



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO®

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

NOMBRE DE LOS ALUMNOS:

RAYMUNDO HIRALES LAZARENO (N. CONTROL: 17212339)

GALAVIZ LONA OSCAR EDUARDO (N.CONTROL: 17212993)

Carrera: Ingeniería Informática

Semestre: 9no

MATERIA: Minería de datos

PROFESOR: JOSE CHRISTIAN ROMERO HERNANDEZ

TRABAJOS: Practica 1

FECHA: 1/11/21

## Desarrollo

Aqui comenzamos con la carga del archivo csv que se utilizara para el analisis de datos de esta practica, una vez cargados procedemos a que los estados sean convertidos a datos categoricos en este caso numeros, despues dividimos el dataframe en dos con un semilla de aleatoriedad para que los datos se repartan de manera aleatoria

```
getwd()
setwd("/home/chris/Documents/itt/Enero_Junio_2020/Mineria_de_datos/DataMining/MachineLearning/MultipleLinearRegression")
getwd()

# Importing the dataset
dataset <- read.csv('50_Startups.csv')

# Encoding categorical data
dataset$State = factor(dataset$State,
                        levels = c('New York', 'California', 'Florida'),
                        labels = c(1,2,3))

dataset
# Splitting the dataset into the Training set and Test set
# Install.packages('caTools')
library(caTools)
set.seed(123)
split <- sample.split(dataset$Profit, SplitRatio = 0.8)
training_set <- subset(dataset, split == TRUE)
test_set <- subset(dataset, split == FALSE)

# Fitting Multiple Linear Regression to the Training set
#regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend + State)
regressor = lm(formula = Profit ~ .,
               data = training_set )

summary(regressor)
```

Por ultimo los resultado del modelo que mediante la regresion

Evidence1

## predicciones

---

Aqui nos muestran las predicciones que tendria cada uno de los campos del dataframe

```
# Prediction the Test set results
y_pred = predict(regressor, newdata = test_set)
y_pred
```



## Optimizacion del modelo para usar backward elimination

---

En este apartado empezamos a optimizar el dataframe para la utilizacion del backward elimination reduciendo el dataframe a solo unos cuantos campos claves del dataframe que se utilizaran para este analisis

```
# Assignment: visualize the simple linear regression model with R.D.Spend

# Building the optimal model using Backward Elimination
regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend +
               State,
               data = dataset )
summary(regressor)
```



```
regressor = lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend,
               data = dataset )
summary(regressor)
```



```
regressor = lm(formula = Profit ~ R.D.Spend + Marketing.Spend,
               data = dataset )
summary(regressor)

regressor = lm(formula = Profit ~ R.D.Spend + Marketing.Spend,
               data = dataset )
summary(regressor)
```



```
y_pred = predict(regressor, newdata = test_set)
y_pred
```



## uso de backwardelimination

---

una vez reducido el dataframe procedemos a realizar el uso de la funcion de backwardelimination. creamos la funcion para que se realice en el dataframe que reducimos

```
# Homework analyse the follow atomation backwardElimination function
backwardElimination <- function(x, sl) {
  numVars = length(x)
  for (i in c(1:numVars)){
    regressor = lm(formula = Profit ~ ., data = x)
    maxVar = max(coef(summary(regressor))[c(2:numVars), "Pr(>|t|)"])
    if (maxVar > sl){
      j = which(coef(summary(regressor))[c(2:numVars), "Pr(>|t|)"] == maxVar)
      x = x[, -j]
    }
    numVars = numVars - 1
  }
  return(summary(regressor))
}

SL = 0.05
#dataset = dataset[, c(1,2,3,4,5)]
training_set
```

 Evidence1

## resultados del backwardelimination

Aqui se nos muestran los resultados como el analisis anterior pero con la funcion del analisis de backwardelimination y nos mostrara una serie de resultados estadisticos como la media, la mediana, el cuartil, el error estandar entre otros que me muestra

```
backwardElimination(training_set, SL)
```

 Evidence1