

Laboratorios en Ciencias Sociales: "Pasar de los conceptos a los índices y documentar las etapas de investigación"

Entrevista a Germán Rosati (Universidad Nacional de San Martín)

*Servent, Mateo**

* Estudiante de Maestría en Ciencia Política, Centro de Investigación y Docencia Económica (CIDE), México. Director de Revista Disputas.



Fuente: Extraído de Tesis Doctoral Germán Rosati [Youtube] <https://www.youtube.com/watch?v=xWVBtZ5fEI0>

German Rosati (Doctor en Ciencias Sociales y Sociólogo por la UBA y Magíster en Generación y Análisis de Información Estadística por la Universidad Nacional de Tres de Febrero) es Investigador Asistente del CONICET. Su línea de investigación gira en torno al estudio de la estructura social y agraria a través de estadística aplicada como Machine Learning y Procesamiento de Lenguaje Natural.

A su vez, se desempeña como coordinador de Factor~Data, un laboratorio en Ciencias Sociales Computacionales, y la Diplomatura en Ciencias Sociales Computacionales y Humanidades Digitales, ambos en la Escuela de Interdisciplinaria de Altos Estudios Sociales de la Universidad Nacional de San Martín. Además, ha sido analista de datos para la administración pública nacional. Dada su trayectoria, en esta entrevista conversamos sobre el porqué de la existencia de laboratorios en ciencias sociales, cuáles son sus objetivos y cuáles son las limitaciones que él encuentra trabajando en un laboratorio desde Argentina.

Sobre el final, nos comenta qué rumbos aconseja a quienes desean empezar un trayecto académico en Ciencias Sociales Computacionales.

RevistaDisputas: Germán, contanos, ¿por qué consideras que existen Laboratorios en Ciencias Sociales? y ¿qué implicancias tienen para las etapas de investigación y la replicabilidad de resultados?

Germán Rosati: Creo que la especificidad está en un énfasis metodológico para la investigación empírica. Más allá del nombre "laboratorio", que podría ser otro, en general estos laboratorios desarrollan líneas de investigación con la mirada puesta en el uso de distintos tipos de técnicas y metodologías que pueden ser computacionales, basadas en datos cuantitativos "tradicionales" o datos cualitativos. Hay un interés por innovar, aplicar y desarrollar técnicas para abordar preguntas y problemas más o menos clásicos de las Ciencias Sociales. Sin embargo, no me parece que haya una diferencia sustantiva entre los equipos de investigación tradicionales y el trabajo de un laboratorio en Ciencias Sociales o que haya algo totalmente novedoso. Sí creo que hay un intento por, como mínimo, documentar mejor las etapas de los procesos de investigaciones. Podrá tener el formato de código en aquellos proyectos que requieran escribir código y generar repositorios de datos o podría tener que ver con un tipo de documentación centrada en el registro de los procedimientos que se fueron aplicando.

Por otro lado, es cierto que uno de los grandes problemas de la investigación social en particular y de la investigación científica en general es el problema de la replicabilidad de los resultados. Y ahí es más fácil que uno pueda replicar una investigación cuantitativa en cualquiera de sus formas antes que una investigación cualitativa. La investigación cuantitativa te obliga a estructurar los datos de cierta forma, a correr ciertos modelos, a dotarte de ciertos procedimientos más o menos estructurados. Generar información cuantitativa no significa garantía de replicabilidad, pero sí, quizás, provee mejores condiciones para hacerlo. Por otro lado, hay perspectivas cualitativas desde las cuales se afirma que este objetivo de replicabilidad no es posible (las visiones más vinculadas a un enfoque hermenéutico, por ejemplo) Pero hay otras líneas que sí: la Teoría Fundamentada y el análisis de contenido cualitativo, por mencionar dos casos, ponen un énfasis en el intento de hacer replicable el proceso, pero es más difícil en términos de que el objeto está menos estructurado. En la investigación que hicimos nosotros sobre comentarios de lectores de noticias de COVID-19, los comentarios podrían llegar a desaparecer de los foros, pero hoy hay un repositorio donde están almacenados. Vos puedes bajar el código, correrlo, construir y analizar la base de datos con los comentarios. Uno de los aspectos importantes de la replicabilidad es la posibilidad de ampliar y profundizar y eventualmente corregir resultados. Eso es lo que está bueno de la replicación. Además de ser "confirmatoria", debiera abrir las puertas a ampliar esos hallazgos y corregirlos. A hacer avanzar en conocimiento.

RevistaDisputas: ¿Cómo nace el caso del Laboratorio Factor~Data de la Universidad Nacional de San Martín en relación a ello?

“
Uno de los grandes problemas de la investigación social en particular y de la investigación científica en general es el problema de la replicabilidad de los resultados.
”

“
 Está muy
 instalado que
 en el mundo
 académico,
 el inglés es
 una especie
 de lengua
 franca y esto es
 efectivamente
 así.”

Germán Rosati: Factor~Data nace a fines de 2019 con el objetivo de instalar el uso de métodos computacionales vinculados a Machine Learning y que se convierta en un espacio de difusión de ese tipo de técnicas para investigaciones en Ciencias Sociales. Desarrollamos 3 líneas de trabajo: la primera la llamamos investigación básica, proyectos de investigación que son propios del equipo. La segunda, investigación aplicada y de vinculación, que abarca consultorías tradicionales, hemos hecho trabajos para CEPAL, para municipios y también hemos establecido relaciones con otros investigadores que se acercan con una pregunta, y nuestro rol allí es darles soporte técnico y metodológico. Y la tercera línea se centra en un aspecto muy importante para nosotros: formación y pedagogía. En esta línea, desarrollamos varias iniciativas: el Diploma en Ciencias Sociales Computacionales y Humanidades Digitales; también un trayecto de grado de cuatro materias que estamos desarrollando para antropología y sociología en UNSAM, y algunos seminarios de posgrado.

RevistaDisputas: ¿Qué limitaciones encontras para investigar en un laboratorio de Ciencias Sociales Computacionales desde Argentina? En términos desde presentación de propuestas de proyectos y publicación de resultados, hasta desafíos como el idioma y restricciones presupuestarias para acceder, por ejemplo, a clusters de procesamiento

German Rosati: Hay problemas que vienen por el lado presupuestario: desde Argentina es difícil acceder a cierto tipo de recursos. Usamos bastante infraestructura en la nube (para tareas computacionales más o menos intensivas). Esto trae algunas dificultades, sobre todo vinculadas a la disponibilidad de divisas para el pago de esos servicios. Para funciones y modelos que tienen que correr mucho tiempo solemos recurrir a ese tipo de herramientas, como AWS, Google y demás. Resuelven bastantes problemas en términos de poder usarlo a demanda y no tener que comprar un servidor, que además envejecen rápido. Dado que somos un laboratorio aún pequeño es cómodo y relativamente rentable utilizar este tipo de tecnología, pero a medida que la cosa vaya escalando no tengo tan claro que esto se mantenga.

Hay un tema importante en términos de restricciones: la cuestión del idioma. Está muy instalado que en el mundo académico, el inglés es una especie de lengua franca y esto es efectivamente así. De hecho, los libros del estado del arte están en inglés, en parte porque gran parte de la investigación en métodos se realiza en inglés. Pero no deja de ser cierto que este hecho deja afuera una gran cantidad de gente. Esto ocurre en diversas etapas, pero seguramente en la etapa formativa es donde probablemente mayor efecto tenga. Por eso es importante publicar papers, documentación y paquetes en castellano, que es un idioma muy grande en términos de comunidad de hablantes. En esa línea hay algunas iniciativas que están en la frontera de lo académico, como las comunidades de usuarios y de programadores, o de científicos que traducen colectivamente.

Una comunidad de usuarios de R tradujo un texto que es como el manual básico de introducción a R para Ciencia de Datos de Hadley Wickham, que está disponible de forma gratuita online, pero hay libros que las editoriales no quieren traducir porque es muy costoso. De todas maneras, está bueno montarse sobre lo que existe, para no reescribir y decir voy a escribir mi libro, si hay una publicación que ya existe está bueno y resulta eficiente pensar en una traducción. En términos de acceso a bibliografía y literatura científica nosotros recurrimos mucho Sci-Hub, en Argentina el circuito informal de producciones tiene bastante peso. Porque ahí hay un problema para resolver con las editoriales académicas y el negocio que tienen. Si bien hay algunos proyectos que dan acceso abierto, trasladan costos a los productores y uno tiene que pagar para publicar, unos dos mil dólares en promedio para revistas de sociología.

Una última cuestión vinculada al tema del idioma que me parece importante: la cuestión de publicar "afuera". Si bien es importante en términos de instalar la producción académica, de tender redes e incrementar el intercambio, en lo personal no estoy seguro de que publicar en el "Norte Global" deba ser un objetivo principal para la investigación científica en países dependientes.

Finalmente, un último obstáculo o problema tiene que ver con el tipo de formación o capacitación. En Argentina, una buena parte de las carreras en Ciencias Sociales tiene un carácter teórico. El correlato, en general, es que la formación en métodos y técnicas tiene menor peso relativo, al menos en muchas de las carreras de Sociología. En muchos casos, seguimos dando materias centrales (como las llamadas "Metodologías de investigación" más o menos con los mismos contenidos, con las mismas técnicas y herramientas de cálculo que las que se usaban en la década del 50-60. Resulta muy llamativo que no hayamos podido incorporar técnicas estadísticas como el bootstrap, cuya implementación está disponible al menos desde mediados de la década del '90. Desde la dictadura se atacó fuertemente a las Ciencias Sociales. Particularmente, a Sociología de la UBA la mandaron a un sótano. Ese tipo de cuestiones impactaron mucho, hubo un cambio de paradigma. Empezó a haber menos énfasis en los métodos y en su interacción con la teoría. Pero quisiera ser claro: no se trata de contraponer lo "cualitativo" y lo "cuantitativo", sino más bien apuntalar eso que suele llamarse "operacionalización": el pasaje de los conceptos a los índices, de la teoría a lo empírico.

Es un punto que, en mi opinión, hay que trabajar. Me parece que hace falta lograr explicar mejor qué son estos enfoques relativamente novedosos ("big data", machine learning, etc.), porque a veces se crea la imagen de un mundo inabarcable o un conjunto de técnicas que son solo utilizables por los "sumos sacerdotes". Mostrar la viabilidad de la aplicación de estas herramientas a las Ciencias Sociales me parece fundamental. Por ejemplo, en una solicitud a un subsidio de investigación una de las evaluaciones que nos dieron consideraba que se trataba de un proyecto muy grande para un equipo pequeño y que era poco probable que se llegara a los resultados planteados en el tiempo de ejecución del proyecto (además había una crítica al enfoque cuantitativo y al uso de datos

secundarios). El caso fue para analizar las modalidades de expansión de la frontera agraria en todo el país. Obviamente que hay un balance entre generalidad y especificidad, pero a veces tenemos que trabajar mejor en comunicar y explicar cómo y por qué lo que decimos que vamos a hacer es factible.

RevistaDisputas: Para finalizar, ¿qué camino recomendarías seguir a quien esté buscando aprender métodos computacionales o bien que esté iniciando una carrera profesional en investigación con ánimos de continuar en esta rama de las Ciencias Sociales?

“
Para el caso específico de estudiantes en Ciencias Sociales, creo que es productivo empezar a trabajar la integración entre método y teoría.
”

Germán Rosati: Bien, se me ocurre que hay dos situaciones: una cosa es estar promediando la carrera de grado y otra cosa es la segunda situación que, entiendo, se trata de alguien que está avanzando en un posgrado o al menos está haciendo sus primeros trabajos de investigación. Allí, creo se abren varios caminos. Primero, creo que el inicio con algún bootcampy con cursos online es una buena opción. Si bien tienen muchas limitaciones, son un muy buen recurso de acercamiento para familiarizarse con este tipo de técnicas. Cursos de análisis de datos y estadística hay un montón. Luego, uno puede pensar en una especie de segundo nivel, que son todas las escuelas de codeo. Como siempre, hay de todo pero muchas tienen la ventaja de estar muy orientadas a la práctica. Finalmente, un tercer nivel más formal, como por ejemplo, un posgrado. Está lleno de maestrías en ciencia de datos, de hecho, hay al menos dos licenciaturas en ciencia de datos en universidades públicas en Argentina (en UNSAM y en la UBA).

Me parece que el punto a tener en cuenta es que en general esos diferentes procesos formativos son agnósticos en términos de la disciplina. Para el caso específico de estudiantes en Ciencias Sociales, creo que es productivo empezar a trabajar la integración entre método y teoría: no es verdad que las técnicas están por un lado y lo conceptual está por otro. Me parece que el Diploma que mencionaba antes es un primer avance en esa dirección. De hecho, los docentes son todos profesionales de las Ciencias Sociales, hay un idioma o un dialecto común, ya que el perfil del estudiante está abrumadoramente vinculado a las Ciencias Sociales. Nosotros creemos que es más provechoso transitar un lugar donde la cuestión disciplinar y el método están de la mano.