

Predicción de COVID-19 en el Perú

Ricardo Manuel Lazo Vásquez

Mayo 2020

Abstract

La predicción de casos del COVID-19 es de suma importancia hoy en día, actualmente se menciona que en el Perú la curva de infección está llegando a su cima, para determinar este y otras estadísticas de la pandemia se utilizaron técnicas de Machine Learning para este fin. El presente trabajo presentara una predicción de los últimos días de la pandemia en el Perú así como en sus regiones desde el último 30% de los días de infección considerando el primer día de inicio la llegada del COVID a cada región, y, de manera general, al país.

1 Introducción

En el presente informe se describe la implementación de predicción de COVID-19 en el Perú y sus regiones. El trabajo está dividido en la siguiente manera: en la Sección 2 se explica los detalles de la implementación, posteriormente en la Sección 3 se describirán los resultados obtenidos y finalmente en la Sección 4 se darán las conclusiones y se hablará de los trabajos futuros.

2 Implementación

2.1 Obteniendo los datos

La implementación fue trabajada con el dataset encontrado en Github [Jmcastagnetto()]; se dividió los datos considerando para la prueba los últimos días de infección teniendo alrededor de 20 días disponibles. Se dividió de la misma manera para trabajar con cada región, siempre teniendo en cuenta el día en el que llegó la pandemia a esa región como día inicial.

Así mismo el código de nuestra implementación está en Github [Wasp(2020)] donde se encuentra el notebook y las imágenes resultantes de su ejecución.

2.2 Visualización de los datos

Tratándose este un problema que no tiene que ver con clasificación no tiene mucho sentido utilizar una visualización por histograma; por el contrario se

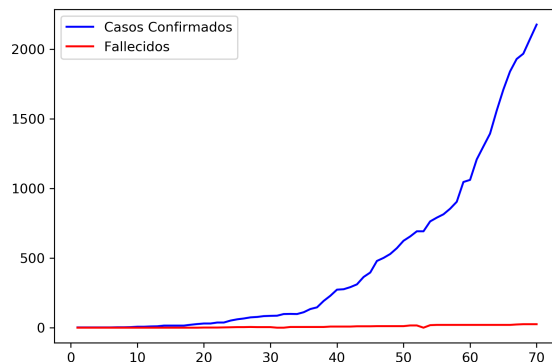


Figure 1: Curva de infectados y de fallecidos en Arequipa.

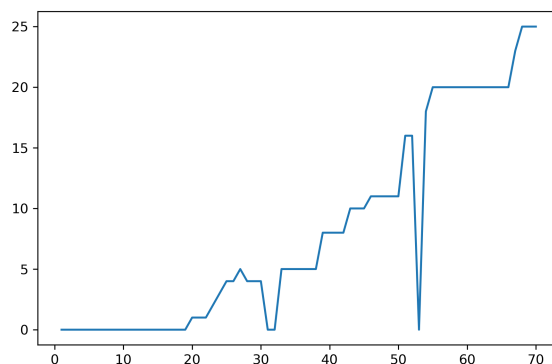


Figure 2: Curva de fallecidos en Arequipa teniendo valores nulos en algunos de sus elementos.

prefiere visualizar las funciones para posteriormente trabajarlas mejor con la gradiente.

Optamos por trabar sobre Arequipa en los primeros experimentos con el fin de obtener resultados preliminares en un dataset de tamaño intermedio para posteriormente trabajar con los datos de todo el país. La visualización de los datos de Arequipa pueden observarse en la Figura 1.

Fue encontrado un conjunto de problemas con el conjunto de datos; entre ellos la correcta división por región, datos intermedios nulos, datos nulos sin posibilidad de análisis. Un ejemplo de cómo afectarían la curva de datos si es que no se corregía este problema sería no poder calcular la predicción de infectados/muertos de cualquier región. Los datos sin trabajar de la región se pueden visualizar en la Figura 2, en ella se pueden observar que no se puede hacer predicciones para calcular los fallecidos ya que esta presenta valores nulos.

2.3 Preparando los datos

Se optó por separar el dataset de todas las regiones en regiones independientes, para empezar se optó por regresión lineal de confirmados/muertos dados los datos de las pruebas y, por otra parte de, una regresión polinomial dada solamente las fechas.

Para ello se separo solo las columnas de interés para obtener los resultados bajo determinados $X's$ para los $Y's = \{Casosconfirmados, Fallecidos\}$.

Finalmente se optó en esta parte predecir los casos confirmados dadas las muertes y las muertes dados los casos confirmados, por lo que se preparó Xs para estos dos casos en el notebook.

2.4 Entrenamiento

Se entrenó primero con gradiente descendiente las regresiones lineal (Datos normales) y polinomial (Fechas), se observó que la regresión polinomial no tenia buenos resultados en su primera implementación por lo que se optó por utilizar la lineal de manera temporal; esto no afectó los resultados ya que había varias columnas de datos que influyen de manera lineal el resultado final a predecir. Sin embargo, se optó por mejorar la implementación de regresión polinomial para obtener resultados más aproximados a la curva de casos confirmados y muertes en trabajos posteriores.

3 Resultados

De la anterior implementación se obtuvieron los resultados en Arequipa que para el día 70 de iniciado el virus en esta región se tenía un estimado de 2183 contagiados y resultados que se pueden comparar con los observados en la Figura 1.

3.1 Otras aproximaciones

Otras maneras de predecir el número de contagiados y fallecidos es utilizando estos mismos para predecirse entre sí, se utilizó gradiente descendiente y el dataset de Perú para lograr este objetivo. Los resultados pueden verse en la Tabla 1

4 Conclusiones y trabajos futuros

En el presente trabajo se predice el COVID-19 en el Perú y sus regiones, se obtuvo resultados basándonos en las pruebas diarias y en la cifra de casos confirmados y fallecidos en el país obteniendo como resultado solo resultados dentro de las fechas anteriores. Para trabajos futuros se fortalecerá la regresión polinomial y se predecirá datos fuera de estos parámetros.

Día de infeccion	Contagiados	Muertes
53	35120.0	906.0
54	39143.0	988.0
55	41862.0	1081.0
56	44692.0	1136.0
57	47895.0	1227.0
58	50055.0	1265.0
59	53779.0	1367.0
60	57094.0	1464.0
61	60595.0	1563.0
62	63835.0	1652.0
63	67559.0	1736.0
64	70353.0	1798.0
65	73034.0	1838.0
66	76610.0	1924.0
67	80781.0	2038.0
68	84431.0	2153.0
69	89086.0	2257.0
70	93965.0	2365.0
71	98620.0	2464.0
72	103872.0	2535.0
73	108527.0	2657.0
74	112624.0	2778.0
75	117242.0	2905.0
76	120817.0	2983.0

Table 1: Predicción de casos confirmados y de muertes en el Perú

References

- [cou()] Deep learning by deeplearning.ai. URL <https://www.coursera.org/specializations/deep-learning>.
- [Jmcastagnetto()] Jmcastagnetto. jmcastagnetto/covid-19-peru-data. URL <https://github.com/jmcastagnetto/covid-19-peru-data/blob/master/datos/covid-19-peru-data.csv>.
- [Wasp(2020)] Wasp. Thereversewasp/tia-covid-peru-prediction, May 2020. URL <https://github.com/TheReverseWasp/TIA-COVID-Peru-Prediction>.