

Reporte: los salarios de los científicos de datos

Antonio David Gutiérrez Páez, Arnold Torres Maldonado

2023-05-03

Resumen

Resumen tipo TT

Introducción

La ciencia de datos es una disciplina que ha estado a la alza recientemente con los avances tecnológicos y con la cantidad de datos que se producen todos los días, siendo actualmente una disciplina indispensable para el análisis y tratamiento de datos en masa. Debido a esto, la relación entre el salario percibido para esta disciplina contra el año en el que una persona se dedicaba a esto va en subida, siendo el último año uno de los mejores pagados para esta profesión, esto también debido a la alta demanda de procesamiento de datos y la poca disponibilidad de personas profesionales que se dediquen a esta disciplina.

ai-jobs.net es un sitio web donde usuarios publican ofertas de trabajo relacionadas con inteligencia artificial, machine learning y ciencia de datos con

Materiales y métodos

Para el flujo de trabajo del proyecto se hizo uso del modelo CRISP-DM (Cross Industry Standard Process for Data Mining) el cual plantea seis fases que tiene que pasar un proyecto de minería de datos: comprensión de problema, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación del modelo generado. Para observar la comparación entre los años y sus salarios percibidos, Utilizamos el método de regresión lineal por mínimos cuadrados así como el modelo de predicción para conocer como es que se comportarán los salarios para esta profesión a lo largo de los años 2020 - 2023.

***explicar que es la reg lineal, describir el dataset, explicar el crispsdm**

Experimentación y resultados

***desglosar el modelo crisp, aplicar todas las secciones del crispsdm** ### Comprensión del problema Empezaremos a analizar los salarios que se perciben a través de los años y sus desviaciones estándar así como su promedio de los profesionistas de data mining, y observaremos si en un futuro dedicarse a esta área seguirá aumentando sus sueldos o en su defecto se mantendrán o incluso si se reducirán los salarios. Así también se realizará un análisis del nivel de experiencia que se necesita contra el salario percibido

Comprensión de los datos

Tenemos un dataset que contiene 11 variables:

- Año trabajado
- Nivel de experiencia
- Tipo de empleo

- Titutlo del trabajo
- Salario
- Divisa del salario
- Salario en dólares
- Residencia
- Trabajo remoto
- Ubicación de la compañía
- Tamaño de la compañía

Debido a las diferencias de divisas utilizaremos la ariable del salario en dólares para normalizar los datos recibidos, año trabajado y el nivel de experiencia, por lo que primeramente harémos dos datas frames con (salario en dólares;año trabajado) y (salario en dólares;nivel de experiencia)

Conclusiones

Referencias bibliográficas

Chaki, A. (2023). Data Science Salaries 2023. [Conjunto de datos]. Kaggle. <https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023>

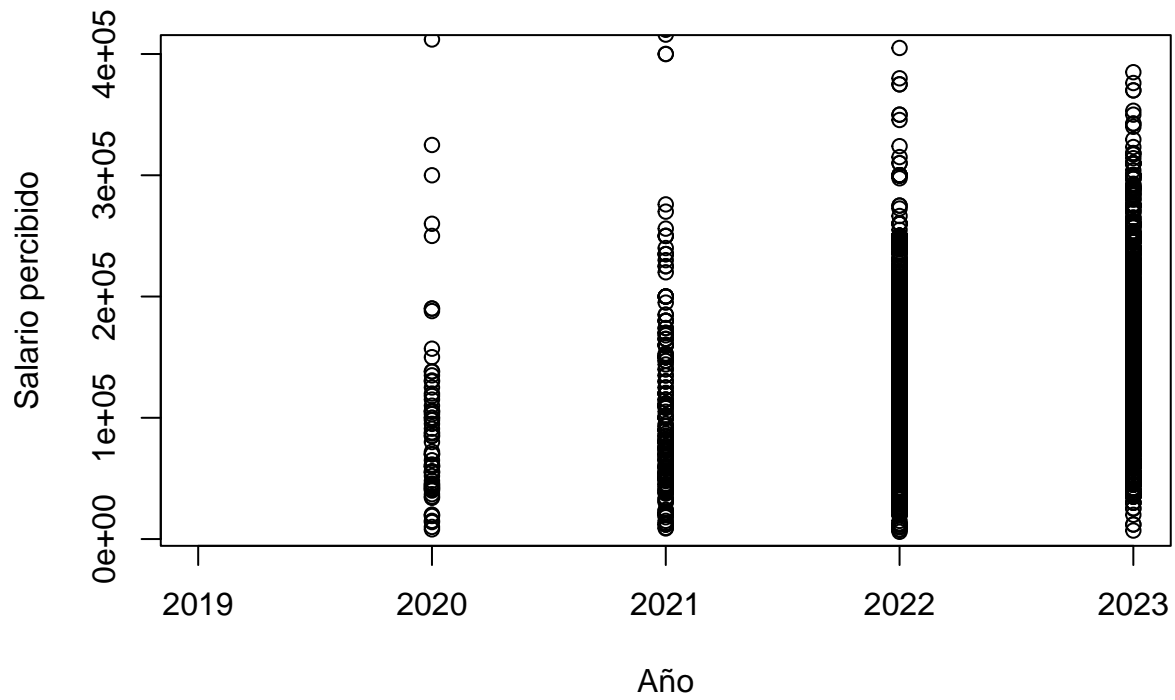
```
summary(cars)
```

```
##           speed           dist
##  Min.      : 4.0    Min.      :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean    : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.     :120.00
```

Including Plots

You can also embed plots, for example:

Relación entre el año en el que se trabaja y el salario recibido



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.