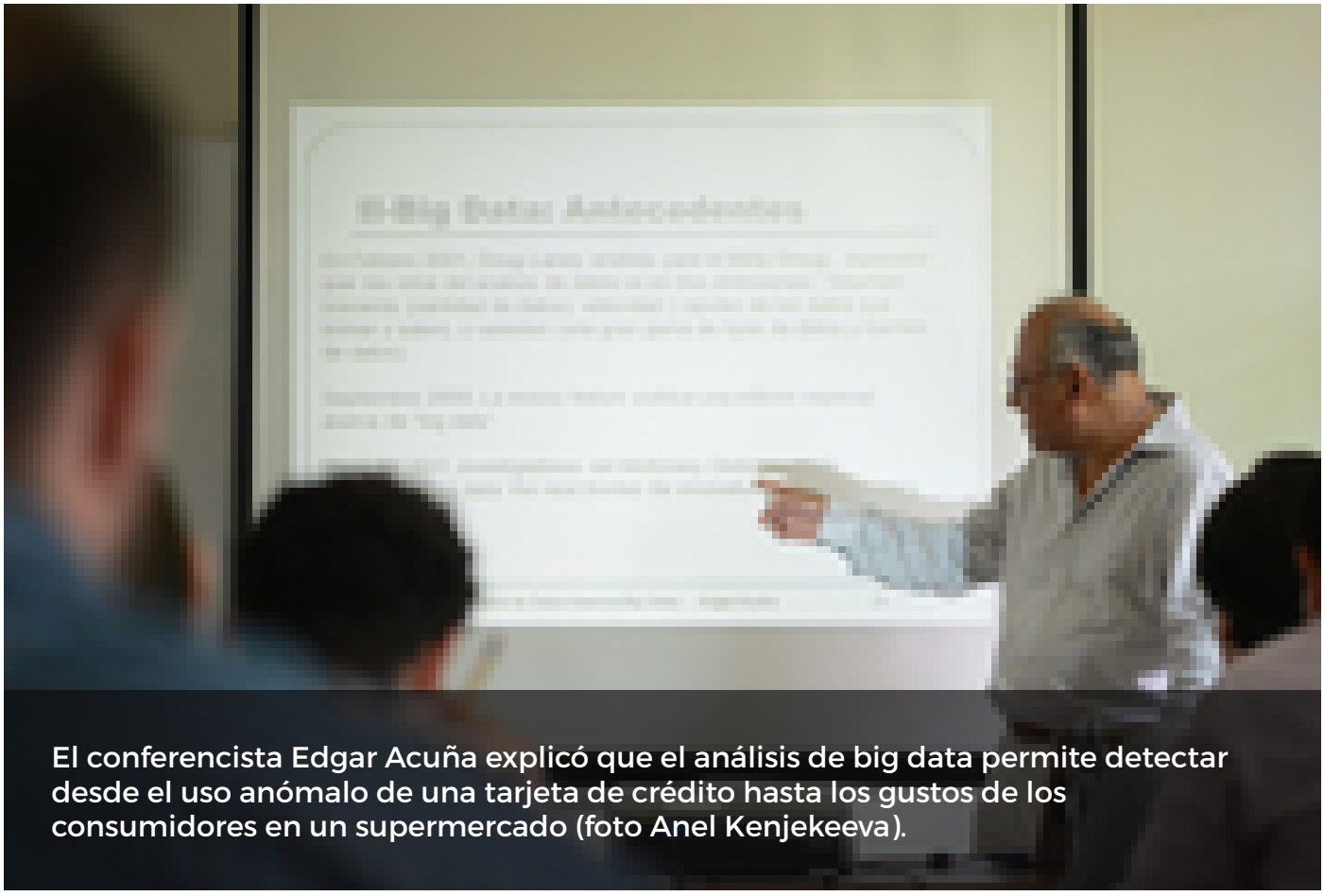
UNIVERSIDAD DE COSTA RICA

Minería de datos y Big data

Experto habla de la evolución y usos del análisis de datos masivos

19 NOV 2015 Innovación



El conferencista Edgar Acuña explicó que el análisis de big data permite detectar desde el uso anómalo de una tarjeta de crédito hasta los gustos de los consumidores en un supermercado (foto Anel Kenjekeeva).

Convertir los datos en conocimiento e información es una de las premisas de todas las ciencias. Hoy la tecnología y la globalización permiten a grandes compañías como Google, Twitter, Facebook, Amazon, Netflix, Linkedin manejar gran cantidad y variedad de datos cada segundo, lo que ha provocado que la estadística se transforme, con la ayuda de la matemática, en **minería de datos primero, hoy por hoy en Big Data y se vislumbra que para el futuro más próximo se denomine Ciencia de datos.**

Las nuevas herramienta computacionales, tendencias científicas, así como el papel que juegan las redes sociales y el negocio millonario detrás de Big data fueron algunos de los temas que desarrolló el especialista en la materia **Dr. Edgar Acuña Fernández, profesor de**

la Universidad de Puerto Rico, en una conferencia que organizó la semana anterior el Centro de Investigación en Matemática Pura y Aplicada (Cimpa) de la Universidad de Costa Rica.

Acuña, quien posee un doctorado en estadística de la Universidad de Rochester estableció algunas diferencias del manejo de los datos bajo el concepto de Big Data y como lo hacen los estadísticos. Una de las grandes diferencias es que **en estadística se utilizan muestras estructuradas y en Big data se analizan millones de datos por medio de un modelo o algoritmo** cuyos resultados permiten la toma de decisiones.

Big data implica el análisis de más de 100 gigabytes si es a nivel empresarial y de diez a 100 gigabytes en el campo académico. Se trata de datos no estructurados en gran volumen, que cambian a alta velocidad y provenientes de diversas fuentes por lo que tienen mucha variabilidad. Cien gigabytes se podría homologar con 25 600 canciones o 13 películas en alta calidad en DVD.

Recomendó Acuña que al igual que en estadística cuando se van a modelar los datos éstos deben estar "bien limpios", porque recordó que "basura que entra, basura que sale".

En años recientes después de migrar de minería de datos a Big data se está hablando de **Ciencia de datos**, esta última se trabaja con equipos interdisciplinarios: estadísticos, informáticos y matemáticos.

El público escuchó atento al especialista Acuña mientras exponía en el auditorio del Centro de Investigaciones en Matemática Pura y Aplicada, ubicado en la Ciudad de la Investigación (foto Anel Kenjekeeva).

Utilidad de un negocio millonario

El estadístico puertorriqueño hizo un recorrido histórico del concepto de Big Data y manifestó que ya en 1962 se publicó un artículo sobre análisis de datos. En 1977 la Asociación Americana de Estadística ofreció una conferencia sobre datos grandes y complejos y más adelante en los años 1994 y 1995 se dictaron conferencias sobre análisis de datos masivos.

El término Big data, dijo el conferencista, se acuñó en un artículo de una revista especializada en Ingeniería Eléctrica y Computación en 1997; y en el 2001 los profesionales en administración de negocios mencionan que para el análisis de Big data se debe tomar en cuenta el volumen, la velocidad y la variabilidad.

En los años 2011 y 2012 el término Big data se popularizó y con el advenimiento de la redes sociales y las posibilidades de guardar datos y realizar procesos en la nube **muchas compañías vieron los réditos del análisis de datos como un buen negocios en campos como el mercadeo, la política, la seguridad bancaria y los grandes buscadores en Internet.**

Equipos robustos

El Dr. Acuña ofreció algunos ejemplos de herramientas computacionales para el trabajo con grandes volúmenes de datos. **Advirtió que ya no es posible trabajar con una sola computadora, se requiere conglomerados (clúster) en una arquitectura de computación en Linux que se conecta a la web, para lograr un trabajo eficiente en este campo.**

También manifestó que se pueden usar varios programas mezclados para lograr mejores resultados y que algunas empresas como Amazon ofrecen cuentas gratis o espacios comerciales en la nube.

Compañías como IBM, Microsoft y Google han establecido sus propios programas para minería y Big data y algunos dan servicio en la nube para quienes realizan investigaciones particulares.

En el caso de la academia ya muchas universidades e institutos de investigación científica cuentan con laboratorios y equipos robustos para el análisis masivo de datos, afirmó el Dr. Acuña.

Elizabeth Rojas Arias
Periodista Oficina de Divulgación e Información
elizabeth.rojas@ucr.ac.cr

Etiquetas: big data, estadistica, mineria de datos, cimpa, edgar acuna.