



Reporte 07

Nombre: Martínez López Andrés

Fecha: 27/06/2021

Referencia bibliográfica	<p>APA L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed. (2020, noviembre). Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. SN Computer Science. https://doi.org/10.1007/s42979-020-00394-7</p> <p>IEEE L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset," SN Computer Science, vol. 2, no. 1, Nov. 2020.</p>
Autor (es)	L. J. Muhammad, Ebrahim A. Algehyne, Sani Sharif Usman, Abdulkadir Ahmad, Chinmay Chakraborty & I. A. Mohammed
Título	Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset
Año	2020
Tipo de publicación	Artículo de investigación
Nombre de la revista, conferencia, Editorial u otro	SN Computer Science
Número de páginas	13 páginas
Problema abordado	La problemática que se trata es que en México no existe realmente algún trabajo relacionado con datos etiquetados entorno a la COVID-19 con respecto al número de casos que ha presentado nuestro país y con ello generar algún modelo de aprendizaje supervisado.
Objetivo	Analizar los datos etiquetados arrojados ante la pandemia de la COVID-19 para poder desarrollar un modelo supervisado con los números de casos registrados tanto positivos como negativos. A partir de esto generar resultados predictivos para poder comprender la situación de México en un presente y futuro cercano.



Justificación	<p>En México es prácticamente nulo este tipo de investigaciones y estudios en donde se aprovechan de buena manera los datos etiquetados, a comparación de otros países. Este tipo de análisis y desarrollos ayudan a entender la situación actual del país y generar acciones preventivas ante los cuidados y medidas sanitarias que se pueden tomar. Es por esto que en México, al ser uno de los países más afectado por la pandemia, es importantísimo el poder analizar este tipo de datos, para poder mejorar y dar nuevas medidas sanitarias que ayuden a controlar y mitigar los casos de COVID-19.</p>
Marco teórico	<p>COVID-19: Es una enfermedad infecciosa causada por el virus recientemente descubierto.</p> <p>Datos etiquetados: Los datos para los que ya conoce la respuesta de destino se denominan datos etiquetados. En la ML supervisada, el algoritmo se enseña a sí mismo a aprender de los ejemplos etiquetados que proporcionamos</p> <p>Red Neuronal artificial: Consiste en un conjunto de unidades, llamadas neuronas artificiales, conectadas entre sí para transmitirse señales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones) produciendo unos valores de salida.</p> <p>Máquinas de vectores de soporte: Conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T. Estos métodos están propiamente relacionados con problemas de clasificación y regresión.</p>
Método utilizado	<p>Primordialmente se utilizaron los algoritmos de Naive Bayes, de regresión logística, de árbol de decisión, máquinas de vectores de soporte (SVM) y Red Neuronal artificial (ANN) para generar un aprendizaje supervisado.</p> <p>Inicialmente, dado que es un aprendizaje automático, se necesitaron datos de entrenamiento, para lo cual, de la muestra total de los datos etiquetados, se trabajo con el 80% de estos para ser procesados y así entrenar el algoritmo para posteriormente ser probado de manera general con el 20% de datos restantes de la COVID-19 en México.</p> <p>De esta manera se obtuvo la correlación entre las características que ofrecían los datos inicialmente y con ello realizar la partición de los datos a partir de lo obtenido. Posteriormente pasar a la etapa de pruebas y la de evaluación para saber si en verdad arrojan resultados esperados y sobre todo predictivos para poder tomar acciones ante la COVID-19 en México.</p>



Fuentes de investigación utilizada	<p>La principal fuente de los datos relacionados con la COVID-19 en México, se obtuvo del Sistema de Vigilancia Epidemiológica de Enfermedades Virales Respiratorias teniendo 41 características (o variables) para su análisis.</p> <p>Además, se consultan artículos semejantes de otros países referentes al cómo afrontar un estudio de estas características.</p>
Herramientas utilizadas	<p>Todos los algoritmos de aprendizaje automático supervisados se desarrollaron haciendo uso del lenguaje de programación Python, esto con el sistema operativo Windows dentro de una computadora HP Branded, con procesador Corei5 y con 8 GB de RAM y velocidad de procesador de 2.8 GHz.</p> <p>Las bibliotecas usadas en el proceso se instalaron en Python Notebook.</p>
Resultados alcanzados	<p>Los modelos obtenidos en este artículo, se clasificaron según su precisión, sensibilidad y especificidad, para lo cual los mejores en su rama respectivamente fueron el árbol de decisiones, máquinas de vectores de soporte y el de Naive Bayes.</p> <p>Así mismo, las correlaciones de importancia que se encontraron fueron varias, en la que se destaca la edad y algunas enfermedades, ya que, al aumentar la edad, los pacientes tenían una mayor probabilidad de contagio, y de la misma forma con las enfermedades, a mayor enfermedades o nivel de enfermo (diferente de COVID-19) era más factible que pudieras contraer el virus.</p> <p>Finalmente se resalta la importancia de estos análisis y el desarrollo de este tipo de estudios los cuales ayudan a tomar acciones preventivas ante situaciones realmente dañinas para una población dentro de ciertos límites como lo puede ser un país, y en este caso, México.</p>
Aspectos de interés	<p>A partir de modelos matemáticos es posible el predecir cómo será que se propague el virus entre las personas.</p> <p>Y que, gracias a las correlaciones obtenidas, podemos conocer los factores que realmente perjudican en las personas al contagiarse por COVID-19, y siendo identificados, podemos ofrecer una mejor atención a aquellas personas que cumplan con los perfiles arrojados en el presente estudio según los factores y variables que se encontraron con los datos etiquetados.</p>