

Universidad Politécnica Salesiana



Nombre:

David Andrés Morales Rivera.

Tema:

Practica 2 – Regresión de los casos de COVID.

Asignatura:

Simulación.

Docente:

Ing. Diego Quisi.

Fecha:

Cuenca, abril de 2021.

Problema.

Generar un modelo de regresión de los casos confirmados de COVID dentro del Ecuador, el mismo que permita predecir el comportamiento y/o predicción de la pandemia, tomar los datos desde el inicio e identificar etapas: Confinamiento, Toques de Queda, Feriados, etc.

Desarrollo.

Para esta practica se implementaron 3 modelos de regresión para poder predecir el número de contagios del COVID-19 en el Ecuador.

La información de contagios fue tomada de:

- https://covid.ourworldindata.org/data/ecdc/new_cases.csv

	date	World	Afghanistan	Albania	Algeria	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina	...
278	2020-10-04	301978	7.0	152.0	148.0	0.0	97.0	0.0	1.0	11129.0	...
151	2020-05-30	123188	623.0	23.0	137.0	1.0	4.0	0.0	0.0	717.0	...
305	2020-10-31	542158	157.0	319.0	319.0	98.0	195.0	0.0	3.0	13955.0	...
76	2020-03-16	12489	6.0	4.0	11.0	3.0	NaN	NaN	NaN	11.0	...
171	2020-06-19	138809	658.0	66.0	117.0	1.0	0.0	0.0	0.0	0.0	...
255	2020-09-11	305692	28.0	156.0	272.0	0.0	59.0	0.0	0.0	12259.0	...
142	2020-05-21	104973	492.0	15.0	165.0	1.0	2.0	0.0	0.0	474.0	...
207	2020-07-25	284289	108.0	104.0	675.0	0.0	39.0	0.0	6.0	6127.0	...

Ilustración 1. Información de contagios de COVID-19

A la fecha de realizar esta practica se cuenta con datos desde el 01-01-2020 hasta el 29-11-2020.

Se debe considerar que la información no es la oficial, debido a esto se tuvo que dar una revisión a los datos y se procedió a eliminar filas que tenían valores nulos o 0, y también eliminar filas que por simple análisis no debían ir.

Los datos validos para Ecuador quedaron de la siguiente manera:

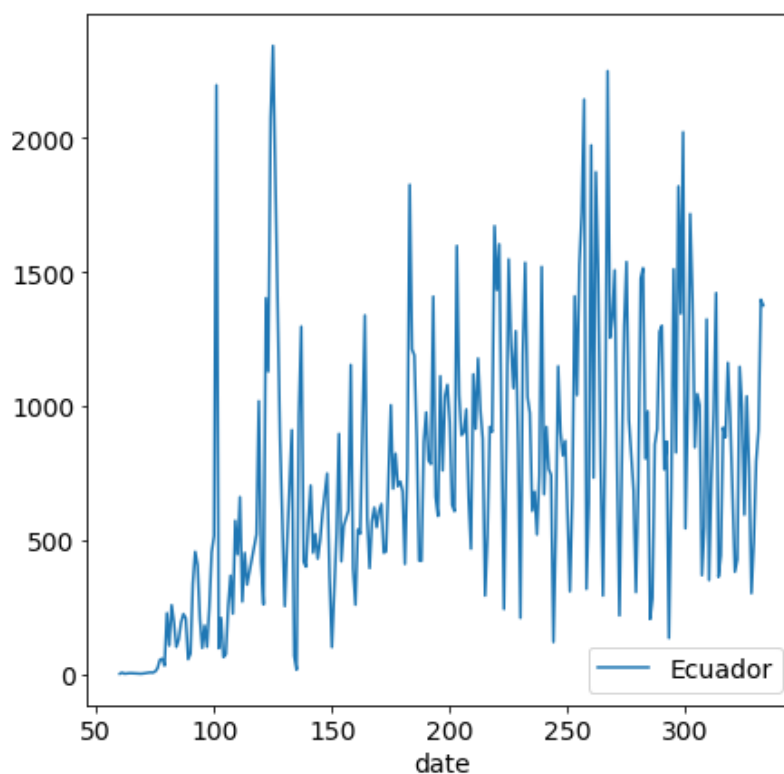


Ilustración 2. Contagios COVID en Ecuador del 01-01-2020 al 29-11-2020

- **Regresión lineal**

La regresión lineal es un algoritmo de aprendizaje supervisado que se utiliza en Machine Learning y en estadística. En su versión más sencilla, lo que haremos es «dibujar una recta» que nos indicará la tendencia de un conjunto de datos continuos.

Recordemos rápidamente la fórmula de la recta:

$$Y = mX + b$$

Donde Y es el resultado, X es la variable, m la pendiente (o coeficiente) de la recta y b la constante o también conocida como el «punto de corte con el eje Y» en la gráfica (cuando $X=0$).

- **Regresión logística**

El modelo logístico se ha utilizado ampliamente para describir el crecimiento de una población. Una infección puede describirse como el crecimiento de la población de un agente patógeno, por lo que un modelo logístico parece razonable. La expresión más genérica de una función logística es:

$$y = a + b * \log(x)$$

En esta fórmula, tenemos la variable x que es el tiempo y tres parámetros: a, b, c .

- **a** se refiere a la velocidad de infección
- **b** es el día en que ocurrieron las infecciones máximas
- **c** es el número total de personas infectadas registradas al final de la infección

- **Regresión polinómica**

Predicción de una variable de respuesta cuantitativa a partir de una variable predictora cuantitativa, donde la relación se modela como una función polinómica de orden n (esto significa que pueden tener de diferentes exponenciales o grados y se debe ir probando). Se puede tener una ecuación con diferentes grados

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n + \epsilon$$

Experimentos.

Para los 3 modelos se realizó una predicción para los siguientes 100 días. (Tomando en cuenta como última fecha el 29-11-2020).

Los resultados fueron los siguientes:

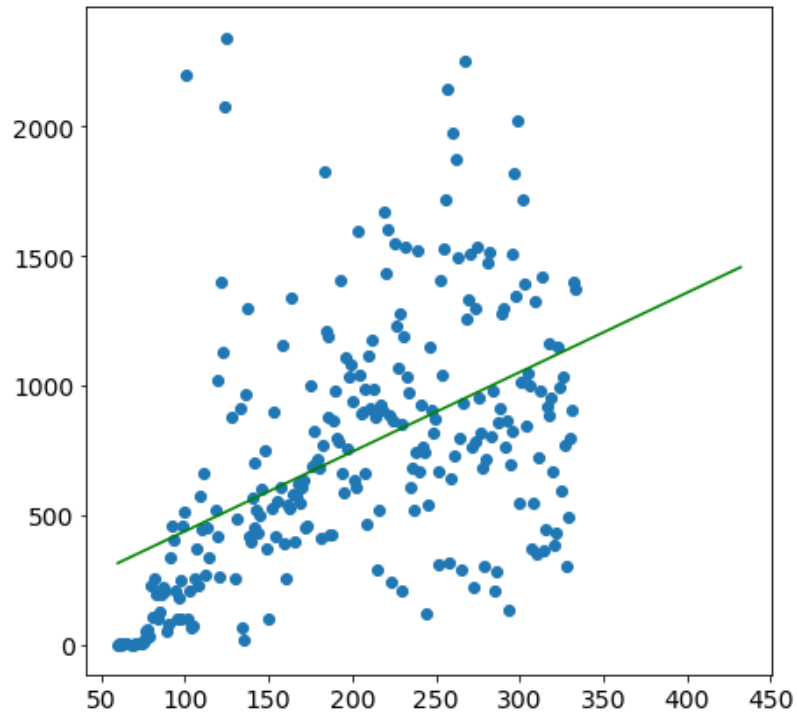


Ilustración 3. Regresión lineal (siguientes 100 días)

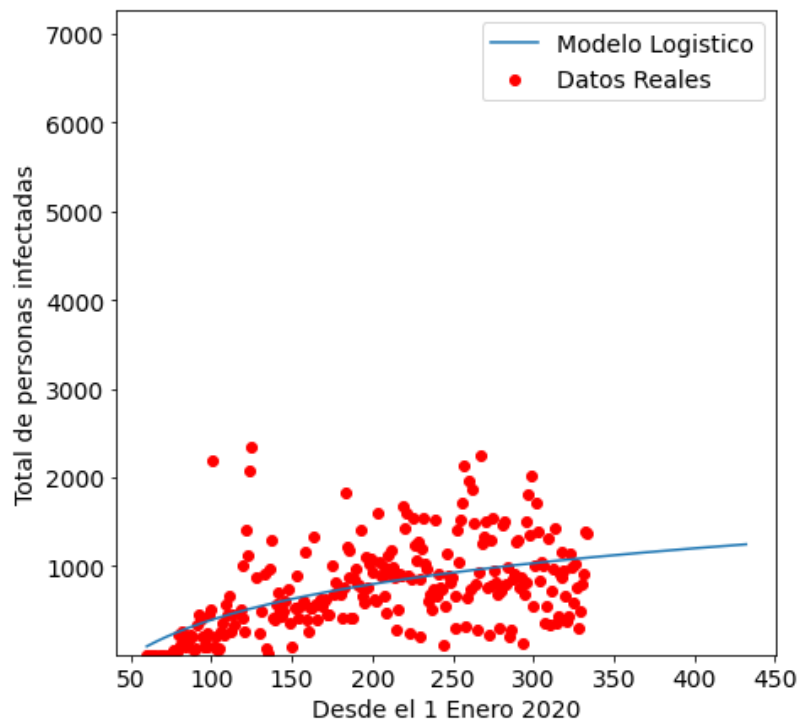


Ilustración 4. Regresión logística (siguientes 100 días)

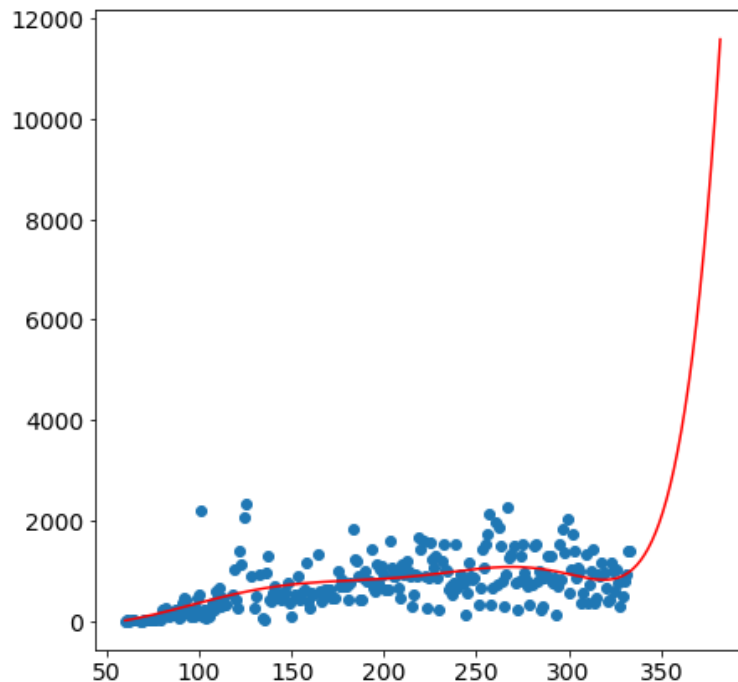


Ilustración 5. Regresión polinómica (siguientes 100 días)

Como se puede apreciar en las ilustraciones, el que tiene un mejor ajuste con los datos es el modelo polinómico, cabe aclarar que este modelo se realizó con grado 8, ya que luego de filtrar la información correcta nos quedaron fechas válidas que equivalían a 8 meses, esto con el fin de que el modelo sea lo más apegado a la realidad de los datos.

Para el siguiente experimento, se busca que los modelos realicen una predicción a la fecha actual (24-04-2021), el resultado será comparado con información oficial del Ministerio de Salud Pública para poder analizar y seleccionar que modelo es el que más se asemeja a la realidad.

Actualmente el Ecuador cuenta con 371 mil casos confirmados de COVID:

Resumen de casos

De [Wikipedia](#) y [JHU CSSE COVID-19 Data](#) · Última actualización: hace 2 horas



Ecuador

Total de casos

371 k

+5913

Personas recuperadas

310 k

Muertes

18,062

+97

Ilustración 6. Casos oficiales de COVID (24-04-2021)

- **Modelo lineal**

```
#Vamos a comprobar:  
# Quiero predecir cuántos "Casos" hay en la fecha actual 24/04/2021,  
# según nuestro modelo, hacemos:  
y_prediccion = regr.predict([[479]])  
print("Prediccion para la fecha 24/04/2021:")  
print(int(y_prediccion))
```

Prediccion para la fecha 24/04/2021:
1600

- **Modelo logístico**

```
#Vamos a comprobar:  
# Quiero predecir cuántos "Casos" hay en la fecha actual 24/04/2021,  
# según nuestro modelo, hacemos:  
prediccion_reg = modelo_logistico(479,log_fit[0][0],log_fit[0][1])  
print("Prediccion para la fecha 24/04/2021:")  
print(int(prediccion_reg))
```

Prediccion para la fecha 24/04/2021:
1306

- **Modelo polinómico (grado 8)**

```
#Vamos a comprobar:  
# Quiero predecir cuántos "Casos" hay en la fecha actual 24/04/2021,  
# según nuestro modelo, hacemos:  
yPoly_prediccion = regresion_lineal.predict(pf.fit_transform([[479]]))  
print("Prediccion para la fecha 24/04/2021:")  
print(int(yPoly_prediccion))
```

Prediccion para la fecha 24/04/2021:
414740

Como se puede observar en las predicciones, para la fecha 24-04-2021 el dato mas cercano al de la informacion oficial, corresponde al modelo polinomico.

Conclusiones

Con este trabajo se puede concluir que es importante realizar varias pruebas con diferentes modelos, con esto podremos asegurar que las predicciones que se realicen sean lo mas precisas y parecidas a la realidad.