esade

EsadeEcPol - Center
for Economic Policy

Open Internet Governance Institute

PAPER SERIES # 3

Ciencia de datos para el bien social: el ejemplo valenciano durante la pandemia de COVID-19

Nuria Oliver, PhD

Cofundadora y directora de la Fundación Unidad ELLIS Alicante, también conocida como Instituto para el Desarrollo de la Inteligencia Artificial centrada en la Humanidad

JUNIO, 2022

Sobre el autor

Nuria Oliver, PhD
Cofundadora y directora de la Fundación Unidad ELLIS Alicante, también conocida como Instituto para el Desarrollo de la Inteligencia Artificial centrada en la Humanidad

Resumen

La ciencia de datos es una disciplina de enorme valor en el sector público por al menos dos razones esenciales

- → permite diseñar políticas públicas basadas en datos y evidencias, frente a intuiciones, información obsoleta o intereses políticos.
- → permite la necesaria evaluación empírica del impacto de las políticas públicas desplegadas, para determinar sus fortalezas y debilidades.

A pesar de su potencial, el uso sistemático de datos para apoyar la elaboración de políticas públicas es todavía poco frecuente en la mayor parte del mundo. En este artículo, ofrezco un resumen de un ejemplo internacionalmente reconocido de aprovechamiento de datos para apoyar la elaboración de políticas durante la pandemia de COVID-19 en España. Presento el trabajo que realizamos entre marzo de 2020 y marzo de 2022 en el "Data Science against COVID-19 Taskforce", una experiencia pionera en la Comunidad Valenciana. Este grupo de trabajo, que yo lideré, estaba compuesto por un equipo multidisciplinar de más de 20 científicos de varias universidades y centros de investigación de la Comunidad Valenciana de España, trabajando estrechamente con los responsables políticos de la región a nivel de Presidencia.

Nos centramos en trabajar en cuatro áreas de impacto para el uso de la ciencia de datos en la lucha contra la pandemia de coronavirus:

(1) modelización de la movilidad humana agregada; lo que nos permitió

- controlar el impacto de las medidas de confinamiento en la movilidad real de los ciudadanos
- identificar las zonas en las que las medidas de confinamiento tuvieron un mayor o menor impacto
- cuantificar el éxito de la campaña #stayathome, así como las medidas adoptadas para restringir la movilidad nocturna y los cierres perimetrales de la Comunidad Valenciana
- modelar el impacto que la reducción de la movilidad tuvo en la progresión del coronavirus
- predecir mejor la propagación y determinar la utilidad de los cierres selectivos
- seguir la vuelta progresiva a la normalidad

(2) desarrollo de modelos epidemiológicos computacionales, dedicados a realizar predicciones sobre la evolución de la pandemia no sólo bajo las condiciones actuales, sino también bajo diferentes escenarios de confinamiento y medidas de control de la pandemia, inmunidad preexistente, vacunación o rastreo de contactos.

(3) desarrollo de modelos predictivos a nivel subregional (departamento de salud) para estimar las variables relevantes para orientar las decisiones políticas.

(4) a aplicación de la ciencia ciudadana a través de una encuesta en línea a gran escala denominada COVID19Impactsurvey, una herramienta clave para arrojar luz sobre cuestiones de gran importancia pública para las que no existen datos recogidos de forma sistemática; por ejemplo, el papel de los contactos cercanos, una primera evaluación del impacto psicológico o la diferencia en la capacidad de seguir los cierres en función de las condiciones sociodemográficas o socioeconómicas.

Este esfuerzo es un ejemplo inspirador de la estrecha colaboración entre la comunidad científica, los ciudadanos y una administración pública para garantizar que las políticas públicas se basen efectivamente en las pruebas y los conocimientos de los expertos.

Introducción

En este artículo, resumo el trabajo llevado a cabo en el Grupo de Trabajo de Ciencias de Datos contra la COVID-19. Este grupo de trabajo estuvo formado por un equipo multidisciplinar de investigadores de la Comunidad Valenciana, que trabajaron muy estrechamente con la Dirección General de Análisis y Políticas Públicas de la Presidencia de la Generalitat Valenciana durante la pandemia de COVID-19 (entre marzo de 2020 y abril de 2022).

El grupo de trabajo lo integraron los siguientes científicos: Alberto Conejero, Miguel Rebollo, Manuel Portolés, Víctor de Elena, Miguel Ángel García-March, Óscar Garibo y Eloy Piñol de la Universitat Politècnica de València; Francisco Escolano, Miguel Ángel Lozano, Juan Carlos Trujillo y Miguel Ángel Teruel de la Universidad de Alicante; Francisco Falcó de la Universidad CEU Cardenal Herrera; Alejandro Rabasa, Aurora Mula, Xavier Barber, Kristina Polotskaya y Elisa Espín de la Universidad Miguel Hernández; Joaquín Huerta, Marina Martínez, Emilio Sansano, Juan Camilo Gómez y Rubén Femenía de la Universitat Jaume I y Adolfo López de FISABIO.

La configuración, la forma de trabajar y los resultados de este equipo son únicos no solo a escala nacional, sino también internacional, como refleja la visibilidad obtenida en medios de comunicación tanto nacionales [1,2,3] como internacionales [4,5].

Además de haber tenido un impacto directo en el apoyo a la elaboración de políticas durante la pandemia de COVID-19 y de haber realizado contribuciones científicas significativas, esperamos que la experiencia descrita en este artículo motive a otros Gobiernos y Administraciones públicas, tanto a nivel nacional como internacional, para hacer una transición hacia una toma de decisiones y una evaluación de políticas públicas basadas en la evidencia.

Nuestro trabajo se dividió en 4 áreas, descritas a continuación y representadas en la figura 1: (1) modelado de la movilidad humana agregada; (2) modelos epidemiológicos computacionales; (3) modelos predictivos; y (4) ciencia ciudadana.

Figura 1. Áreas de trabajo del Grupo de Trabajo de Ciencia de Datos contra la COVID-19.]



1.
Análisis de la movilidad humana agregada

La primera línea de trabajo se centró en el análisis de la movilidad humana. La movilidad humana es clave para la propagación de las enfermedades infecciosas. En la bibliografía existente, hay ejemplos del uso de la movilidad agregada y anonimizada de la red de telefonía móvil para ayudar a combatir enfermedades como el ébola [6], el zika [7] y la malaria [8].

Cuando una enfermedad infecciosa que se transmite entre personas se encuentra en una fase de transmisión comunitaria, la contención de la movilidad humana es una de las intervenciones no farmacológicas (INF) más implementadas para limitar la propagación de la enfermedad. En Europa, la mayoría de los países limitaron la movilidad de su población, en mayor o menor grado, durante las primeras semanas de la pandemia de COVID-19 (desde mediados de marzo hasta mayo de 2020) y, posteriormente, en respuesta a aumentos significativos de la incidencia de la enfermedad durante todo el periodo 2020-2022.

En España se declaró el estado de alarma el 14 de marzo de 2020 para aplicar estas medidas de contención de la movilidad, que comenzaron el 16 de marzo y acabaron cuando se volvió a la «nueva normalidad», el 21 de junio de 2020.

Nuestro trabajo de análisis de la movilidad en el contexto de la primera ola de la pandemia de COVID-19 (marzo-junio de 2020) fue un proyecto piloto en la Comunidad Valenciana, como declaró la vicepresidenta Nadia Calviño el 23 de marzo de 2020.

Durante la primera ola de contagios por SARS-CoV-2, empezamos a analizar datos agregados y anonimizados extraídos de la red de telefonía móvil y compartidos por el Instituto Nacional de Estadística (INE), gracias a un acuerdo de colaboración entre el INE y los tres mayores operadores de telefonía móvil de España (Telefónica, Vodafone y Orange), correspondientes al periodo entre el 16 de marzo de 2020 y el 30 de junio de 2020 [9]. El proyecto cumple con las condiciones de anonimización que establece el Reglamento General de Protección de Datos, tal como las describe el INE [10]. Los datos de movilidad compartidos por el INE se hicieron posteriormente públicos en su página web.[1]

También analizamos los datos de movilidad compartidos por el Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA)[2] y los datos de movilidad agregados y anonimizados compartidos por Facebook y Google.

1 https://www.ine.es/covid/covid_movilidad.htm

2 https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad

Un análisis cuantitativo de la movilidad humana a gran escala permite:

(a) monitorizar el impacto de las medidas de contención en la movilidad real de los ciudadanos;

(b) identificar las zonas en las que las medidas de confinamiento tuvieron un impacto mayor o menor, y monitorizar su comportamiento a lo largo del tiempo;

(c) cuantificar el éxito de la campaña #quédateencasa, así como de las medidas adoptadas para restringir la movilidad nocturna y de los cierres perimetrales en la Comunidad Valenciana;

(d) modelar el impacto que tuvo la reducción de la movilidad en la progresión del coronavirus;

(e) identificar comunidades basadas en la movilidad de la población para predecir mejor la propagación del coronavirus y determinar la utilidad de posibles cierres selectivos, si se produjeran nuevos brotes posteriores al confinamiento.

(f) cuantificar el regreso progresivo a los niveles de movilidad previos a la pandemia a medida que se levantan las medidas de confinamiento.

El volumen de la información sobre los datos de movilidad fue uno de los primeros retos que tuvimos que abordar. Para procesar, visualizar y analizar datos de una magnitud tan grande, utilizamos diferentes tecnologías de *big data* (representadas en la figura 2) tanto en el backend como en el *frontend* (visualización).

Figura 2. Soluciones tecnológicas utilizadas en el área de trabajo sobre análisis de la movilidad.



Permitir una visualización intuitiva de datos tan complejos es vital para apoyar la elaboración de medidas políticas. En consecuencia, desarrollamos varias visualizaciones utilizando diferentes herramientas, según el público objetivo al que queríamos presentar los resultados. Desarrollamos diferentes *dashboards* que hicieron más fácil que diferentes perfiles entendieran una información que, debido a su complejidad, es difícil de presentar de otra manera, como se ilustra en la figura 3.

Figura 3. Dashboard de visualización del análisis de la movilidad.



En las siguientes secciones, resumimos las principales conclusiones de nuestro análisis sobre el impacto que tuvieron en la movilidad las diferentes medidas de contención implementadas durante la pandemia.

Cumplimiento de la campaña «Quédate en casa» durante el confinamiento de la primera ola (marzo-mayo de 2020)

Basándonos en los datos del INE, cuantificamos la movilidad diaria real de los ciudadanos desde el inicio de la pandemia y la comparamos con datos de movilidad de referencia proporcionados también por el INE para un día "normal" o *baseline* anterior al comienzo de la pandemia (noviembre de 2019).

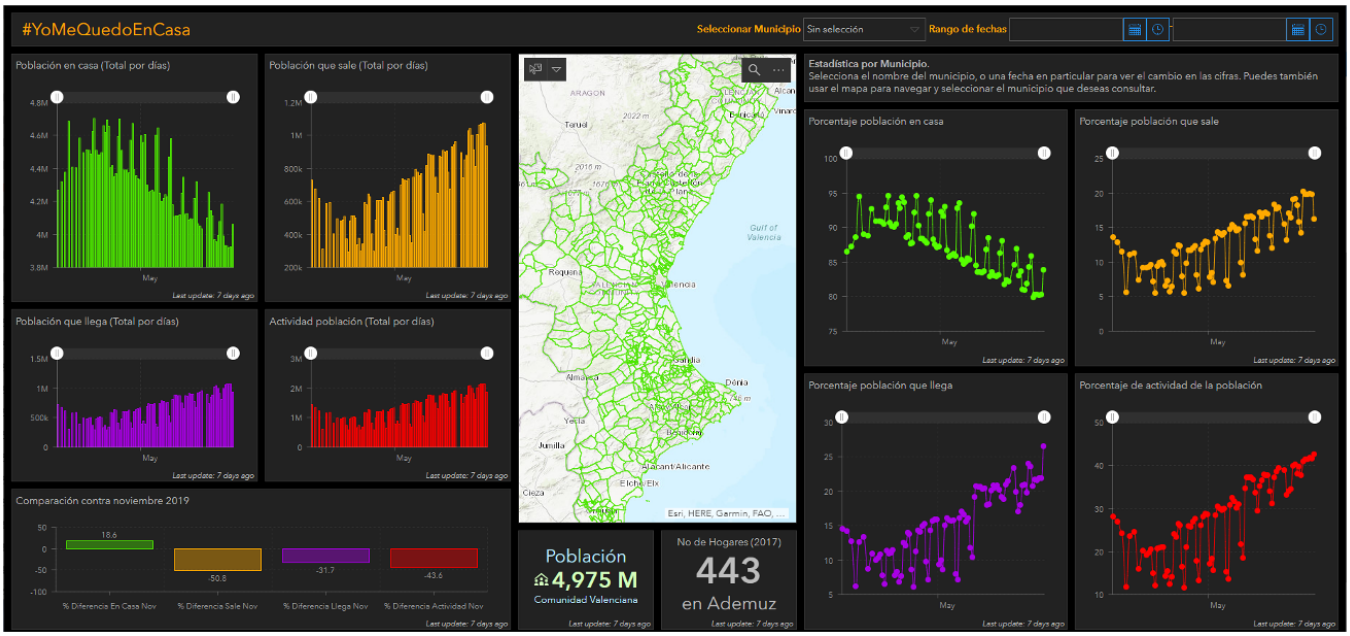
Hallamos un alto grado de cumplimiento de la campaña «Quédate en casa». En el periodo comprendido entre el 16 de marzo y el 27 de abril de 2020, de media, el 88% (días laborables) y el 92% (fines de semana y festivos) de la población de la Comunidad Valenciana permaneció en su zona de residencia, lo que demuestra un alto cumplimiento de las medidas de confinamiento.

Movilidad asociada al trabajo durante el confinamiento de la primera ola (marzo-mayo de 2020)

También estudiamos los niveles de movilidad laboral en comparación con un periodo laboral de noviembre previo a la COVID-19. De media, entre el 16 de marzo y el 27 de abril, observamos una reducción del 59% en el número de personas que pasaron al menos 2 horas fuera de su zona de residencia durante el horario laboral.

La figura 4 muestra una visualización de los datos y el análisis de movilidad que hemos realizado. Puede encontrarse un resumen de este trabajo sobre la movilidad en dos informes publicados por la Generalitat Valenciana en abril [11] y mayo [12] de 2020.

Figura 4. Visualizador de los datos de movilidad (http://t.ly/9riC)



Comunidades de movilidad

Además del análisis del flujo de movimientos en sí, los datos de movilidad son la entrada para un módulo de análisis de grafos. Los grafos son estructuras que están presentes en numerosos fenómenos naturales y artificiales y se han utilizado para modelar, por ejemplo, los flujos migratorios o el tráfico aéreo. Nosotros usamos esos flujos para identificar patrones en la movilidad de las personas. Este método nos permite detectar qué zonas están interrelacionadas, con independencia de los límites administrativos que dividen el territorio. Para hacer esto, efectuamos la detección automática de las comunidades que surgen de los flujos utilizando el algoritmo propuesto por Newman [13].

Una *comunidad de movilidad* está compuesta por zonas geográficas con una alta densidad de movimientos internos dentro de ellas, y muy pocos movimientos con las demás. Del análisis de las comunidades basado en los datos de movilidad descritos antes, y de su evolución durante el confinamiento, obtuvimos las siguientes conclusiones:

→ Identificamos una estructura clara de comunidades que es estable y no coincide exactamente con los límites administrativos. Desde el punto de vista de la movilidad, existe un gran flujo de movimiento de población entre el norte de Castellón y el sur de Tarragona, así como en las localidades limítrofes entre Murcia y Alicante. Se observa el mismo fenómeno en la movilidad entre provincias. Por esta razón, sería necesario coordinar las políticas de contención entre las diferentes Comunidades Autónomas del país.

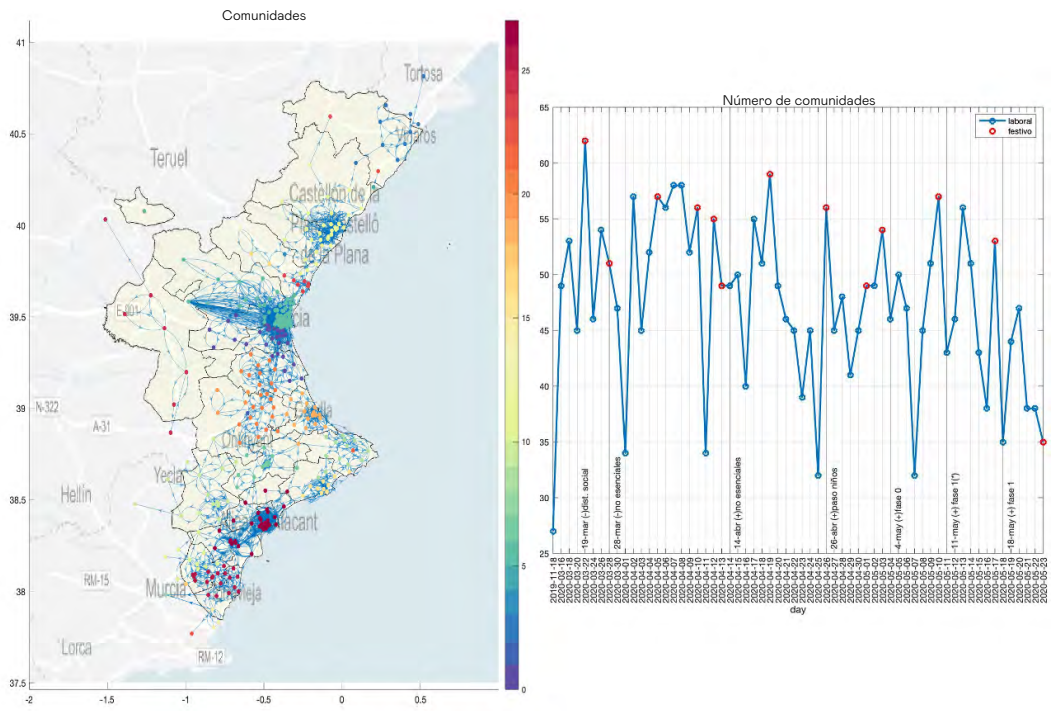
→ El número de comunidades aumentó a medida que se ampliaron las restricciones de movilidad, pasando de 27 comunidades en la semana de referencia de noviembre de 2019 a entre 50 y 60 durante el

confinamiento de la primera ola de contagios (véase la parte derecha de la figura 3). Luego observamos un descenso gradual del número de comunidades, con picos durante los días festivos comparables al periodo de confinamiento más estricto. Durante este tiempo, el número de comunidades aumentó al menos un 50% en comparación con la semana de referencia. El hecho de tener muchas comunidades pequeñas e independientes ayudó a contener la propagación del SARS-CoV-2. Esta misma reducción de tamaño se observó también durante los festivos y los fines de semana.

→ La división en comunidades es coherente con la división en zonas de salud de la comunidad autónoma. Por lo tanto, los datos respaldan la toma de decisiones basada en las zonas de salud.

Este análisis de comunidades basado en los flujos de movilidad es importante para apoyar la toma de decisiones sobre el confinamiento selectivo de determinadas zonas geográficas en función de su situación epidemiológica. Cuanto más autocontenida sea la movilidad de una zona geográfica, menor será el impacto epidemiológico de un confinamiento en esa zona, ya que la mayor parte de su movilidad es interna, no con origen/destino en otras zonas geográficas.

Figura 5. División en comunidades de movilidad (izquierda) y evolución de su tamaño a lo largo del tiempo (derecha)



Vuelta a la movilidad

Cuando empezaron a levantarse las medidas de confinamiento, los niveles de movilidad de la población aumentaron. Durante 2021, analizamos los datos agregados de movilidad compartidos por Facebook y Google para monitorear ese aumento de los niveles de movilidad. Observamos un aumento significativo —hasta niveles incluso superiores a los previos a la pandemia— en las actividades relacionadas con espacios al aire libre, el ocio y los parques, lo que probablemente refleja el deseo de la población de estar al aire libre después de meses de confinamiento.

2.
Modelos epidemiológicos computacionales

En esta área de trabajo desarrollamos tres tipos de modelos epidemiológicos computacionales: un modelo metapoblacional basado en el SEIR, un modelo individual basado en agentes y un modelo basado en redes neuronales profundas.

Los modelos epidemiológicos nos permiten hacer predicciones sobre la evolución de la pandemia no solo bajo las condiciones actuales, sino también en diferentes escenarios de confinamiento y de medidas de control de la pandemia, de inmunidad preexistente, vacunación o rastreo de contactos.

Modelo metapoblacional basado en el SEIR

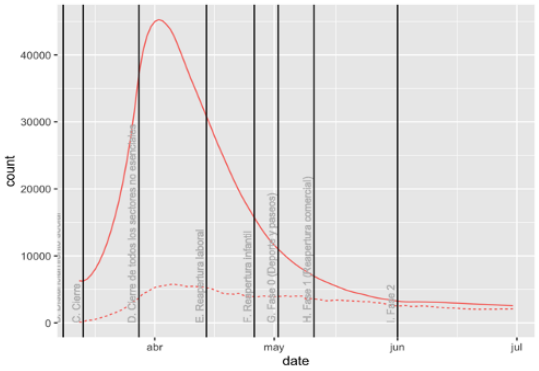
El modelo de metapoblación que desarrollamos se basó en un modelo SEIR clásico [14], en el que cada letra del acrónimo representa un estado diferente en el que pueden encontrarse los miembros de una población: S representa el número de individuos susceptibles de contraer la enfermedad, E el número de individuos de la población expuestos a contraer la enfermedad, I denota el número de individuos infectados que pueden transmitir la enfermedad y, por último, R representa el número de individuos que se retiran del sistema, bien porque se han recuperado o han muerto. Los parámetros del modelo representan las tasas de transferencia de individuos entre las cuatro subpoblaciones diferentes en que se ha dividido la población, de modo que podemos determinar cuántos individuos componen cada clase en cualquier instante.

Este modelo determinista, introducido por Joan L. Aron e Ira B. Schwartz [14] en 1984, se presenta como un paradigma científico del comportamiento dinámico de los diferentes grupos en los que dividimos la población. Desarrollamos una extensión estocástica de este modelo determinista con el fin de introducir la incertidumbre en el modelo original.

La figura 6 muestra las estimaciones realizadas por el modelo SEIR comparadas con los datos reales de los casos notificados en la Comunidad Valenciana. Obsérvese que un alto porcentaje de los casos de COVID-19 fueron asintomáticos o tuvieron síntomas leves. Esto significa que, sobre todo en la fase inicial de la pandemia, cuando la disponibilidad de test era limitada, hubo un gran número de casos no notificados. El modelo nos permite estimar el número subyacente de contagios totales y compararlo con el número de casos notificados. Al hacerlo, se observa cómo ha ido aumentado progresivamente la tasa de casos notificados, desde alrededor del 10% al principio de la pandemia hasta casi todos los casos en el mes de julio de 2020 (véase la figura 6, a), cuando ya había disponibles suficientes test de COVID-19.

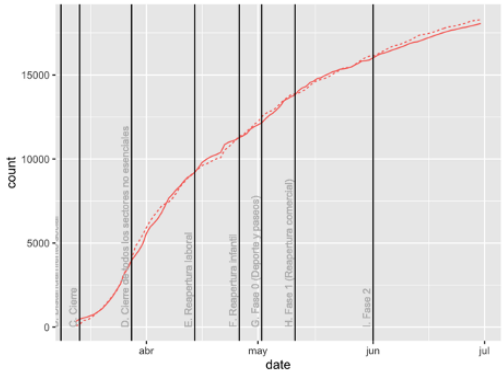
Figura 6.

a



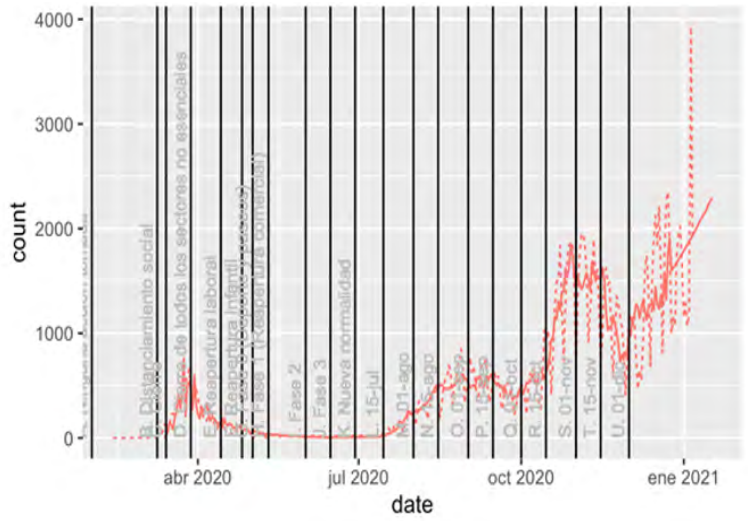
Comparación entre los casos activos subyacentes predichos por el modelo y los casos activos notificados durante la primera ola (marzo-mayo de 2020). Obsérvese el alto porcentaje de casos subyacentes no notificados (debido a la falta de test, porque eran asintomáticos, etc.)

b



Total de casos acumulados que el modelo estima que habrían sido notificados, y total de casos acumulados notificados en la serie de datos reales.

c



Predicciones del modelo SEIR (línea continua) y datos reales observados (línea discontinua) en la Comunidad Valenciana.

Modelo individual basado en agentes y el impacto del rastreo de contactos

Una de las medidas más eficaces para reducir la tasa de contagio es el rastreo de contactos. Esta técnica se basa en la detección precoz del mayor número posible de casos positivos con el fin de aislarlos y cortar así la cadena de transmisión. Los casos positivos detectados se aíslan y se identifica a quienes han mantenido un contacto estrecho con ellos. Lo ideal es que los contactos estrechos en riesgo de haberse contagiado

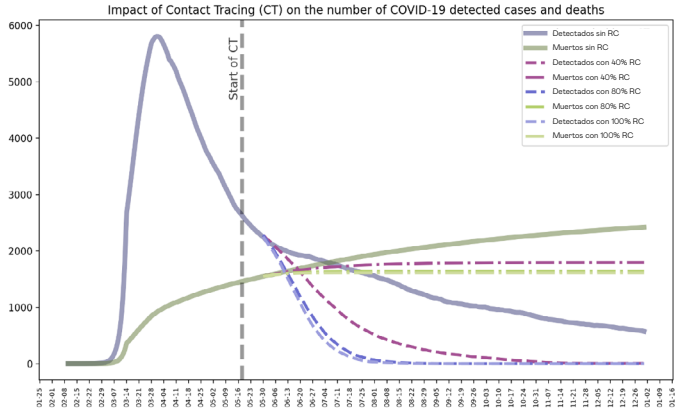
se hagan los test necesarios para determinar si lo están y se aíslen hasta que conozcan los resultados de los test. A todos los contactos estrechos que dan positivo en infección por coronavirus se les pide que se aíslen y se identifica a sus contactos estrechos, en un intento de romper la cadena de transmisión lo antes posible.

En nuestro estudio, modelamos la interacción entre la capacidad existente para realizar un rastreo efectivo de los contactos y la curva epidemiológica resultante basada en esa capacidad, con el fin de estimar diferentes escenarios en la evolución de la pandemia. Simulamos una serie de escenarios que contemplaban diferentes contextos en cuanto a la eficacia del rastreo de contactos a partir de mediados de mayo de 2020, momento en el que las medidas de confinamiento empezaron a relajarse, dando paso a una desescalada gradual.

Los datos usados para las simulaciones o bien estaban disponibles públicamente o los compartió con nosotros la Generalitat Valenciana. Las estimaciones se hicieron utilizando el modelo REINA [16], que adaptamos a las características de la Comunidad Valenciana. Se trata de un modelo epidemiológico individual basado en agentes que modela la dinámica de la población a través de las interacciones sociales entre individuos. Una de las características distintivas de este tipo de modelo es que los agentes no son entidades homogéneas, sino que se crean con ciertos rasgos que representan el comportamiento de diferentes poblaciones. A continuación, se ejecuta la simulación mediante una serie de reglas e interacciones entre los agentes que intentan reflejar las características de la población que se estudia. Así, la propagación de la epidemia no es la misma para todos los agentes, sino que depende de sus características y sus interacciones con otros agentes. Por ejemplo, un agente de edad que se contagie tendrá más probabilidades de necesitar servicios médicos que un agente más joven. En general, los modelos basados en agentes requieren una cantidad considerable de información demográfica, sociológica y conductual para desarrollar interacciones realistas. Asimismo, suelen ser costosos desde un punto de vista computacional, porque requieren hacer simulaciones del comportamiento individual de millones de agentes (en nuestro caso, alrededor de 5 millones de habitantes de la Comunidad Valenciana).

En cuanto al rastreo de contactos, los posibles escenarios van desde su inexistencia hasta un rastreo de contactos del 100%, lo que implicaría que todos los contactos de cada caso positivo se detectaran y aislaran por completo durante dos semanas. Como era de esperar (véase la figura 14), este último escenario, aunque imposible de llevar a cabo en el mundo real, conlleva una rápida reducción del número de casos detectados. Los resultados de nuestra simulación muestran que un rastreo de contactos cuya eficacia sea del 40% reduce considerablemente el número de casos y, en consecuencia, el número de hospitalizaciones y muertes.

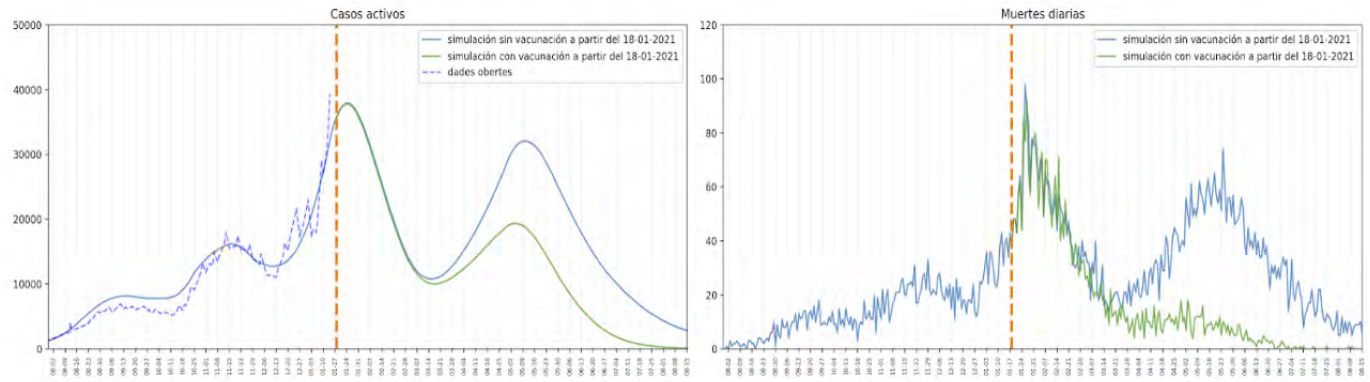
Figura 7.
Predicciones de un modelo basado en agentes de acuerdo con diferentes escenarios de rastreo de contactos. Las líneas continuas representan el número de casos detectados (azul) y de muertes (verde) si no se implementa ningún tipo de rastreo de contactos. Las líneas discontinuas representan los diferentes escenarios de rastreo de contactos simulados (40%, 80% y 100%).



Tras la llegada de las vacunas contra el SARS-CoV-2, en la primavera de 2021 ampliamos el modelo REINA para incluir diferentes escenarios de vacunación y su impacto en el número esperado de casos de COVID-19, hospitalizaciones y muertes.

La figura 8 representa las predicciones que realizó este modelo a partir de enero de 2021. Obsérvese cómo el modelo predice correctamente la cuarta ola de contagios que tuvo lugar en el verano de 2021, con un impacto significativamente menor en cuanto a muertes (gráfico de la derecha, curva verde) que si no hubiera habido ninguna vacunación.

Figura 8. Predicciones del modelo basado en agentes que incluye la vacunación.

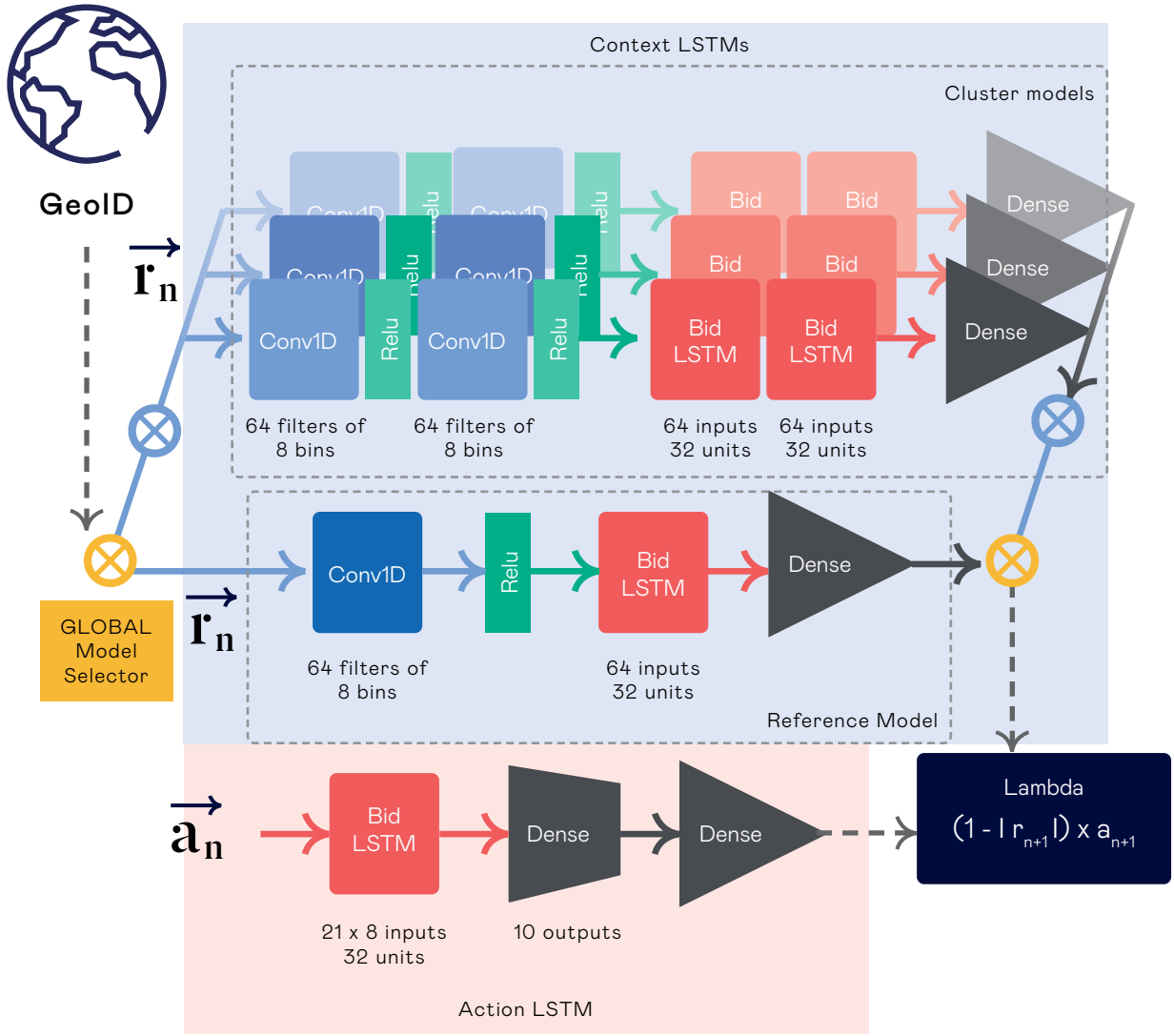


Modelo basado en redes neuronales profundas

El tercer enfoque metodológico que desarrollamos para modelar la progresión de la pandemia se basa en redes neuronales profundas. El objetivo principal es contar con un modelo predictivo del número de casos que tenga en cuenta las intervenciones no farmacéuticas (INF) aplicadas en cada país/región. Este modelo se desarrolló en el contexto del concurso mundial XPRIZE Pandemic Response Challenge, para el que tuvimos que crear modelos epidemiológicos computacionales para 236 regiones/países del mundo.

La figura 9 muestra la arquitectura de este modelo. Consiste en 2 LSTM entrenadas en paralelo, cuyos *outputs* se combinan en un módulo Lambda. La primera LSTM se llama *LSTM de contexto* y modela el número de casos a lo largo del tiempo. La segunda LSTM se llama *LSTM de acción* y modela la serie temporal de intervenciones no farmacológicas (INF) implementadas en cada momento para contener la pandemia. Desarrollamos un banco de 9 LSTM de contexto aplicadas a diferentes países/regiones del mundo.

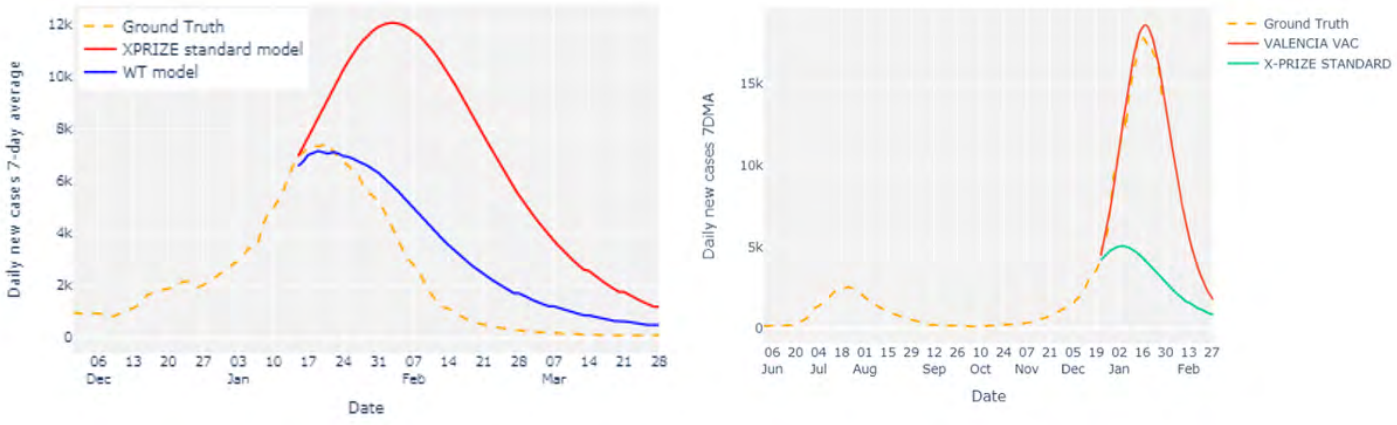
Figura 9. Modelo epidemiológico basado en redes neuronales profundas



El modelo que acabo de describir fue parte de la solución que presentamos al concurso XPRIZE. Nuestro equipo, ValencIA4COVID, fue declarado ganador mundial del reto XPRIZE. Fue la primera vez que un equipo español ganaba un concurso XPRIZE. Los detalles técnicos de nuestro modelo se describen en esta publicación [25], ganadora del Premio al Mejor Artículo sobre Ciencia de Datos Aplicada en el ECML-PKDD 2021.

Este modelo se utilizó ampliamente para hacer predicciones diarias de los casos de COVID-19 en la Comunidad Valenciana a partir de finales de diciembre de 2020. La figura 10 ilustra las predicciones que hizo el modelo antes de la tercera y la sexta olas de casos de COVID-19 en enero de 2021 y de 2022, respectivamente. Obsérvese lo precisas que son las predicciones cuando se comparan con los datos reales (la línea de puntos amarilla en las figuras).

Figura 10. Predicciones de los casos de COVID-19 según el modelo epidemiológico basado en redes neuronales profundas para la Comunidad Valenciana.

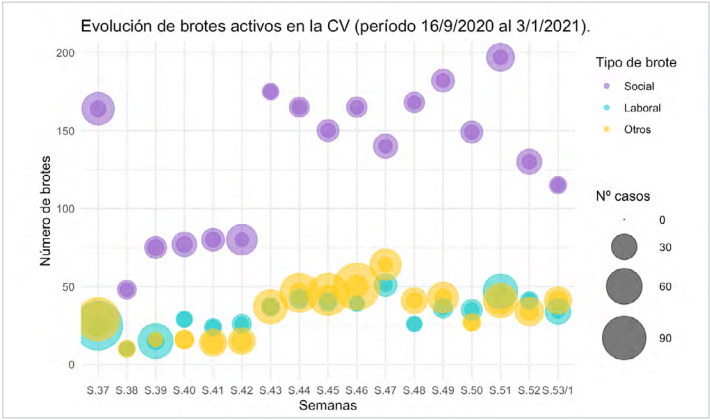


Modelo de evolución de los brotes epidémicos

En septiembre de 2020, gracias a una mayor disponibilidad de datos, fuimos capaces de estudiar las dinámicas de los brotes de SARS-CoV-2 identificados y de analizar sus características por lo que respecta a su origen. Esta información es de interés para poder evaluar a posteriori si las INF implementadas a escala local influyen en la contención de la enfermedad. Después de un primer análisis de los brotes, observamos que seguían un patrón basado en distribuciones libres de escala, lo cual es esperable dado que esos modelos se apoyan en la movilidad humana [18]. Después de este análisis, establecimos como medida de control el tamaño máximo del 90% de los brotes de cada uno de los siguientes tipos: social, laboral y otros tipos.

La figura 11 ilustra la evolución temporal del tamaño y el número de brotes de tres tipos diferentes (social, laboral y otros) en la Comunidad Valenciana entre septiembre y diciembre de 2020. Como se ve en la figura, en esta comunidad el tipo de brote más común fue el de carácter social. Observamos una reducción del número y el tamaño de los brotes de origen social a partir de finales de 2020, lo que podría ser indicativo de un aumento de la transmisión comunitaria y de un descenso del número de casos ligados a brotes, dado que la tercera ola de contagios empezó en ese momento.

Figura 11.
Evolución semanal de los brotes de COVID-19.
El círculo exterior indica el tamaño del mayor brote de ese tipo y el círculo interior el tamaño del percentil 90 de los brotes de ese tipo.



Este análisis nos permite determinar si la evolución temporal de los brotes activos se corresponde con la progresión de la enfermedad observada o si una parte significativa de los casos nuevos tienen su origen en la transmisión comunitaria y, por lo tanto, no se clasifican como brotes.

3.
Modelos predictivos

Modelo predictivo del número de casos positivos por departamento de salud

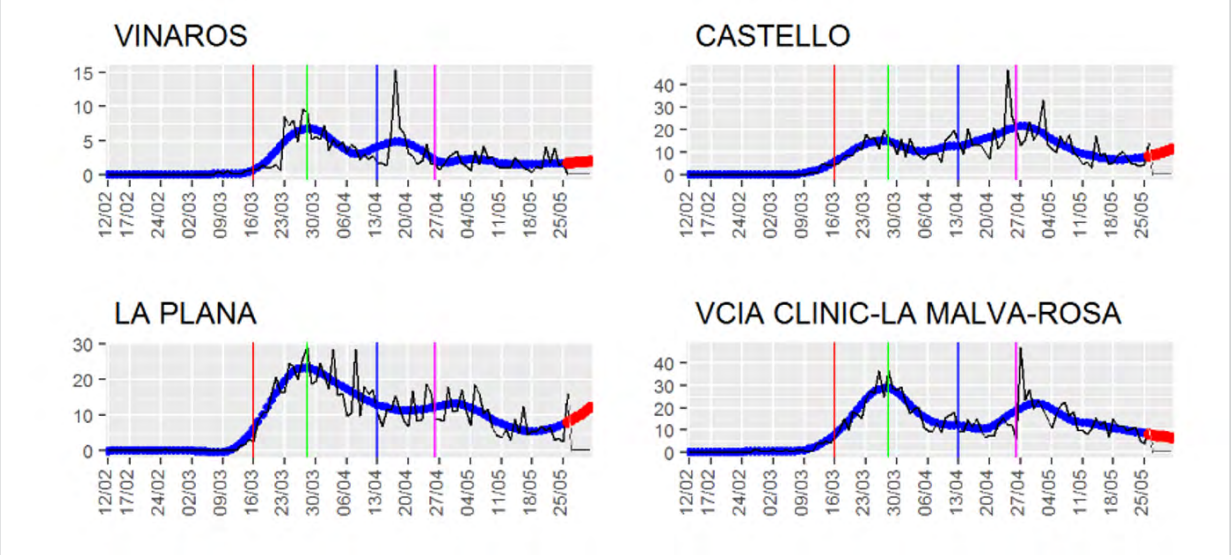
En esta área de trabajo, desarrollamos modelos predictivos para estimar las siguientes variables de relevancia cotidiana para cada uno de los 24 departamentos de salud de la Comunidad Valenciana: el número de casos de COVID-19, el número de hospitalizaciones, los casos activos, las personas ingresadas en cuidados intensivos y los fallecidos en un horizonte temporal de 5 días.

Para hacer predicciones de los casos de COVID-19 utilizamos un modelo de suavizado basado en un método de regresión no paramétrica ajustado al vecindario en una ventana temporal de 5 días para detectar tendencias y ciclos. Los modelos predictivos a corto plazo son modelos tipo ARIMA, que tienen diferentes estructuras para cada departamento de salud, y modelos GAM que utilizan como predictores las observaciones históricas de cada departamento de salud.

Aplicamos diariamente estos modelos a las series temporales del número de casos de COVID-19, el número de hospitalizaciones, tanto en el hospital como en las unidades de cuidados intensivos, y los fallecidos.

La figura 12 ilustra las predicciones para cuatro departamentos de salud de la Comunidad Valenciana a partir del 20 de mayo de 2020. La línea continua negra representa el número de casos de COVID-19, los puntos azules corresponden a los datos suavizados al pasado, los puntos rojos son las predicciones de nuestro método. Las líneas verticales indican las diferentes fases de la desescalada: la línea roja, la fecha de inicio del estado de alarma (16/03/2020); la línea verde, 10 días después del estado de alarma (26/03/2020); la línea azul, el día en que se levantó parcialmente el confinamiento (13/04/2020); la línea morada indica el inicio de la desescalada de las medidas de confinamiento.

Figura 12. Predicciones de casos de COVID-19 por departamento de salud
(ejemplo ilustrativo para 4 departamentos de salud)



Los modelos epidemiológicos computacionales que predicen el número de casos de COVID-19 son, sin duda, de enorme importancia. Sin embargo, los modelos predictivos de otras variables —en especial, el número de hospitalizaciones y de unidades de cuidados intensivos— y con otras escalas geográficas —por ejemplo, a nivel de departamento de salud— son de vital importancia para apoyar la elaboración de políticas públicas. Estos modelos permiten a las autoridades sanitarias preparar recursos para evitar un colapso del sistema. La metodología que desarrollamos para construir este tipo de modelos es diferente de los modelos epidemiológicos computacionales tradicionales, como el modelo SEIR descrito antes.

Por lo tanto, desarrollamos tres familias de modelos para predecir el número de **hospitalizaciones** y el uso de las unidades de **cuidados intensivos** en cada uno de los 24 departamentos de salud de la Comunidad Valenciana.

Modelo espaciotemporal del número de hospitalizaciones diarias por departamento de salud

Para predecir el número de hospitalizaciones, los modelos lineales bajo un supuesto de normalidad no son apropiados: los modelos lineales generalizados con distribuciones Poisson, binomial o binomial negativa son más adecuados, dada la naturaleza y las características del fenómeno subyacente.

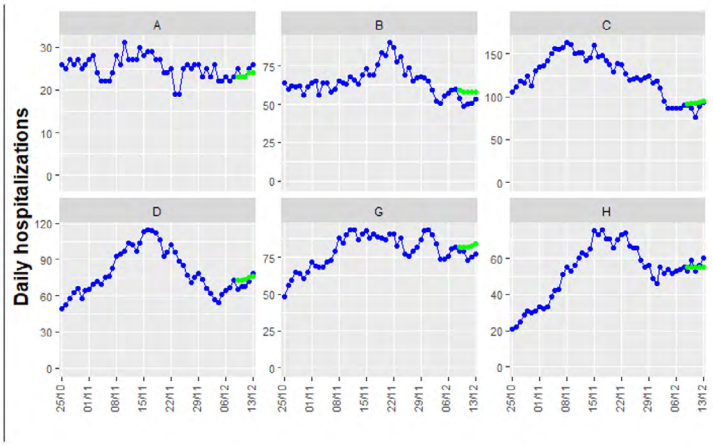
Como hemos descrito antes, la movilidad humana da lugar a una dinámica espaciotemporal en la propagación de la enfermedad infecciosa que también debe considerarse en el modelado. Esta es la razón por la que son necesarios los modelos espaciotemporales.

En nuestro caso, cada zona espacial es la zona geográfica que atiende un departamento de salud. Cada departamento de salud D tiene K vecinos, que son los departamentos de salud adyacentes. En nuestro modelo, suponemos que los casos y las hospitalizaciones observados en D no solo dependen de lo que ocurre en D, sino también de lo que sucede en sus K vecinos, dada la movilidad entre los departamentos de salud. Un reto habitual cuando se modelan este tipo de datos es la autocorrelación, es decir, que las observaciones en áreas geográficamente próximas y en períodos temporalmente cercanos tienden a tener valores más parecidos que las observaciones en zonas y períodos más separados.

El modelo que desarrollamos tiene en cuenta este factor para predecir el número de hospitalizaciones que se producirán en los próximos 5 días en cada departamento de salud.

Figura 13.

Predicciones de hospitalizaciones diarias por departamento de salud (datos reales en azul; predicciones a 5 días en verde). Ejemplo ilustrativo para 6 departamentos de salud.



Modelo de predicción del porcentaje de ocupación de las ucis a 7 días por departamento de salud

A partir de mediados de marzo de 2020, a medida que aumentaba el número de casos de COVID-19, también lo hizo el número de pacientes hospitalizados, lo que llevó a unos niveles de ocupación hospitalaria sin precedentes, próximos a la saturación. Por lo tanto, tener acceso a predicciones fiables y precisas de la ocupación de los cuidados intensivos era de gran importancia para optimizar la gestión de los departamentos de salud.

Para predecir la ocupación futura de las ucis, desarrollamos un modelo basado en redes neuronales profundas.

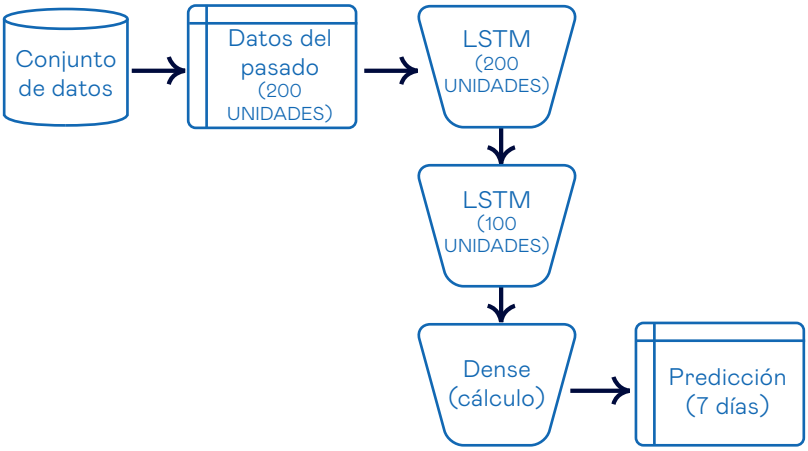
El objetivo es predecir, con 7 días de antelación, el porcentaje de ocupación de las camas de las unidades de cuidados intensivos en cada departamento de salud de la Comunidad Valenciana. La información utilizada para hacer esa predicción, además del historial de ocupación de las ucis, es el número de casos de COVID-19 y de hospitalizaciones por coronavirus en cada departamento de salud.

Modelo basado en redes neuronales profundas

Utilizamos redes neuronales recurrentes (LSTM) para predecir el porcentaje de ocupación de las ucis. Estas redes integran bucles de retroalimentación y son muy apropiadas para modelar series temporales.

La figura 14 muestra la arquitectura del modelo basado en LSTM. Como puede verse, se trata de un modelo LSTM apilado con dos capas, de 200 y 100 neuronas LST respectivamente. La predicción a 5 días es dada por la tercera capa, formada por una única neurona de salida. En total, este modelo tendrá 287-701 parámetros que se entrenan con datos del pasado.

Figura 14.
Arquitectura del modelo basado en LSTM para la predicción de la ocupación de las ucis.



Entrenamos el modelo con todos los datos históricos disponibles sobre la ocupación de las ucis y el número de casos de COVID-19 para cada departamento de salud de la Comunidad Valenciana. Utilizamos como función de pérdida el error el cuadrático medio (ECM) y el optimizador Adam. Aplicamos este modelo cada día para predecir el porcentaje de ocupación de las ucis en los siguientes 5 días. Este modelo realizó predicciones competitivas, que se utilizaron cada día para apoyar la planificación en los departamentos de salud.

4.
Ciencia ciudadana

Ante el crecimiento exponencial del número de casos confirmados de SARS-CoV-2 que se produjo en España en marzo de 2020, el hecho de poder hacer una evaluación rápida de la situación y la percepción de los ciudadanos era de vital importancia en el contexto del diseño, la priorización y la toma de decisiones de políticas públicas. Así pues, en marzo de 2020 diseñamos y lanzamos una gran encuesta ciudadana *online* llamada *Covid19impactsurvey* [19], que, con más de 470.000 respuestas de España y más de 220.000 de otros países del mundo, se ha convertido en una de las mayores encuestas ciudadanas sobre la COVID-19 hasta la fecha.

La encuesta es anónima y se estructuró originalmente en 26 preguntas que incluyen información demográfica sobre los participantes (país, edad, sexo, código postal), su situación en el hogar (tipo de hogar, número de cohabitantes en el hogar y sus edades), su comportamiento social (medidas de protección individual, percepción de seguridad en diferentes actividades, número estimado de contactos cercanos, si han tenido contacto estrecho con alguien infectado con coronavirus, etc.), su percepción de las medidas adoptadas para paliar la pandemia, su situación laboral y económica, el impacto emocional que la pandemia está teniendo en su vida y su estado de salud. En junio de 2021, añadimos 6 preguntas relacionadas con el aislamiento social mediante la escala de aislamiento social de Lubben de 6 ítems.

De acuerdo con nuestra experiencia, esta encuesta es una herramienta valiosa para clarificar temas de gran importancia pública para los que no existe una recogida de datos sistemática. Dado el éxito y el valor del análisis de las respuestas de los primeros días tras su lanzamiento (del 28 de marzo al 4 de abril de 2020), decidimos mantener la encuesta *online* y analizar los datos en oleadas semanales. Las preguntas, la metodología de análisis y las principales conclusiones del análisis de las respuestas de la primera oleada se describen en un artículo científico publicado en *JMIR* [20].

En esta sección resumimos, brevemente, los resultados más importantes.

En primer lugar, descubrimos que los contactos estrechos desempeñan un papel crucial en la transmisión del coronavirus. Entre los encuestados que declararon haber dado positivo en SARS-CoV-2 durante la primera ola, el 81% también dijo haber mantenido contacto estrecho con una persona contagiada de coronavirus a la que conocía. Es decir, en la mayoría de los casos el origen probable del contagio era conocido. Es probable que este porcentaje fuera tan alto porque a finales de marzo y principios de abril de 2020 nos encontrábamos en una situación de confinamiento. A partir de la suspensión del estado de alarma y el establecimiento de la "nueva normalidad", este porcentaje disminuyó hasta el 65%. Sin embargo, sigue siendo un porcentaje elevado. Los miembros de un hogar son los contactos estrechos que se citan con mayor frecuencia, seguidos de familiares, amigos y compañeros de trabajo.

En segundo lugar, identificamos diferencias de género estadísticamente significativas en el impacto que tuvo la pandemia, que sitúan a las mujeres en una situación de mayor vulnerabilidad o exposición en comparación con los hombres. Asimismo, las mujeres declaran los niveles más altos de ansiedad, estrés, tristeza y soledad en comparación con los hombres del mismo grupo de edad.

En tercer lugar, la edad también es un factor importante. Identificamos diferencias estadísticamente significativas en el comportamiento de contacto social entre los participantes mayores de 60 años y los más jóvenes. Las personas más mayores eran casi el doble de propensas a declarar que se quedaban en casa sin salir que los participantes más jóvenes. También identificamos diferencias en la edad con respecto a la actitud hacia las medidas de confinamiento, siendo los jóvenes los que apoyaban la adopción de medidas más estrictas para paliar la pandemia y, al mismo tiempo, tenían una menor capacidad para permanecer confinados.

En general, los encuestados solían demandar más actuaciones por parte del Gobierno a medida que aumentaba la incidencia acumulada de casos de COVID-19. Observamos este comportamiento durante toda la pandemia.

El impacto económico de la pandemia es evidente. Según la encuesta, los sectores más afectados son la hostelería, la construcción, el servicio doméstico y el comercio.

El impacto psicológico también es evidente, sobre todo entre las mujeres jóvenes (18-29 años), que declaran los niveles más altos de estrés (50%), ansiedad (46%) y un uso excesivo de la tecnología (57%) durante toda la pandemia. Los hombres jóvenes (18-29 años) declaran los niveles más altos de discusiones en casa, consumo excesivo de alcohol y drogas, tristeza y soledad, con niveles superiores a los declarados por los mayores de 60 años. También son llamativos los altos niveles de un uso excesivo de la tecnología en los niños de las familias con hijos, que superaron el 70% en algunos momentos de la pandemia.

Respecto a las medidas de protección personal, observamos que el uso de mascarillas, la desinfección de manos y la intención de vacunarse fueron generalizados (90% o más). La medida que se adoptó menos, sobre todo entre los jóvenes, fue la ventilación interior. Los encuestados jóvenes (18-29 años) adoptaron en un grado significativamente menor las medidas de distanciamiento físico (evitar abrazos, besos y dar la mano; mantener el distanciamiento físico; limitar los contactos estrechos) en comparación con los adultos más mayores.

El deporte individual, las compras en tiendas pequeñas y la asistencia a lugares con cita previa, por ejemplo peluquerías, son las actividades que los participantes de la encuesta consideraban más seguras, por lo que respectaba a la probabilidad que conllevan de contraer el coronavirus. Las que se consideran menos seguras son viajar en avión, asistir a servicios religiosos y utilizar el transporte público. Alrededor de un tercio de los encuestados cree que es posible acudir a escuelas e institutos con un riesgo bajo de contagiarse de coronavirus.

Un aspecto preocupante durante toda la pandemia es el alto porcentaje de participantes (alrededor del 50%) que afirman no ser capaces de hacer una cuarentena efectiva en caso de necesidad. Observamos

diferencias muy significativas según la edad, siendo entre los jóvenes significativamente menos probable que sean capaces de hacer una cuarentena efectiva en comparación con los adultos más mayores. El principal motivo es el hecho de compartir casa, seguido del cuidado de los hijos o de otras personas, razones psicológicas (entre ellas el miedo a la estigmatización) y razones económicolaborales.

Por último, durante la primera oleada de la encuesta, desarrollamos un modelo de regresión logística para estimar la prevalencia del SRAS-CoV-2 a partir de tres preguntas de la encuesta (síntomas, contacto estrecho con alguien del hogar que hubiera dado positivo por coronavirus y sexo). Este modelo fue importante porque en aquel momento había escasez de PCR. En abril de 2020 el modelo estimó una prevalencia del 5% a nivel nacional, una estimación muy alineada con la proporcionada por el estudio de seroprevalencia realizado por el Instituto Carlos III.

A partir de las respuestas de la encuesta ciudadana, construimos modelos para predecir dos variables de especial relevancia en la toma de decisiones durante el confinamiento: la disposición declarada por los ciudadanos a seguir confinados y la satisfacción con las medidas adoptadas por el Gobierno. Estos modelos se basaron en árboles de clasificación [21, 22] y en algoritmos de extracción de patrones [23]. Estos destacan qué combinaciones de las respuestas de la encuesta conducen (y con qué probabilidad lo hacen) a determinados niveles de disposición a permanecer confinado y a la valoración de las medidas del Gobierno. Este análisis se publicó en *Nature Scientific Reports* [26].

Los patrones extraídos mediante árboles y reglas de clasificación nos permitieron proponer una serie de acciones comunicativas dirigidas a la población muy concretas, en función de su situación demográfica, socioeconómica y de salud. Algunas de estas recomendaciones apuntaban a la necesidad de una desescalada progresiva, sin que se relajaran los mensajes de concienciación a los jóvenes; la conveniencia de ofrecer soluciones de aislamiento alternativas a las familias intergeneracionales con dificultades para el confinamiento en el hogar y la promoción activa del teletrabajo para reducir la movilidad, así como la importancia de detectar a los pacientes asintomáticos en la población en edad de trabajar, entre otras. También hicimos un análisis comparativo de la eficacia de las estrategias de control de la pandemia TTI (las siglas en inglés de rastreo-test-aislamiento) en España e Italia desde junio de 2020 hasta enero de 2021, para clarificar qué factores podrían haber contribuido a la aparición de la segunda ola en otoño de 2020 [27].

Por último, preocupados por el impacto social que podía tener la aplicación continuada de las medidas de confinamiento, analizamos la prevalencia del aislamiento social en España en la segunda mitad de 2021 [28]. Obtuvimos resultados preocupantes: una prevalencia estimada de aislamiento social de casi el 26% en la población general. Observamos diferencias según la edad y el género, con la mayor prevalencia de aislamiento entre los individuos de mediana edad; una fuerte relación entre el impacto económico y el aislamiento social; y diferencias en el aislamiento social, en función del número de medidas de protección contra la COVID-19 y de la percepción del riesgo de contagio por coronavirus por parte de los participantes en el estudio.

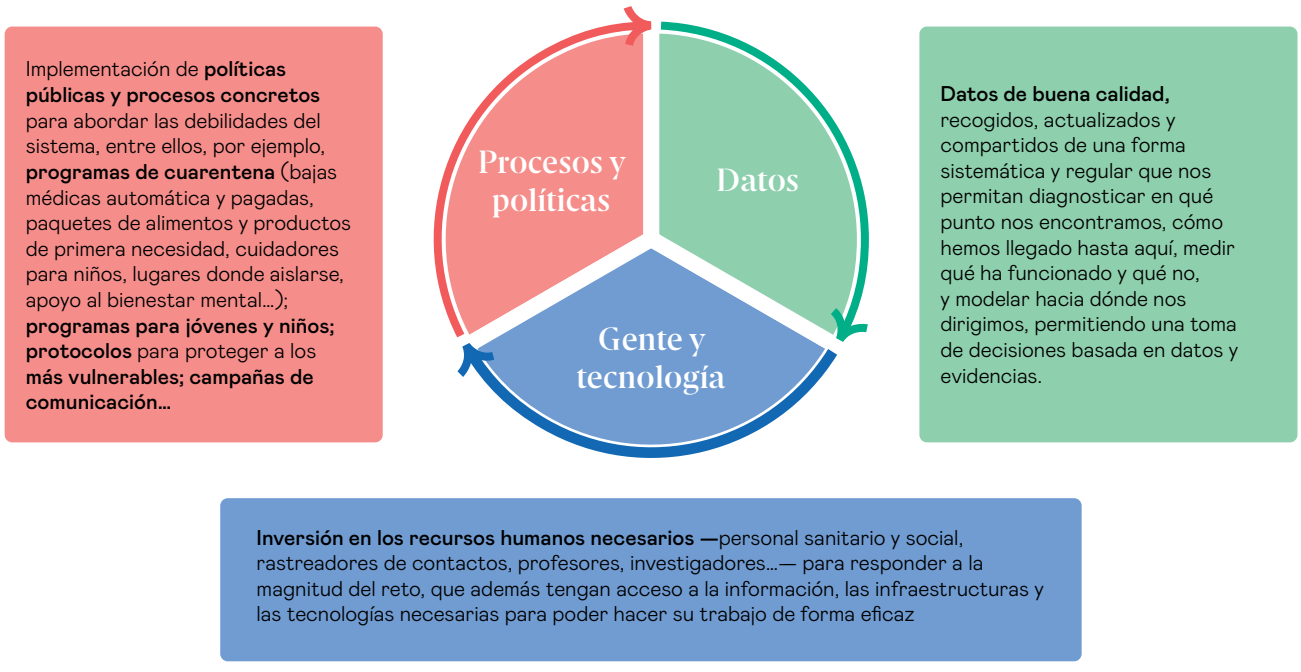
5. Conclusiones

El trabajo que hemos realizado durante los dos años de pandemia de COVID-19 nos ha enseñado que una pandemia no es solo un problema de salud pública, sino un reto social que requiere soluciones holísticas y multidisciplinares: existen oportunidades para desarrollar sinergias entre las herramientas y los procesos tecnológicos, sanitarios y sociales y responder mejor a las complejidades y los desafíos que supone una pandemia.

Nos hemos enfrentado a una pandemia que ha durado dos años y que nos ha enseñado la importancia de lograr soluciones que sean sostenibles desde varias perspectivas, entre ellas las dimensiones psicológica, económica, medioambiental y social.

Este enfoque holístico no solo implica el desarrollo de una visión clara, en la que colaboren los diferentes agentes, sino, más importante aún, una ejecución urgente de la misma que incluya los siguientes tres aspectos clave que forman un círculo virtuoso, ilustrado en la siguiente figura.

Figura 15. Círculo virtuoso de datos, gente, tecnología, procesos y políticas en un enfoque holístico para luchar contra la pandemia de SARS-CoV-2.



1. La disponibilidad de **indicadores específicos** y **datos** de calidad, recogidos, actualizados y compartidos de manera sistemática y regular, que nos permita hacer un diagnóstico del punto en que nos encontramos, analizar las causas, determinar qué ha funcionado y qué no, y modelar hacia dónde nos dirigimos, permitiendo así una toma de decisiones basada en evidencias y conocimiento, entre ellas la decisión de si utilizar o no una app para el rastreo de contactos.

2. Invertir en los **recursos humanos** —personal sanitario y social, rastreadores, profesores, investigadores, etc.— necesarios para abordar la magnitud del reto, que además tengan acceso a la información, las infraestructuras y las tecnologías necesarias para poder realizar su trabajo de forma eficaz.

3. La implantación de **políticas públicas y procesos concretos** que aborden las debilidades del sistema, entre ellos programas que faciliten las cuarentenas (por ejemplo, la baja médica pagada e inmediata, paquetes de alimentos y productos de primera necesidad, cuidadores para los niños o adultos que los necesiten, lugares donde la gente pueda aislarse si en su casa no es posible, garantías de no perder el trabajo, apoyo psicológico, etc.); campañas de comunicación para fomentar una cultura en la que las personas no se relacionen con los demás si tienen la más mínima sospecha de estar contagiadas por coronavirus; protocolos para proteger a los grupos más vulnerables; regulaciones para minimizar el riesgo de contagio en lugares y actividades proclives a los brotes (más ejemplos, plantas de procesado de alimentos, discotecas, celebraciones familiares, etc.) y un conjunto de acciones específicas para niños, adolescentes y jóvenes, que, no podemos olvidar, son quienes están sufriendo con mayor intensidad la carga emocional de la pandemia.

Trabajemos juntos, las personas y la tecnología, la sociedad civil, las empresas y las administraciones, en la lucha contra el virus. El trabajo descrito en este artículo es un ejemplo en esa dirección. La unión, sin duda, es lo que nos hace fuertes.

Agradecimientos

El trabajo descrito en este artículo ha sido financiado parcialmente por la Generalitat Valenciana (Decreto 202/2020 y Convenio Singular 2021 entre la Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital y la Fundación Unidad ELLIS Alicante), la Fundación BBVA (proyecto IA4COVID19) y el fondo Supera COVID del Banco de Santander y la CRUE (proyecto CD4COVID19).

Bibliografía

[1] https://www.muyinteresante.es/salud/articulo/una-foto-semanal-de-la-situacion-del-coronavirus-en-espana-811585915653

[2] https://www.lavanguardia.com/politica/20200424/48693659917/una-encuesta-recoge-que-42-de-espanoles-podria-estar-un-mes-mas-confinado.html

[3] https://www.rtve.es/alacarta/audios/por-tres-razones/tres-razones-inteligencia-artificial-para-combatir-pandemia-24-04-20/5564188/

[4] https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-ai-is-helping-fight-a-pandemic-europes-coronavirus-app-insights-from-valencia/

[5] https://www.youtube.com/watch?feature=youtu.be&v=A06j9Nv8-CA

[6] A. Wesolowski A, C.O. Buckee, L. Bengtsson, E. Wetter, X. Lu, A.J. Tatem. "Commentary: containing the ebola outbreak - the potential and challenge of mobile network data". *PLoS Curr*. 2014;6: currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e. Published 2014 Sep 29. doi:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e

[7] C.M. Peak, A. Wesolowski, E. zu Erbach-Schoenberg, A. J Tatem, E. Wetter, X. Lu, D. Power, E. Weidman-Grunewald, S. Ramos, S. Moritz et al. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. International Journal of Epidemiology, Volume 47, Issue 5, October 2018, Pages 1562–1570, https://doi.org/10.1093/ije/dyy095

[8] A. Wesolowski, N. Eagle, A.J. Tatem, D.LSmith, A.M. Noor, R.W Snow, C.O. Buckee (2012). "Quantifying the impact of human mobility on malaria". Science 338:267-270.

[9] https://www.ine.es/covid/covid_movilidad.htm

[10] https://www.ine.es/covid/exp_movilidad_covid_proyecto.pdf

[11] http://infocoronavirus.gva.es/documents/170024890/170025022/Informe+Movilidad+gva+Abril+2020.pdf/2d728f25-f202-4e7d-81ca-cc25eef9e7d4

[12] http://infocoronavirus.gva.es/documents/170024890/170025022 Informe+Movilidad+gva+-Mayo+2020.pdf/5b043319-eed9-4a66-8477-d214ffe11c39

[13] M. E. J. Newman. "Modularity and community structure in networks". PNAS 103(23) 8577-8582. 2006

[14] Joan L.Aron and Ira B.Schwartz. "Seasonality and period-doubling bifurcations in an epidemic model". Journal of Theoretical Biology, Volume 110, Issue 4, 21 October 1984, Pages 665-679. https://doi.org/10.1016/S0022-5193(84)80150-2

[15] F. Escolano, P. Suau, B. Bonev. "Information Theory in Computer Vision and Pattern Recognition", Springer 2009.

[16] J.T. Tuomisto, J. Yrjölä, M. Kolehmainen, J. Bonsdorff, J. Pekkanen, T. Tikkanen. "An agent-based epidemic model REINA for COVID-19 to identify destructive policies". medRxiv 2020.04.09.20047498; Doi: https://doi.org/10.1101/2020.04.09.20047498https://www.medrxiv.org/content/10.1101/2020.04.09.20047498v2

[17] Miikkulainen, R., FranconOr. Meyerson, E., Qiu, X., Canzani, E., & Hodjat, B. (2020). From Prediction to Prescription: AI-Based Optimization of Non-Pharmaceutical Interventions for the COVID-19 Pandemic. arXiv preprint arXiv:2005.13766.

[18] A.L. Barabasi, The origin of bursts and heavy tails in human dynamics. Nature, 435:7039, 207-211 (2005).

[19] https://covid19impactsurvey.org

[20] N. Oliver, X. Barber, K. Roomp, K. Roomp. "Assessing the Impact of the COVID-19 Pandemic in Spain: Large-Scale, Online, Self-Reported Population Survey". J Med Internet Res 2020;22(9):e21319. DOI: 10.2196/21319

[21] J.R. Quinlan. "Induction of decision trees". Mach Learn 1, 81–106 (1986). https://doi.org/10.1007/BF00116251

[22] L. Breiman, J. Friedman, C.J. Stone, , R.A. Olshen. "Classification and Regression Trees". Ed. Taylor & Francis. The Wadsworth and Brooks-Cole statistics-probability series (1984).

[23] M. Almiñana, L.F. Escudero, A. Pérez-Martín, A. Rabasa, L. Santamaría, "A classification rule reduction algorithm based on significance domains". TOP, 22. 397-418 (2012)

[24] F. Escolano, M.A. Lozano, E.R. Hancock. "The Entropy of Graph Embeddings: A proxy of Potential Mobility in Covid19 Outbreaks. S+SSPR2020: IAPR Joint International Workshops on Statistical Techniques in Pattern Recognition (SPPR 2020) and Syntactic and Structural Pattern Recognition (SSPR 2020).

[25] M.A. Lozano, O. Garibo, E. Piñol, M. Rebollo et al, Open Data Science to fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge, Proc. European Conf. on Machine Learning and Data Mining (ECML-PKDD), 2021 **Best Paper Award (Data Science track)**

[26] M. Martinez-Garcia, To. Rabasa, X. Barber, K. Polotskaya, K. Roomp, N. Oliver, Key factors affecting people's unwillingness to be confined during the COVID-19 pandemic in Spain: a large-scale population study, Nature Scientific Reports, 2021

[27] M. De Nadai, K. Roomp, B. Lepri, N. Oliver, The Impact of Control and Mitigation Strategies during the Second Wave of COVID-19 Infections in Spain and Italy, medRxiv, 2021

[28] M. Martinez-Garcia, E. Sansano-Sansano, A. Castillo-Hornero, R. Femenia, K. Roomp, N. Oliver, Social isolation during the COVID-19 pandemic in Spain: a population study, MedRxiv, 2022

esade

EsadeEcPol - Center
for Economic Policy

Open Internet Governance Institute

The OIGI is EsadeEcPol's effort to shape debates on Internet, data & digital governance both in Spain and across the European Union, while simultaneously contributing to a better understanding of how best use new data and AI-related tools to support and improve policymaking.

We intend to contribute in a balanced and evidence-based manner, departing from the delimitation of weighed dilemmas to focus on offering viable solutions. Our ultimate goal is to help building a system of global and open internet governance, fostering the best possible digital environment among the many future worlds that open before us.

Supported by

Meta