

Instituto Tecnológico de Tijuana

Nombre de Facultad

Ingeniería Informática



Proyecto / Tarea / Practica:

Practica 1 Unidad 3

Materia:

Minería de datos

Facilitador:

JOSE CHRISTIAN ROMERO HERNANDEZ

Alumnos:

Erik Saul Rivera Reyes

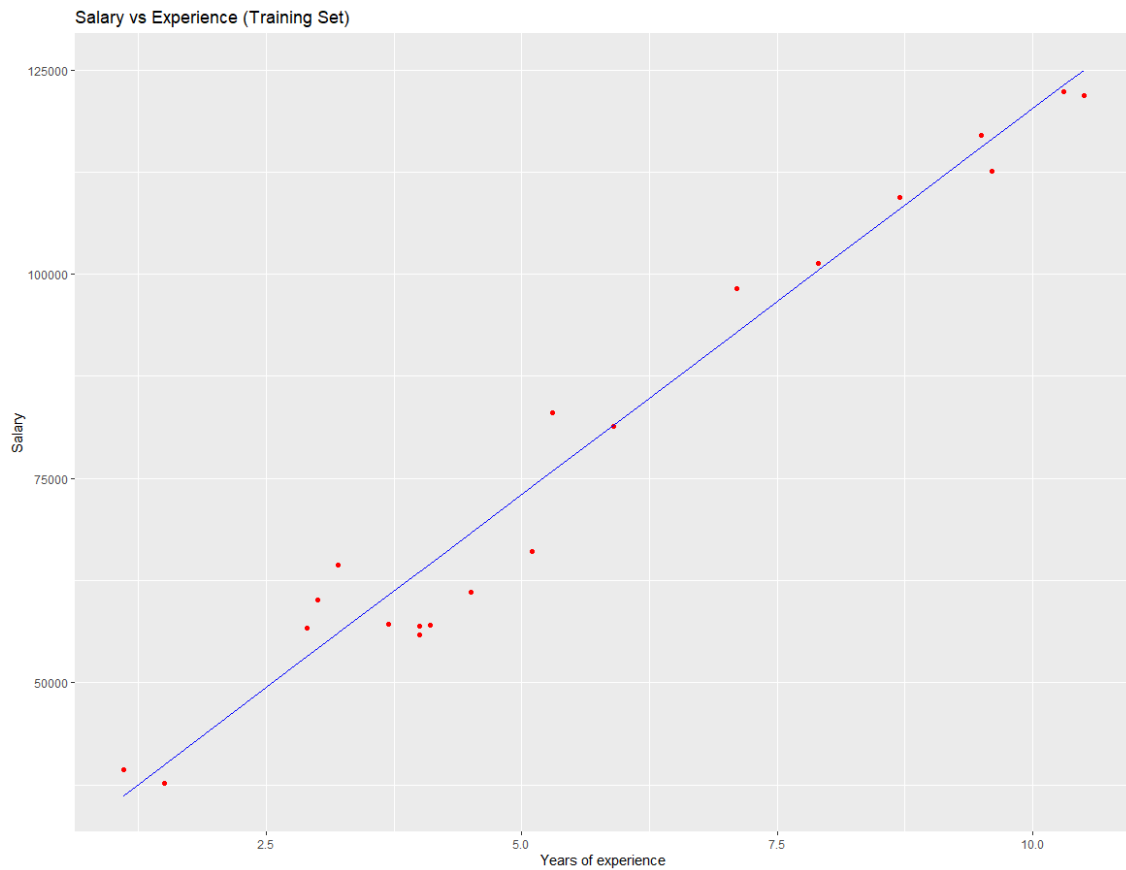
Fecha:

Tijuana Baja California a 26 de 11 2021

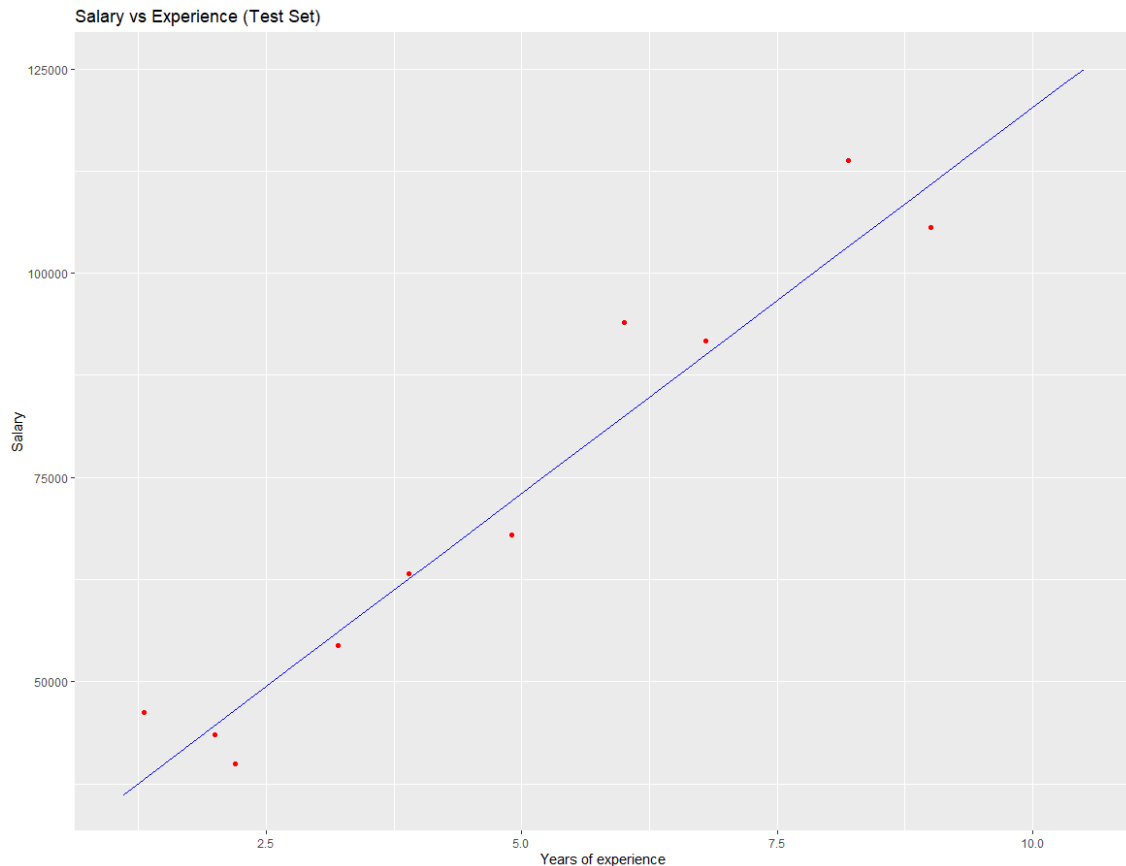
Código

```
1. # Importing the dataset
2. dataset <- read.csv('Salary_Data.csv')
3.
4. # Splitting the dataset into the Training set and Test set
5. # Install.packages('caTools')
6. library(caTools)
7. set.seed(123)
8. split <- sample.split(dataset$Salary, SplitRatio = 2/3)
9. training_set <- subset(dataset, split == TRUE)
10. test_set <- subset(dataset, split == FALSE)
11.
12. # Fitting Simple Linear Regression to the Training
    set
13. regressor = lm(formula = Salary ~ YearsExperience,
14.                 data = dataset)
15. summary(regressor)
16.
17. # Predicting the Test set results
18. y_pred = predict(regressor, newdata = test_set)
19.
20. # Visualising the Training set results
21. library(ggplot2)
22. ggplot() +
23.   geom_point(aes(x=training_set$YearsExperience,
24.                  y=training_set$Salary),
25.              color = 'red') +
26.   geom_line(aes(x = training_set$YearsExperience,
27.                 y = predict(regressor, newdata = training_set)),
28.             color = 'blue') +
29.   ggtitle('Salary vs Experience (Training Set)') +
30.   xlab('Years of experience') +
31.   ylab('Salary')
32.
33. # Visualising the Test set results
34. ggplot() +
35.   geom_point(aes(x=test_set$YearsExperience,
36.                  y=test_set$Salary),
37.              color = 'red') +
38.   geom_line(aes(x = training_set$YearsExperience,
39.                 y = predict(regressor, newdata = training_set)),
40.             color = 'blue') +
41.   ggtitle('Salary vs Experience (Test Set)') +
42.   xlab('Years of experience') +
43.   ylab('Salary')
```

En la primera grafica se nos muestran las diferencias de que hay entre el salario y experiencia, se puede observar que entre mas años de experiencia se gana mas salario



En la segunda grafica tenemos algo similar a la primera, pero en esta ocasión estamos usando datos diferentes, los datos de “Test Set” los cuales al observar nos damos cuenta que a pesar de que entre más años de experiencia más dinero se gana obtenemos unos resultados un poco más dispersos y variados lo cual nos da a entender que no en todos los casos será lo mismo, aunque claramente la tendencia es que entre más experimentado eres más salario te deberían de pagar.



Explicacion del código

```
40. library(caTools)
41. set.seed(123)
42. split <- sample.split(dataset$Salary,
  SplitRatio = 2/3)
43. training_set <- subset(dataset, split == TRUE)
44. test_set <- subset(dataset, split == FALSE)
```

En esta fracción del código utilizamos la librería “catTools” para poder dividir nuestros datos como lo queremos, en este caso en base al salario y en la proporción la cual es 2/3.

```
45. regressor = lm(formula = Salary ~ YearsExperience,
46. data = dataset)
47. summary(regressor)
48.
49. # Predicting the Test set results
50. y_pred = predict(regressor, newdata = test_set)
51.
52. # Visualising the Training set results
53. library(ggplot2)
54. ggplot() +
55.   geom_point(aes(x=training_set$YearsExperience,
  y=training_set$Salary),
56.   color = 'red') +
57.   geom_line(aes(x = training_set$YearsExperience,
  y = predict(regressor, newdata = training_set)),
58.   color = 'blue') +
59.   ggtitle('Salary vs Experience (Training Set)') +
60.   xlab('Years of experience') +
61.   ylab('Salary')
```

En este caso se utiliza la formula regresión lineal simple para poder predecir los datos de nuestro interés así como el visualizar lo que nos interesa lo cual fue mostrado en las dos graficas previamente mostradas y explicadas.