

Computación Blanda

Soft Computing

Autor: Daniel Giraldo Muñoz

IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia

Correo-e: daniel.giraldo2@utp.edu.co

Resumen— Este documento presenta un resumen sobre un programa de machine learning básico, basado en el histórico de casos positivos contagiados con el virus COVID-19. Se pretende hacer un estudio, tomando en cuenta el histórico de personas que en algún momento estuvieron contagiadas desde el inicio de la emergencia sanitaria hasta el día 1 de octubre, a partir de esto podremos predecir en que momento se llegara a contagiar el 2% de población total de Colombia ya que las estadísticas oficiales de la Organización Mundial de la Salud (OMS) prevé que por lo menos entre el 1% y el 5% de la población total de una determinada región será contagiada con este virus.

Palabras clave— algoritmo, arreglo, COVID-19, contagios machine learning, solicitud, virus.

Abstract— This document presents a summary of a basic machine learning program, based on the history of positive cases infected with the COVID-19 virus. It is intended to carry out a study, taking into account the history of people who at some point were infected from the start of the health emergency until October 1, from this we can predict when 2% of patients will be infected. Colombia's total population, since official statistics from the World Health Organization (OMS) foresee that at least between 1% and 5% of the total population of a certain region will be infected with this virus.

Key Word— algorithm, array, COVID-19, infection, machine learning, request, virus.

I. INTRODUCCIÓN

Debido a la crisis sanitaria por la cual se encuentra cursando la humanidad, Desde los comienzos de la pandemia, la OMS ha trabajado sin descanso para apoyar a los países a prepararse y responder a este nuevo virus.

El 30 de enero se declaró que la epidemia constituía una emergencia de salud pública de importancia internacional, el mayor nivel de alarma con arreglo al derecho internacional.

Tan solo cuatro días después, la OMS publicó su primer Plan Estratégico de Preparación y Respuesta.

Incluso antes de ello, una de nuestras prioridades principales era identificar rápidamente este nuevo virus y poner a disposición de los países pruebas para detectarlo.

En un plazo de dos semanas desde la notificación de los primeros casos a la OMS, elaboro y preparo el primer protocolo para el desarrollo de una prueba RCP.

El 2 de febrero se fabricaron las primeras pruebas y se distribuyeron por todo el mundo.

Desde entonces han distribuido millones de pruebas entre más de 150 países. Ninguna prueba de laboratorio se ha desarrollado y desplegado jamás con tanta rapidez. La OMS no aisló el virus, ni secuenció ni publicó el genoma, no fabricó las pruebas ni construyó las máquinas para las RCP. Con todo, fue la OMS la que estableció los vínculos necesarios para proporcionar a los países la capacidad de detectar este nuevo virus y combatirlo, Colombia no fue la excepción, país en el cual este virus llegó aproximadamente el 10 de marzo del 2020.

Contando actualmente con un registro histórico aproximado de 886.000 personas contagiadas con dicho virus en el país

Hace diez meses, este virus era completamente desconocido. Ahora, han publicado más de 400 documentos de orientación dirigidos a particulares, las comunidades, las escuelas, las empresas, el sector industrial, el personal de salud, los centros sanitarios y las administraciones públicas.

I.1 METODOLOGIA

Se comienza recopilando los datos, para esto nos dirigimos a un canal gubernamental oficial con todas las estadísticas y casos reportados como positivos en diferentes fechas recopiladas a lo largo de lo que va corrido del año 2020.

Encontramos mucha información detallada sobre cada caso, pero esta información no era global, es decir, nos describía

detalladamente un solo caso de un día cualquiera del plazo anteriormente mencionado.

Por lo tanto, se hizo una selección y agrupación ya que cada caso tena un ID, reunimos los ID de cada caso para hacer una analogía del total de casos partiendo del ID.

Para efectos de este estudio necesitábamos conocer cuántos casos en total se han registrado hasta una fecha estipulada, es decir, no era de nuestro interés conocer el orden de estos casos o las fechas en las cuales fueron detectados, queremos conocer el recorrido histórico de pacientes positivos para dicho virus.

El estudio se realizo con una fecha limite establecida la cual es 1 de octubre.

Conocemos que el año tiene 365 días, contrastando esta información con los días recorridos hasta el 1 octubre obtenemos 277 días.

Los días serán nuestra primera columna y los casos por días serán nuestra segunda columna.

Queremos analizar según el histórico de casos en que semana se llegara al 2% de casos históricos contagiados ya que entre el 1% y el 5% de la población serán los contagios totales por determinada región según la Organización Mundial de la Salud (OMS).

Partimos de un paquete de datos a ser procesados, tenemos una columna que nos indica el número días, otra columna que nos indica el número de contagios acumulados históricos

Haciendo uso de la librería numpy asignamos los valores a una variable llamada “data”, con las herramientas que nos ofrece numpy dividimos el arreglo en dos vectores, columnas x,y.

Con las siguientes instrucciones que nos permiten utilizar todas las filas con la columna cero y viceversa “x= data[:,0], y=[:,1]”

A continuación, verificamos la dimensión de los vectores con la función “n.dim” la cual nos proporciona el número de columnas, seguido de función “shape” la cual nos muestra el número de filas.

Sabemos que en un conjunto de datos se pueden presentar datos erróneos o fuera de unos parámetros establecidos, por lo tanto, debemos conocer cuántos valores erróneos existen en nuestros vectores para que no afecten los siguientes procedimientos.

Con las siguientes funciones nos permite conocer el número exacto de datos erróneos “(np.sum(np.isnan()))”

Np.sum nos devuelve una sumatoria de los valores encontrados por la función np.isnan que nos dice verdadero si un dato esta fuera de sus parámetros o falso si se encuentra dentro de sus parámetros

Ahora deseamos eliminar los datos erróneos de los vectores, negando la función “np.isnan()”

X= x[~np.isnan(y)]

y= y[~np.isnan(y)]

lo que me produce es que ya los valores erróneos son negativos, ¿por qué solo se hace esto en el vector y?, porque los valores erróneos solo se encuentran en el vector y.

importamos la librería matplotlib.pyplot para realizar las gráficas, dibujamos los puntos x, y, con un radio de 10 “plt.scatter(x,y, s=10)” ponemos los respectivos nombres a sus ejes y procedemos a realizar el cálculo de las horas por semanas.

plt.xticks([w*7 for w in range(50)], ['semana %i' % w for w in range(10)])

Esta función nos permite conocer las semanas hasta la semana 50, mientras a su vez nos da la etiqueta por semana para realizar la respectiva gráfica.



Podemos observar es una grafica casi que lineal con una pendiente.

Con la siguiente función podemos obtener los coeficientes para un polinomio de cualquier grado, para posterior mente trazar una ruta la cual describa muy bien todos los puntos encontrados en la gráfica y cuál será su próximo rumbo “polyfit()”.

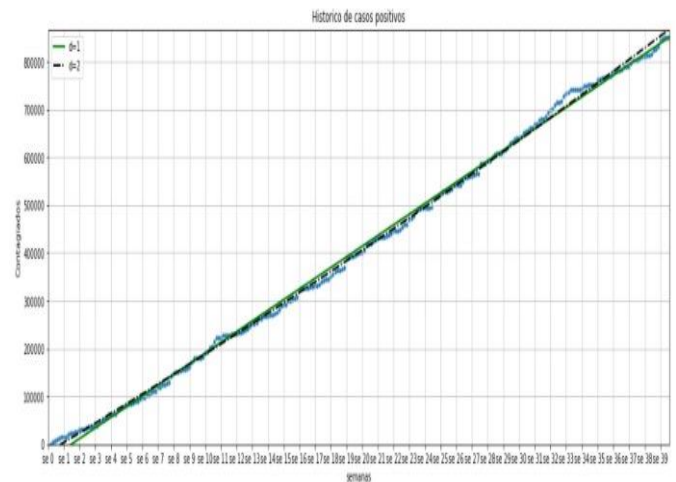
Realizamos la misma acción pero esta vez con un polinomio de grado dos, grado 3, grado 10 y luego de grado 100 para mirar cual de todos se comporta mejor.

Necesitamos hallar un punto de inflexión, para efectos de este estudio decidimos ubicar el punto de inflexión en la semana 23 ya que entre la semana 20 y la semana 25 se encuentra una pequeña variación, es un poco difícil ya que esta grafica se asemeja mas a una pendiente que a una curva.

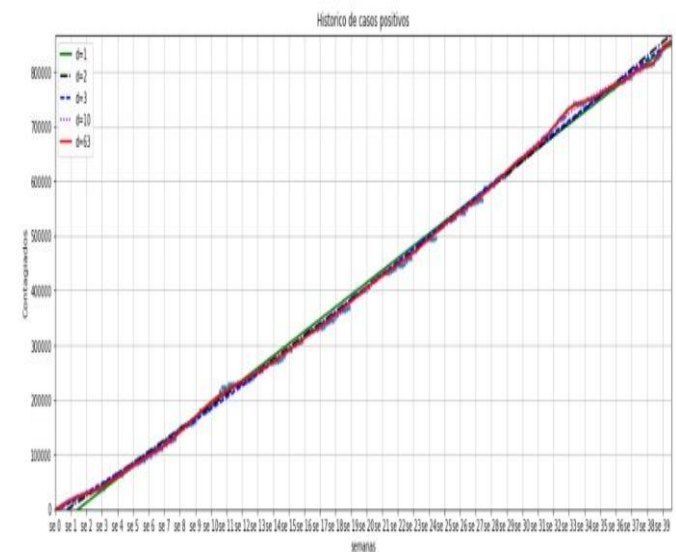
Graficamos los polinomios correspondientes de acuerdo al grado de los polinomios.



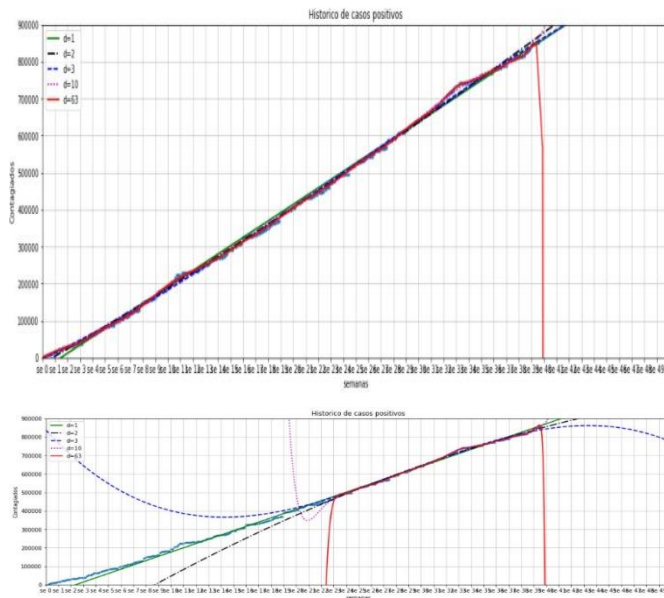
Polinomio de grado 2



polinomio de grado 2, 3 y 10



Polinomio de grado 10 y 63



Después de haber tenido las graficas realizamos en el entrenamiento de los datos para poder predecir el comportamiento de las rectas descritas por los polinomios.

I.2 CONCLUSIONES

Con este análisis y tratamiento que le dimos al conjunto de datos principales tenemos una base para partir y realizar las posibles deducciones, aunque a simple vista se puede ver como una función lineal o de crecimiento exponencial los cual nos dificulta un poco al momento de la predicción.

La función tiene una tendencia al alza, pero sabemos que en algún momento tendrá un cambio ya bien sea leve o brusco.

Para efectos de este análisis en este estudio pusimos un límite basándonos en estadísticas y predicciones hechas por la OMS.

Efectivamente encontramos que estos contagios tendrán un cambio ya que su crecimiento no seguirá de una forma exponencial, por lo contrario, tendra a ser lineal.

Gracias al machine learning pudimos encontrar que el total de casos históricos esperados siguiendo los lineamientos y estadísticas dadas por la OMS son 1.000.000 de casos históricos en la semana 90, es decir, mas del doble del tiempo que llevamos en la emergencia sanitaria, lo cual es un gran indicador de que esta función no seguirá una tendencia exponencial, si no que en algún momento entre el tiempo de la semana 40 y la semana 90 tendrá un cambio, que en cuestiones de salud es positivo ya que nos indica que la crisis sanitaria por la cual atraviesa el país podrá terminar en muy poco tiempo.

REFERENCIAS

Referencias en la Web:

[1]

<https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr>

[2]

<https://www.who.int/es/dg/speeches/detail/who-director-general-s-opening-remarks-at-executive-board-meeting>



Góiraldo Daniel. Nació en Pereira, Colombia en 2000. Bachiller del colegio La Anunciación (2016), estudiante becado de la Universidad Tecnológica de Pereira (2016) cursando actualmente pregrado en Ingeniería en Sistemas y Computación.