



El Desafío de Descubrir Conocimientos en Bases de Datos

Uno de los mayores desafíos de la medicina actual es procesar y analizar ingentes cantidades de datos generados a partir de múltiples fuentes, en el momento de atención del paciente. A modo de ejemplo, en las unidades de cuidados intensivos (UCI), los datos pueden provenir de los monitores de signos vitales, las bombas de infusión, los dispositivos de asistencia respiratoria mecánica, las historias clínicas electrónicas, los sistemas de prescripción electrónica, los sistemas de imágenes y laboratorio, etc. Estos datos se registran en forma casi continua durante toda la estadía del paciente en la UCI. A nivel poblacional, los Registros nacionales o internacionales también constituyen importantes fuentes de datos; diseñados con propósitos científicos, clínicos o administrativos los registros son en esencia grandes bases de datos que contienen información de la práctica clínica, los resultados en los pacientes, la seguridad y la efectividad comparativa, proveniente del mundo real.

En este escenario, es evidente la necesidad de desarrollar procesos automáticos o semiautomáticos que permitan encontrar conocimiento útil y entendible en los datos. Este proceso, conocido como KDD (del inglés, *Knowledge Discovery in Databases*) involucra el análisis exploratorio y la generación de modelos a partir de grandes y complejos repositorios de datos con la finalidad de identificar patrones válidos, novedosos, útiles y comprensibles. La relación entre los datos, la información y el conocimiento se describe generalmente como la pirámide del conocimiento (Figura 1) que expresa que una gran cantidad de datos puede transformarse en información si se incluye en un contexto; esta información agregada y analizada puede llevar al descubrimiento de patrones en los datos que representen un conocimiento nuevo en el campo estudiado.

La esencia de KDD es transformar una gran cantidad de datos crudos en información útil y conocimiento que pueda ser interpretado con facilidad y que colabore con la toma de decisiones clínicas o administrativas, a nivel del paciente o de la población. En este sentido, los Registros son considerados uno de los 3 pilares de infraestructura en tecnología de la información, necesarios para mejorar la salud de la población (junto con las historias clínicas electrónicas y los portales de pacientes).

En la actualidad, la mayoría de los registros son liderados por profesionales médicos e informáticos, altamente especializados en el campo de interés, con habilidades técnicas como *expertise* en salud pública o gestión. Sin embargo, la necesidad actual y sin duda futura es disponer de expertos capaces de gestionar el flujo de los datos y su análisis para encontrar conocimiento.



Para construir mejores registros, es necesario el trabajo colaborativo de un equipo interdisciplinario, integrado tanto por médicos como por científicos de datos. Desde el Programa SATI-Q, compartimos plenamente esta visión.

En esta línea desde el año 2017, iniciamos un trabajo colaborativo estrecho con el Dr. Leo Celis, investigador principal del Laboratorio de Medicina Computacional del Massachusetts Institute of Technology (MIT) y nos sumamos al Proyecto GOSSIS (Global Open Source Severity Score Project). Esta iniciativa tiene como objetivo desarrollar un score abierto y sin costo a nivel mundial, para evaluar la severidad de los pacientes adultos internados en Unidades de Cuidados Intensivos (<https://www.youtube.com/watch?v=j7sPB6VNFwk>). Actualmente, nos encontramos trabajando en la fase de preprocesamiento de la base de datos, especialmente en los mapeos diagnósticos (<https://www.satiq.net.ar/investigacion>).

En el año 2018, el Dr. Antonio Gallesio y Ariel L Fernández (coordinador médico y responsable informático del Programa) participaron en el Critical Care Data Summit y Datathon organizado en Tarragona (España), donde tomaron contacto con referentes internacionales expertos en ciencias de datos e investigación clínica (<https://www.satiq.net.ar/investigacion>). En ese evento integraron el equipo que recibió el Segundo Premio del Datathon con el proyecto Measuring eICU data quality: an open and standardized approach, donde trabajaron en un modelo para evaluar la calidad de los datos en Big Data. (<https://www.satiq.net.ar/reconocimientos>)

Estos lazos de colaboración internacional, con el objetivo común de encontrar conocimiento a partir del análisis de grandes bases de datos, nos han permitido sumarnos a otros eventos internacionales como el Datathon Virtual realizado en Melbourne (Australia) en el contexto del Congreso Mundial de Terapia Intensiva (octubre 2019) en el que participaron la Dra. María del Pilar Arias y Ariel Fernández, en representación de SATI-Q (<https://www.satiq.net.ar/post/qu%C3%A9-es-undatathon>). También hemos participado de reuniones de especialización como la realizada en Uruguay en enero 2020 "An awakening in Medicine; humanity and Machines" (<https://www.satiq.net.ar/post/charla-del-dr-leo-celis-mit-an-awakening-in-medicine-humanity-and-machines-en-montevideo>) y en el Congreso Virtual de Inteligencia Artificial aplicada a Medicina (Colombia, octubre 2020; <https://www.acrvirtual.org/congreso-virtual-ia>)

A nivel regional, en el año 2019 nos hemos unido al Consorcio LOGIC (Linking of Global Intensive Care). Esta plataforma permite disponer de información agregada y actualizada en forma anual de distintos Registros Nacionales, Grandes bases de datos y Redes que integran el consorcio (ANZICS, CUBREA, CRITICAL CARE ASIA, UTIS BRASILERAS, NICE, etc). (<https://www.satiq.net.ar/post/somos-parte-de-logic>). Al mismo tiempo se trata de una red de investigación que ha generado diversas publicaciones colaborativas, cuyo objetivo central es la mejora del cuidado del paciente crítico a partir del Benchmarking (<https://www.sciencedirect.com/science/article/pii/S0883944120306729>).



Sin duda, el análisis inteligente de grandes bases de datos como la que dispone hoy el Programa SATI-Q (más de 200000 registros de internaciones en UCI adultos y 70000 en UCI pediátricas) tiene un gran potencial para generar conocimientos de resultados y prácticas en la vida real. Sin embargo, sabemos que no está libre de desafíos, como asegurar una adecuada calidad de los datos, protección de la información, manteniendo la representatividad con participación de unidades de distintos ámbitos y con distinta infraestructura. En ese sentido, continuamos trabajando junto con HL7 Argentina, bajo los estándares definidos por la Organización HL7 Internacional en la interoperabilidad del software SATI-Q, de forma que todas las unidades puedan participar del Programa independiente de su herramienta de registro de datos.

Por último, la necesidad de especialización en nuevas técnicas de análisis inteligente de grandes volúmenes de datos no es un tema menor. Estamos convencidos que el trabajo interdisciplinario y la generación de redes colaborativas a nivel local como internacional es el camino que nos permitirá descubrir conocimiento valioso en los datos que llevará a un mejor resultado en el tratamiento de los pacientes críticamente enfermos en el país.

Invitamos a los que quieran sumar y compartir experiencias con la gestión de datos, a participar de las reuniones mensuales del comité y avanzar en sinergia en este camino.