EDITORIAL

Ciencia de Datos: hacia la automatización de las decisiones

Analizar datos hoy en día es una tarea sumamente relevante por cuanto es un activo de gran valor tanto para investigadores como para organizaciones. Gracias al avance en tecnologías de la información y comunicación, hoy en día existen muchas plataformas que permiten registrar datos. Por ejemplo, los sistemas de información que apoyan a las organizaciones en la gestión de sus procesos registran todos los eventos en bases de datos tales como compras, traspasos de dinero, visitas médicas, entrega de productos, trámites notariales, etc. El comercio electrónico mueve millones de dólares diariamente en transacciones y con ello todos el detalle de cada transacción. Las redes sociales generan petabytes de datos diarios. Solo la red social Facebook tiene más de 1 billón de usuarios activos quienes aportan con cientos de petabytes de datos.

A comienzos del 2000, el autor Doug Laney introduce el términos de las 'V' que caracterizan a Big Data y que son los desafíos a estudiar: Volumen, Velocidad y Variedad de los datos. Ya en la primera década del 2000 se proyectaba el aumento en datos que produciría como resultado de la incorporación de nuevas tecnologías de información. Al analizar los datos que se generan con estas tecnologías, permite a las organizaciones tener mayor conocimiento permitiendo una mejor tomar de decisiones. Extraer conocimiento de los datos no es una tarea sencilla, ya que requiere de un proceso que depende las 'V' antes mencionadas. El procesar los datos genera información con un valor agregado que es el insumo para el descubrimiento de conocimiento que se basa en la información, valores y también en experiencia y cómo se hacen la cosas. Esto es un ciclo que se repite ya información y conocimiento a su vez generan más datos.

Hoy en día el mayor obstáculo que presentan las organizaciones no son los datos sino la comprensión de éstos para aplicar, análisis y mejorar los negocios¹ (LaValle, Lesser, Shockley, Hopkins y Kruschwitz, 2011), es por ello que en las últimas décadas han surgido áreas de investigación aportando con herramientas y técnicas de procesamiento y análisis de datos. Es así como Ciencia de Datos es una área interdisciplinaria que tiene como propósito transformar datos en valor² para poder reportar, diagnosticar, predecir y también recomendar soluciones o mejoras en productos, servicios y/o procesos.

Enmarcado dentro de Ciencia de Datos, Inteligencia Artificial y Machine Learning, desde que emergen como disciplina, han tenido un fluctuante impacto en desarrollo y conocimiento. Sin embargo, hoy en día, gracias a mayor madurez en tecnología y avances en la Ciencia de Datos, es posible avanzar con mayor firmeza mediante el cumplimiento de ciertas condiciones. Por ejemplo, mejoras en arquitectura de computadores que permiten mayor capacidad de almacenamiento y computo, librerías de código abierto de gran complejidad que permiten a los investigadores centrarse en análisis de procesar datos más que en la infraestructura y algo muy importante, la disponibilidad de datos abiertos³ que son la principal entrada de modelos matemáticos para comprender mejor la realidad.

No hay duda alguna que la pasada década se caracterizó por avances en Inteligencia Artificial. Entre los muchos proyectos que más relevancia han tenido en la comunidad destacan Zippedi, robot chileno

S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins and N. Kruschwitz. "Big data, analytics and the path from insights to value". MIT sloan management review. Vol. 52, Issue 2, pp. 21-32. 2011.

² W. Van Der Aalst. "Data science in action. In Process mining". Springer, Berlin, Heidelberg, pp. 3-23. 2016.

³ https://datos.bcn.cl/es/informacion/mapa-global-de-datos-abiertos

que permite escanear pasillos de supermercados para reposición stock y ahora en proceso creación de una plataforma que permite realizar compras online, sistemas de navegación que usan GPS en optimización de rutas, asistentes de voz como Google Home o Amazon Echo, entre muchos otros.

Es de vital importancia que las organizaciones puedan dar respuesta oportuna a los requerimientos de sus clientes, ya sea en productos o servicios. A través de la implantación de modelos de análisis de datos, es posible reducir el esfuerzo cognitivo que implica recuperar y procesar información para luego tomar decisiones. El esfuerzo debe estar puesto en gestionar información de manera automatizada y generar modelos que permitan tomar decisiones automatizadas que entregan respuestas precisas y oportunas. En este sentido las líneas de investigaciones deben estar orientadas en la extracción de conocimiento de datos no estructurados (área de estudio conocida como Procesamiento del Lenguaje Natural) que permita tomar requerimientos de clientes, procesarlos automáticamente y dar respuesta inmediata, el procesamiento de imágenes que busca patrones usando redes neuronales como modelo computacional que permite encontrar patrones con un porcentaje de probabilidad logrando grandes avances en áreas como visión por computador y reconocimiento de voz⁴.

Los desafíos que se presentan en decisiones automatizadas es saber el límite de usabilidad de los datos y poder responder preguntas como ¿quién es el dueño de los datos?, ¿se puede garantizar la integridad de los datos?, ¿cuál es límite en el uso de datos por parte de las organizaciones?, ¿qué datos son privados?. Son varias las preguntas que se podrían seguir planteando, sin embargo, la discusión central es cómo generar mejores servicios automatizados a través de aprendizaje automático y un análisis de datos tanto descriptivo, como predictivo donde las decisiones tengan un bajo costo y riesgo.

Mauricio Arriagada-Benítez, PhD. Facultad de Ingeniería, Departamento de Ingeniería en Computación e Informática, Universidad de Tarapacá. Arica, Chile.

E-mail: mauricio.arriagada@academicos.uta.cl

557

J.J. López Murphy y G. Zarza. "La Ingeniería del Big Data: Como Trabajar Con Datos". Print. Tecnología. Vol. 534. 2017.