**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>



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 tuve 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:

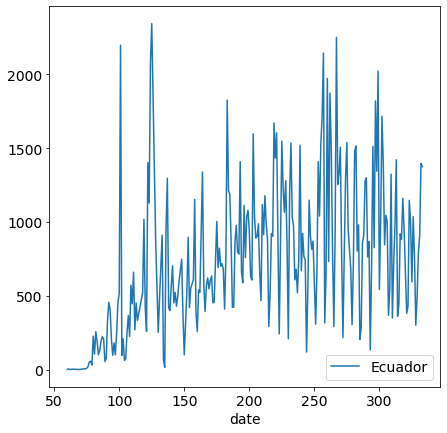


Ilustración 2. Contagios COVID en Ecuador del 01-01-2020 al 29-11-2020

* **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 = a0 + a1x + a2x^2 + a3x^3 + … + anx^n + ε**

**Experimentos.**

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

Los resultados fueron los siguientes:

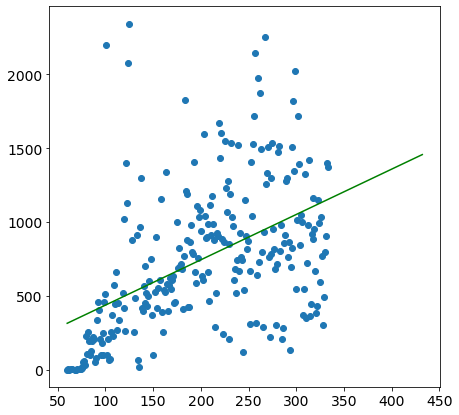


Ilustración 3. Regresión lineal (siguientes 100 días)

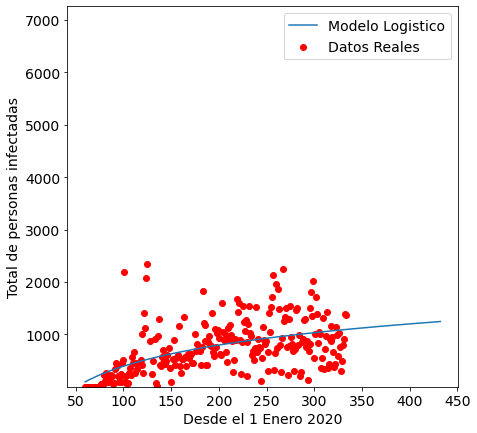


Ilustración 4. Regresión logística (siguientes 100 días)

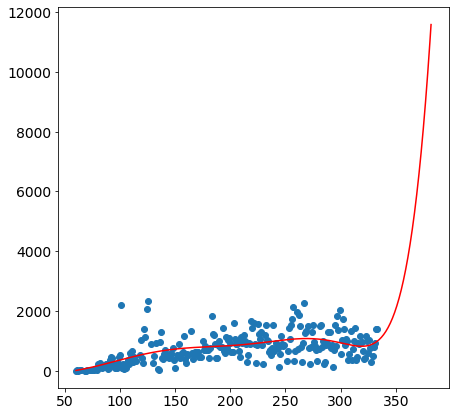


Ilustración 5. Regresión polinómica (siguientes 100 días)

Como se puede apreciar en las ilustraciones, el que tiene un mejor ajuste con los datos es el modelo polinomico, cabe aclarar que este modelo se realizo con grado 8, ya que luego de filtrar la informacion correcta nos quedaron fechas validas que equivalian a 8 meses, esto con el fin de que el modelo sea lo mas apegado a la realidad de los datos.

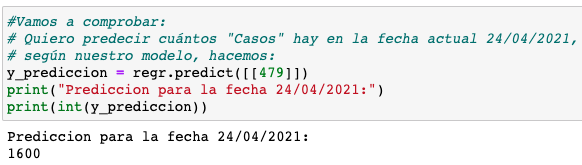
Para el siguiente expermento, se busca que los modelos realicen una prediccion a la fecha actual (24-04-2021), el resultado sera comparado con informacion oficial del Ministerio de Salud Publica para poder analizar y seleccionar que modelo es el que mas se asemeja a la realidad.

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

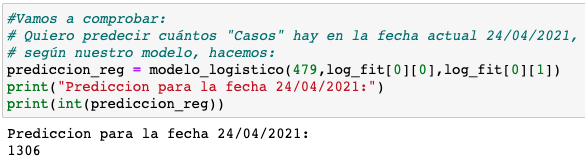


Ilustración 6. Casos oficiales de COVID (24-04-2021)

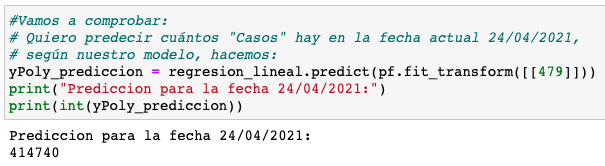
* **Modelo lineal**



* **Modelo logistico**



* **Modelo polinómico (grado 8)**



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.