

Big data para tomar decisiones inteligentes

Desde que Colombia se conectó a internet en 1994, EAFIT ha sido líder en el tema y en la actualidad participa en el liderazgo de la apropiación de nuevas tendencias en la región a través de iniciativas como Caoba.

La Universidad le apunta al uso de técnicas de inteligencia de datos mediante formación del talento humano, investigación aplicada y desarrollo de productos. Además, desde 2016, EAFIT participa en la Alianza Caoba, un Centro de Excelencia y Apropiación en Big Data y Data Analytics.

Mónica María Vásquez Arroyave
Colaboradora

Aunque el concepto *big data* (datos masivos, macrodatos o inteligencia de datos) está tan en boga en la actualidad, esta tecnología comenzó a usarse hace más de 20 años para solucionar problemas relacionados con almacenamiento y gestión de grandes volúmenes de información.

En este sentido, Juan Guillermo Lalinde Pulido, profesor del Departamento de Informática y Sistemas de EAFIT, afirma que este es más un término de moda, cuyo sentido y tema de fondo radica en el trabajo



Foto: Shutterstock

con analíticas, que consisten en procesar información para poder hacer una de tres operaciones: predecir, clasificar y prescribir. (Ver 'Aplicaciones de *big data*').

Debido a que las técnicas tradicionales de análisis de datos demandan mucho tiempo para procesar grandes volúmenes de información, las que aparecen con *big data* facilitan un análisis estadísticamente significativo para obtener lo mismo en menos tiempo: información de base para procesos eficientes de toma de decisiones.

Helmuth Trefftz Gómez, jefe del Departamento de Informática y Sistemas de EAFIT, explica que en *big data* se utilizan algoritmos como los basados en técnicas *MapReduce* –un modelo de programación para procesar grandes volúmenes de información–, así como en *machine learning* y *data mining* que permiten encontrar el sentido de los datos.

En comparación con los países desarrollados, en Colombia el principal problema es la falta de datos. Sin embargo, el profesor Trefftz sostiene: “Tenemos acceso a las máquinas e investigadores especializados en el tema con un conocimiento profundo de los algoritmos utilizados. Además, las empresas ya se preguntan cómo hacer *big data* e internet de las cosas. Así que vamos a vivir una etapa de ajuste, desarrollo y apropiación de la tecnología”.

EAFIT no es ajena a esta tendencia global y trabaja en iniciativas relacionadas con *big data* y analíticas, a través de la Alianza Caoba –un Centro de Excelencia y Apropiación en Big Data y Data Analytics–, además de apostarle a otros temas relacionados con ciencias y tecnologías de la información y la comunicación como el Centro de Computación Científica Apolo y el internet de las cosas.

“La visión de EAFIT es ser parte del proceso de apropiación de la inteligencia de datos en la región para contribuir con el desarrollo del *big data* y la analítica desde el punto de vista académico y de la generación de conocimiento. La meta es desarrollar capacidades y tener conocimientos de punta en esos temas para aportar conjuntamente soluciones al sector productivo social, dentro de la relación Universidad-Empresa-Estado”, afirma Félix Londoño González, director de Investigación de EAFIT.

Análisis de información

En opinión de Helmuth Trefftz, “los adelantos tecnológicos y los nuevos usos de estos se potencian mutuamente en el tiempo y crean un círculo virtuoso”.

so". Por ejemplo, en el sector salud es posible buscar datos sobre la cantidad de personas en el mundo con una enfermedad, información que se procesa luego con algoritmos de *machine learning* (aprendizaje de máquinas) para obtener hallazgos.

De hecho, Mark Mattingley-Scott, director de IBM Alemania, manifiesta que las iniciativas y aplicaciones en *big data* se relacionan con la salud, la industria, la logística y el comercio virtual.

"Hace cuatro o cinco años, el *big data* se enfocaba en tecnología, luego fue un elemento generador de capacidades de negocio y ahora estamos en una etapa en la que es también un factor específico para industrias y sectores particulares y especializados", opina Mattingley-Scott.

Para el directivo de IBM, el impacto de la tecnología es hacer todo más 'líquido' (flexible y adaptable), dinámico y mucho más cercano para el consumo. "El fin de la tecnología será alcanzar esa confluencia para anticipar lo que se necesita y proveer los medios para alcanzarlo en el momento y lugar requeridos, y eso redefinirá completamente la sociedad", asegura.

Por su parte, el investigador Juan Guillermo Lalinde señala que una empresa tiene que tener fuentes de información sobre sus clientes, el mercado, las interacciones en redes sociales, entre otra variedad de datos.

En este contexto de los negocios, los datos masivos permiten conocer al detalle la oferta y la demanda de cualquier mercado para tomar decisiones, acota Sergio Alonso Castrillón Orrego, jefe del pregrado en Administración de Negocios de EAFIT.

"De alguna manera se hace realidad la mano invisible de la que hablaba Adam Smith a finales del siglo XVIII. Lo que significa que hay mejores elementos y, si esos datos se convierten en información, es posible tomar mejores decisiones. Estos datos se capturan con distintos instrumentos y luego, para que tengan sentido, se procesan con mecanismos de inteligencia de mercados, razonamiento y conexión de variables", puntualiza el profesor Sergio Castrillón.

"La visión de EAFIT es fortalecer y consolidar las capacidades en generación de conocimiento en *big data* y analítica de datos de manera articulada con los programas de formación": Félix Londoño González, director de Investigación de EAFIT.

No obstante, para analizar cada tipo de información hay técnicas especiales. "No es lo mismo predecir riesgos financieros asociados con una persona, que identificar quiénes son los usuarios que jalonan las tendencias en las redes sociales. De ahí que *big data* no es una técnica en sí misma, sino que reúne todas las técnicas que se están desarrollando para poder analizar grandes volúmenes de información", indica el docente Juan Guillermo Lalinde.

Una vez obtenida la información, esta se debe procesar utilizando la capacidad computacional y de almacenamiento disponible. Mientras algunas empresas envían discos duros al proveedor de capacidad de cómputo para que los copie directamente en su máquina y procese la información, otras almacenan la información directamente en la nube para procesarla posteriormente.

Para procesar la información, Juan Guillermo indica que existen herramientas matemáticas para manejar altas dimensionalidades, es decir, problemas



Foto: Robinson Henao

Desde el segundo semestre de 2016 comenzó la planeación y el diseño de un nuevo laboratorio de IoT en la Institución.

+



Foto: Robinson Henao

Uno de los objetivos del Centro de Excelencia Caoba es tener su propia infraestructura para desarrollar proyectos internos y externos.

matemáticos que se resuelven en cientos o miles de dimensiones. También, hay técnicas de aprendizaje de máquinas, aprendizaje profundo, técnicas de inteligencia artificial, técnicas de optimización heurísticas y análisis econométrico que se pueden utilizar para analizar información.

“Es un tema transversal y no hay una herramienta específica para *big data* sino técnicas y herramientas matemáticas para resolver problemas, dependiendo de la complejidad de cada uno de estos”, afirma el profesor Lalinde.

Alianza Caoba

Desde el primero de enero de 2016 EAFIT participa en la Alianza Caoba, un Centro de Excelencia y Apropiación en Big Data y Data Analytics financiado por el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) y gestionado desde el Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias).

Se trata del primer acuerdo público-privado que promueve el *big data* y la analítica en Colombia, que hace parte de la estrategia del Gobierno Nacional de crear dos centros estratégicos para el país –el otro centro es de internet de las cosas– para fortalecer la generación de soluciones de análisis de información y generar estrategias en diversos sectores industria-

les, gubernamentales y académicos, a través de la formación del talento humano, la investigación aplicada y el desarrollo de productos.

Así lo explica Edwin Nelson Montoya Múnera, jefe del pregrado en Ingeniería de Sistemas de EAFIT, sobre esta alianza de la que también hacen parte los grupos Bancolombia y Nutresa, IBM de Colombia, SAS Institute Colombia, EMC Information Systems Colombia, Cluster Creativ, el Departamento Nacional de Planeación (DNP) y las universidades Icesi, EAFIT, Los Andes y Javeriana.

Por parte de EAFIT, a través de la relación interinstitucional, en la Alianza Caoba participan ocho investigadores de las escuelas de Ingeniería, Ciencias y Economía y Finanzas, así como representantes en la junta directiva y el comité científico, quienes aportan en el desarrollo de proyectos y en la formación de talento humano de alta calidad en el país a través de programas de maestría y doctorado de la Institución.

De hecho, gracias a la Alianza, en la actualidad se forman 30 estudiantes de maestría y nueve de doctorado en *big data*, una cuota importante en el país para que estas personas participen en proyectos en las empresas, manifiesta el docente Montoya Múnera. De esos 30 estudiantes, por parte de EAFIT se forman cinco estudiantes de maestría y tres de doctorado. Además, un investigador ocupa una posición posdoctoral en Caoba. Dicha cualificación se da en programas ya establecidos en las tres escuelas mencionadas.



Foto: Robinson Henao

La propuesta de valor de la Alianza Caoba se basa en la generación de soluciones en torno a las tecnologías de análisis de datos.

Caoba tiene también como objetivo el desarrollo de proyectos, inicialmente en los primeros dos años, con las entidades ancla (Bancolombia, Nutresa y DNP), y se espera que en tres años, luego de tener una mayor consolidación, se ofrezcan servicios de asesoría, consultoría y proyectos conjuntos para diferentes compañías nacionales o internacionales.

“Es importante ser innovador en el ámbito tecnológico y en el de negocios. No se trata siempre de ser el primero, pues, por ejemplo, Google no fue el primer buscador, sino Altavista”: Mark Mattingley-Scott, director de IBM Alemania.

+

Otra de las apuestas de esta Alianza es crear una estructura administrativa y técnica, a partir de un modelo de gestión, operación y negocio, para consolidar dentro de EAFIT un centro con una infraestructura tecnológica y con un director general, un director de proyectos o un director de arquitectura que lo materialicen.

“Queremos crear un centro dentro de la Universidad para responder a proyectos de investigación, desarrollo e innovación, así como unificar los esfuerzos y la capacidad de la Institución en el tema. Por eso, como *big data* es un área interdisciplinar, nos unimos cinco escuelas –Ingeniería, Ciencias, Administración,

Humanidades y Economía y Finanzas– comprometidas con este tema e iniciamos el proyecto de crear una maestría en Big Data”, precisa Edwin Montoya.

Por ahora, el trabajo de Caoba se centra, principalmente, en análisis financiero, análisis transaccional y hábitos de consumo, en el caso de Bancolombia; y en el tipo de mercado y el público objetivo de los productos, en el caso de Nutresa.

A escala nacional, con el DNP trabajan en proyectos de movilidad inteligente en Bogotá y se tienen planteados otros como evaluación de la base de impuestos, detección de fraude en entidades y proyectos públicos, y mejoramiento de la eficiencia del sistema de Sisbén. Esto, apunta el profesor Montoya, requiere cruzar mucha información en todo el país para conocer si los afiliados son quienes realmente deben estar. En el sector de educación se pretende medir el impacto de la política pública y del nivel de inversión de los proyectos nacionales.

“EAFIT ha estado trabajando también en proyectos en el área de salud con la IPS Universitaria, donde exploran alternativas de cómo pueden, a partir de información médica y clínica y de ayudas diagnósticas, extraer información y conocimiento para generar valor y mejorar la eficiencia del sistema en salud”, puntualiza el jefe del pregrado en Ingeniería de Sistemas de la Institución.



Aplicaciones de big data

Predecir: en medicina permite establecer correlaciones entre factores genéticos y enfermedades específicas para predecir el desarrollo de algunas dolencias si una persona cumple con determinadas categorías o propiedades. Las empresas de seguros lo utilizan para calcular la tarifa o la tasa del seguro que le van a cobrar a una persona, dependiendo de la forma como conduce su auto.

Clasificar: las empresas lo usan para identificar el tipo de clientes y los segmentos de mercado para establecer cómo impactar, de manera específica, cada segmento.

Prescribir: permite tomar acciones para cambiar tendencias indicadas por los datos. Cuando se estudia la tasa de crimen en una ciudad y los factores que influyen en esto, dicha información sirve no solo para predecir los crímenes, sino para tomar medidas activas, por ejemplo, ubicar agentes de policía o cambiar rutas de patrullas.

La Escuela de Ciencias, a través del grupo de investigación en Modelado Matemático liderado por la profesora Lucía Quintero Montoya, trabaja en procesamiento de señales y análisis de información de voz.

Otros proyectos se relacionan solo con el tema de analítica, por ejemplo, recoger la información de los sensores del carro solar EAFIT-EPM, la cual se procesa para optimizar variables, dispositivos, entre otros.

+

“La visión de EAFIT es ser parte del proceso de apropiación de la inteligencia de datos en la región para contribuir con el desarrollo del big data y la analítica desde el punto de vista académico y de la generación de conocimiento”: Félix Londoño González, director de Investigación de la Institución.

Internet de las cosas

La posibilidad de interconectar objetos cotidianos con internet, tener una casa inteligente o programar tareas de manera remota es una realidad que en la actualidad es posible gracias al internet de las cosas (IoT, por su sigla en inglés de Internet of things).

Sobre este concepto, el investigador Helmuth Trefftz Gómez menciona que los países desarrollados llevan una ventaja. Por ejemplo, muchos de los carros de la compañía alemana Opel (filial de General

Motors) tienen sensores que, sin que el dueño tenga que esforzarse, le envían datos a la empresa con información referente al estado de las piezas. Incluso puede notificar que se ha pasado de los kilómetros para los que están diseñadas.

También se puede conocer sobre el tipo de cajas que tienen que diseñar para un país montañoso como Colombia, el porcentaje de usuarios que transitan por lomas, la manera en que emplean los cambios o el uso que un conductor le da a un camión para determinar las implicaciones en la reducción de la vida útil del vehículo.

En el caso de Colombia, el docente Trefftz considera que primero se vivirá una etapa de ajuste –por la que ya han pasado los países desarrollados– para establecer, a partir de las empresas y sobre todo de los consumidores, qué funcionalidades sirven, cuáles se deben descartar, cuáles se deben adoptar, cuáles adaptar y cuáles desarrollar.

Privacidad de la información

Como al analizar datos se puede descubrir información privada de una persona, es necesario que el *big data* tenga en cuenta la Ley de *Habeas Data* sobre protección de datos personales.

Por eso, cuando se recopilan –señala el docente Juan Guillermo Lalinde– se deben tomar todas las medidas necesarias para que un tercero no pueda inferir información confidencial de personas o instituciones mediante el análisis de los datos.

“Esto es especialmente sensible en temas como datos personales, historia clínica, propiedades o historia crediticia. Así mismo, los resultados obtenidos mediante técnicas de aprendizaje de máquinas se perciben como objetivos, en muchos casos, cuando estas técnicas conducen a sistemas que ‘aprenden’ de los prejuicios sociales y los replican, pero al no ser ciertos pueden causar mucho daño a la sociedad”, concluye Juan Guillermo.

Con técnicas de análisis de datos, el área de Mercadeo de EAFIT caracteriza el entorno cultural de Medellín a partir de las interacciones que se dan en las redes sociales.

+ Industria 4.0

Aunque EAFIT no hace parte del Centro de Excelencia en IoT, el profesor Edwin Montoya plantea que dentro de la Institución se requieren unas capacidades conjuntas para poder realizar proyectos de asesoría, consultoría y formación en este tema.

En este sentido, desde el segundo semestre de 2016 comenzó la planeación y el diseño de un nuevo laboratorio de IoT, para lo cual han establecido contactos nacionales e internacionales para configurar esa área en la Universidad, mediante la conformación de una red de investigadores y académicos que respondan a dichas necesidades.

En el Simposio Internacional en Tecnologías Informáticas (SITI), realizado en EAFIT el 20 y 21 de septiembre de 2016, se discutió “la integración de IoT con *big data*, es decir, cómo se combinan en un concepto conocido como Industria 4.0 [industria inteligente], que es la nueva visión de integrar una fábrica con sistemas de monitoreo y análisis de datos”, comenta el investigador Edwin.



Foto: Shutterstock

Apolo complementa la computación de alto rendimiento

En la actualidad hay dos tipos de arquitecturas en computación de alto rendimiento, una para procesar *big data* y otra para computación científica, pero ninguna permite hacer ambas cosas, explica el profesor Juan Guillermo Lalinde Pulido.

Por eso, como complemento, EAFIT cuenta con el Centro de Computación Científica Apolo que, aunque no es una infraestructura para manejar grandes volúmenes de información —es decir, para *big data*—, sí es una máquina que sirve para resolver problemas matemáticos como la evaluación de modelos y la ejecución de simulaciones, afirma el investigador Lalinde.

El nuevo Apolo tiene una capacidad de almacenamiento de 27 teras (27.648 gigas) y está diseñado “para la función que siempre ha cumplido en la Universidad: facilitar a los investigadores el uso de computación científica, por ejemplo, en procesos intensivos en cómputo en sus investigaciones”, manifiesta el catedrático Lalinde.

+

Investigadores responsables de Caoba

Departamento de Ingeniería de Sistemas: Edwin Nelson Montoya Múnera (investigador principal), Juan Guillermo Lalinde Pulido, Marta Silvia Tabares Betancur y Rafael David Rincón Bermúdez.

Departamento de Economía: Andrés Ramírez Hassan y Diego Alexander Restrepo Tobón.

Departamento de Ciencias Matemáticas: Lucía Quintero Montoya.

Departamento de Ciencias Biológicas: Javier Correa Álvarez.

Ocupa una posición posdoctoral en Caoba: Daniel Sierra Sosa.

Estudiantes del doctorado en Ingeniería: Leandro Fabio Ariza Jiménez y Santiago Rodríguez Cadavid.

Estudiante del doctorado Finanzas: Carolina Crispín Fory.

Estudiantes de la maestría en Ingeniería: Santiago Angee Agudelo y Camilo Torres Botero.

Estudiante de la maestría en Finanzas: Jim Anderson Sánchez González.

Estudiante de la maestría en Economía: Edwar Antonio Londoño Zapata.

Estudiante de la maestría en Matemáticas Aplicadas: David Ortiz Puerta.