

V Jornadas de Investigadorxs en Formación
Instituto de Desarrollo Económico y Social (IDES)
Ciudad de Buenos Aires, 7, 8 y 9 de octubre de 2020

EJE 7. Ciencia, Tecnología y Democracia

**Predecir la curva: Las declaraciones de científicos de datos frente a la publicación de modelos
y análisis sobre SARS-CoV-2 en redes sociales.**

María Belén Lopez Castro¹

Resumen

La pandemia provocada por SARS-CoV-2 representa un desafío para las democracias modernas. La alta contagiosidad de la enfermedad desafía la capacidad de respuesta y resiliencia de los sistemas de salud en todo el mundo. Estimar el posible comportamiento de la enfermedad en cada jurisdicción se vuelve fundamental para informar la toma de decisiones en diversos niveles.

El amplio acceso a los datos y a los recursos para generar modelos predictivos permitió que se multiplicaran en las redes sociales gráficos que mostraban contagios y muertes esperables en Argentina desde la declaración del primer caso. Frente a la publicación de predicciones en el espacio público, varios “científicos de datos” y referentes del ámbito manifestaron su preocupación y desalentaron la visibilización de datos no generados en espacios académicos o gubernamentales. En este trabajo nos proponemos analizar las declaraciones de los referentes de la “ciencia de datos” expresadas en el espacio público, principalmente en la red social Twitter durante los meses de marzo y abril con relación a las predicciones sobre la evolución de SARS-CoV-2 en Argentina. Sostendremos que uno de los principales argumentos presentados por la comunidad de datos en esta situación se tensiona con argumentos utilizados por los mismos actores en otros contextos al describir sus tareas evidenciando la complejidad de la comunicación con el público masivo y del perfil profesional de la comunidad.

Palabras clave: Antropología de la Ciencia y la tecnología - Informática en Salud - Ciencia de Datos.

¹ Lic. Ciencias Antropológicas. Orientación Sociocultural. UBA, FFyL, ICA, lopezcastromb@gmail.com.

Introducción

El presente trabajo se realiza gracias una Beca Doctoral UBACyT en el marco de un equipo pionero en la región en los estudios de la Antropología de la Ciencia y la Tecnología², con amplia trayectoria en la realización de etnografías de comunidades científicas, instituciones académicas y redes interdisciplinarias de diversos campos- como biotecnología (Stagnaro, 2015), física nuclear (Gaggioli, 2003; Marzorati, 2011), neurociencias (Smulski, 2014), ciencias de la atmósfera (Carabajal, 2020)-. Siguiendo esta tradición, el centro de nuestro interés es la novel comunidad de *informáticos en salud* del ámbito nacional y uno de los objetivos del proyecto doctoral es dar cuenta del proceso de profesionalización de la disciplina.

La *informática en salud* surgió a mediados del siglo XX gracias al impulso de las ciencias de la computación con el objetivo de mejorar la gestión de la información clínica. Inicialmente llamada *informática médica*, en los '90 devino en *informática en salud* para reflejar la diversidad de los perfiles profesionales (reconociendo no son sólo médicos e ingenieros, sino también enfermeros, odontólogos, antropólogos, psicólogos, entre muchos otros). Este cambio de denominación también permitió ampliar el campo de acción y la naturaleza del dato relevante para la disciplina. Es decir, la noción de *informática en salud* permitió incluir no solo el dato clínico gestionado por los médicos sino por otros profesionales de la salud; y los datos de procesos administrativos asociados a la atención. La *informática en salud* tiene como punto de partida el desarrollo de sistemas de registros de los datos de salud y como meta de llegada los sistemas de soporte a la toma de decisiones. Así, las Historias Clínicas Electrónicas y los Portales Personales de Salud alimentan sistemas de soporte que pueden asistir a profesionales clínicamente en el momento de la consulta, y/o a los tomadores de decisiones a nivel de los centros de salud y de la políticas públicas -permitiendo entre otras cosas la vigilancia epidemiológica.

El registro, almacenamiento, recuperación y uso óptimo de los datos para la resolución de problemas y la toma de decisiones es uno de los objetivos y quehaceres de la *informática en salud*. Esta línea se yuxtapone, alimenta y dialoga con otras disciplinas como son la bioestadística y/o la *ciencia de datos*. La primera es una rama de la estadística aplicada a las ciencias de la vida que gracias a la tecnificación del cálculo ha realizado grandes aportes a la disciplina. Quienes se han formado o realizan sus tareas en este campo desde hace años y se han actualizado en el uso de las nuevas tecnologías suelen participar de la comunidad de *ciencias de datos*, algunos otros de la *informática en salud*; y otros profesionales mantienen relación con ambas comunidades. A diferencia de la larga y legitimada

² Proyecto: Co-producción de conocimiento: nuevos formatos asociativos y materialidad de la creatividad científica (UBACyT 0593BA) dirigido por la Dra. Cecilia Hidalgo.

trayectoria de la bioestadística en salud, la *ciencia de datos* es un campo interdisciplinario cuyo perfil e identidad profesional es tan heterogéneo que está en un proceso de construcción y reconfiguración en su interfaz con el ámbito sanitario ya que este argumenta a favor de una especificidad y particularidad de los datos que demanda conocimiento y pericia en el dominio para realizar análisis.

Durante la pandemia generada por SARS-CoV-2 se multiplicaron los gráficos analizando datos epidemiológicos y modelos de predicción de la enfermedad que se amplificaron al ser fuente de consulta de los medios de comunicación y usuarios de redes sociales. La mayoría de estas iniciativas fueron generadas por desarrolladores independientes o departamentos de ciencias de la computación con mayor o menor trayectoria en el campo de la salud. Esta proliferación de modelos llevó a que en la discusión pública se analizaran las fuentes que alimentaban a estos modelos y la dificultad de medir e incluir como variables algunos aspectos de la acción humana. En este trabajo en particular, nos centraremos en un tercer eje de la discusión: la noción de autoridad experta de quienes generaron y publicaron visualizaciones de datos y predicciones. Nos proponemos analizar las declaraciones de los profesionales de las *ciencias de datos* en el espacio público. Para ello, nos centramos especialmente en los intercambios generados en la red social Twitter por un grupo de profesionales de la *comunidad de datos* ante la publicación de modelos y análisis de Covid-19 en estos entornos durante marzo y abril de 2020. Argumentaremos que, en este contexto, en lugar de apelar a la inducción estadística como heurística de descubrimiento, una parte de la autodenominada *comunidad dataera* apeló al principio precautorio y a la experiencia en el dominio desde una mirada normativa de la profesión para criticar a los colegas que decidieron publicar predicciones sobre la enfermedad y la capacidad de respuesta del sistema de salud argentino ante la pandemia.

Informática en salud: especialidad médica vs campo de aplicación

En la intersección entre (a) las ciencias de la salud, (b) la administración y (c) las tecnologías de la información y la comunicación, surgió la *informática en salud* como una interdisciplina que se nutrió de y acogió a diversos perfiles de profesionales. Si bien existe desde hace varios años ofertas de formación de grado (en otros países), la mayoría de los profesionales se han especializado en el ejercicio de su trabajo o en cursos de posgrado. Como señalamos, el cambio de denominación de la especialidad de *informática médica* a *informática en salud* guarda relación con la ampliación del perfil de los profesionales que se desempeñan en el ámbito. Desde su inicio, los proyectos que tuvieron lugar en los ámbitos universitarios contemplaron la participación de diversos perfiles profesionales. La coordinación y conducción estuvo principalmente liderada por ingenieros y médicos, aunque, representantes de otras disciplinas han desarrollado un rol importante en la

producción científico-tecnológica de la especialidad y en su consolidación como un campo de conocimiento (Ackerman et al., 2012; Forsythe, 1999).

Desde los inicios ha existido una diversidad de perfiles implicados en los primeros proyectos que se mantuvo con la institucionalización de propuestas educativas. La consolidación y legitimidad de las ofertas se enfrentaron a dos debates: (a) el primero relacionado con la definición de los bordes de la especialidad y las relaciones con otras disciplinas (como la bioingeniería y la bioinformática); (b) el segundo de eje implica discutir si se entiende a la *informática en salud* como una especialidad médica, o bien como un ámbito de aplicación de los sistemas de información. Este segundo eje es el que genera mayores controversias. Todavía es posible encontrar de modo subyacente estas tensiones en algunas publicaciones y congresos internacionales -a la vez que se reedita cotidianamente en centros de salud privados y públicos-.

La vigencia del debate sobre la *informática en salud* como especialidad médica o campo de aplicación de los sistemas de información es fácilmente rastreable al identificar la heterogeneidad de estrategias de digitalización de los registros de salud. En el ámbito público, hay proyectos de digitalización provinciales llevados a cabo por los ministerios de salud dirigidos por médicos especialistas (CABA), por analistas de sistemas (Neuquén), ingenieros (San Juan) y hay proyectos fuera del ámbito estricto de la salud con el apoyo de grandes empresas de software (Salta). En el ámbito privado, es posible encontrar grandes departamentos de sistemas bajo el mando estratégico de médicos con una mirada sanitarista, como también pequeñas áreas lideradas por médicos bajo el mando de departamentos de sistemas.

La ampliación de la oferta educativa durante los últimos 5 años en el ámbito nacional y la gran demanda laboral del sector generó un aumento exponencial de profesionales en el campo y la conformación de una comunidad diversa y heterogénea de especialistas que hasta el momento se había mantenido circunscripta a unos pocos espacios académicos. Además de quienes decidieron formarse o especializarse en la disciplina, otros profesionales han optado por actualizarse dentro de las propuestas de la *informática en salud* e incluso, reconfigurar sus identidades profesionales y reclamar un lugar en la emergente comunidad de especialistas (entre ellos podemos nombrar a quienes trabajan justamente en la bioestadística y la estadística para las ciencias de la salud). La comunidad local es todavía muy joven y se encuentra en un proceso de crecimiento y profesionalización. Sin embargo, debemos señalar que sus miembros ocupan posiciones importantes en el sistema de salud ya que al estar involucrados en todo proceso que contemple datos, su tarea es transversal y clave.

Ante el avance de la Covid-19, el trabajo de los *informáticos en salud* se vio sobredemandado y reconfigurado. Este contexto impulsó y jerarquizó muchos de los proyectos que tenían en agenda y

les permitió que algunos de sus reclamos (ley de recetas digitales y telemedicina) tuvieran repercusión en las agendas política, pública y mediática. Sin embargo, ninguno de los referentes se manifestó públicamente frente a las predicciones de evolución del SARS-CoV-2 que circularon en redes sociales y medios de comunicación. Tampoco, desde este sector impulsaron desarrollos de visualización pública de datos locales. Las expresiones de los miembros de la comunidad continúan centradas en las posibilidades y beneficios de la telemedicina y de la aplicación de inteligencia artificial a procesos muy específicos -como el diagnóstico por imágenes o el triage de pacientes-.

Ciencia de datos: experiencia en el dominio vs inducción estadística

Como señalamos al inicio, la *ciencia de datos* se construye como una interdisciplina en la intersección de la matemática (estadística) y la ciencia computacional (aprendizaje automático). La comunidad local de *científicos de datos* presenta una gran diversidad de perfiles profesionales y de filiaciones institucionales; es más numerosa que la de *informáticos en salud* y tiene una mayor presencia y distribución en el territorio nacional. Al igual que la *informática en salud*, en este campo se mantienen las discusiones en torno a las trayectorias educativas, las fronteras de su quehacer e incluso la denominación del campo.

La alta demanda del mercado por *científicos de datos* genera una brecha entre quienes han accedido a capacitaciones de rápida salida laboral en espacios no formales y quienes se especializaron en minería y explotación de datos o Big Data como instancia de posgrado. En este sentido, el segundo perfil está compuesto principalmente por economistas, analistas de sistemas y, en tercer lugar, por científicos sociales. En el ámbito nacional desde hace algunos años este campo se presenta como de desarrollo estratégico nacional y cuenta con becas de formación y financiamiento para proyectos de investigación.

Una particularidad de la *ciencia de datos* es su dependencia de las innovaciones tecnológicas en el ámbito de la computación (capacidad de almacenado y velocidad de respuesta). La velocidad vertiginosa en que los avances se producen genera que las ofertas de educación formal e incluso de actualización tiendan a ir por detrás de la demanda de capacitación de la industria. La comunidad de profesionales se hace fundamental para que los miembros puedan mantenerse informados y adquirir las nuevas habilidades. En este sentido, funciona como una *comunidad de prácticas* (Brown & Duguid, 1991), generosa que comparte conocimientos, recursos y proyectos. Como resultado, un amplio público puede acceder a conocimiento siguiendo alguna de las tantas explicaciones disponibles sin tener experiencia previa en dominios o lenguajes de programación.

La *ciencia de datos* y su ubicuidad han reposicionado la inducción estadística. Esta situación ha permitido la reedición de viejos debates sobre el análisis empírico orientado por los datos o por la teoría. En este sentido, algunos autores han señalado los *dilemas epistemológicos* que surgen en el contexto de los desarrollos y análisis de los grandes volúmenes de datos (Ekbja et al., 2015), mientras que otros han recuperado la necesidad de no radicalizar las posiciones (Maass, Parsons, Pura, Storey, & Woo, 2018). El resurgimiento del enfoque orientado por los datos en un momento en que las herramientas de análisis de datos estructurados y no estructurados se han diseminado y permeado en todos los aspectos de la cotidianeidad ha permitido que varios describan a este momento como el del *fetichismo de los algoritmos* (Thomas, Nafus, & Sherman, 2018).

Lo novedoso de estas discusiones epistemológicas de larga data son las formas en que se relaciona con la construcción de la identidad de los *científicos de datos*. En la cotidianeidad, estos suelen reivindicar las ventajas de la posición orientada por los datos (*data driven*), ya que este enfoque ha obtenido resultados positivos en campos como la salud y las ciencias de la atmósfera (Dhar, 2013). En este sentido, hacia el exterior de la comunidad se argumenta que no es necesario ser especialista en el campo o dominio de donde provienen los datos, sino que deben desarrollarse habilidades como programador y buen analista de los resultados que emergen de las exploraciones de los datos. Esta afirmación suele ser una estrategia de posicionamiento del campo y de los profesionales hacia sus clientes.

La formulación de modelos explicativos implica una serie de problemas (como ser (a) las objeciones sobre la inclusión de variables y sesgos, (b) los problemas del muestreo y (c) los problemas para identificar causalidades y relaciones a partir de datos estadísticos) que han sido advertidos históricamente desde diferentes disciplinas; y en los últimos tiempos se han señalado con relación a las aplicaciones de Big Data y Data Analytics. La ubicuidad del análisis de los datos y la consolidación de este perfil profesional altamente requerido hace necesario analizar la construcción de su identidad y sus formas de organización. Hulme (2001), por ejemplo, advierte sobre la consolidación de los modeladores como una comunidad epistémica que domina el discurso sobre el cambio climático. Otros autores han señalado el mismo proceso donde los modeladores se consolidan como un grupo social que tiene no solo poder científico, sino también político para orientar la toma de decisiones en distintos ámbitos (O'Neil, 2016).

La ciencia de datos en salud y en el contexto Covid-19

A las limitaciones generales señaladas a los modelos explicativos que resultan solamente de la inducción estadística se suman, en este contexto cuestiones particulares del ámbito sanitario. Los

profesionales de la salud argumentan constantemente sobre la especificidad del tipo de dato con el que trabajan destacando principalmente su sensibilidad (Ornelas, 2007). En este sentido, existe un discurso homogéneo sobre la particularidad de los datos y el tipo de daño que puede generar en las personas particulares y a las comunidades el mal manejo de los datos personales de salud (Cristea Uivaru, 2018). Es por ello que no solo existen recomendaciones sobre las buenas prácticas de seguridad y cuidado de la privacidad (como la deidentificación de pacientes) por parte de los profesionales de la salud; sino que, además, el dato en salud suele contar con legislación especial en la mayoría de los países.

La sobreprotección del dato en salud se tensiona con activismos como los *datos abiertos*, que demandan la disponibilidad de bases de datos públicas. Una de las aspiraciones de los activistas es que los análisis sirvan como contralor de las políticas públicas e informen una mejor toma de decisiones. De este modo, es común ver como práctica política para la movilización de reclamos y la sensibilización pública la publicación de visualizaciones de datos de diferentes dominios. La disponibilidad de bases de datos epidemiológicas permitió que esta práctica se replicara ante la pandemia por coronavirus.

Si analizamos las iniciativas de visualización con más visitas -y que fueron principalmente consultadas por los medios de comunicación-, vemos que sus desarrolladores provienen del ámbito de los sistemas de información, con mayor o menos inserción en instituciones académicas y muchos de ellos sin una trayectoria ligada al análisis de datos del campo de la salud. A estas iniciativas de desarrolladores se sumaron varios ciudadanos y políticos independientes que hicieron lo propio. Además, como si fuera poca la cantidad de información producida y circulando en espacios públicos -sin la evaluación y revisión por pares que caracteriza al ámbito científico- la comunidad académica también eligió comunicar en medios masivos para alcanzar a un público amplio generando una gran proliferación de modelos y gráficos.

La predicción del avance y la disponibilidad de camas por Covid-19 en Argentina

La comunidad argentina de datos se hizo eco de los debates en las redes por la proliferación de análisis relacionados al Covid-19; varios referentes manifestaron su preocupación ante la publicación de modelos por parte de profesionales sin trayectoria en el ámbito de la salud. Lo que prevaleció en la mayoría de sus declaraciones fue la crítica esta práctica: la falta de autoridad experta en el dominio y el daño que podía generar a la ciudadanía.

El corpus de declaraciones para el presente análisis se construyó de manera intencionada a partir de la selección de perfiles de profesionales que participan activamente en redes sociales y en diferentes

espacios de la *comunidad de datos* local. Hace aproximadamente dos años se seleccionó mediante el método bola de nieve una serie de usuarios/perfiles como parte del trabajo de campo. El punto inicial de la construcción de este listado fueron los profesionales de estadística que trabajan con los *informáticos en salud*. Luego ese círculo se amplió a profesionales que tienen relación con los profesionales de datos del primer grado de contacto y participan de modo activo en redes sociales compartiendo diferentes proyectos de educación, investigación y desarrollo. También se incluyeron los perfiles de comunidades de usuarios de software, instituciones educativas, organizaciones y asociaciones relacionadas a la autodenominada “comunidad dataera”. Recapitulando, entonces, las características de los perfiles de usuarios que seguimos en redes sociales son: (a) que estudian o desempeñan su trabajo en algún ámbito relacionado a la *ciencia de datos*, (b) son activos en las redes compartiendo información sobre su trabajo de investigación, desarrollo y/o las actividades de educación y capacitación y; (c) tienen interacción con los *científicos de datos* y estadísticos que se trabajan en el ámbito de la salud. Analizando este listado de perfiles que se conformó a partir de relaciones y piezas de información relevantes para el campo del proyecto doctoral, reconocemos que la mayoría de los profesionales desarrollan sus tareas en el ámbito de la Ciudad de Buenos Aires y el Gran Buenos Aires; y es posible reconocer algún tipo de relación con la Maestría de Minería y Explotación de Datos de la Facultad de Ciencias Naturales y Exactas, UBA.

La principal red social utilizada por la comunidad -y el entorno del cual se recuperaron la mayoría de las declaraciones- es Twitter ya que habilita una comunicación y relación transversal que no se produce en otras redes sociales especializadas (como GitHub o Slack) donde los intercambios se dan de modo parcializados, ya sea por la temática de discusión o por el software utilizado. Por esta razón, realizamos un seguimiento de las interacciones especialmente en esta red donde la comunidad suele compartir información y recursos, así como debatir.

Cuando comenzó a circular la información sobre el stress que estaban sufriendo los sistemas de salud de los países asiáticos y europeos frente al brote de coronavirus, dos referentes locales se mostraron preocupados en que se tradujera al español el lema *flatten the curve*. Argumentaban que es una idea estadística que no colabora a que la ciudadanía entienda y adopte políticas de cuidado individuales. En poco tiempo la idea de *aplanar la curva* devino en un lema mundial que organizó tanto políticas públicas como iniciativas ciudadanas. La proliferación de modelos y visualizaciones de datos comenzó a ser necesaria para poder evaluar la situación y el impacto de las medidas sobre la gestión de recursos; en especial, las camas de terapia intensiva.

La publicación de gráficos y predicciones son, entonces, parte de la narrativa general de la pandemia SARS-CoV-2 y varios profesionales decidieron generar los suyos de modo independiente y publicarlos en redes sociales como Twitter y LinkedIn. Además de estas publicaciones independientes

realizadas por profesionales, empresas de software, asociaciones civiles, universidades e incluso gobiernos han convocaron a encuentros como hackatones para idear y potenciar desarrollos que mitiguen los efectos negativos de la pandemia. Los usuarios que pertenecen al grupo de referentes locales construido por su relación con los *informáticos en salud* salieron a criticar esta práctica expresándose en esta dirección³:



En las dos interacciones citadas se puede ver que la idea de conocimiento sobre el campo y la experiencia en el ámbito es considerada un condicionante en la autoridad para analizar datos de salud en el contexto de pandemia. Como se desarrolló anteriormente, la inducción estadística es un pilar de la *ciencia de datos* que sirve como heurística de interpretación y como punto de inicio en la búsqueda de explicaciones causales -que suelen construirse en equipos interdisciplinarios. También comentamos que esta posición orientada por los datos, juega un papel importante en el discurso que hace a la identidad profesional. Por esta razón, nos resultaron significativas estas declaraciones.

Otro elemento que emerge en los dos tweets es la tensión entre el conocimiento sobre el uso de la herramienta y su aplicación. Frente a un reclamo de acceso a más datos para aportar de manera activa como ciudadanos al debate sobre la toma de decisiones políticas, estos usuarios señalan que el contexto no es el adecuado para esta práctica. Se puede ver que consideran que los modelos no van a ser discutidos estadísticamente y que el impacto en un público lego puede generar desinformación.

³ Si bien los perfiles de las redes son públicos se eligió anonimizar a los profesionales.

En esta línea la fundación Sadosky fue clara y explícita al apelar al principio precautorio en una declaración pública:

Resumiendo, la inteligencia artificial es una herramienta con mucho potencial y aplicable a las más diversas áreas, pero requiere acceso a datos confiables, revisiones y pruebas para evitar sesgos, y más aún, como cualquier dispositivo aplicado a la salud, requiere de homologaciones y aprobación por los organismos de control que aseguren su idoneidad y fiabilidad. Además, debemos utilizarla de manera crítica. De no ser así, se podría incurrir en conclusiones falaces, toma de decisiones erróneas, crear angustia y preocupaciones innecesarias, por lo que, nuevamente, desde la Fundación Sadosky exhortamos a la población, y los medios, a minimizar su uso y difusión, y en tal caso a no tomar sus predicciones como certeras bajo ningún aspecto.⁴

Discusión

Al inicio se presentó como marco de este trabajo un proyecto cuyo objetivo es dar cuenta del proceso de profesionalización de la *informática en salud* en el país. El enfoque que utilizamos para analizar este proceso entiende que el contexto de producción de conocimiento y la representación de la realidad son constitutivas de la vida social, de modo que el mundo social y natural se producen mutuamente (Jasanoff, 2004). Así, se vuelve relevante reparar en cómo la ciencia y los científicos en la cotidianeidad construyen conocimientos que atraviesan elementos centrales de las sociedades como son las identidades, las instituciones, los discursos y las representaciones. Además, entendemos que el contexto científico actual se caracteriza por un *giro colaborativo* donde el hacer ciencia implica la colaboración interdisciplinaria, intersectorial e interinstitucional entorno a problemas relevantes y complejos. En este escenario entonces, las formas en que los *informáticos en salud* y los *científicos de datos* configuran sus identidades, relaciones y las fronteras entre sus tareas es un aporte al proyecto de investigación; y los debates sobre la publicación de modelos predictivos durante la pandemia resultó ser una coyuntura para analizar esta relación.

Uno de los ejes programáticos de la *informática en salud* hace a la planificación estratégica de los sistemas de información en salud- sea para la atención clínica, la gestión administrativa y/o epidemiológica. Los *informáticos en salud* poseen autoridad para pronunciarse sobre las consecuencias de la publicación de modelos predictivos en las redes sociales. Sin embargo, es necesario recordar que, en la organización interna de los equipos de trabajo, el desarrollo de estos modelos y visualizaciones suele recaer en *científicos de datos*. En el contexto de la pandemia por

⁴ Fundación Sadosky, 30 de marzo de 2020 “Inteligencia Artificial en épocas de pandemia”. Disponible en: <http://www.fundacionsadosky.org.ar/inteligencia-artificial-en-epocas-de-pandemia/>

coronavirus, los *informáticos en salud* vieron reconfigurado y sobredemandado su trabajo. Públicamente sus declaraciones apuntaron hacia la sanción de la ley de telemedicina, la extensión de la prescripción digital y las consecuencias generales de la infodemia.

Por su parte, los *científicos de datos* reaccionaron con crítica frente a la publicación de modelos, en especial por parte de los colegas que incurrieran en esta práctica. Uno de los argumentos principales fue la falta de profesionalismo. Cuando la noción de profesionalismo es esgrimida dentro de la misma ocupación, permite identificar las formas en que se construye la identidad de la profesión; así como las relaciones con otros: como los clientes y los estados (Evetts, 2003). En este caso, la dimensión normativa que atraviesa la idea de profesionalismo para los *científicos de datos* marca los límites de su quehacer resaltando la necesidad de un trabajo colaborativo donde la validación de los modelos predictivos no depende sólo de la pericia estadística, sino que recae en equipos que incluyan a profesionales del campo de la salud. Esta frontera que se traza con los profesionales de la salud para generar modelos predictivos -sobre la disponibilidad de camas de terapia y el avance de la enfermedad-, invita a repensar las formas tradicionales para superar los obstáculos y sesgos en el contexto del “giro colaborativo” como forma de producción de conocimientos y de organización del trabajo para resolver problemas sociales (Eker, 2020; Saltelli et al., 2020).

Igualmente, el argumento del profesionalismo como sistema de valores, apunta a la responsabilidad por las consecuencias que puede generar no solo la intromisión en un campo donde no se tiene experiencia, sino en un momento de conmoción de la totalidad de la sociedad. Siguiendo esta línea: una parte importante de la *ciencia de datos* yace en las promesas del desarrollo computacional y en la potencialidad de la Inteligencia Artificial. El llamado de estos profesionales es a limitar las expectativas de la sociedad y de los medios de comunicación y; especialmente, a identificar las certificaciones de las instituciones -siempre estatales- que demuestren la seguridad en el uso de dispositivos y algoritmos. Así, vemos no solo el reconocimiento de los *científicos de datos* hacia las ciencias de la salud sino también hacia los organismos de control.

Evetts (2003) argumentó que, a diferencia de otras categorías de análisis sobre las ocupaciones, la noción de profesionalismo permite articular las tres escalas (macro-meso-micro) y dar cuenta de las relaciones y los sistemas de valores e ideología que cada comunidad busca construir, a la vez que se desempeña como un mecanismo de control. La apelación al profesionalismo de los *científicos de datos* permite ver desde donde eligen construir su identidad y credibilidad pública (Larson, 1979) marcando el límite de su autoridad sobre los temas y señalando las alianzas necesarias para legitimar sus producciones. Como señalamos, la *comunidad de datos* en Argentina es muy amplia y heterogénea. La diversidad de ofertas de cursos y la gran demanda del mercado -que deja trunca algunas trayectorias formativas en espacios formales- colaboran a que sea difícil la construcción de

una identidad a partir de experiencias educativas. Incluso, esto dificulta la conformación de sociedades, asociaciones profesionales y gremios con poder para regular las prácticas y consolidar el control sobre el mercado de trabajo. Por eso, frente a quienes muestran sus habilidades para “fitear una curva en LinkedIn” y exacerban el discurso *data driven* colaborando a la fetichización de los datos; nuestro clúster apeló a un discurso normativo. Este discurso normativo que emerge en la construcción de una identidad que hacia “otros”; tiene a los *informáticos en salud* como potenciales compañeros de desarrollo e investigación (aunque también como clientes: los *informáticos en salud* suelen decidir qué perfil o perfiles profesionales cumplen el rol de desarrollo de tableros de control y análisis estadístico). En este punto se vuelve relevante recordar entonces, dos aspectos del grupo de usuarios que construimos: tienen relación más o menos estrecha con *informáticos en salud* y sus trayectorias educativas incluyen la formación de posgrado universitaria. Esta última característica es la que identificamos como significativa al momento de reconocer la autoridad que otorga la experiencia en cada campo de conocimiento para evitar reduccionismos y sesgos por parte de los *científicos de datos*.

Conclusión

El presente trabajo se propuso analizar las declaraciones de los *científicos de datos* en redes sociales frente a las predicciones sobre el avance del coronavirus en el territorio nacional. La intención fue analizar sus interacciones a la luz de la conformación de su identidad como comunidad y en relación con otras, como podría ser la de los *informáticos en salud*, quienes también poseen autoridad para participar de la discusión pública sobre el tema.

En primer lugar, identificamos que los *científicos de datos* seleccionados criticaron la publicación en cuentas personales de predicciones y señalaron poca profesionalidad por parte de quienes lo hacían sin conocimiento del dominio. En este sentido, señalamos que se tensiona con un discurso de las *ciencias de datos* que construye la identidad del profesional desde la idea del buen analista que domina la inducción estadística sin la necesidad de contar con experiencia en el campo.

En segundo lugar, esperamos haber mostrado que los *científicos de datos* que alzaron las críticas comparten trayectorias formativas que colaboran en la construcción de una identidad profesional que promueve formas de producción del conocimiento desde el dialogo y colaboración con otro. En un contexto donde la *ciencia de datos* se expande y busca ampliar el quehacer de los profesionales, los *científicos de datos* de nuestra muestra, por el contrario, marcar los límites de su trabajo y la necesidad de construir autoridad sobre un tema de coyuntura en relación son otros actores, buscando superar las limitaciones tradicionales de los modelos predictivos.

Bibliografía

- Ackerman, S. L., K. Tebb, J. C. Stein, B. W. Frazee, G. W. Hendey, L. A. Schmidt & R. Gonzales (2012). Benefit or burden? A sociotechnical analysis of diagnostic computer kiosks in four California hospital emergency departments. *Social Science & Medicine*, 75 (12): 2378–2385.
- Brown, J. S. & P. Duguid (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science*, 2 (1).
- Carabajal, M. I. (2020). Producción, circulación y uso de la información climática: Contribuciones de la Antropología al estudio de los Servicios Climáticos. *Runa*, 41.
- Cristea Uivaru, L. (2018). *La protección de datos de carácter sensible: Historia Clínica Digital y Big Data en Salud*. Barcelona: JB Bosh.
- Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, 56 (12): 64–73.
- Ekbia, H., M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, ... C. R. Sugimoto (2015). Big data, bigger dilemmas: A critical review. *Big Data, Bigger Dilemmas: A Critical Review*, 66 (8): 1523–1545.
- Eker, S. (2020, December 1). Validity and usefulness of COVID-19 models. *Humanities and Social Sciences Communications*, Vol. 7: 1–5.
- Evetts, J. (2003). The Sociological Analysis of Professionalism Occupational Change in the Modern World. *International Sociology* 2003 18: 395, 18(2), 395–455.
- Forsythe, D. E. (1999). “It’s Just a Matter of Common Sense”: Ethnography as Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8 (1–2): 127–145.
- Gaggioli, N. N. G. (2003). *La comunidad nuclear. Una mirada antropológica sobre el desarrollo nuclear argentino*. Ciudad Autónoma de Buenos Aires: Sociedad Argentina de Antropología.
- Hulme, M. (2001). Reducing the Future to Climate: A Story of Climate Determinism and Reductionism. *Klima* (2011), 26 (1): 245–266.
- Jasanoff, S. (2004). *States of Knowledge: The Co-Production of Science and the Social Order* (S. Jasanoff, Ed.). New York: Routledge.
- Larson, M. S. (1979). Professionalism: Rise and Fall. *International Journal of Health Services*, 9 (4): 607–617.

- Maass, W., J. Parsons, S. Purao, S. Storey & C. Woo (2018). Data-Driven Meets Theory-Driven Research in the Era of Big Data: Opportunities and Challenges for Information Systems Research. *Journal of the Association for Information Systems*, 19 (12): 1253–1273.
- Marzorati, Z. del V. (2011). *Plantear utopías. La conformación del campo científico-tecnológico nuclear en Argentina (1950-1955)*. Ciudad Autónoma de Buenos Aires: Ciccus-CLACSO.
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Books.
- Ornelas, L. (2007). Manejo de la información de salud: Derecho de acceso y protección de datos personales. *CONAMED*, 12 (1): 31–34.
- Saltelli, A., G. Bammer, G., I. Bruno, E. Charters, M. Di Fiore, E. Didier, E., ... P. Vineis (2020, June 1). Five ways to ensure that models serve society: a manifesto. *Nature*, Vol. 582: 482–484.
- Smulski, M. (2014). La antropología ante los nuevos avances de la neurociencia. El desafío del diálogo interdisciplinario. *Kula*.
- Stagnaro, A. (2015). *Ciencia a pulmón: etnografías de laboratorios argentinos de biotecnología*. Ciudad Autónoma de Buenos Aires: Ciccus-CLACSO.
- Thomas, S., D. Nafus, & J. Sherman (2018). Algorithms as fetish: Faith and possibility in algorithmic work. *Big Data & Society*, 1–11.