones))
```

```
## Warning in train$SalePrice - predicciones: longer object length is not a
## multiple of shorter object length
```

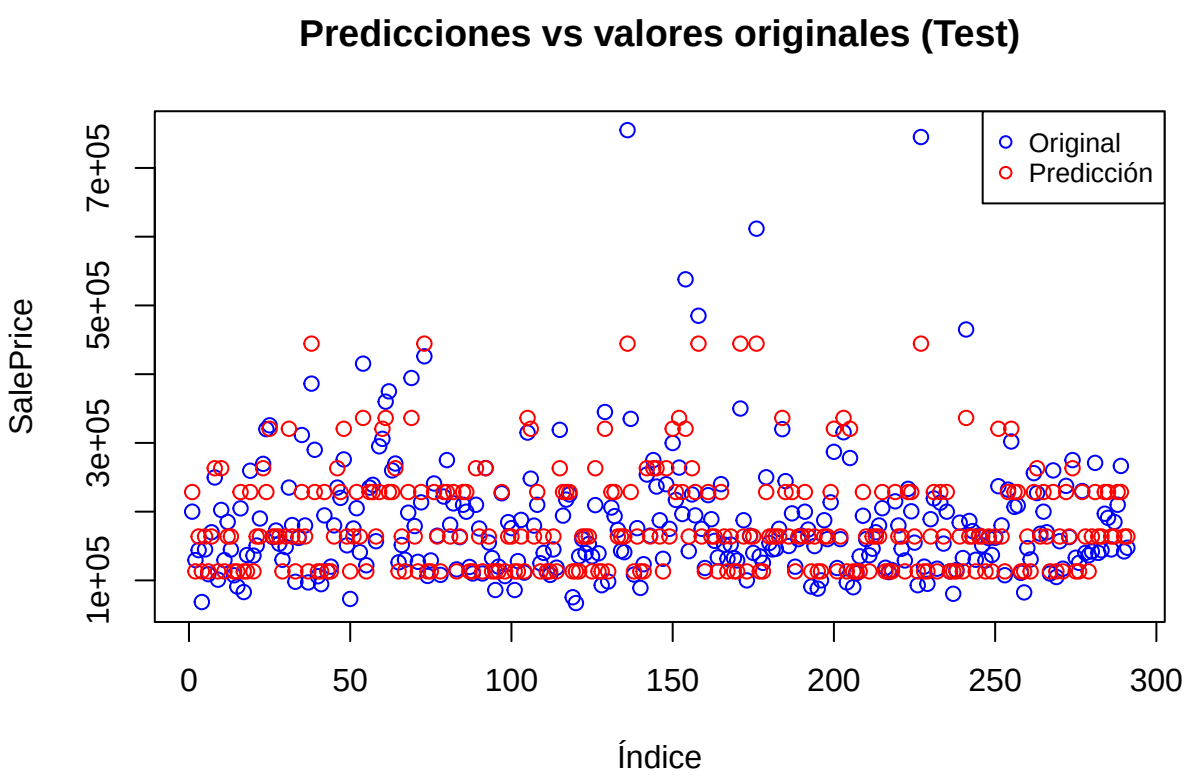
```r
# Mostrar los resultados
cat("MSE:", mse, "\nMAE:", mae)
```

```
## MSE: 11607730572
## MAE: 80238.62
```

```r
# Graficar los valores originales del conjunto de prueba
plot(test$SalePrice, col = "blue", main = "Predicciones vs valores originales (Test)",
     xlab = "Índice", ylab = "SalePrice")

# Agregar las predicciones al gráfico
points(predicciones, col = "red")

# Agregar la leyenda
legend("topright", legend = c("Original", "Predicción"),
       col = c("blue", "red"), pch = 1, cex = 0.8)
```



El model tiene un MAE y un MSE altos, la predicción es muy burda.

4. Haga, al menos, 3 modelos más, cambiando el parámetro de la profundidad del árbol. ¿Cuál es el mejor modelo para predecir el precio de las casas?

```
# Modelo original (sin especificar maxdepth, usa el máximo por defecto)
arbol1 <- rpart(SalePrice ~ ., data = train)

# Modelos con diferentes profundidades
arbol2 <- rpart(SalePrice ~ ., data = train, control = rpart.control(maxdepth = 4))
arbol3 <- rpart(SalePrice ~ ., data = train, control = rpart.control(maxdepth = 3))
arbol4 <- rpart(SalePrice ~ ., data = train, control = rpart.control(maxdepth = 2))

# Función para calcular MSE y MAE
calcular_errores <- function(modelo) {
  pred <- predict(modelo, newdata = train)
  mse <- mean((train$SalePrice - pred)^2)
  mae <- mean(abs(train$SalePrice - pred))
  return(c(MSE = mse, MAE = mae))
}

# Calcular errores para cada modelo
errores1 <- calcular_errores(arbol1)
errores2 <- calcular_errores(arbol2)
errores3 <- calcular_errores(arbol3)
errores4 <- calcular_errores(arbol4)

# Mostrar los resultados
resultados <- data.frame(
  Modelo = c("Original (sin maxdepth)", "maxdepth = 4", "maxdepth = 3", "maxdepth = 2"),
  MSE = c(errores1[1], errores2[1], errores3[1], errores4[1]),
  MAE = c(errores1[2], errores2[2], errores3[2], errores4[2])
)

print(resultados)
```

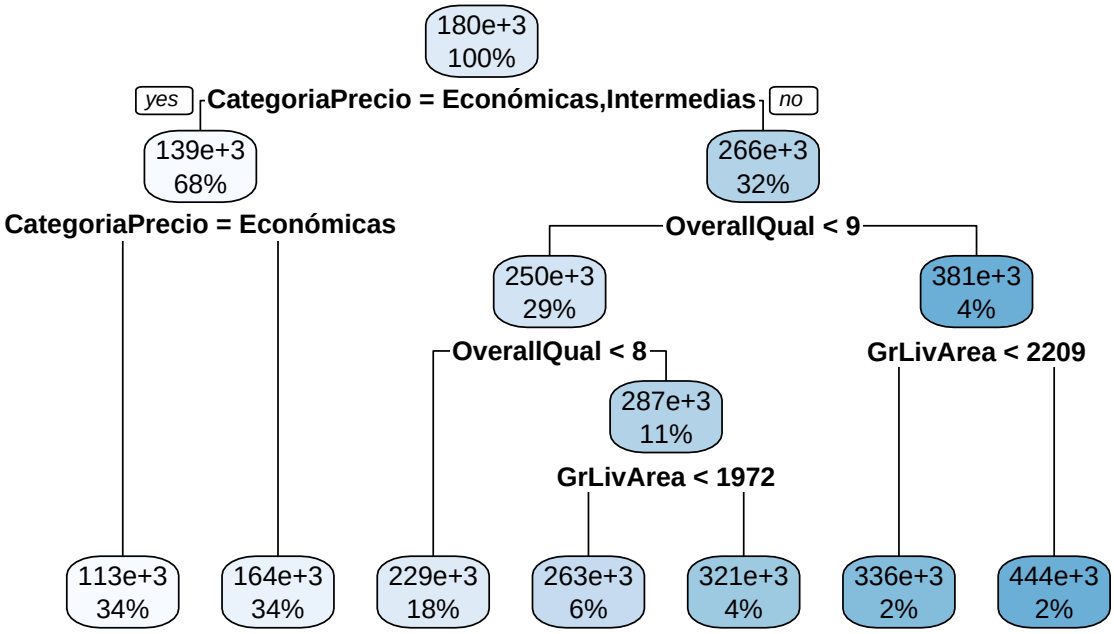
```
##                    Modelo        MSE      MAE
## 1 Original (sin maxdepth)  847777122 20836.61
## 2            maxdepth = 4  847777122 20836.61
## 3            maxdepth = 3  933933647 21532.44
## 4            maxdepth = 2 1268814335 24223.60
```

```
# Identificar el mejor modelo (menor MSE)
mejor_modelo <- resultados[which.min(resultados$MSE), "Modelo"]
cat("\nEl mejor modelo según el MSE es:", mejor_modelo, "\n")
```

```
##
## El mejor modelo según el MSE es: Original (sin maxdepth)
```

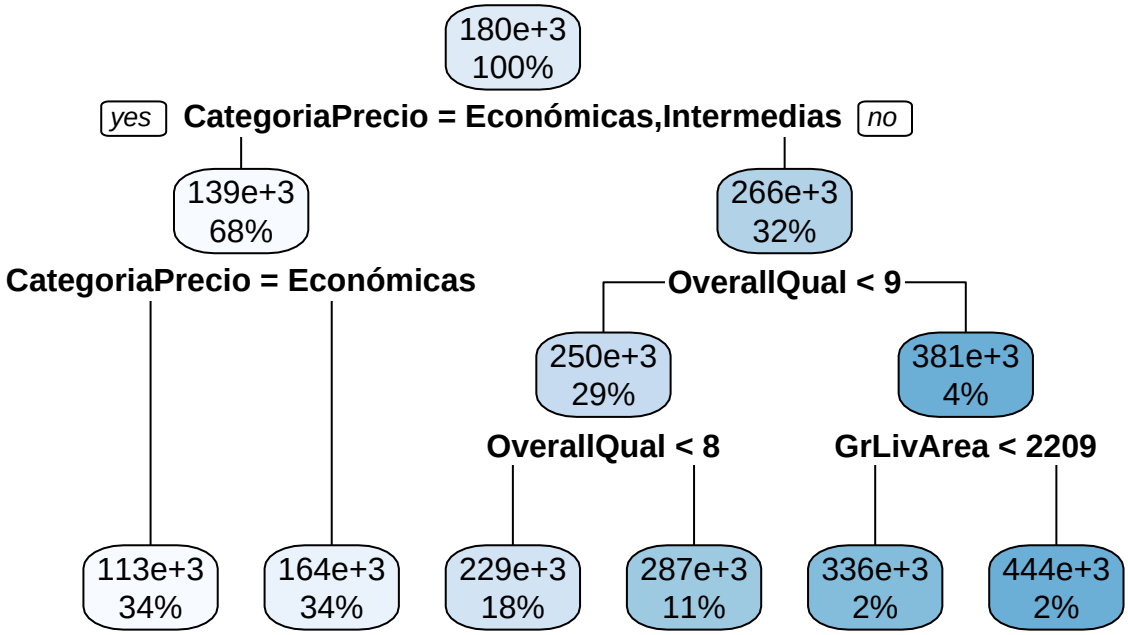
```
rpart.plot(arbol2, main = "Árbol con maxdepth = 4")
```

Árbol con maxdepth = 4

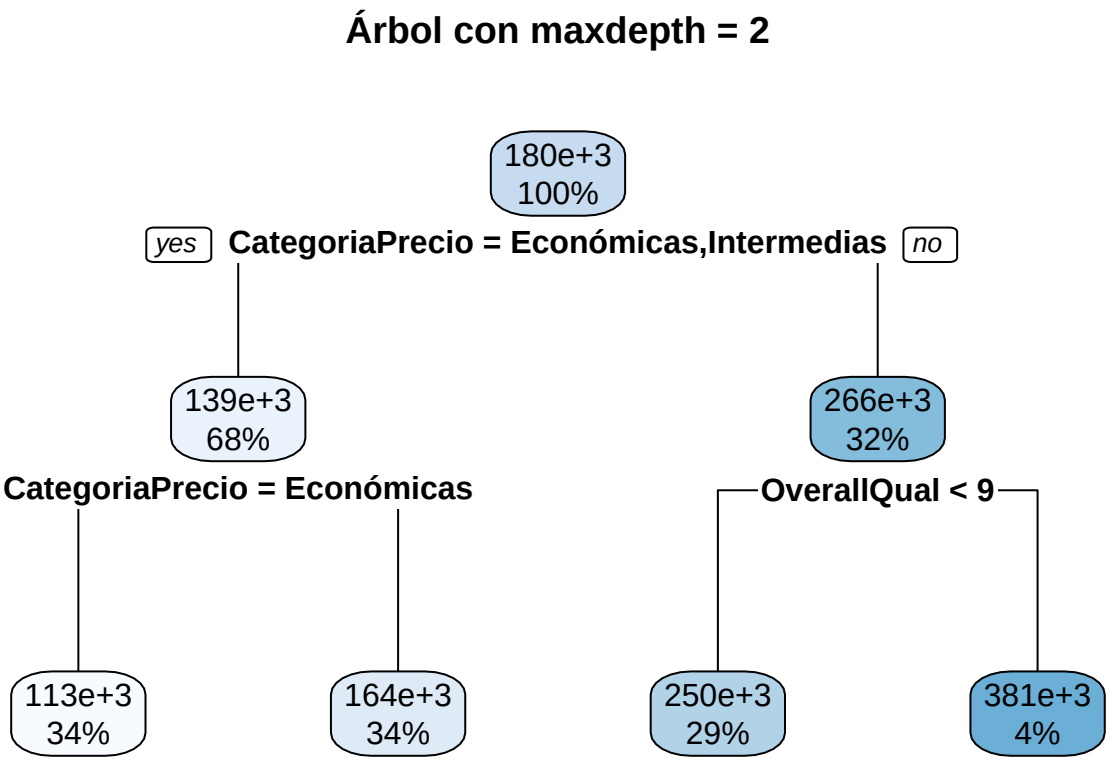



```
rpart.plot(arbol3, main = "Árbol con maxdepth = 3")
```

Árbol con maxdepth = 3




```
rpart.plot(arbol4, main = "Árbol con maxdepth = 2")
```




```
# Resetear la ventana gráfica
par(mfrow = c(1, 1))
```

Como es de esperar a mayor profundidad mayor error

5. Compare los resultados con el modelo de regresión lineal de la hoja anterior, ¿cuál lo hizo mejor?

6. Dependiendo del análisis exploratorio elaborado cree una variable respuesta que le permita clasificar las casas en Económicas, Intermedias o Caras. Los límites de estas clases deben tener un fundamento en la distribución de los datos de precios, y estar bien explicados

7. Elabore un árbol de clasificación utilizando la variable respuesta que creó en el punto anterior. Explique los resultados a los que llega. Muestre el modelo gráficamente. Recuerde que la nueva variable respuesta es categórica, pero se generó a partir de los precios de las casas, no incluya el precio de venta para entrenar el modelo.