Librerías

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import numpy as np
```

Divisón del data set en entrenamiento y prueba

```
1 # Función que divide los datasets en entrenamiento y prueba
2 import random
4 def split_dataset(porcentaje, dataset):
    columna0 = dataset.columns[0]
    columna1 = dataset.columns[1]
    X = dataset[columna0].values
    Y = dataset[columna1].values
10
    cantidad = int((len(X) * porcentaje) / 100)
    lst = list(range(0,len(X)))
12
    arr train X = np.array([])
    arr train Y = np.array([])
14
15
    for i in range(0, cantidad):
16
      indice = random.choice(lst)
17
18
19
      # Creación de DataFrame de entrenamiento de X
      arr_train_X = np.append(arr_train_X, X[indice])
20
      train_X = pd.DataFrame(arr_train_X, columns = [columna0])
21
22
23
      # Creación de DataFrame de entrenamiento de Y
      arr_train_Y = np.append(arr_train_Y, Y[indice])
24
```

```
15/9/22, 19:23
```

```
train_Y = pd.DataFrame(arr_train_Y, columns = [columna1])

train_Y = pd.DataFrame(arr_train_Y, columns = [columna1])

train_Y = pd.DataFrame(x[lst], columns = [columna0])

test_Y = pd.DataFrame(Y[lst], columns = [columna1])

return train_X, train_Y, test_X, test_Y
```

Construcción de modelo de regresión

```
y = \beta 1 * x + \beta 0
```

→ Fit del modelo

```
1 def fit(dataset_X, dataset_Y):
 3
    columna0 = dataset X.columns[0]
    columna1 = dataset_Y.columns[0]
    X = dataset_X[columna0].values
    Y = dataset_Y[columna1].values
    # Cálculo de los coeficientes
    suma x = X.sum()
    suma y = Y.sum()
10
11
    suma_xy = (X*Y).sum()
12
13
    suma x2 = (X**2).sum()
14
    suma y2 = (Y^{**}2).sum()
15
16
    promedio_x = np.mean(X)
17
    promedio y = np.mean(Y)
18
```

```
19
20    n = len(X)
21
22    b0 = (suma_y*suma_x2 - suma_x*suma_xy) / (n*suma_x2 - (suma_x**2))
23    b1 = (n*suma_xy - suma_x*suma_y) / (n*suma_x2 - (suma_x)**2)
24    coeficientes = [b0, b1]
25    train_y = b1*X + b0
26
27    return X, Y, coeficientes, train_y
```

▼ Entrenamiento

```
1 def regresion(dataset_X, dataset_Y, fit):
2    X, Y, coeficientes, train_y = fit
3
4    X_test = dataset_X[dataset_X.columns[0]].values
5    Y_test = dataset_Y[dataset_Y.columns[0]].values
6
7    test_y = coeficientes[1]*X_test + coeficientes[0]
8    resultado = pd.DataFrame({test_X.columns[0] : test_X.iloc[:, 0], test_Y.columns[0]: test_Y.iloc[:, 0], 'Predicción '+ toggette...'
9
10    return test_y, resultado
```

→ Evaluación

```
SSE = resultado['Squared Error'].sum()
    MSE = SSE / n
10
    MAE = resultado['Error Absoluto'].sum() / n
11
    RMSE = MSE**(1/2)
12
    MAPE = (abs(resultado['Error']/resultado.iloc[:,1]).sum()) / n
    R2 = (((resultado.iloc[:,2] - media_y)**2).sum()) / (((resultado.iloc[:,1] - media_y)**2).sum())
14
15
16
    print('Mean Squared Error: ', MSE)
    print('Mean Absolute Error: ', MAE)
18
    print('Root Mean Squared Error', RMSE)
    print('Mean Percent Absolute Error', MAPE)
    print('Sum of Squared Errors', SSE)
    print('Coeficiente de determinación', R2)
21
22
23
    return resultado, MSE, MAE, RMSE, MAPE, SSE, R2
```

Cargar datos

```
1 data = pd.DataFrame()
 2 data['income'] = np.array([3.86264742, 4.97938138, 4.92395694, 3.21437244, 7.19640925,
3
         3.72964348, 4.67451739, 4.49810382, 3.12163053, 4.63991444,
         4.63283951, 2.7731789 , 7.11947859, 7.4666532 , 2.11774233,
4
5
         2.55916582, 2.35479322, 2.38815725, 4.75568027, 1.99427505,
6
         7.31091603, 3.52831896, 2.42875167, 3.54274787, 5.22720124,
7
         6.69199314, 3.90040994, 2.29105548, 2.38051271, 2.54960878,
8
         6.93329583, 1.85564517, 3.58902314, 6.82647791, 2.07060188,
9
         5.22420527, 2.24311363, 7.07616637, 4.1906725, 1.95648612,
10
         5.06175818, 3.98218993, 3.06505862, 3.68287749, 3.78942929,
         5.35871566, 5.19611978, 5.24118957, 7.10161959, 3.42402109,
11
12
         2.25339902, 5.3703369 , 6.225606 , 5.48286223, 4.03417211,
13
         6.51021872, 6.02921386, 6.94911289, 7.19503729, 2.7573385,
         6.95607948, 4.67019258, 6.36829267, 6.16668117, 6.07415829,
14
15
         5.48471897, 1.58957474, 1.68047393, 5.49994796, 4.0438909,
16
         5.00509324, 4.86358173, 1.50627504, 2.86466385, 5.87790629,
17
         6.48398379, 4.93803659, 5.62543442, 7.22826521, 5.33746029,
18
         2.82582659, 5.93136667, 3.5202552, 3.23994059, 3.49838637,
          7 10/1100/ / 7101//07 0 50/0010/ 0 0000/10 1 51/15000
```

1.29542729, 1.85334761, 5.3672361, 2.46629234, 1.72359356,

4.30728842, 3.15636744, 1.69045809, 5.08552301, 2.48162387,

173

174

```
15/9/22, 19:23
                            Momento de Retroalimentación: Módulo 2 Implementación de una técnica de aprendizaje máquina sin el uso de un framework - Colaboratory
  175
             1.35331734, 5.27529415, 3.67029153, 3.69639041, 2.67988732,
             2.36356467, 1.4385961, 5.36244826, 2.32530923, 1.52956094,
  176
  177
             4.03572702, 3.55033718, 3.55460032, 4.19972853, 3.13547515,
  178
             1.14529581, 3.53985413, 3.82220396, 3.83557765, 5.68375056,
  179
             2.51740088, 4.32466038, 2.9586968, 1.68099713, 4.2535013,
  180
             2.26039569, 3.61547105, 1.50221163, 3.77648463, 3.20104243,
  181
             4.05831162, 4.10522452, 3.72209764, 3.82227031, 3.70666098,
  182
             4.90968847, 2.15573363, 5.74373608, 4.33757145, 2.58241664,
  183
             4.28430556, 2.96216821, 3.99041589, 5.8901487, 3.68853952,
  184
             3.21784736, 2.22047033, 5.91028488, 4.85303252, 3.80297363,
  185
             3.09216465, 3.23331375, 5.26547896, 4.81153338, 5.36804117,
  186
             1.7987183 , 4.76001433, 3.41580648, 5.04017458, 1.96563593,
  187
             1.87195785, 5.32350739, 4.50115834, 2.43509486, 4.55321655,
             4.36170927, 1.4787872 , 2.76852313, 3.96063637, 5.9157814 ,
  188
             3.98019535, 4.70820299, 1.41666106, 4.3284171, 6.19629555,
  189
  190
             1.94596105, 0.69845996, 5.13545532, 1.56035538, 4.23499122,
  191
             0.68890923, 0.68584887, 1.79175357, 3.22747217, 5.15933124,
  192
             3.80297067, 4.21605759, 4.93240998, 2.23344058, 4.37825945,
  193
             3.48392568, 2.03029704, 3.59639832, 3.2611821, 3.48113777,
             3.39343915, 4.22966335, 4.3213643, 4.40045515, 3.35759738,
  194
  195
             5.335718 , 5.58838934, 2.99420428, 2.41338175, 4.98536659,
             4.98198112, 6.3842745 , 3.26747544, 1.31275898, 4.03954799,
  196
             4.42599508, 3.3427494, 3.49231019, 3.75122546, 5.38655933,
  197
  198
             5.96342209, 4.98525493, 2.79816073, 2.96732327, 1.22877318,
  199
             3.96172964, 4.64914365, 4.7541684 , 2.96195666, 4.63344105,
  200
             1.29413658, 0.54836519, 2.11513648, 4.56870456, 2.53500211,
  201
             4.39745128, 2.07066377, 3.71019298])
```

Pruebas

```
1 # Prueba 1
2 train_X, train_Y, test_X, test_Y = split_dataset(70, datos) # División de dataset de prueba y entrenamiento
3 test_y, resultado = regresion(test_X, test_Y, fit(train_X, train_Y)) # Predicción de y del dataset de prueba
4 resultado, MSE, MAE, RMSE, MAPE, SSE, R2 = evaluacion(resultado)
5 resultado
```

Mean Squared Error: 0.4784252664607663
Mean Absolute Error: 0.5471487170441045
Root Mean Squared Error 0.6916829233548898
Mean Percent Absolute Error 0.27988871657177916

Sum of Squared Errors 71.76378996911494

Coeficiente de determinación 0.7151342963811769

	income	happiness	Predicción happiness	Error	Error Absoluto	Squared Error
0	4.639914	3.737942	3.531611	0.206330	0.206330	0.042572
1	7.466653	5.960547	5.521604	0.438943	0.438943	0.192671
2	3.528319	2.546525	2.749060	-0.202536	0.202536	0.041021
3	2.428752	1.200786	1.974977	-0.774191	0.774191	0.599372
4	3.542748	3.078293	2.759218	0.319075	0.319075	0.101809
145	7.448117	5.963422	5.508555	0.454867	0.454867	0.206904
146	7.225192	4.985255	5.351618	-0.366363	0.366363	0.134222
147	4.243873	4.754168	3.252802	1.501366	1.501366	2.254101
148	5.249209	4.568705	3.960548	0.608157	0.608157	0.369855
149	3.440847	2.070664	2.687481	-0.616817	0.616817	0.380463

150 rows × 6 columns

En la segunda prueba se modificó el tamaño de los dataset de entrenamiento y prueba y el resultado fue que los valores de los errores disminuyó

```
1 # Prueba 2
```

² train_X, train_Y, test_X, test_Y = split_dataset(90, datos) # División de dataset de prueba y entrenamiento

³ test_y, resultado = regresion(test_X, test_Y, fit(train_X, train_Y)) # Predicción de y del dataset de prueba

⁴ resultado, MSE, MAE, RMSE, MAPE, SSE, R2 = evaluacion(resultado)

⁵ resultado.head()

Mean Squared Error: 0.5413035373328922
Mean Absolute Error: 0.6061574065890772
Root Mean Squared Error 0.7357333330309918
Mean Percent Absolute Error 0.22451272185720014

Sum of Squared Errors 27.06517686664461

Coeficiente de determinación 0.7544831690314232

	income	happiness	Predicción happiness	Error	Error Absoluto	Squared Error
0	4.755680	2.666116	3.591761	-0.925645	0.925645	0.856819
1	1.994275	2.584729	1.620063	0.964666	0.964666	0.930581
2	3.542748	3.078293	2.725703	0.352590	0.352590	0.124320
3	3.065059	3.407980	2.384624	1.023357	1.023357	1.047259
4	5.358716	3.752659	4.022341	-0.269681	0.269681	0.072728

1

1

Colab paid products - Cancel contracts here

✓ 0s completed at 7:14 PM

×