



Universidad Politecnica Salesiana

Nombre: Paulo Gonzalez

Carrera: Ingeniera en Computación

Materia: Simulación

Tema: Practica 2. Regreción Covid

Cuenca - Ecuador
2021-2021

Generar un modelo de regresión (polinomial y lineal) sobre los casos de COVID dentro del Ecuador, demostrando el comportamiento y la predicción de la emergencia sanitaria, tomando datos desde el inicio de la emergencia hasta 1500 días después, tomando en cuenta confinamientos y feriados.

1. Leemos el dataset filtrando los datos, es decir detectamos los datos con los cuales son relevantes para generar la predicción; este es el resultado

```
In [238]: #seleccionamos el dataset
dataSetRead = pd.read_csv("covid-regresion.csv")
dataSetRead = dataSetRead.fillna(0)
dataSetRead.head()
```

Out[238]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	gdp_per_capita
0	ECU	South America	Ecuador	2020-03-01	6	6	0.0	0.0	0.0	0.0	...	10581.936
1	ECU	South America	Ecuador	2020-03-02	6	0	0.0	0.0	0.0	0.0	...	10581.936
2	ECU	South America	Ecuador	2020-03-03	7	1	0.0	0.0	0.0	0.0	...	10581.936
3	ECU	South America	Ecuador	2020-03-04	10	3	0.0	0.0	0.0	0.0	...	10581.936
4	ECU	South America	Ecuador	2020-03-05	13	3	0.0	0.0	0.0	0.0	...	10581.936

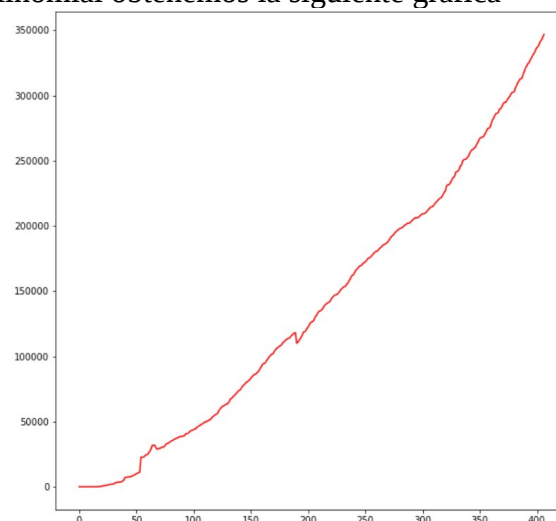
5 rows × 59 columns

2. Ahora lo que se hace es obtener las fechas de forma numerica, ya que eso nos permite trabajar para demostrar en las graficas, y el sistema de predicción tenga valores para comparar entre las fechas y la cantidad de casos que se han ido generando.

```
In [ ]: #Obtenemos la data de las fechas relacionadas
dataSetRead['date'] = pd.to_datetime(dataSetRead['date'])
dataSetRead['date'] = (dataSetRead['date'] - dataSetRead['date'].min()) / np.timedelta64(1, 'D')
```

```
In [ ]: #Estructura para la predicción
FinalTotalDate = dataSetRead['date'].values.reshape(-1,1)
FinalTotalCases = dataSetRead['total_cases'].values.reshape(-1,1)
rg_lineal = LinearRegression()
rg_lineal.fit(FinalTotalDate, FinalTotalCases)
prediction = rg_lineal.predict(FinalTotalCases)
```

3. Para la regresión polinomial obtenemos la siguiente gráfica



4. Ahora generamos un modelo de entrenamiento para obtener las predicciones, en cuanto al avance del covid, para ello generamos un test y un train, tomando en cuenta el grado del polinomio establecido

```

In [242]: #Se define el grado del polinomio
rg_polinomial = PolynomialFeatures(degree = 5)

In [243]: #Se realiza un modelo y entrenamiento para que las características existentes sean de mayor grado
X_train_rg_polinomial = rg_polinomial.fit_transform(X_train_p)
X_test_rg_polinomial = rg_polinomial.fit_transform(X_test_p)

rg_polinomial_model = linear_model.LinearRegression()
rg_polinomial_model.fit(X_train_rg_polinomial, y_train_p)
Y_prediction_rg_polinomial = pr.predict(X_test_rg_polinomial)

```

5. Generando el siguiente resutado:

Valor del coeficiente "a" :

```
[ 0.00000000e+00  2.25132000e+02  2.60841709e+00  3.12348831e-03
 -4.09269854e-05  6.66048066e-08]
```

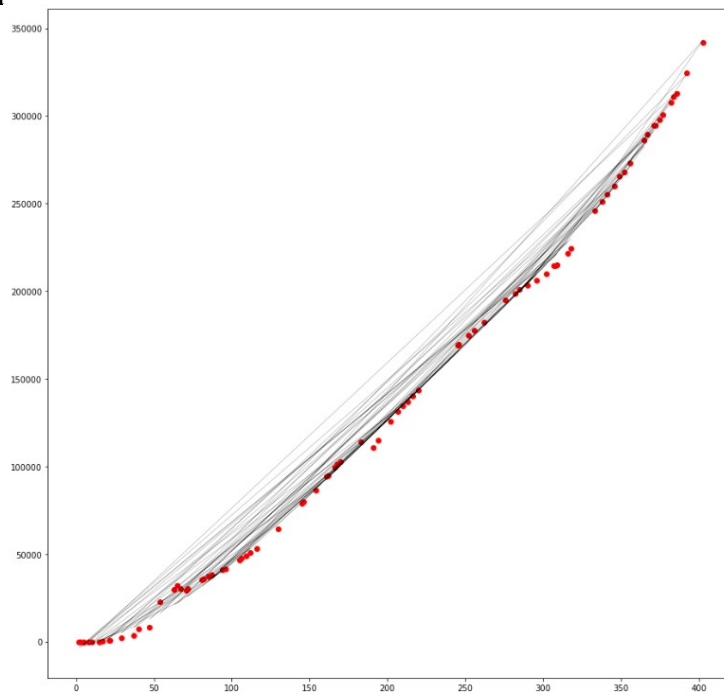
Valor del coeficiente "b":

```
-3800.8914346888487
```

Precisión del modelo:

```
0.9991072841881748
```

6. Grafica de la predicción

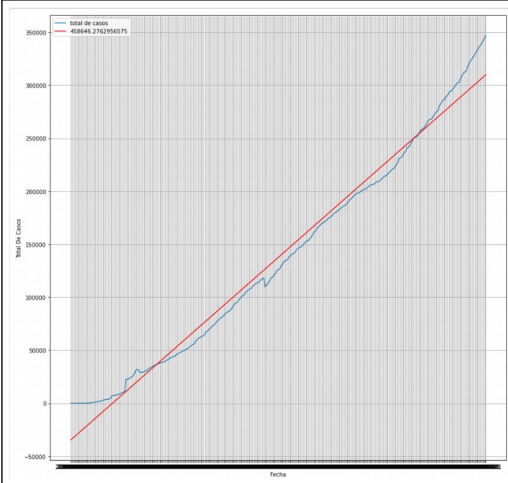


7. Para la regresion lineal con el modelo establecido se obtuvo los siguientes resultados

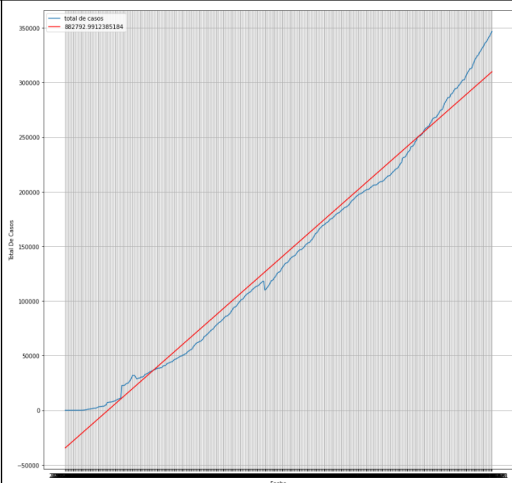
Valor del coeficiente "a":
[[848.29342989]]

Valor del coeficiente "b":
[-34499.5613528]

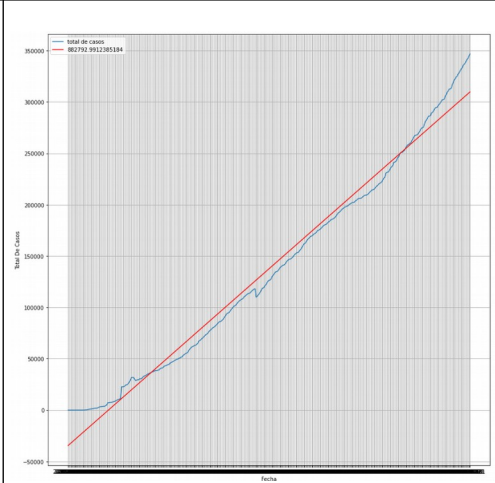
8. Ahora las graficas resultantes se toma en cuenta en base a cada 500 dias de la ultima estadistica realizada



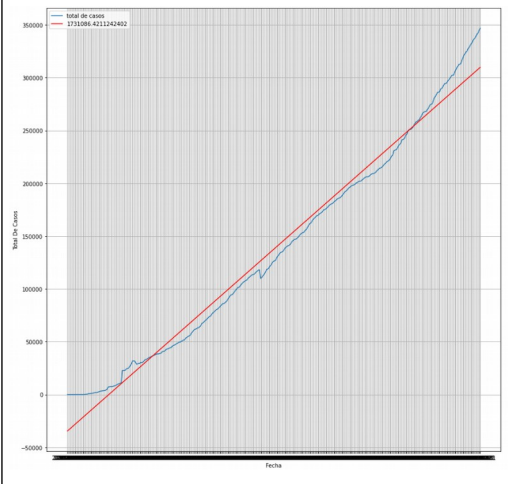
Despues de 500 días, la preccion resultante es : 458646.2762956575



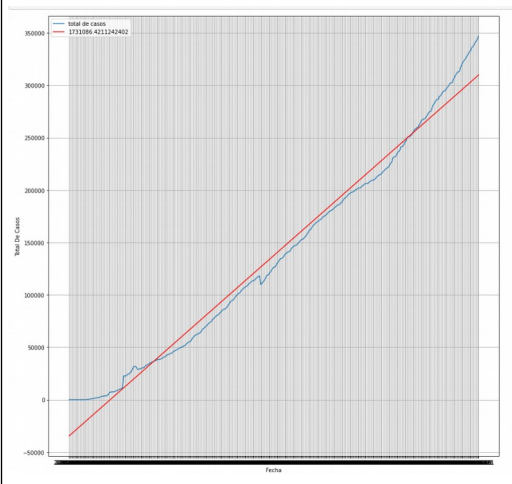
Despues de 1000 días, la preccion resultante es : 882792.9912385184



Despues de 1000 días, la preccion resultante es : 882792.9912385184



Despues de 1500 días, la preccion resultante es : 1306939.7061813793



Despues de 2000 días, la preccion resultante es : 1731086.4211242402

Observaciones

Los datos del Ecuador muchos al parecer son alterados ya que la cantidad que contiene son pocas a relacion con otros países, además de ello en los feriados se nota una alza significativa en cuanto a los casos que se registraron al inicio de esta emergencia.

Conclusiones

El modelo de predicción se hizo lo mas eficiente posible a pesar de la cantidad de datos que no fueron los mas acertados, además de ello el aumento casi exponencial de los casos de COVID dentro del país es muy notorio, y el control que se tiene no es el mas eficiente, en los días establecidos ha crecido muy notoriamente.