



TECNOLÓGICO NACIONAL DE MEXICO
INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA
DEPARTAMENTO DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES

SEMESTRE FEBRERO-JUNIO 2022

MATERIA:

Minería de datos.

UNIDAD 3

Practica 1

Regresión lineal

DOCENTE:

JOSE CHRISTIAN ROMERO HERNANDEZ

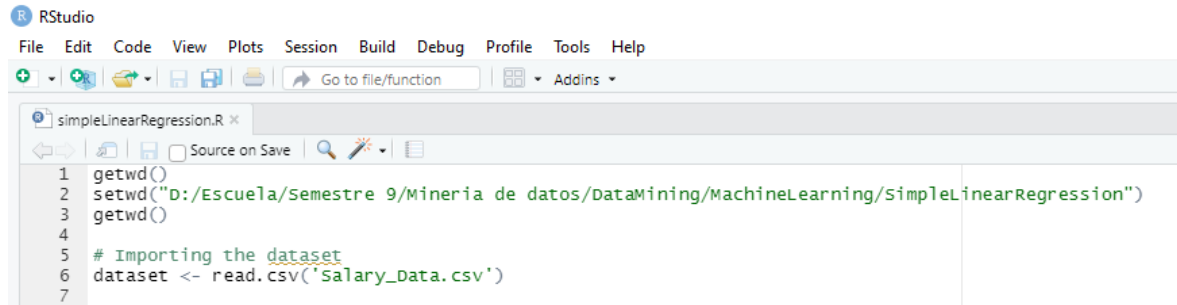
ALUMNO:

López Higuera Saúl Alfredo #18210493

Munguía silva Edgar Geovanny #17212344

Tijuana BC 17 de mayo del 2022

Importamos el archivo csv.



```
1 getwd()
2 setwd("D:/Escuela/Semestre 9/Mineria de datos/DataMining/MachineLearning/simpleLinearRegression")
3 getwd()
4
5 # Importing the dataset
6 dataset <- read.csv('Salary_Data.csv')
7
```

Instalamos paquetería caTools para tomar datos aleatorios no secuenciales.

Al igual debemos de poner la semilla para poder crear los números.

```
# Splitting the dataset into the Training set and Test set
# Install.packages('caTools')
library(caTools)
set.seed(123)
split <- sample.split(dataset$Salary, splitRatio = 2/3)
training_set <- subset(dataset, split == TRUE)
test_set <- subset(dataset, split == FALSE)
```

Realizaremos una regresión del modelo lm o lineal con la relación entre salario y años de experiencia.

Variable de predicción de datos:

```
# Fitting Simple Linear Regression to the Training set
regressor = lm(formula = Salary ~ YearsExperience,
               data = dataset)
summary(regressor)

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

Importación de nuestra librería de ggplot para nuestra graficación de los datos de

Salary vs Experience de Training Set y de Test Set:

Salary vs Experience de Training Set:

```
# Visualising the Training set results
library(ggplot2)
ggplot() +
  geom_point(aes(x=training_set$YearsExperience, y=training_set$Salary),
             color = 'red') +
  geom_line(aes(x = training_set$YearsExperience, y = predict(regressor, newdata = training_set)),
            color = 'blue') +
  ggtitle('Salary vs Experience (Training Set)') +
  xlab('Years of experience') +
  ylab('Salary')

# Visualising the Test set results
ggplot() +
  geom_point(aes(x=test_set$YearsExperience, y=test_set$Salary),
             color = 'red') +
  geom_line(aes(x = training_set$YearsExperience, y = predict(regressor, newdata = training_set)),
            color = 'blue') +
  ggtitle('Salary vs Experience (Test Set)') +
  xlab('Years of experience') +
  ylab('Salary')]
```



