

Una evaluación de modelos de pronóstico para casos COVID-19 en Cantabria

Alberto Lezcano Lastra
Instituto Cántabro de Estadística
lezcano_a@cantabria.es

Alejandro López Cagigas
Instituto Cántabro de Estadística
lopez_ale@cantabria.es

Francisco Parra Rodríguez
Instituto Cántabro de Estadística
parra_fj@cantabria.es

Gonzalo Llamosas García
Instituto Cántabro de Estadística
llamosas_g@cantabria.es

PALABRAS CLAVE

Machine learning, COVID-19, SARS-CoV-2, Inteligencia artificial, Predicción

RESUMEN

La pandemia del COVID-19 ha supuesto un reto sanitario, económico y social sin precedentes en la historia de la humanidad con más de 170 millones de casos en todo el mundo. Las restricciones sanitarias planteadas por gobiernos de todo el planeta se han mostrado efectivas para controlar los brotes, aunque perjudicando al mismo tiempo la actividad económica y social. Con el proceso de vacunación avanzado, se espera una mejoría epidemiológica que aproxime una normalidad previa a la crisis pandémica.

El creciente número de casos activos por SARS-CoV-2 que ha dado lugar a varios brotes de contagios preocupantes, ha propiciado el interés generalizado tanto en el pronóstico de casos positivos como en el número reproductor de cada contagio. Los resultados académicos llevados a cabo hasta el momento muestran que las técnicas basadas en Inteligencia Artificial (AI) pueden ser una herramienta poderosa para anticipar la evolución de un shock epidemiológico, al menos a corto plazo. Diferentes metodologías de series temporales basadas en “Machine Learning”, por ejemplo, las redes neuronales, los árboles de decisión y las máquinas de soporte de vectores, utilizan sofisticados algoritmos basados en aprendizaje adaptativo que han producido resultados positivos en el pronóstico de casos COVID-19 en comparación con otras metodologías de series temporales más tradicionales como ARIMA.

En este contexto, el Instituto Cántabro de Estadística (ICANE), en colaboración con el Servicio Cántabro de Salud (SCS), ha llevado a cabo un seguimiento diario de casos COVID-19, así como de otros indicadores epidemiológicos, incluyendo el número de reproducción, económicos y demográficos a través de un aplicativo web diseñado específicamente para tal propósito. Al mismo tiempo, aprovechando los datos diarios recogidos en el aplicativo web, se han utilizado diferentes modelos de series temporales basados en aprendizaje adaptativo para realizar proyecciones de infectados por SARS-CoV-2 a 14 días.

El principal objetivo de este trabajo es el de plantear un análisis comparativo de los principales modelos de predicción utilizados por el ICANE (ARIMA, regresión local, red neuronal, bosque aleatorio, etc.) con el propósito de anticipar el incremento de individuos infectados en Cantabria. Para ello se comparan los pronósticos realizados por los diferentes modelos a través de un análisis de métricas de rendimiento, verificando al mismo tiempo el cumplimiento de las hipótesis estadísticas planteadas. Este trabajo muestra además el desarrollo del aplicativo web diseñado por el equipo informático del ICANE, así como la experiencia diaria en la recogida, procesamiento y actualización de datos del SCS para la construcción de indicadores COVID-19.