Regresion - Ropa de mujeres

willandru

2024-10-02

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
data.train <-read.csv("train.csv")
str(data.train)
```

```
## 'data.frame':    1600 obs. of  12 variables:
##  $ idloc    : int  1648 1259 674 1558 1626 939 176 782 163 1703 ...
##  $ edadloc  : int  11 3 13 7 22 8 22 8 4 4 ...
##  $ correo   : int  10042 11909 10669 6981 10940 12689 9773 13018 9546 13050 ...
##  $ paginas  : int  68 107 75 65 74 110 73 106 84 112 ...
##  $ telefono : int  35 42 33 19 37 55 34 45 39 46 ...
##  $ impresa  : num  22609 24562 38914 39205 23947 ...
##  $ servicio : int  30 35 52 30 46 57 50 41 25 39 ...
##  $ nomina   : int  66307 29107 9519 28651 30654 30528 27890 11107 111305 34739 ...
##  $ idmercado: int  8 8 4 7 7 7 8 3 7 9 ...
##  $ tamamer  : chr  "Grande" "Grande" "Median" "Grande" ...
##  $ promo    : int  2 2 3 1 2 1 3 1 2 1 ...
##  $ ropamujer: num  43748 37484 57523 50338 45676 ...
```

```
dummies <-dummyVars(~ tamamer ,data=data.train)
dummies <- as.data.frame(predict(dummies,newdata=data.train))
dummies <- dummies[,-1]
```

```
data.train <- data.train[,-10]

data.train <- cbind(data.train, dummies)
str(data.train)
```

```
## 'data.frame':    1600 obs. of  13 variables:
##  $ idloc         : int  1648 1259 674 1558 1626 939 176 782 163 1703 ...
##  $ edadloc       : int  11 3 13 7 22 8 22 8 4 4 ...
##  $ correo        : int  10042 11909 10669 6981 10940 12689 9773 13018 9546 13050 ...
##  $ paginas       : int  68 107 75 65 74 110 73 106 84 112 ...
##  $ telefono      : int  35 42 33 19 37 55 34 45 39 46 ...
##  $ impresa       : num  22609 24562 38914 39205 23947 ...
```

```
##  $ servicio      : int  30 35 52 30 46 57 50 41 25 39 ...
##  $ nomina        : int  66307 29107 9519 28651 30654 30528 27890 11107 111305 34739 ...
##  $ idmercado     : int  8 8 4 7 7 7 8 3 7 9 ...
##  $ promo         : int  2 2 3 1 2 1 3 1 2 1 ...
##  $ ropamujer     : num  43748 37484 57523 50338 45676 ...
##  $ tamamerMedian : num  0 0 1 0 0 1 1 1 0 0 ...
##  $ tamamerPequeño: num  0 0 0 0 0 0 0 0 0 0 ...
```

MODELO SIMPLE

```
modelo1<-lm(ropamujer~.,data=data.train)
summary(modelo1)
```

```
##
## Call:
## lm(formula = ropamujer ~ ., data = data.train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31816  -7801   -630   7157  52982
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.156e+04  2.619e+03 -15.871  < 2e-16 ***
## idloc          -3.066e+00  4.947e-01  -6.198 7.25e-10 ***
## edadloc         8.887e+00  4.971e+01   0.179   0.8581
## correo          1.864e+00  1.888e-01   9.876  < 2e-16 ***
## paginas         1.822e+02  2.476e+01   7.361 2.91e-13 ***
## telefono        8.424e+01  4.656e+01   1.809   0.0706 .
## impresa         9.349e-01  4.529e-02  20.640  < 2e-16 ***
## servicio        6.820e+02  2.777e+01  24.555  < 2e-16 ***
## nomina         -8.425e-04  2.301e-02  -0.037   0.9708
## idmercado      -2.776e+00  1.091e+02  -0.025   0.9797
## promo           7.220e+01  4.387e+02   0.165   0.8693
## tamamerMedian   2.310e+02  8.573e+02   0.269   0.7877
## tamamerPequeño  5.108e+02  1.084e+03   0.471   0.6377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11290 on 1587 degrees of freedom
## Multiple R-squared:  0.5897, Adjusted R-squared:  0.5866
## F-statistic:   190 on 12 and 1587 DF,  p-value: < 2.2e-16
```

MODELO STEP

```
modeloStep<-step(modelo1,direction="both",trace=0)
summary(modeloStep)
```

```
##
```

```
## Call:
## lm(formula = ropamujer ~ idloc + correo + paginas + telefono +
##     impresa + servicio, data = data.train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31942  -7855   -668   7181  53312
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.124e+04  2.214e+03 -18.628  < 2e-16 ***
## idloc       -3.049e+00  4.891e-01  -6.234 5.83e-10 ***
## correo       1.863e+00  1.883e-01   9.893  < 2e-16 ***
## paginas      1.812e+02  2.466e+01   7.348 3.19e-13 ***
## telefono     8.512e+01  4.644e+01   1.833    0.067 .
## impresa      9.363e-01  4.519e-02  20.719  < 2e-16 ***
## servicio     6.848e+02  2.135e+01  32.074  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11270 on 1593 degrees of freedom
## Multiple R-squared:  0.5894, Adjusted R-squared:  0.5879
## F-statistic: 381.2 on 6 and 1593 DF,  p-value: < 2.2e-16
```