



La ciencia de datos en tiempos de la COVID-19



La pandemia por el virus SARS-CoV-2 ha generado un cambio radical en nuestra sociedad. Nuestros hábitos de convivencia, trabajo, interacción social, comercialización y consumo han sido alterados y, quizá, nunca vuelvan a ser los mismos. El último registro de una disrupción similar se remonta a los años 1918 y 1919 con la llamada gripe española. De forma más reciente, en 2009 se generó la pandemia por la influenza AH1N1, con una escala de afectación mucho menor.

Un distintivo de la era marcada por COVID-19 es la alta conectividad de la sociedad a través de servicios como internet, redes celulares, cómputo móvil, almacenamiento en la nube, internet de las cosas y redes sociales. Las comunicaciones de cuarta y quinta generación (4G y 5G) permiten que este intercambio de información sea casi inmediato. Así, en la sociedad actual los datos han llegado a ser el nuevo petróleo, y por su manejo, exploración y aprovechamiento en la toma de decisiones se ha acuñado el término de ciencia de datos. Pero, ¿cómo puede ser la ciencia de datos una herramienta para enfrentar los desafíos asociados con COVID-19?

La ciencia de datos permite generar herramientas de visualización con el fin de analizar —prácticamente en tiempo real— la progresión de la enfermedad de forma mundial, así como sus afectaciones a la población. Una de las herramientas más utilizadas fue desarrollada por el Centro de Ciencia e Ingeniería de Sistemas de la Johns Hopkins University, a través de la cual se visualiza la progresión del número de infecciones, decesos, casos activos y pacientes recuperados (<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>). Otro ejemplo de cómo la ciencia de datos construye herramientas para enfrentar los desafíos por COVID-19 es la iniciativa Nextstrain, que permite analizar la progresión evolutiva mundial de los virus asociados con el SARS-CoV-2 (<https://nextstrain.org/ncov/global>). Estas iniciativas no sólo han sido lideradas y desarrolladas por la comunidad académica internacional, también los gobiernos se han preocupado por difundir el avance local y las afectaciones por COVID-19 por área geográfica, estado e, incluso, municipio. Por ejemplo, la herramienta generada por el Ministerio de Ciencia e Innovación en España (<https://cncovid.isciii.es/covid19/>) o la elaborada por el Gobierno mexicano (<https://datos.covid-19.conacyt.mx/>). A partir de estas herramientas, cualquier usuario puede descargar las bases de datos para un análisis posterior o, incluso, generar gráficas de la evolución de la pandemia.

Otras aplicaciones de la ciencia de datos se han enfocado en construir modelos de predicción que estimen el número de infecciones por SARS-CoV-2 y la mortalidad asociada. Esto permite dimensionar los sistemas de salud y generar asignaciones presupuestales por los gobiernos. Dentro de la ciencia de datos, la inteligencia artificial y el aprendizaje de máquina son técnicas claves para dirigir la toma de decisiones y, en

este contexto, lograr un diagnóstico temprano que permita evitar o contener la propagación de la enfermedad. Así, el aprendizaje de máquina y la inteligencia artificial se han utilizado para el diagnóstico y revisión médica de pacientes a través de técnicas de imagenología, como rayos X y tomografía computarizada de tórax.

En específico, se han utilizado de forma exitosa técnicas de inteligencia artificial para detectar la opacidad en las imágenes de pacientes con COVID-19. También se ha empleado el aprendizaje de máquina para un diagnóstico rápido a partir de índices obtenidos del análisis de sangre. Asimismo, el seguimiento de contactos es fundamental para frenar la evolución de infecciones por SARS-CoV-2, lo cual puede realizarse al emplear información de GPS, enlaces Bluetooth, historial de transacciones con tarjetas de crédito y seguimiento de la ubicación de teléfonos celulares en conjunto con aprendizaje de máquina. Aunque, si bien este tipo de seguimiento puede ser clave, también es importante estar atentos a los posibles riesgos sociales que este monitoreo focalizado pudiese generar.

El desarrollo que ha tenido la ciencia de datos en los últimos 10 años, en conjunto con la disponibilidad inmediata de información y el aumento en el poder de cómputo de los ordenadores han permitido generar diversas herramientas para analizar las causas, la progresión, el desarrollo y la detección de COVID-19. Todo ello ha permitido impulsar una conciencia colectiva del avance de la enfermedad y promover una detección temprana en algunos escenarios. Aunque la aprobación y comercialización de vacunas que ataquen el SARS-CoV-2 permiten visualizar el fin de la pandemia, las herramientas basadas en ciencia de datos han abierto nuevas opciones de visualización, seguimiento y diagnóstico para futuras enfermedades. 

