

Inteligencia Artificial Nivel Intermedio

Fecha 8-7-2024

Taller #1

A. Modelo de regresión Lineal

Instrucciones de Ayuda para la primera parte del taller

Datos de vuelos de aviones de estados Unidos 2015

Se debe subir a GoogleColab el DataSet: On_Time_On_Time_Performance_2015_2ac.csv

Desarrolle las siguientes actividades

1. Importe las bibliotecas de manejo de datos, matrices y traficación
2. Suba la data a GoogleColab
3. Visualice toda la data

	Year	Quarter	Month	DayOfMonth	DayOfWeek	FlightDate	UniqueCarrier	AirlineID	Carrier	TailNum	...	Div4Wheelsoff	Div4TailNum	Div5Airport	Div5AirportID	Div5AirportSeqID	Div5Wheelsoff	Div5TotalGTime	Div5LongestGTime	Div5Wheelsoff	Div5TailNum
0	2015	1	2	23	1	23/02/2015	AS	19930	AS	N796AS	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2015	1	2	15	7	15/02/2015	AS	19930	AS	N798AS	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2015	1	2	22	7	22/02/2015	AS	19930	AS	N706AS	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	2015	1	2	10	2	10/02/2015	OO	20304	OO	N612EV	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2015	1	2	8	7	8/02/2015	AS	19930	AS	N785AS	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
22134	2015	1	2	12	4	12/02/2015	UA	19977	UA	N477UA	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22135	2015	1	2	5	4	5/02/2015	OO	20304	OO	N719SK	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22136	2015	1	2	20	5	20/02/2015	NK	20419	NK	N602NK	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22137	2015	1	2	22	7	22/02/2015	AA	19806	AA	N4YRAA	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22138	2015	1	2	5	4	5/02/2015	US	20355	US	N289AY	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

22139 rows x 100 columns

Para trabajar seleccione las columnas: `DepTime`, `DepDelay`, `ArrDelay`, `ArrDelayMinutes`

`DepTime`=Tiempo de Salida del avión

`DepDelay`= Tiempo de Retraso en la salida del Avión

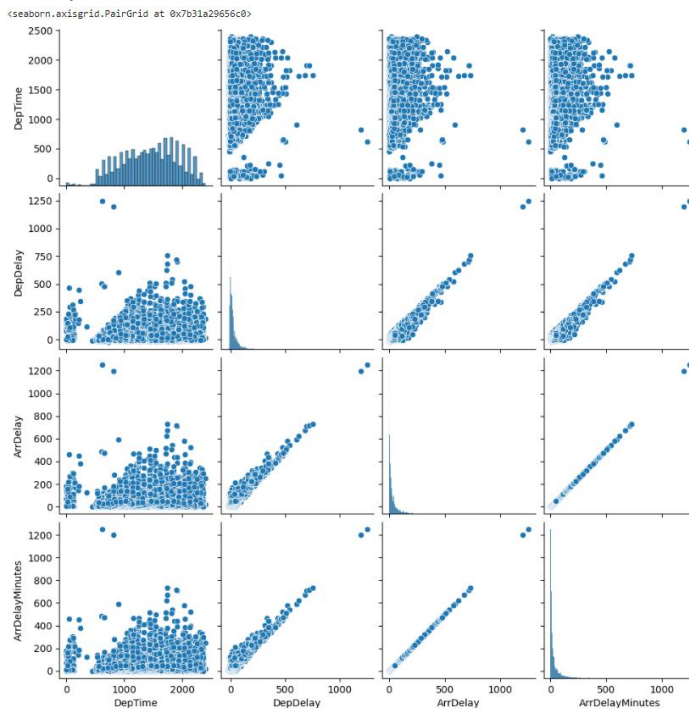
`ArrDelay`=Retraso en la llegada del avión

`ArrDelayMinutes`= minutos de retraso del avión

4. Visualice las columnas seleccionadas

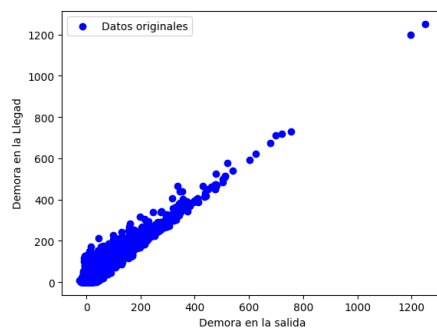
	DepTime	DepDelay	ArrDelay	ArrDelayMinutes
0	1052	-8	3	3
1	1130	30	26	26
2	1540	16	22	22
3	2220	75	87	87
4	1530	6	2	2
...
22134	933	54	47	47
22135	1852	57	48	48
22136	1459	38	49	49
22137	1658	57	51	51
22138	1020	40	51	51

5. Grafique con Seaborn las columnas seleccionadas



6. Realice la descripción de los datos seleccionado y explique los resultados

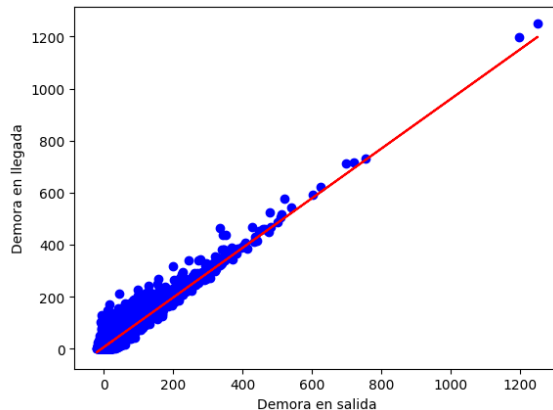
7. Realice una regresión Lineal simple con los datos: X= DepDelay= Tiempo de Retraso en la salida del Avión, Y= ArrDelay=Retraso en la llegada del avión
Con matplotlib, grafique las dos variables seleccionadas



8. Divida el conjunto de datos en 80% para datos de entrenamiento y 20% para pruebas
9. Defina el modelo con regresión lineal simple
10. Con los datos de entrenamiento, entrene el modelo, si es necesario, cambie la estructura del vector X. de entrenamiento

```
X_train  
X_test =  
  
(17711,)
```

11. Realice la predicción con el modelo entrenado con los datos de entrenamiento
12. Grafique los datos de entrenamiento de y, vs los datos de predicción



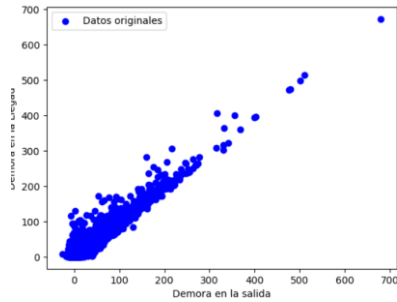
13. Realice el cálculo de R2_score

```
0.9141731928561244
```

14. Cree el programa para utilizar el modelo, donde pide demora en la salida, y calcula los minutos de demora en la llegada, donde pida por teclado el valor de demora en la salida

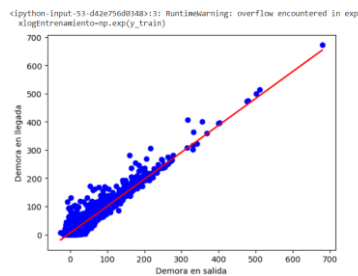
```
Ingrese la demora en salida5  
Demora en llegada [10.24148768]
```

15. Realice el proceso anterior con los datos de prueba X_test, recuerde cambiar la forma del vector de prueba.
16. Grafique los datos de prueba con Matplotlib



17. Realice la predicción con los datos de prueba

18. Grafique la predicción vs los datos de prueba



19. Calcule el r^2_{score} de prueba

→ 0.9097125927606629

20. Compare los dos resultados, el de prueba y el de entrenamiento

B. Depuración de datos

1. Borre las filas con valores negativos de la data de vuelos de aviones en las columnas: `DepTime`, `DepDelay`, `ArrDelay`, `ArrDelayMinutes`
2. Borre las columnas `Div4WheelsOff`, `Div4TailNum`, `Div5Airport`, `Div5AirportID`, `Div5AirportSeqID`, `Div5WheelsOn`, `Div5TotalGTime`, `Div5LongestGTime`, `Div5WheelsOff`, `Div5TailNum`
3. Filtre los datos de la columna `Carrier`, datos OS
4. Borre datos vacíos de una columna que usted seleccione.
5. Ahora seleccione los datos con origen `Origin`

	DepTime	DepDelay	ArrDelay	ArrDelayMinutes	Origin
0	1052	-8	3	3	WRG
1	1130	30	26	26	WRG
2	1540	16	22	22	PSG
3	2220	75	87	87	RHI
4	1530	6	2	2	PSG
...
22134	933	54	47	47	EWR
22135	1852	57	48	48	IAH
22136	1459	38	49	49	LAX
22137	1658	57	51	51	BWI
22138	1000	40	51	51	BWI

6. Cuento los datos por origen

	DepTime	DepDelay	ArrDelay	ArrDelayMinutes
Origin				
ABI	69	69	69	69
ABQ	6	6	6	6
ABR	13	13	13	13
ABY	22	22	22	22
ACT	30	30	30	30
...
VPS	30	30	30	30
WRG	16	16	16	16
XNA	1	1	1	1
YAK	5	5	5	5
YUM	54	54	54	54

7.