

Innovaciones en la prestación de servicios públicos

Número 4

El análisis predictivo: impulsando la mejora a partir de los datos

Stephen Goldsmith, Susan Crawford
y Benjamin Weinryb Grohsgal

Instituciones para
el Desarrollo

División de Capacidad
Institucional del Estado

DOCUMENTO PARA
DISCUSIÓN N°
IDB-DP-440

Innovaciones en la prestación de servicios públicos

Número 4

**El análisis predictivo:
impulsando la mejora a partir de los datos**

**Stephen Goldsmith, Susan Crawford
y Benjamin Weinryb Grohsgal**

Julio de 2016

<http://www.iadb.org>

Copyright © 2016 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Contacto: Pedro Farias, pfarias@iadb.org.

Innovaciones en la prestación
de servicios públicos
Número 4
El análisis predictivo:
impulsando la mejora a partir de los datos

Autores: Stephen Goldsmith, Susan Crawford
y Benjamin Weinryb Grohsgal



Prólogo	1
Introducción	5
Cómo crear un enfoque empresarial para el análisis predictivo.....	11
Ámbitos de políticas clave para el análisis predictivo.....	21
Anexo A. Guía	33
Referencias.....	41

La complejidad es el rasgo que define a un mundo cada vez más interconectado y da cuenta de los desafíos apremiantes que enfrenta el mundo en el siglo XXI: cambio climático, estabilidad de los mercados, disponibilidad de energía y recursos, pobreza, y conflicto, entre otros.¹ Por otra parte, es la nueva realidad de operación de los gobiernos: ciudadanos más exigentes y más conscientes de sus derechos con un mejor acceso a la información a través de la tecnología y con expectativas de servicios de alta calidad; restricciones presupuestarias cada vez mayores a la vez que aumenta la demanda de servicios públicos; y una cambiante composición demográfica, que repercute en la provisión de servicios. En este contexto, las autoridades se ven obligadas a diseñar soluciones innovadoras que permitan hacerle frente a los desafíos.

Esta complejidad se refleja en los cuantiosos volúmenes de datos que producen tanto ciudadanos como organismos estatales y que los gobiernos deben encauzar mediante el análisis predictivo si quieren tener una idea más acabada de las necesidades de los ciudadanos y una asignación de recursos públicos más eficiente.

¿De qué se trata el análisis predictivo? En pocas palabras, es el uso de datos históricos para identificar patrones, lo que permite anticipar necesidades futuras. Algunos países de América Latina ya están utilizando este método innovador para apoyar procesos de toma de decisiones más efectivos a través de la concientización sobre situaciones en tiempo real: Montevideo está siguiendo el ejemplo de Los Angeles, Santa Cruz y Atlanta al utilizar datos básicos sobre delito (tipo, momento y lugar). El objetivo es analizar patrones de comportamiento delictivo y así colaborar con la policía de la ciudad en la mejora de los cronogramas de patrulla de los barrios sobre la base de la identificación de puntos de delito cruciales.² El gobierno de México ha utilizado datos relativos al consumo de electricidad a fin de elaborar previsiones en tiempo casi real (*nowcasts*) del PIB trimestral, lo que permite satisfacer la necesidad de contar con previsiones actualizadas sobre la actividad económica para formular y evaluar políticas.³ A su vez, Colombia está utilizando datos de búsquedas en Google para inferir la actividad económica a nivel sectorial; los indicadores, que se obtienen de manera mucho más oportuna, son una alternativa a las estadísticas tradicionales.⁴

Más allá de estos logros, la penetración del análisis predictivo en el sector público de América Latina y el Caribe (ALC) ha sido lenta. Poner los datos a trabajar en función de los gobiernos de la región requiere un proceso y una transformación cultural profundos dentro de los organismos. Si bien muchos gobiernos de ALC están modernizando la gestión pública, la fragmentación institucional y los procedimientos administrativos complejos continúan siendo los principales obstáculos hacia un

¹Véase http://www3.weforum.org/docs/WEF_GAC_PerspectivesHyperconnectedWorld_ExecutiveSummary_2013.pdf.

²Véase <http://www.elpais.com.uy/informacion/predpol-software-policia-predecir-delitos.html>.

³Previsiones en tiempo casi real del PIB de México sobre la base del consumo de electricidad (2015).

⁴Véase <http://cea.cepal.org/sites/default/files/lacbigdatansopapernov10finaldraft.pdf> p.25.

gobierno efectivo, eficiente y abierto. Durante años, el Banco Interamericano de Desarrollo (BID) ha estado trabajando con los gobiernos de ALC en pos del fortalecimiento de las capacidades de gestión y del mejoramiento de la calidad de los servicios que estos suministran. El apoyo incluye asistencia técnica y financiera y la generación e intercambio de conocimiento sobre políticas a fin de comprender más cabalmente los factores impulsores del cambio institucional y la modernización del gobierno que mejorarán la prestación de servicios públicos y favorecerán la innovación en el sector público.

Como parte de esta labor, el BID ha creado la serie de documentos de discusión denominada “Innovaciones en la prestación de servicios públicos”. Entre los casos previos y las lecciones aprendidas que se documentan, analizan y comparten dentro de la serie figuran: “Can 311 Call Centers Improve Service Delivery? Lessons from New York and Chicago”; “Los servicios en línea como derecho ciudadano: El caso de España”; y “Descomplicar para avanzar: O caso Minas Fácil”. El presente documento examina la forma en que los funcionarios pueden utilizar datos históricos debidamente analizados para rastrear patrones y tendencias con el fin de reorganizar la forma en que prestan servicios, anticipan acontecimientos futuros y, en muchos casos, incluso previenen eventuales problemas. Enumera ejemplos de uso del análisis predictivo y sus implicaciones en las acciones futuras del gobierno a partir de dos casos de estudio pormenorizados: los proyectos piloto de análisis predictivo que actualmente se están llevando a cabo en Chicago a fin de potenciar los resultados operacionales en determinados departamentos de la ciudad, y el proceso seguido por el estado de Indiana en la creación de un Centro de Administración y Desempeño que permitirá aplicar el análisis en las bases de datos estatales. El documento ofrece un catálogo de uso del análisis predictivo por parte de gobiernos de todo el mundo, organizado por ámbito de políticas, que puede aportar lecciones importantes para los gobiernos de ALC que procuren adoptar soluciones similares. Áreas clave de las políticas públicas entre las que se cuentan salud, seguridad pública, servicios sociales e impuestos podrían beneficiarse del uso del análisis predictivo. Por último, los autores presentan una hoja de ruta resumida que marca el rumbo hacia la implementación del análisis predictivo en la administración pública.

Stephen Goldsmith, uno de los tres autores del presente documento, es Profesor “Daniel Paul” de la Práctica de Gobierno y Director del Programa de Innovaciones en el Gobierno de Estados Unidos de la Escuela de Gobierno Kennedy de la Universidad de Harvard. En la actualidad, tiene a cargo el proyecto Data-Smart City Solutions, que tiene como objetivo destacar las labores de los gobiernos locales destinadas a utilizar nuevas tecnologías que conecten los avances decisivos relativos al uso del análisis de datos masivos con los apartes de la comunidad, a fin de redefinir la relación entre el gobierno y los ciudadanos. Susan Crawford es Profesora Invitada “John A. Reilly” de Propiedad Intelectual en la Escuela de Leyes de Harvard. Trabajó para el Presidente como Asistente Especial en la elaboración de políticas relativas a ciencia, tecnología e innovación en 2009, codirigió el equipo de transición entre las administraciones Bush y Obama de la Comisión Federal de Comunicaciones (FCC, por sus siglas en inglés), y fue miembro del Consejo Asesor del Alcalde Michael Bloomberg en asuntos de tecnología e innovación. Ben Weinryb Grohsgal es Ingeniero en Software en BuzzFeed. Anteriormente se desempeñó como Consultor Senior para Booz Allen Hamilton, la iniciativa Data-Smart City Solutions, la Oficina del Alcalde de Innovación Cívica en la ciudad de San Francisco, y la Oficina de Desarrollo Económico en Littleton (Colorado).

Este documento pudo realizarse gracias a la colaboración entre el BID y el Centro Ash para la Gobernabilidad Democrática e Innovación del Kennedy School, con la coordinación del Profesor Goldsmith. Su importancia reside en que sienta las bases para la serie de documentos de discusión que coordina Pedro Farías, Especialista Principal de Modernización del Estado en la División de Capacidad Institucional del Estado del BID (ICS).

Los autores desean destacar que la presente publicación fue posible gracias al apoyo técnico y financiero del Fondo de Fortalecimiento de la Capacidad Institucional del BID (ICSF), que cuenta con el financiamiento del Gobierno de la República Popular China, y de la República de Corea a través del Fondo de Fortalecimiento de Capacidad. Asimismo, desean expresar su agradecimiento a Pablo Valenti y Alejandro Pareja por sus valiosos comentarios durante el proceso de revisión, así como a Ana Catalina García de Alba y Florencia Cabral por sus aportes, sin los cuales el documento no habría prosperado.

Miguel Porrúa

*Especialista Líder en Gobierno Electrónico
División de Capacidad Institucional del Estado
Sector de Instituciones para el Desarrollo
Banco Interamericano de Desarrollo*

Aunque los gobiernos tienen a su disposición gran cantidad de datos electrónicos tanto actuales como históricos, el conocimiento y las tendencias que la inspección sistemática de esos datos podrían revelar suelen permanecer ocultos para los funcionarios públicos. En su mayor parte, los sistemas gubernamentales de tecnología de la información se desarrollan como medio para procesar transacciones aisladas. Los datos almacenados en formatos incompatibles y en equipamientos obsoletos se utilizaban para llevar a cabo esas operaciones aisladas, y le permiten al gobierno seguir haciendo lo que venía haciendo, pero no relacionar esos datos con otros. Como resultado, el conocimiento que podría adquirirse recurriendo a la información generada mediante los diversos sistemas no estaba al alcance del gobierno; ciertamente, los gobiernos no sabían lo que sabían.

Hoy en día, el hardware es cada vez menos costoso, el tiempo de procesamiento de datos disminuye rápidamente, las técnicas de minería de datos pueden eludir los problemas de interoperabilidad con los sistemas tradicionales y la administración de los gobiernos es cada vez más profesional y contrata a personal que se siente más cómodo con el uso de datos para gestionar el rendimiento y predecir las tendencias. Están surgiendo dirigentes que tienen la visión y la energía personal necesarias para instar a los organismos a que trabajen juntos a fin de sacar provecho de los datos y lograr mejores resultados para los ciudadanos. En consecuencia, una nueva era del análisis predictivo está surgiendo en todo el mundo. Este documento proporciona un conjunto de herramientas y marcos que serán de utilidad para aquellos empleados del sector público que consideren la posibilidad de recurrir al análisis predictivo para aumentar la eficacia de las operaciones gubernamentales.

¿Qué es el análisis predictivo?

El análisis predictivo es el uso de datos históricos para identificar patrones y tendencias que pueden ayudar a anticipar el futuro. Las empresas utilizan tales prácticas constantemente: por ejemplo, Amazon predice qué libros querrán leer a continuación sus clientes y les facilita la compra; CVS, una gran cadena de farmacias de Estados Unidos, sabe cuándo sus clientes necesitan reponer sus multivitaminas antes que ellos mismos, y a partir de los datos de ventas anteriores les envía cupones para fomentar que vuelvan a comprarlas; mientras que Walmart almacena una mayor cantidad de agua embotellada, cinta adhesiva y aperitivos cuando se prevén importantes fenómenos meteorológicos adversos. La campaña de Obama utilizó el análisis predictivo para identificar a votantes indecisos con el fin de persuadirlos. El Departamento de Seguridad Nacional de Estados Unidos está desarrollando un software que puede predecir qué individuos pueden suponer un riesgo de seguridad atendiendo a su comportamiento mientras esperan en la cola de la inspección de

seguridad, mediante la utilización de datos sobre el nivel de inquietud, la sudoración, los movimientos oculares y la frecuencia cardíaca.

Predecir el futuro no es una nueva tendencia. Durante décadas, los encargados del pronóstico meteorológico de la televisión han pronosticado si iba a haber sol o llovería, y los encuestadores pronosticaban quién iba a ganar las elecciones. Lo novedoso actualmente es el gran volumen y la ubicuidad de los datos que se pueden utilizar para realizar predicciones. Ahora es posible aprovechar la potencia de procesamiento necesaria para analizar una enorme cantidad de datos y combinar y comparar bases de datos complejas, todo en tiempo real y con dispositivos cada vez más pequeños.

Al igual que lo vienen haciendo desde hace algún tiempo empresas de Internet como Google y Facebook, los gobiernos pueden ahora recurrir al análisis predictivo para orientar mejor sus actividades. La idea es sencilla: a partir de datos históricos sobre características de un tema determinado (por ejemplo, una red de transporte, un restaurante, niños obesos, edificios que podrían sufrir incendios), los estadísticos aíslan las características que parecen correlacionarse de forma clara con ciertos resultados. También pueden buscar anomalías o valores atípicos en los datos acerca de esas características. A continuación, llevan a cabo rigurosas pruebas —primero mediante el uso de datos exclusivamente y luego sobre el terreno— para evaluar si al darle prioridad a determinadas actividades en relación con esas características (cambiar el horario de los trenes, clasificar a los restaurantes de otra manera, proporcionar mejores instalaciones para hacer ejercicio físico, inspeccionar primero determinados edificios en relación con el riesgo de incendio) daría lugar a una mejora de la situación actual respecto de un resultado deseado. Una vez que se validan tales predicciones, se obtienen motores operacionales y de política que previenen las consecuencias no deseadas que se identificaron. Entonces se repite el proceso. No obstante, no todos los problemas se adaptan a este método; a veces lo que se necesita para resolver un problema no es más que la visibilidad de datos de distintas bases de datos, y no una predicción. Los problemas que implican la necesidad de priorizar una lista de tareas gubernamentales son los más susceptibles al uso del análisis predictivo.

Los problemas que deben resolver las empresas y los gobiernos son similares y el análisis predictivo puede resultar de ayuda. Ciertas tareas que podrían parecer imposibles se tornan posibles, tal como se ilustra en el cuadro 1.

Esta descripción del análisis predictivo es a la vez excesiva y escasamente inclusiva. No capta el dinamismo de los sistemas o la complejidad de un mundo en que se recogen a diario miles de millones de datos. La sociedad actual vive en un ambiente hiperconectado; por lo tanto, es posible que las predicciones de los resultados —y los mecanismos de aplicación (o “tratamientos”) que pueden ser probados para abordarlos— tengan que cambiarse sobre la marcha como reacción a pequeños cambios en el mundo que acaban teniendo efectos no lineales. Al mismo tiempo, algunos déficits básicos en los recursos gubernamentales hacen que hasta los más sencillos experimentos de análisis predictivo sean extraordinariamente difíciles de llevar a cabo.

Es importante que el análisis predictivo se utilice como herramienta para mejorar el criterio y la eficacia de los empleados públicos, y no como sustituto de la intuición. Como señala IBM, los mejores profesionales que trabajan con jóvenes “hacen juicios perspicaces sobre cuál de las personas que atienden está en riesgo”, y “el análisis predictivo simplemente proporciona pruebas empíricas robustas para apoyar sus conclusiones profesionales” (IBM, 2011a). Al mismo tiempo,

Cuadro 1. Análisis predictivo en los sectores público y privado

Aplicación en la empresa	Aplicación en el sector público
Un proceso conlleva un gran número de decisiones similares, pero cada decisión requiere considerar una amplia gama de variables.	Inspecciones
Tanto el problema como el resultado implican a un sistema y no solo a un organismo (es decir, abordar la manifestación del problema puede hacer pasar por alto buena parte de la causa subyacente).	Absentismo escolar (seguridad pública y educación)
Hay disponible una abundante cantidad de información en formato de datos electrónicos que sirve de base para tomar decisiones y medir resultados.	Detección de fraudes
Es posible introducir un modelo de cálculo en el proceso institucional, ya sea para automatizar decisiones o para apoyar a los que las toman.	Asistencia social individual
Se tienen en cuenta explícitamente las condiciones de privacidad, normalización, seguridad de los datos y transparencia.	Recaudación de impuestos

estas técnicas pueden ser útiles en la detección de valores atípicos que podrían escapar a la atención incluso de los profesionales. El Grupo CGI (2013) ha comprobado que los modelos predictivos pueden ayudar a producir valor permitiendo a los trabajadores llegar a la conclusión correcta de forma más rápida, económica y eficaz; en algunos casos, se pueden automatizar procesos tradicionales enteros con la ayuda del análisis. En definitiva, el análisis predictivo no debe sustituir al conocimiento y a la experiencia local; al contrario, tiene mayor utilidad como apéndice o complemento de las prácticas existentes.

En el sector privado, los accionistas de las empresas se benefician cuando el análisis predictivo impulsa mejoras en las prácticas empresariales y un aumento de la cuota de mercado. Cuando el gobierno utiliza estas mismas técnicas, quienes se benefician son los ciudadanos.

¿Por qué ahora?

Los gobiernos de todos los niveles se enfrentan a demandas de servicios diarias que superan con creces los recursos disponibles para producirlos; por lo tanto, los funcionarios se ven obligados a ofrecer servicios gubernamentales de mejor calidad, más económicos y más rápidos. El análisis predictivo permite a los empleados del sector público orientar los recursos a las áreas donde más se necesitan, mejorando la capacidad de toma de decisiones de los empleados públicos, al tiempo que abre oportunidades de reestructuración de los procesos institucionales. Utilizando datos, los gobiernos pueden comprender los riesgos y tomar las precauciones necesarias con mucha mayor precisión, ya que los recursos se dirigirán a las áreas donde la mitigación del riesgo es más importante. En pocas palabras, el análisis predictivo permite a los gobiernos ser más reactivos a las necesidades de los ciudadanos en una época de recursos públicos limitados.

No es la tecnología lo que nos impide alcanzar esta nueva normalidad de un gobierno reactivo; el problema suelen ser más bien las estructuras de gobierno desfasadas. Los países con sistemas avanzados de administración pública han logrado importantes mejoras en la reducción de la corrupción, exigiendo a los empleados operar dentro de un estrecho conjunto de normas que acotan la discrecionalidad, y controlando, inspeccionando y anticipando el cumplimiento de esas normas. Lo anterior tenía sentido en un mundo sin análisis predictivo; los gobiernos no podían distinguir entre el buen ejercicio de la discrecionalidad y su uso corrupto, y trataban de poner límites a ambos.

Sin embargo, las herramientas digitales actuales permiten a los gobiernos la posibilidad de otorgar una mayor discrecionalidad a la vez que siguen más de cerca el tiempo empleado en las tareas. Las mismas herramientas que ayudan a la administración a encontrar pérdidas de agua antes de que ocurran y a identificar a los contribuyentes que pagan menos de lo que deberían también pueden identificar valores atípicos entre los empleados públicos. Los datos sobre qué exactamente está haciendo un día un empleado, el número de infracciones que pone un inspector y a quién o cuándo, y qué funcionarios u organismos públicos recurren a proveedores de servicios privados también ayudarán a identificar valores atípicos. Creemos que los gobiernos no pueden permitirse el lujo de no aprovechar estas oportunidades.

¿Qué obstáculos hay que superar?

Para utilizar de forma satisfactoria los datos con vistas a predecir y resolver problemas es fundamental superar una serie de obstáculos que no son técnicos. El primero de tales desafíos implica una falta de claridad respecto de los valores públicos a los que el gobierno intenta servir mediante los datos. Las oportunidades más interesantes se centran en mejorar más bien los resultados que los productos: mejorar la salud pública en lugar de aumentar el número de hospitales públicos; reducir el número de personas sin hogar en lugar de aumentar el número de camas en albergues; hacer que el país sea más seguro en lugar de arrestar a más delincuentes. Sin embargo, definir el problema que debe resolverse es una tarea difícil.

El segundo desafío organizativo radica en la estructura de gobierno. Los funcionarios bien intencionados que traten de utilizar los datos para resolver un problema sistémico que se extiende a través de múltiples departamentos y organizaciones no gubernamentales tendrán problemas si operan en un ámbito estrecho definido por una actividad.¹ Intercambiar datos y poner en marcha soluciones suele exigir acciones que abarquen a todo el gobierno.

Cuando los científicos de datos de la ciudad de Nueva York descubrieron cómo predecir qué edificios presentaban el mayor riesgo de sufrir un incendio mortal, el jefe de operaciones del municipio necesitaba autoridad tanto para armar equipos multidisciplinarios de funcionarios de incendios,

¹ Según *Unified Incident Command and Decision Support* (UICDS, sin fecha): "Hoy en día, en una operación de emergencia —ya sea en respuesta a un incidente real o a indicaciones y advertencias procedentes de una agencia de inteligencia— cada persona, equipo y organización conoce su propia información. La policía sabe sobre las cuestiones de la policía, los bomberos, sobre las de los bomberos, y la lucha contra el terrorismo, sobre la lucha contra el terrorismo. Sin embargo, esas fuerzas de respuesta operan virtualmente en compartimentos de información aislados... hasta que cada uno comparte su información, según pueden y se les permite. Esto da lugar a lagunas, solapamientos e incoherencias respecto de quién sabe qué y cuándo, lo que produce información aislada y un proceso de toma de decisiones enturbiado".

salud y construcción, y aprovechar su experiencia colectiva, como autoridad de ejecución para resolver el problema. Resultó que los propietarios sin escrúpulos y los residentes en esos edificios de riesgo respondían con más frecuencia a un cargo público de los bomberos en uniforme que a un inspector de edificios, aunque este último fuera el que tenía la verdadera solución. Hacer operativa la respuesta que los datos predicen suele requerir la creación de nuevas estructuras entre organismos para producir un valor real.

El tercer reto consiste en superar la situación actual, ya sea en relación con la administración, la cultura o las barreras legales percibidas entre los organismos de la administración pública. El análisis predictivo desbloquea respuestas a los problemas cuando se trabaja con los datos entre diversos organismos, y no sólo con los datos dentro de un único organismo. Sin embargo, estos no cooperan entre sí de forma natural. Si no hay un fuerte impulso por parte de un alto funcionario elegido o designado, el poder del análisis nunca alcanzará su potencial. Los ejemplos que se exponen a continuación ponen de manifiesto que cualquier persona que dirija una iniciativa predictiva necesita tener la autoridad real y aparente para despertar en su empresa el deseo de lograr las soluciones que se derivan del análisis.²

En cuarto lugar, se necesitarán inversiones en personal y en tecnología para impulsar un ahorro futuro mediante el uso del análisis predictivo. (El anexo A proporciona un conjunto detallado de recomendaciones en este sentido.) Tal como puso de manifiesto un reciente estudio: “los datos son abundantes y normalmente fáciles de extraer, pero los recursos (por ejemplo, los recursos humanos) necesarios para transformar los datos en información útil suelen ser escasos” (Van Barneveld, Arnold y Campbell, 2012).

El quinto obstáculo estructural implica tanto adquirir datos utilizables como superar la idea de que los datos que se usen han de ser perfectos. Los problemas que se elijan solucionar primero —ya sean grandes o pequeños— deben referirse a áreas que cuenten con una cantidad razonable de datos que sean de fácil extracción. Si se combinan datos abiertos, transparencia y uso de las mejores bases de datos existentes es posible lograr avances. Para encontrar una respuesta, los funcionarios públicos interesados no necesitan extraer cada conjunto de datos que posiblemente sea aplicable. (Además, tendrán que contar con la inteligencia empresarial y las herramientas de visualización necesarias para traducir sus resultados de una manera que motive y ayude a los usuarios. El anexo A proporciona más detalles.)

Estos cinco obstáculos estructurales, junto con un hardware inadecuado para el procesamiento de datos, disputas territoriales entre organismos que no están dispuestos a proporcionar datos relevantes, un personal que no dispone de los conocimientos suficientes y requisitos perjudiciales en materia de adquisiciones y contratación, han hecho que los gobiernos de todo el mundo estén todavía en los albores de la historia en lo que al uso del análisis predictivo se refiere. Sin embargo, cada vez hay más líderes públicos que ofrecen buenos ejemplos de la utilización eficaz de esas técnicas.

² Según LaVall *et al.* (2011): “El principal obstáculo para la adopción generalizada del análisis es la falta de comprensión sobre cómo usarlo para mejorar el funcionamiento, según declararon casi cuatro de cada diez encuestados”.

Por ejemplo, el máximo tribunal de Brasil, el Supremo Tribunal Federal (STF), llevó a cabo una reforma donde los datos permitieron mejorar el rendimiento, mediante la predicción de necesidades futuras y una reasignación más eficaz de los recursos existentes. Ante un problema considerable en relación con la planificación de los recursos, Almir Antonio da Costa —Jefe de la Administración de Datos de la Secretaría de Tecnología de la Información— decidió que en primer lugar tenían que centrarse en la calidad y la normalización de los datos. Antes de comenzar, no existía información unificada acerca de las prácticas del tribunal ni datos normalizados sobre el número de casos en tramitación. Con el fin de producir el cambio cultural necesario, creó un comité interno para normalizar las definiciones y los procedimientos, y adoptó herramientas de inteligencia empresarial que podrían utilizar esos datos para identificar los casos que plantearan problemas similares. Como resultado, ahora el tribunal puede tomar mejores decisiones sobre su operación, usar el análisis para medir los indicadores clave y evaluar las tendencias de los datos que indican necesidades de recursos en el futuro.

Este documento presenta otros ejemplos del uso del análisis predictivo y sus implicaciones para la acción gubernamental futura, a partir de dos estudios de caso detallados: los proyectos piloto actuales de la Alcaldía de Chicago para mejorar los resultados operacionales en determinados departamentos de la ciudad, y el proceso seguido por el estado de Indiana en la creación de un Centro de Administración y Desempeño que permitirá aplicar el análisis en diversas bases de datos estatales. Finalmente, el documento ofrece un catálogo de uso del análisis predictivo por parte de gobiernos de todo el mundo, organizado por ámbito de políticas.

Chicago: resultados impulsados por el análisis predictivo

En Chicago, el liderazgo del Alcalde Rahm Emanuel ha sido esencial en los esfuerzos municipales para llevar a la práctica el análisis predictivo. Como dice Brenna Berman, directora de información municipal: “él viene hablando del impulso a través de los datos desde el primer día, porque así es como toma las decisiones, así es como habla de todo. Siempre quiere contar con los números y cree en el análisis predictivo”. A su vez, la firmeza del Alcalde impulsa a su personal: “queremos ser los mejores en esto”, dice Berman.

En el verano de 2014, la planificación deliberada y cuidadosa de la Alcaldía de Chicago para utilizar el análisis predictivo comenzó a dar frutos. El Departamento de Innovación y Tecnología (DIT) buscó ampliar los proyectos piloto de análisis predictivo a más departamentos de la ciudad en 2015. Cada uno de los proyectos está dirigido a un problema operacional bien definido (por ejemplo, ¿cómo distribuir mejor los cebos para ratas?) y cada experimento está sujeto a pruebas y validación rigurosas. Los empleados municipales esperan que, gracias al análisis predictivo, todo este trabajo constante conduzca a una clara mejora de los resultados.

Estos proyectos piloto, junto con un ambicioso portal de datos abiertos, un extenso trabajo con departamentos municipales para alentarlos a compartir datos adicionales con el DIT, y el desarrollo —promovido por una donación de US\$1 millón proveniente de Bloomberg Philanthropies— de una plataforma SmartData de código abierto que será compartida con otras ciudades, forman parte de una larga lista de proyectos del DIT. Este departamento, sin embargo, no actúa solo: alianzas formales e informales con universidades locales, la comunidad civil local, fundaciones de Chicago como MacArthur y Chicago Community Trust, y muchas otras personas y grupos, han incrementado la capacidad de la ciudad para emprender proyectos difíciles y han incentivado a la ciudad a que examine el uso de datos constantemente.

Guiada por el Alcalde, Chicago se ve a sí misma, en comparación con otras ciudades del mundo, como líder en el uso de la tecnología, resolviendo los problemas que las grandes urbes tienen y seguirán teniendo a medida que aumente la población urbana en el futuro. La experiencia de Chicago en el uso del análisis predictivo para, en particular, lograr mejores resultados es informativa. A decir de Berman, el resultado deseado para el análisis predictivo es “convertirlo en el modo de vida habitual” en las operaciones de la administración municipal. Como señala Tom Schenk, Director de Análisis de la Ciudad de Chicago desde fines de 2012: “nuestros proyectos están muy centrados en las necesidades operacionales. No emprendemos ningún proyecto a menos que podamos ponerlo en práctica sobre una base operacional”.

Para lograr la meta, Berman y su predecesor, Brett Goldstein (quien fuera el primer Director Municipal de Datos del país) necesitaban tomar algunas medidas

básicas. En primer lugar, Goldstein logró que la alcaldía invirtiera en lo que Schenk denomina “hardware fantástico”, unas máquinas que pueden procesar datos de forma extremadamente rápida. La ciudad necesitaba urgentemente contar con esa capacidad cuando Goldstein asumió como Alcalde; como resultado de la inversión, pueden elaborarse informes en 5% del tiempo que solía utilizarse, ahorrando un número incalculable de horas de trabajo de los empleados municipales y aumentando la capacidad de la ciudad para emplear un análisis avanzado.

Luego, Berman y sus colegas trabajaron con los departamentos de la ciudad mediante el uso de casos de negocio a fin de demostrarles que valía la pena migrar sus datos al entorno del DIT. Uno de los argumentos convincentes del DIT a tal efecto fue la velocidad: según Schenk, “la gente migrará” cuando entienda que puede ahorrar 95% del tiempo de procesamiento. Si se cuestiona a los empleados públicos se conseguirán respuestas a las preguntas del tipo “qué pasaría en caso de” de forma más frecuente, en un entorno de TI rápido y reactivo.

Primer proyecto piloto operacional

El DIT emprendió su primer piloto operacional de análisis predictivo: cebos para ratas de mayor alcance. Este proyecto nació de la minería de datos de todos los datos del servicio telefónico 311, que contempla 500 tipos de llamadas, desde quejas sobre restaurantes faltos de higiene hasta problemas de baches. El equipo de ingenieros del DIT observó un vínculo entre los tipos de llamadas del 311 relacionados con problemas de plagas de ratas, y presentó esa información al Departamento de Calles y Saneamiento (DSS).

Poco después, Schenk formó un equipo para colaborar con Josie Cruz, Directora de Control de Roedores del DSS. Al conocer la propuesta de Schenk, Cruz se resistía a creer que ese proyecto piloto pudiera tener éxito, y manifestó: “al principio, me preguntaba si sería realmente posible”. Nuestras operaciones se basan en las denuncias presentadas al 311. ¿Ayudaría realmente utilizar este tipo de datos para lograr resultados?”

Para probar su eficacia, Berman y Schenk presentaron a Cruz el modelo predictivo del DIT. A partir de 31 tipos de llamadas al 311 que sirvieron como indicadores que suponían plagas de ratas, el sistema desarrolló un mapa de la ciudad que definía zonas específicas donde intervenir. Al superponer esos puntos con los de Cruz, Berman demostró que 80% de los lugares identificados por el DIT eran puntos adonde el equipo de Cruz ya planeaba ir. Berman añade que “luego había otro 20% donde Cruz nunca había estado, así que no tenía la menor idea de que pudiera haber problemas. Eso le intrigó sobremanera, porque estábamos alineados en la gran mayoría de los sitios, lo cual da mucha confianza. Ese 20% despertó su interés, lo que nos puso en sintonía para luego crear un proyecto piloto y explorar el origen del porcentaje”.

Berman y Charlie Williams —Comisionado del Departamento de Calles y Saneamiento— organizaron un proyecto piloto con la ayuda de expertos informáticos de la Universidad Carnegie Mellon. Era fundamental que fuera lo menos invasivo posible, y que ofreciera mejoras sin reinventar la rueda. El DSS y el DIT estuvieron abiertos a colaborar desde el principio y mantuvieron diversas reuniones para discutir a fondo los detalles; asimismo, comprendieron los beneficios que su cooperación produciría para la ciudad.

El único cambio en las operaciones del DSS fue el diseño experimental mínimamente invasivo del proyecto piloto: los colocadores de cebos de un grupo de control recibirían listas de solicitudes de intervención como de costumbre, mientras que un grupo experimental recibiría listas optimizadas con datos. Afortunadamente, el modelo funcionó según lo esperado: el análisis predictivo no reemplazaba a los métodos tradicionales del DSS, sino que los mejoraba. Así, durante el período de prueba, el sistema del DIT envió un equipo de control de roedores a una casa desde donde no se había realizado ninguna llamada al 311 relacionada con ratas. Al llegar, descubrieron una infestación nunca vista por el Departamento de Calles y Saneamiento. Resultó que, pese a que no había datos históricos que indicaran la presencia de ratas en el lugar —por ejemplo, denuncias telefónicas— había efectivamente otros indicadores de sospecha, como por ejemplo llamadas para denunciar la presencia de basura, que históricamente se habían vinculado con la presencia de ratas y al ser utilizadas para orientar iniciativas de inspección dieron resultado.

Las ratas eran una opción perfecta para el proyecto piloto: como no pertenecen a un determinado distrito electoral, no hubo presión política alguna para no estudiarlas. Y el éxito del programa supone un buen presagio para el futuro del uso del análisis en otros organismos. “Los resultados del proyecto piloto me han sorprendido e impresionado mucho”, afirma Cruz. “He pasado a creer en el análisis predictivo, y me parece que esta metodología podría crear una herramienta que optimizará las operaciones de nuestro departamento a largo plazo”.

Berman sostiene que el DIT puede llevar a cabo varios proyectos piloto operacionales de análisis predictivo cada año; el DIT tiene ahora unos cuantos proyectos listos para implementar que la oficina del Alcalde y otras partes interesadas están analizando. Berman señala que es crucial trabajar con las partes interesadas en torno a los proyectos piloto: “informamos a la oficina del Alcalde sobre lo que estamos haciendo desde el principio. Cuanto más conozcan y comprendan lo que hacemos, mejor, porque a veces hay implicaciones de lo que hacemos que ni siquiera nosotros conocemos. También informamos en detalle al departamento jurídico”.

Con el tiempo, todos estos proyectos piloto operacionales individuales se irán fundiendo en un conjunto general de cuestiones de funcionamiento a las que la alcaldía les podrá aplicar el análisis predictivo. Berman considera que algún día podrán formularse preguntas del tipo “qué pasaría en caso de”; a su vez, tiene planes futuros para poner a prueba la capacidad de lanzar consultas al sistema sobre problemas que no estén expresamente definidos.

La plataforma no será capaz de responder a esas preguntas hipotéticas durante un par de años; necesita datos para poder responder y hay un montón de “incógnitos que aún se desconocen”, señala Danielle DuMerer —Directora de Planificación, Políticas y Gestión—, quien ha dirigido un sinnúmero de proyectos de datos abiertos, inteligencia empresarial y análisis predictivo en el DIT, y cree que cada vez que el DIT trabaja en un proceso debe tratar de producir tantos datos como sea posible (“sin que resulten gravosos”) para su entorno.

En Chicago aún no se llevan a cabo análisis interdepartamentales. Al respecto, hay todavía problemas de inteligencia empresarial. En suma, los organismos necesitan entender mejor lo que están haciendo otros organismos. “En general, no se trata del análisis —dice Schenk—. Se trata más bien de la transparencia respecto de lo que está pasando entre estos dos organismos”.

La plataforma SmartData

Chicago está preparando una plataforma de SmartData de código abierto para el análisis predictivo, con el respaldo de una donación de US\$1 millón proveniente de Bloomberg Philanthropies. El plan es que otras ciudades adopten la plataforma. Según Danielle DuMerer, a las ciudades que deseen replicar la plataforma de Chicago se les podría entregar el código limpio que sirve de base para el entorno de información del DIT, “una vez borrada nuestra información específica”. El paquete que se prepare para que lo adopten otras ciudades incluiría asesoramiento y algoritmos: “no se trata solo de poner el código en Github y decir ‘ahí lo tenéis’ —dice DuMerer—. No habría muchas posibilidades de que alguien lo adoptara de esa forma”. En cambio, a las ciudades se les entregará documentación sobre todos los aspectos, desde la relación con las partes interesadas hasta la opción de las cuestiones. “Estas son las mejores prácticas. Así es como conviene desarrollar el proyecto en el marco del plan”, dice. Por su parte, Brett Goldstein sostiene que “no funcionará automáticamente: cada ciudad tendrá que construir sus propias conexiones con sus sistemas transaccionales. Es una tarea difícil”.

Uno de los principales objetivos de Berman en términos de la plataforma SmartData es que sea fácil de usar. “Lo fundamental de la plataforma SmartData es construir una aplicación que le permita al gestor experto —ese que hoy es capaz de ejecutar un informe sorprendente que le dice al comisariado exactamente qué tiene que hacer— lanzar consultas y aprovechar el análisis sin la participación de un ingeniero de datos”, apunta. No todas las ciudades cuentan con varios científicos de datos en su personal. Para ello, el código de la plataforma SmartData incluirá la interfaz WindyGrid, que es fácil de usar y permite la representación de datos históricos y en tiempo real mapeados según la ubicación.

La plataforma no lo hace todo. Ya existen sistemas que dan prioridad, por ejemplo, a las solicitudes de intervención; Chicago no tiene la intención de volver a construir tales sistemas como parte de la plataforma SmartData. Sin embargo, en cuanto herramienta básica, que avanza hacia un análisis predictivo replicable, escalable y sostenible que pueden adoptar ciudades más pequeñas que no cuenten con los recursos de Chicago, la plataforma SmartData será una importante contribución a la vida ciudadana. En última instancia, las ciudades serán capaces de cargar sus datos en la plataforma y comparar sus indicadores —de bienestar, contaminación y cualquier cosa que pueda construirse a partir de los datos de la ciudad— con los de otras ciudades.

Chicago quiere que el análisis predictivo se convierta en la norma. El proceso de evangelización y de debate sobre el modelo de funcionamiento con sus propios organismos municipales continúa. Según Schenk, aún no han terminado completamente la migración de datos, y puede que hasta fines de 2014 no estén todos los datos operacionales municipales bajo el control central del DIT y, por tanto, disponibles para su rápido procesamiento. También ha supuesto un reto el hecho de que algunos organismos externalicen el alojamiento de sus datos a proveedores privados, para quienes ayudar al DIT a añadir más datos a su entorno no es necesariamente una gran prioridad; Schenk espera superar este obstáculo con prontitud.

Otro recurso útil que utilizó el DIT con los departamentos municipales es lo que Schenk denomina “Desarrollo Rápido de Aplicaciones”. Por ejemplo, un analista de bases de datos del DIT

es experto en el uso de APEX de Oracle, una ligera herramienta de desarrollo de aplicaciones web que se utiliza con bases de datos de Oracle. Si un departamento necesita una interfaz web sencilla —por ejemplo, para poder mostrarles datos a los ciudadanos—, el DIT construirá una aplicación APEX para ese departamento. La condición es que el departamento migre sus datos. Por último, el motor predictivo del DIT está incorporando datos abiertos que no proceden de la alcaldía, tales como datos meteorológicos y tuits públicos georreferenciados. Chicago viene recopilando tuits desde principios de 2012 para aprovechar las indicaciones que tales datos puedan proporcionar.

Como indica DuMerer, el análisis ha de ser “solo una parte de nuestro trabajo en el municipio. Así es como trabajamos para mejorar los servicios. El análisis ayuda a dirigir nuestra energía no solo para reaccionar a los problemas, sino para lograr anticiparnos a ellos”. Poner lo importante por delante de lo urgente, por una vez, puede ayudar enormemente a las operaciones de la administración pública y mejorar los resultados para los ciudadanos.

Indiana: construir un centro de análisis

En Indiana, Paul Baltzell —Oficial Jefe de Información del Gobernador Mike Pence— supervisa las bases de datos que almacenan toda la información del estado. Mientras que otros en su puesto tendrían que convencer a los organismos o departamentos para que autoricen al Oficial Jefe de Información a alojar sus datos, Baltzell —que asumió en enero de 2013— corre con ventaja: ya cuenta con todos los datos gracias al modelo centralizado del estado.

Desde que asumió, y bajo el liderazgo y estímulo del Gobernador Mike Pence, la Oficina de Administración y Presupuesto (OMB), en colaboración con la Oficina Tecnológica de Indiana (IOT), ha estado trabajando estrechamente con Chris Atkins —Director de la Oficina de Administración y Presupuesto— a fin de crear un centro de análisis ubicado en la cámara estatal de Indiana y que se espera que empiece a funcionar en el otoño de 2014. Este moderno centro, denominado Centro de Administración y Desempeño (CAD), tendrá como objetivo hacer frente a algunos de los problemas más acuciantes de Indiana, extrayendo datos almacenados en las bases de datos de Baltzell e incorporándolos en un motor analítico, así como contratando científicos y analistas de datos para evaluar los patrones que se encuentren y así abordar las intervenciones específicas del estado. Todos los organismos del estado podrán usar ese recurso del CAD. Creemos que la experiencia reciente de Baltzell y Atkins para llegar a este punto puede ser de utilidad para otros gobiernos que estén considerando la posibilidad de hacer del análisis predictivo una herramienta central de su gestión.

Cuando una empresa privada quiere instruir a sus unidades para cooperar con el uso de datos en pos de buenos resultados, resulta relativamente simple: las direcciones de Amazon o de Priceline.com, por ejemplo, no tienen más que publicar una orden. En la administración pública, las cosas son diferentes: existen leyes, conflictos burocráticos por prerrogativas entre organismos, límites estrictos en los recursos e incompatibilidades que representan barreras para una amplia colaboración en torno a los datos. Para superar la inercia del *statu quo*, ayuda tener un problema inicial compartido —un “caso de uso” para el análisis de datos que obligue a cooperar— y una señal clara de un líder ejecutivo de alto nivel. Baltzell tiene ambos.

Indiana tiene una de las tasas más altas del país en términos de mortalidad de menores de un año y de obesidad y tabaquismo: 7,7 muertes de lactantes por cada 1.000 nacimientos, por encima de la tasa nacional, que es de 7. Indiana ha elegido la mortalidad infantil como el primer caso de uso para el Centro de Administración y Desempeño. Y el Proyecto del CAD cuenta con el apoyo de las más altas instancias: el Gobernador ha encomendado a la OMB y a la IOT que la iniciativa sea un éxito. “El Gobernador [Mike] Pence nos apoya —observa Baltzell—. Él va donde va el análisis de datos”.

La correlación entre mortalidad infantil e indicadores tales como una mala alimentación, una vivienda inadecuada y la falta de acceso a la asistencia sanitaria es bien conocida y fácil de comprender. Históricamente, sin embargo, ha sido difícil actuar sobre esa correlación mediante el incremento específico de fondos para apoyar los programas y los esfuerzos de asistencia social que podrían abordar la mortalidad infantil de manera más eficaz. No obstante, la plataforma de análisis del CAD permite que el centro integre en su análisis los datos de la administración referentes a familia, salud, finanzas, empresas y organismos de empleo, revelando conexiones menos obvias.

El objetivo del proyecto es utilizar tales resultados para capacitar a los trabajadores sociales en la toma de decisiones basadas en datos. Mediante el acceso a las herramientas desarrolladas por el CAD, un trabajador social podrá comparar la información de una determinada familia en relación con los datos pasados y presentes de todas las familias en situación de riesgo en la base de datos. El resultado sería una clara estimación del riesgo que le permitiría al trabajador social determinar la probabilidad de que un niño se encuentre sometido a un riesgo de daño futuro. Por ejemplo, el algoritmo puede determinar que “existe un 80% de probabilidad de que a este niño le suceda algo realmente malo —explica Baltzell—. Hay que sacarlo de allí ya mismo”. Por otra parte, una estimación del riesgo de 10% puede indicar una posibilidad de problemas futuros pequeña pero significativa. La plataforma, a su vez, le permitirá al agente mitigar el riesgo de forma más directa, ya sea poniéndose en contacto con el oficial de libertad condicional correspondiente, un abogado defensor o cualquier otro especialista en otro organismo gubernamental.

A fin de centrar su atención en la mortalidad infantil, la IOT tuvo que hacer algo más que alojar —o almacenar— la información de las bases de datos de los organismos estatales: fue necesario construir vínculos entre las bases de datos de 92 organismos del Poder Ejecutivo y un motor de análisis que fuera capaz de ingerir datos, masticarlos y generar informes en el menor tiempo posible. (El Comité Directivo del CAD optó por utilizar la plataforma de datos de SAP HANA como motor de análisis.) Baltzell decidió no construir esos vínculos simplemente por decreto, sino trabajar en estrecha colaboración con los organismos cuyos datos ya estaba almacenando para fundamentar la conveniencia del análisis.

Baltzell encontró cierto rechazo por parte de los funcionarios de nivel medio en los organismos que estaban acostumbrados a hacer las cosas de una manera particular. No se trataba solamente de preocupaciones de índole cultural, sino también de inquietudes sobre la posible violación de reglamentos federales y sobre la seguridad. El Comité Directivo del CAD examinó cuidadosamente todas esas preocupaciones con la ayuda de un asesor jurídico externo. Existe un gran respeto por las preocupaciones legales importantes de los organismos. Una orden ejecutiva emitida el

16 de marzo por el Gobernador Pence proporciona una vía para que los organismos presenten sus preocupaciones legales ante la OMB y la IOT.

La seguridad resultó ser la principal tarea previa al lanzamiento para Baltzell, lo cual no le sorprendió. Pese a que su centro de datos centralizado gestiona los sistemas y las bases de datos operativos utilizados por los organismos estatales, las aplicaciones utilizadas para extraer los datos a menudo están bajo el control de organismos concretos (como suele ser el caso cuando una aplicación especializada es utilizada por un único organismo). Los organismos estimaron que había un riesgo de que el proyecto sobre mortalidad infantil en cierta forma pusiera los datos en las manos equivocadas.

Baltzell afianzó la seguridad desde el principio. Indiana nunca antes había tenido una operación de análisis de datos en todo el estado, por lo que Baltzell estaba operando en un área totalmente nueva, lo cual facilitó las cosas; podría planificar cuidadosamente la seguridad. Como él mismo dice, “estamos haciendo todo a nuestro alcance”.

En primer lugar, su equipo separó física y lógicamente la operación de análisis de datos —el Centro de Administración y Desempeño— del entorno donde se almacenan los datos del organismo. Los datos del organismo ya estaban protegidos por un cortafuegos, que los resguardaba del mundo, pero había “otra capa de la cebolla”, que hacía que “los datos importantes no estuvieran disponibles para ellos”. Por otro lado, el centro no está conectado a Internet, y los empleados sin autorización no pueden acceder a la habitación donde están los científicos de datos; es un ambiente controlado y monitoreado.

Los datos no se traen al centro (o, más exactamente, no pasan a formar parte del centro) a menos que estén etiquetados con metadatos que muestren de dónde vienen, cuándo se crearon y cuáles son los requisitos legales para su concesión (por ejemplo, leyes federales sanitarias, tributarias o educativas que limitan su uso). Esto permitirá que el equipo de Baltzell realice un seguimiento de la utilización y combinación de datos. El requisito de insertar metadatos significa que las restricciones legales irán con los datos donde quiera que estos vayan dentro del centro.

Los empleados de la Oficina de Tecnología serán responsables de los paquetes de algoritmos matemáticos que dan sentido a los datos del centro. Tras avanzar por un detallado proceso de aprobación, esas personas tendrán acceso a una sala reservada donde se ejecutan algoritmos con herramientas de software SAP HANA. Los programas clasifican los datos almacenados temporalmente en el ordenador y permiten un procesamiento inmediato, lo que hace posible la toma de decisiones en tiempo real. Para acceder a la sala reservada, los empleados tienen que obtener un nivel de habilitación adecuado y cumplir con una serie de promesas contractuales. Los arquitectos de Baltzell que trabajan en el centro han hecho más estricta la verificación de antecedentes, que permite el acceso administrativo a los datos; los empleados que se encargan de las bases de datos tienen un menor nivel de acceso y no tienen acceso administrativo. A los científicos de datos que van a trabajar en algoritmos de análisis predictivo se los controla de manera estricta y no se les permite trabajar a distancia o ingresar a la sala reservada con llaves USB o sus propios dispositivos. Deben trabajar físicamente en el entorno controlado.

Los datos extraídos del centro en forma de informes para su visualización no pueden incluir información personal identificable. Los trabajadores sociales que se ocupan de los riesgos de mortalidad infantil sobre el terreno no sabrán más detalles que los necesarios para poder hacer su trabajo

—tendrán acceso a información privada, como los números de la seguridad social—. No obstante, tendrán acceso a la información de contacto de los programas y organismos que puedan ser de utilidad para las familias que atienden.

A fin de garantizar que el centro cumple con los más altos estándares de seguridad, Baltzell ha pasado por varios exámenes de seguridad realizados por proveedores externos antes de iniciar las operaciones del centro. Está en proceso de realizar modificaciones que surgieron de las recomendaciones, y antes de comenzar a funcionar prevé pasar por más “pruebas de penetración”, en manos de expertos de seguridad de “guante blanco” que operan desde el exterior (así como otros que se hacen pasar por empleados descontentos) con el propósito de intentar atravesar las barreras electrónicas. Todas las prácticas de seguridad que se están adoptando están asentándose en un documento — muy extenso— que Baltzell tendrá a mano en caso de que surja alguna pregunta. Por otro lado, tiene un registro de auditoría listo para, entre otras cosas, cumplir con los estándares de seguridad cibernética del NIST, adoptados a nivel federal.

Baltzell está orgulloso de lo que ha hecho en materia de seguridad en pos del centro y quiere continuar extendiendo la protección: “creo que hemos introducido todas las medidas de seguridad posibles para proteger los datos privados de las personas”. Reconoce que ha ido más allá de lo que otros podrían haber hecho con respecto a la seguridad, aunque el trabajo con los datos no sea del todo cómodo: “algunas de las medidas que hemos tomado, las hemos llevado al extremo, pero lo hemos hecho porque no queremos que exista ningún riesgo de que suceda algo malo. De hecho, afirma que las medidas les han dificultado el trabajo con datos a los matemáticos y científicos, pues no pueden trabajar de forma remota”.

El espacio y los empleados

A partir del otoño de 2014, una deslucida oficina en el sótano del Parlamento se ha transformado en el Centro de Administración y Desempeño con aires que recuerdan a Silicon Valley, con el objetivo de involucrar a las personas y fomentar la colaboración, y con pantallas interactivas, un diseño de espacios abiertos para los empleados, salas de reuniones y paredes llenas de pizarras. El espacio será útil para formación, capacitación y divulgación, así como para el trabajo diario de los científicos y analistas de datos.

Entre 10 y 15 empleados trabajarán allí a tiempo completo. Los analistas se combinarán con los científicos de datos, y estas parejas de trabajadores podrán traer al espacio a líderes del organismo o a actores de otros organismos que ya estén haciendo un trabajo de inteligencia empresarial; juntos, dice Baltzell, trabajarán en problemas que afecten a todo el estado. Los científicos de datos entrarán y saldrán cuando sea necesario.

Uno de los desafíos clave con los que se topó Baltzell para que el centro comenzara a funcionar fue trabajar con científicos de datos muy inteligentes pero que tienen menos experiencia en comunicar los resultados de su investigación. Como él mismo dice, “les digo ‘oye, realmente tengo que escoger algunos puntos clave para salvar a unos niños’”. Es muy difícil encontrar a científicos de datos que sepan comunicar: “ya es bastante difícil encontrar a alguien con el conjunto de habilidades necesarias simplemente para hacer la parte matemática del trabajo, la parte algorítmica.

Y, de todas formas, si se pudiera encontrar a alguien que tuviera los dos tipos de habilidades, probablemente no podría solventar su salario". Su esperanza es que logrará crear alianzas entre el centro y las universidades de Purdue e Indiana a fin de construir una vía de entrada de científicos de datos que trabajen en el centro. (Baltzell está fomentando la participación de investigadores de la Universidad de Indiana, que se encarga del centro estatal de copias de seguridad para la restauración de datos, así como de académicos y científicos de datos de la Universidad de Purdue.) "Estoy dispuesto a contratar a un chico joven y trabajar con él durante un par de años, aunque esté recién salido de la universidad y aunque se vaya a ir luego a ganar el doble a una de las costas". Hasta ahora, sin embargo, ha tenido que subcontratar todo el trabajo, y es consciente de que contratar va a seguir siendo difícil: "ya sé que esos tipos pueden ganar mucho más en el sector privado".

Casos de uso futuros del Centro

Baltzell planea llevar a cabo reuniones de lluvia de ideas sobre casos de uso. Según dice, "el Gobernador tiene una hoja de ruta sobre las cosas que quiere encarar" que suponen un problema para Indiana. Baltzell traerá grupos que podrían tener datos relevantes acerca de determinados problemas; también está interesado en detectar otros problemas que podrían ser sensibles al análisis predictivo. Lo que estudiarían luego sería la reincidencia. Baltzell cree que los datos podrían revelar qué áreas de determinadas instalaciones o qué delincuentes (o grupos de delincuentes) están vinculados con las mayores tasas de retorno a prisión.

La justicia criminal, en general, puede ser un área madura para el uso del análisis en Indiana: Baltzell cree que el análisis puede ser útil para determinar qué delincuentes tienen un menor riesgo asociado de violencia, quienes, por tanto, podrían ser supervisados con bandas de tobillo de GPS o con otros tipos de monitoreo en lugar de que se los envíe de vuelta a la cárcel. Indiana ha cambiado recientemente su código penal, con una revisión de la forma en que se clasifican los delitos, lo que significa que se destinarán más prisioneros a las cárceles de condado que a las estatales. El análisis predictivo puede ayudar a evaluar qué número de camas y de personal necesitará en el futuro una cárcel de condado determinada.

Otra cuestión clave para Indiana es el desarrollo económico. Baltzell espera ser capaz de identificar los factores impulsores de la creación de empresas en el estado. "Si nos falta hacer algo desde el punto de vista legislativo o político que pueda ayudar a aumentar los ingresos de las empresas o a atraer nuevos negocios a Indiana, lo haremos", dice Baltzell.

Una de las consecuencias más prometedoras del centro puede radicar en el aumento de la eficacia de la propia administración pública de Indiana. En su orden ejecutiva de marzo de 2014, en la que instaba a la Oficina Estatal de Administración y Presupuesto y a la Oficina de Tecnología a crear el Centro de Administración y Desempeño, el Gobernador Mike Pence afirmó que el centro sería una "herramienta para la mejora continua de los procesos del estado de Indiana", y pidió a la OMB que le proporcionara recomendaciones sobre "oportunidades para utilizar los datos recogidos por los organismos estatales con el fin de impulsar la innovación y la eficiencia en todos los organismos estatales".

A tal efecto, el personal de Baltzell ha estado convirtiendo los indicadores clave de rendimiento que existen en los organismos estatales en material apto para el motor de análisis de datos. Eso ha supuesto un reto; anteriormente, los estadísticos empleados en cada organismo calculaban sus propios indicadores y gran parte de este trabajo histórico se encuentra únicamente en hojas de cálculo de Excel y visualizaciones propias. El personal de Baltzell ha estado automatizando los indicadores clave de rendimiento; y una vez tratados por los analistas del centro, se les presentarán a los jefes de los organismos y se mejorarán mediante herramientas de visualización de alta calidad y la velocidad de la computación en memoria.

Una ventaja importante de centralizar los indicadores clave de rendimiento y automatizarlos es que, con el tiempo, el centro contará con la capacidad de aplicar en los organismos mejores prácticas de datos, gracias a las mediciones del programa subyacentes a los indicadores clave de rendimiento global del organismo. A la fecha, un tema importante en lo que se refiere a los datos del organismo ha sido la calidad; gracias al análisis, en el largo plazo, el equipo de Baltzell será capaz de llevar a cabo lo que él denomina la “gestión de datos maestros”. En sus palabras: “podemos correlacionar los datos para mejorar su calidad y después alimentar el sistema con datos limpios extraídos de sistemas antiguos a fin de corregir defectos”, en un bucle que se reforzaría de forma constante. Al respecto, el centro está todavía a cierta distancia de conseguirlo, pero Baltzell vislumbra las posibilidades.

Generar apoyo

20

Los anuncios sobre el centro vendrán junto con el lanzamiento del portal de datos abiertos de Indiana. Usando Socrata, y siguiendo el ejemplo de muchas ciudades y países de todo el mundo, Indiana publicará datos y una API que le permitirá al público crear aplicaciones que utilicen esos datos. En Indiana e Indianápolis ya se han llevado a cabo eventos públicos de creación de software (*hackathones*), que, según Baltzell, han sido fructíferos. Por otro lado, está planificando becas de corta duración que implican el trabajo con datos públicos.

Paul Baltzell y Chris Atkins persisten en su labor de hacer realidad el Centro de Administración y Desempeño que solicitara el Gobernador Pence, y se muestran cautelosamente optimistas. Tras meses de coordinación y atención constante a la seguridad, la privacidad y el caso de uso de la reducción de las tasas de mortalidad infantil, el centro está casi listo para funcionar.

El estado de Indiana sitúa su unidad de análisis predictivo dentro de su grupo de desempeño e innovación. Los conocimientos obtenidos a partir de los datos no sólo orientarán al equipo altamente interactivo respecto de los cambios que producirán el mayor valor sino que le brindarán la confianza que necesitan para emprender un cambio audaz. Armados con los datos, el equipo lo tiene más fácil para apartar a los burócratas reacios y defender un esfuerzo que podría no conducir a los éxitos previstos. El uso de datos le permite al equipo del Gobernador de Indiana reducir el riesgo así como generar un entorno favorable para la innovación.

El análisis predictivo fue utilizado por primera vez por empresas comerciales para las cuales merecía la pena hacer grandes inversiones en tecnología y personal y así obtener ligeras mejoras en el rendimiento. Por ejemplo, dado que los modelos de negocio de las compañías de tarjetas de crédito y aseguradoras dependen de la detección del fraude y de otorgarles puntos a los clientes sobre la base de predicciones de capacidad de pago futura derivadas de la actividad pasada, esas empresas vieron desde el principio los beneficios de automatizar las tareas estadísticas que solían realizarse —laboriosamente— a mano. A medida que el hardware y las herramientas de gestión de datos se hicieron menos costosos y más accesibles, el uso de las tecnologías de predicción se ha disparado. Hoy en día, la elaboración de informes, que antes podría haber llevado horas o que simplemente no se realizaba —porque los procesadores eran demasiado lentos y muy caros— puede realizarse en minutos si se dispone del hardware adecuado. Como resultado, en casi todos los rincones de la industria privada se aprovecha el análisis predictivo.

Los gobiernos han sido más lentos en adoptar estas técnicas. Una vez más, un sector en particular ha llevado la delantera: el “modelo de negocio” de la comunidad comprometida con la seguridad pública depende de la detección rápida y la predicción de la delincuencia con el fin de proteger al público, razón por la cual se ha mostrado dispuesta a invertir en paquetes de software fáciles de implementar y hardware más rápido. Como resultado, la detección de zonas conflictivas —es decir, conocer de antemano dónde es probable que suceda el próximo delito— es algo que está muy extendido, particularmente en Estados Unidos.

En esta sección se exponen ejemplos del uso del análisis predictivo por parte de los gobiernos, organizados por ámbitos de política. Los ejemplos oscilan entre usos de la tecnología posibles y estudiados. Algunas de las técnicas aquí descritas se desarrollaron internamente en los organismos gubernamentales, mientras que otras se adquirieron de proveedores externos. Esperamos que este catálogo inspire nuevas experiencias en el análisis predictivo. El uso de datos no sólo mejorará los resultados de los programas gubernamentales, sino que reducirá los riesgos para quienes innovan en el gobierno. El riesgo es a menudo el obstáculo más importante dentro de los gobiernos —que tradicionalmente se han mostrado reacios a probar cosas nuevas— pero las predicciones basadas en datos aumentarán la confianza en el cambio y, por tanto, crearán condiciones más favorables para la innovación. En otras palabras, las innovaciones basadas en datos ayudarán a los innovadores a tomar las decisiones correctas y proporcionarán apoyo (en forma de análisis) en caso de fallo.

Seguridad pública

Hace años que las prisiones utilizan datos y estadísticas detalladas en sus predicciones sobre qué presos deberían salir en libertad condicional. Muchas jurisdicciones aplican modelos predictivos para determinar quién debe ser detenido antes del juicio

y a quién se puede permitir de forma segura estar en libertad a la espera de un juicio. Sin embargo, las herramientas de evaluación no suelen acceder a la amplia variedad de datos disponibles; además, demasiado a menudo se basan en datos generados por el propio infractor. Por otro lado, no se cuenta con información sobre los factores que afectan a la conducta de una persona en particular.

El análisis predictivo es un paso natural para los departamentos de policía que usan prácticas policiales basadas en servicios de inteligencia o en datos. Décadas atrás, con la llegada de registros computarizados, sistemas de gestión y sistemas automatizados de recepción de llamadas se crearon grandes almacenes de datos sobre las operaciones policiales básicas. En la década de 1990, el desarrollo de CompStat utilizó las estadísticas informáticas como herramienta de gestión para asignar recursos a las rondas policiales con problemas específicos y servir como herramienta de gestión y rendición de cuentas. Otros desarrollos simultáneos en los sistemas de información geográfica y en herramientas de presentación automatizada de datos han permitido a los departamentos identificar y mapear las “zonas críticas” de la criminalidad y dedicar recursos a esas zonas. Tales innovaciones han impulsado una filosofía basada en la medición del rendimiento que proporciona una base sólida para el análisis predictivo.

Los departamentos de policía pueden utilizar las predicciones de varias formas: para predecir patrones delictivos, para predecir qué delincuentes podrían cometer más delitos y requieren una vigilancia más cuidadosa, y para predecir qué individuos, grupos o barrios están en riesgo de ser víctimas de la delincuencia. Cabe repetir, las prácticas policiales predictivas no están destinadas a sustituir a las técnicas policiales existentes, sino a complementarlas. El análisis predictivo toma como base la actuación policial comunitaria, orientada a los problemas, basada en evidencia o inteligencia, y las prácticas policiales en zonas críticas, así como otras estrategias de eficacia comprobada. La actuación policial predictiva es más eficaz cuando forma parte de estrategias proactivas más amplias donde se forjan relaciones sólidas entre los departamentos de policía y las comunidades para resolver los problemas de la delincuencia.

Los análisis predictivos no pueden reemplazar a una sólida infraestructura de datos. Los datos de calidad son un insumo importante para el análisis predictivo, y las predicciones no pueden sustituir a la gestión básica de datos. Los líderes de la policía siguen necesitando informes de gestión, como paneles, mapas de la delincuencia, indicadores clave de rendimiento, etc.

La actuación policial predictiva utiliza algoritmos estadísticos para anticipar probables eventos delictivos, lo que permite a los departamentos preparar y prevenir ese tipo de delitos. Las predicciones pueden centrarse en diversas variables, como lugares, personas, grupos o incidentes. Las tendencias demográficas, las poblaciones en libertad condicional y las condiciones económicas pueden afectar a todos los índices de criminalidad en áreas particulares. El objetivo del análisis predictivo es similar a las metas de muchas innovaciones tecnológicas anteriores en la policía: reducir el trabajo administrativo y dar a los funcionarios e investigadores más tiempo para la patrulla proactiva, la interacción con la comunidad, la respuesta a incidentes y la investigación activa de los delitos.

Las agencias de seguridad pública grandes y pequeñas están utilizando el análisis predictivo para abordar diversos desafíos. Los departamentos desarrollan algoritmos predictivos para analizar las relaciones entre los diversos elementos, tales como la relación entre el absentismo escolar y un aumento en los robos en el vecindario, edificios abandonados y puntos de venta de drogas ilegales,

etc. Otros posibles factores para examinar son los parques públicos, los datos de desalojos, los datos escolares, la información de zonificación y los datos de salud mental.

- Los Ángeles fue una de las primeras ciudades que experimentó con el análisis predictivo, aprovechando el trabajo de los académicos de la cercana Universidad de California en Los Ángeles. El profesor Jeff Brantingham, del Departamento de Antropología, desarrolló un modelo matemático —sobre la base de predicciones de réplicas de terremotos— que permitió predecir delitos futuros. El sistema calcula y genera predicciones basadas en datos simples, en constante actualización, tales como la ubicación, la hora y el tipo de delito. Entonces, el sistema crea cuadros de predicción de 500 pies cuadrados y los muestra, con código de colores, en un mapa de patrulla. En una de las primeras áreas donde se implementó la actuación policial predictiva (Foothill Division), los delitos descendieron 13% en los cuatro meses siguientes, en comparación con un aumento de 0,4% en el resto de la ciudad donde no se había realizado el despliegue. En palabras del líder del Departamento de Policía de Los Ángeles respecto del proyecto de análisis predictivo, “hemos impedido que cientos y cientos de personas llegaran a casa y se encontraran con que les habían robado” (CBS Los Ángeles, 2012).
- El Departamento de Policía de Chicago (CPD) tiene un largo historial de liderazgo en términos de innovación tecnológica. El CPD fue seleccionado por el Departamento de Justicia de Estados Unidos para la aplicación y financiación de la investigación para avanzar en el análisis predictivo. El proyecto piloto en curso en el CPD evaluará la capacidad del software de reconocimiento de patrones modificados utilizado actualmente para fines de diagnóstico médico con el fin de predecir patrones de delincuencia. También evaluará la eficacia de un software que cuantifica y mapea la actividad de pandillas para predecir áreas emergentes de conflictos entre pandillas, aprovechando la investigación sobre las redes sociales de Andrew Papachristos, sociólogo de Yale. El Dr. Papachristos ha descubierto que la actividad violenta se limita a un pequeño número de individuos concentrados en una zona particular. Al identificar no sólo las “zonas críticas” de la delincuencia, sino también las personas “críticas” y sus redes, los agentes de policía pueden acercarse de forma proactiva a los miembros de la red. Al establecer contacto con los miembros de la red, la policía y los trabajadores sociales les animan a dejar atrás sus estilos de vida de pandillas violentas, ayudándolos a encontrar trabajo. Como resultado, las estadísticas de la delincuencia en la zona muestran una disminución temprana. Como declaró a *The Verge* el comandante Jonathan Lewin: “este programa se convertirá en la mejor práctica nacional. Informará a los departamentos de policía de todo el país y de todo el mundo sobre la mejor manera de utilizar la actuación policial predictiva para resolver problemas. Se trata de salvar vidas” (Stroud, 2014).
- El comisario de Policía de Nueva York, Bill Bratton, ha defendido durante décadas el uso de datos como herramienta de gestión y goza de reconocimiento por el desarrollo y avance de CompStat como herramienta de gestión policial. Él puso en práctica la actuación policial predictiva en Los Ángeles. Ahora, en su segundo período de servicio como comisario de Policía de Nueva York planea hacer lo mismo, para que sus agentes puedan predecir y prevenir delitos. “El departamento está analizando las horas previas a cada disparo para analizar qué

quejas al 311 o llamadas al 911 por cuestiones menores —volumen alto de música, reuniones multitudinarias— precedieron a los episodios de violencia, las cuales, si se abordan en el futuro, podrían ayudar a impedirlos”.

- La herramienta de análisis predictivo del Departamento de Policía del Condado de Miami-Dade se centra en la resolución de casos aún no resueltos y la captura de delincuentes reincidentes. Cuando se comete un delito, un oficial puede acceder rápidamente a una lista de sospechosos potenciales en función de la probabilidad de coincidencia de los datos existentes en el sistema. La lista les permite a los investigadores acotar el enfoque, de miles de delincuentes conocidos a sólo aquellos con la mayor probabilidad de haber cometido el delito. El sistema aprovecha patrones de criminalidad e información sobre el *modus operandi* del delincuente extraídos de grandes volúmenes de datos históricos.
- El Departamento de Policía de Shreveport llevó a cabo un proyecto piloto en seis de los distritos con tasas de criminalidad más altas para analizar si un modelo predictivo podía reducir la delincuencia. El modelo utiliza datos sobre quejas relativas a personas jóvenes, alto volumen de música, personas que cometen ofensas al orden público, actividades sospechosas, personas que deambulan sospechosamente, riñas y merodeadores. Los analistas de delitos desarrollaron mapas de localización probable de los delitos contra la propiedad. Un experimento que buscaba determinar si las predicciones podrían reducir la delincuencia no fue concluyente, debido a la falta de potencia estadística y a la fidelidad incompatible con el modelo del programa. Sin embargo, las áreas en las que se utilizó el modelo predictivo gastaron menos (entre 6% y 10% menos en costos laborales) para obtener los mismos resultados. Durante los primeros cuatro meses del período experimental, las áreas que utilizan modelos predictivos redujeron los delitos contra la propiedad en 35%.
- El Departamento de Policía de Charlotte-Mecklenberg (CMPD) tiene un sistema de análisis predictivo que utiliza cuadros de mando operacionales, tácticos y predictivos para visualizar en un mapa las zonas con mayor probabilidad de que se cometa un delito en un período indistinto de cuatro horas. Los datos del modelo predictivo incluyen patrones históricos de delincuencia, delitos recientes (en las últimas 24 y 48 horas), llamadas recientes para intervenir, motivos, ubicación física, día de la semana, hora, clima, factores políticos, factores económicos y calendarios escolares, períodos de pago de salarios, eventos especiales y muchos otros factores. El conocimiento que se obtiene de esta forma ayuda al personal de mando a desplegar recursos con mayor conocimiento y permite a los oficiales gestionar más eficazmente sus áreas de responsabilidad para disuadir la delincuencia.
- Gracias al uso del análisis predictivo, el Departamento de Policía de Memphis (MPD) redujo la delincuencia en más de 30%, incluida una reducción de 15% en delitos violentos durante un período de cuatro años. El MPD es ahora capaz de evaluar los patrones de incidencia en toda la ciudad y pronosticar “zonas críticas” delictivas, y así asignar recursos y desplegar personal de manera proactiva, especialmente vigilancia dirigida, refuerzos localizados para el tráfico, grupos de trabajo, operaciones, patrulla de gran visibilidad e investigaciones dirigidas. El sistema recopila grandes volúmenes de datos sobre delincuencia, que incluyen fuentes de datos de las patrullas en función del tipo de delito, hora, día de la semana, o

varias características sobre la víctima o el agresor. Ahora, el MPD puede orientar mejor las decisiones diarias para hacer frente a la actividad delictiva y para que los funcionarios estén en una mejor posición estratégica para responder a un delito que se esté fraguando.

- El Departamento de Policía de Arlington en Texas utiliza datos sobre robos residenciales para identificar zonas críticas y luego comparar esas ubicaciones con las zonas donde se registran violaciones al código penal. Según el comisario jefe Theron Bowman, los agentes encontraron una conexión directa entre la decadencia física y los robos residenciales. Entonces, Arlington desarrolló una fórmula para ayudar a identificar las características de los barrios “frágiles” de forma tal que les permitiera trabajar activamente con otras administraciones públicas municipales para ayudar a prevenir la delincuencia (Pearsall, 2010).
- La ciudad de Lancaster, en California, desplegó con éxito un modelo predictivo que redujo la delincuencia en 37% durante un período de tres años. Lo primero que hicieron fue contratar a un analista de datos, quien fue capaz de construir un modelo predictivo, y luego mostró los patrones de la delincuencia en mapas de calor y con códigos de color. Los mapas permitieron que los datos fueran utilizados de manera proactiva para asignar los recursos policiales en toda la ciudad al tiempo que se logró una drástica disminución de la delincuencia y un beneficio anual en la productividad de US\$1 millón (IBM, 2011b).

Las redes sociales han brindado a los departamentos de policía una nueva forma de compartir información con el público en tiempo real. Como quedó demostrado por el Departamento de Policía de Boston en su respuesta al atentado ocurrido durante la maratón de esa ciudad, el uso de Twitter para mantener actualizado al público puede ser una manera eficaz de proporcionar información importante de seguridad en forma directa, así como de disipar rumores y ofrecer información precisa.

Los datos de las redes sociales también pueden ayudar a predecir las tendencias de la delincuencia. Los datos públicos de Twitter ofrecen una vasta cantidad de información. Un estudio reciente en Chicago utilizó 1,5 millones de tuits geocodificados durante un período de tres meses, el cual mostró que en 19 de los 25 tipos de delitos estudiados, la incorporación de datos de Twitter a los modelos de predicción de la delincuencia existentes mejoró la precisión de las predicciones (Gerber, 2014).

Seguridad pública

Los limitados fondos públicos sumados al impulso constante de darle prioridad a los problemas urgentes de hoy en vez de a las cuestiones importantes de largo plazo hacen que los puentes, los túneles y las carreteras que necesitan inspección y mantenimiento presenten un gran peligro. Sería útil que los administradores de los gobiernos que se interesan por maximizar los beneficios de las costosas reparaciones y mejoras conozcan qué segmentos de la infraestructura física están por sufrir un deterioro. Sin embargo, una inspección física perfecta requeriría de un gran número de inspectores que podrían no informar de manera plenamente correcta.

A medida que los sensores de todo tipo son cada vez más económicos y más fáciles de implementar, pueden ayudar a los administradores de infraestructura a orientar sus esfuerzos de actualización y reparación. Por ejemplo, los sensores que “llaman a casa” con los datos que recogen de forma automática pueden proporcionar información sobre humedad, temperatura, contaminación

y otros problemas ambientales que hacen que la infraestructura se degrade. La “reparación predictiva” es ahora una posibilidad que utiliza datos para predecir la probabilidad de degradación, y luego establece un umbral que pone en marcha las reparaciones antes y no después de que un sistema se dañe. La programación del mantenimiento predictivo reduce los costos de largo plazo y la vida de los ciudadanos se ve menos afectada que cuando hay que salir a responder a una crisis.

El mantenimiento preventivo también puede aplicarse a los sistemas de transporte público. Por ejemplo, en el metro de Londres, una recopilación exhaustiva de datos reveló indicadores clave que se correlacionan fuertemente con fallos en equipos. Este trabajo también identifica las partes mecánicas en las escaleras mecánicas del sistema que necesitan ser reemplazadas con el fin de evitar futuras averías (Shueh, 2014). Como resultado de este trabajo, las reparaciones registraron entre 50% y 70% de eficacia en la primera visita, lo que se traduce en un ahorro considerable para el futuro y facilita el desplazamiento de los pasajeros.

Las redes de sensores pueden convertirse en infraestructuras productoras de conocimiento que permiten un despliegue mucho más eficaz de los servicios públicos. En Rotterdam, París, Lovaina y Londres, por ejemplo, el proyecto RainGain está cubriendo estas ciudades con sensores de lluvia (Rain Gain, 2012). Los datos de captación procedentes de los sensores se pueden combinar con datos de mapeo de radar para ayudar a las ciudades a predecir de forma dinámica los efectos de las inundaciones calle por calle. (Las imágenes de satélite, tomadas aisladamente sin los datos del sensor, no proporcionarían esas predicciones a un nivel tan fino.)

En Lovaina, los sensores detectan el nivel del agua del río, así como la velocidad del movimiento del agua dentro de la red de alcantarillado y la altura de las crecidas. El sistema aún se encuentra en plena construcción, pero la idea es que los responsables de los pronósticos puedan diseñar de forma más inteligente la infraestructura urbana para gestionar las lluvias periódicas y reducir los riesgos de desastres futuros. Rotterdam, una ciudad que está en su mayor parte por debajo del nivel del mar, cuenta con una red de bombas de agua móviles y depósitos de almacenamiento que permite mantener secas calles y sótanos. Las predicciones en tiempo real del sistema de radares ayudarán a que la ciudad pueda desplegar bombas donde más se necesiten y reducir el riesgo de inundación.

En Estados Unidos, por la carretera de peaje New Jersey Turnpike pasan 200 millones de vehículos al año, en una de las regiones más densas del país. El tráfico sigue creciendo, causando congestión en horas pico. Por ello, la Autoridad de Turnpike ha construido más carriles para facilitar el flujo de tráfico. Al mismo tiempo, ha comenzado a instalar sensores a intervalos regulares que recopilan información sobre el volumen de tráfico, la ocupación por carril y la velocidad. Con esa información y las predicciones que facilita, los controladores de tráfico pueden desviar los coches a lo largo de la calzada y señalar a los camiones para evitar accidentes (Goldsmith, 2014a). Las redes de sensores y el análisis predictivo permiten que los datos puedan ser utilizados para tornar más útil la costosa infraestructura física, lo que facilita una gestión de recursos más inteligente.

DetECCIÓN DE FRAUDES

Los grandes programas de ayuda social y servicios sociales sufren fraudes y abusos, con la consiguiente carga para las arcas públicas y el peligro para la financiación destinada a personas necesitadas.

Sin embargo, el coste prohibitivo de las auditorías y de la detección del fraude en manos de personas supone una gran pérdida de recursos estatales. Entidades como la Dirección General de Servicios de Integridad de Service Canada —que se ocupa de desembolsar los fondos del seguro de desempleo— están usando el análisis predictivo de riesgos para detectar fraudes y abusos (Oficina del Comisionado de Protección de la Privacidad de Canadá, 2012). El Departamento de Servicios Sociales Públicos del Condado de Los Ángeles hace lo mismo: analiza sus datos para encontrar anomalías, que luego pueden ser utilizadas para priorizar las tareas de los investigadores responsables de detectar posibles fraudes dentro del programa de cuidado de niños del estado (Goldsmith, 2014b). Los patrones de estos datos revelan indicadores que inducen a una correlación histórica con la presencia de fraude. Al centrarse en lo previsible de esos indicadores en lugar de utilizar otro método para priorizar tareas, los auditores pueden aumentar la probabilidad de que su trabajo dé frutos y obtenga una mayor tasa de éxito. La presencia de tales indicadores se convierte en una señal de alerta de que puede haberse cometido un abuso, y las auditorías resultantes conservan recursos. En 2012, este enfoque produjo un ahorro sustancial: se señalaron 200 casos con 85% de precisión, lo que supuso un ahorro para el departamento de US\$6,8 millones en los casos de fraude eliminados (Heaton, 2012). Los organismos que utilizan los datos pueden ser proactivos: dejar de depender de indicaciones de líneas telefónicas o de corazonadas para descubrir el fraude, y más bien saber exactamente qué patrones fácilmente detectables pueden estar asociados con el fraude e investigar esos hechos directamente.

Recursos humanos

Como las prácticas presupuestarias han mejorado, dejando al descubierto las verdaderas fuentes de costos en el funcionamiento de grandes entidades, los responsables de finanzas se han dado cuenta de lo costoso que resulta atraer y mantener al capital humano. El análisis predictivo puede ayudar a mejorar el éxito de las prácticas de contratación (Fitz-Enz, 2009), y su aplicación tendría buenos resultados en los sectores militares y en otros que invierten fuertemente en sus recursos humanos, pero corren el riesgo de perder la inversión si los nuevos empleados no perduran. El análisis predictivo puede utilizarse para mejorar la contratación desde el comienzo y fomentar la retención de los trabajadores. También pueden utilizarse indicadores que predigan el éxito y la satisfacción en el trabajo a largo plazo y así centrar los escasos recursos de la administración en materia de contratación y asesoría en aquellos soldados que tengan más probabilidades de permanecer en las filas a largo plazo (Web Builders, 2011).

Varias empresas de punta —por ejemplo, Google, Cisco y General Electric— están utilizando el análisis predictivo para orientar su contratación en lugar de confiar en su intuición acerca de los candidatos (Human Capital Media, 2014). La Oficina de Administración de Personal del Gobierno de Estados Unidos está aplicando el análisis predictivo para aportar en el proceso de toma de decisiones respecto de puestos a cubrir e individuos más adecuados para desempeñar un trabajo. Además, esta oficina está utilizando diversos datos sobre los trabajadores que guían la toma de decisiones, como la evaluación de habilidades, la diversidad en el trabajo, los requisitos de jubilación, la ubicación de los empleados, los patrones de contratación y jubilación y las estadísticas de renovación de personal (Ward, Tripp y Maki, 2012).

Servicios sociales

Cuando se trata de la prestación de servicios tales como protección de la infancia, la diversidad de los casos existentes y la dificultad e importancia de trabajar con poblaciones en riesgo dificultan el logro de resultados positivos. La cuestión se complica porque, aunque los proveedores experimentados desarrollan una habilidad especial a lo largo de los años para conectar a los niños adecuados con los servicios pertinentes y mejorar los resultados, es difícil transmitir el conocimiento entre los trabajadores de manera formal. El conocimiento se pierde y debe ser reconstruido cada vez que un trabajador social o un gerente dejan de trabajar en el servicio.

Indiana —como se vio anteriormente— tiene la intención de utilizar el análisis para ayudar a reducir la mortalidad infantil (Howard, 2013). Aunque los funcionarios eligieron ese objetivo en particular debido a su viabilidad política —al ser una causa que todo el mundo puede apoyar—, el esfuerzo no es excesivamente caro. El estado ha recibido una donación de US\$500.000 de una organización filantrópica privada para llevar a cabo este proyecto, así como otras mejoras tecnológicas; la mayor parte del dinero proviene del presupuesto regular de TI (Hughes, 2014).

La Fundación para la Juventud de Medway ha comenzado a utilizar el análisis predictivo en lugar del laborioso proceso utilizado anteriormente para encontrar e identificar a adultos jóvenes en situación de alto riesgo. Al analizar sus registros digitales y la búsqueda de correlaciones históricas entre las características particulares de los jóvenes y los resultados, la Fundación es ahora capaz de predecir qué jóvenes tendrán problemas en el futuro. Ha sustituido un proceso manual que implicaba el examen de miles de registros por una forma mucho más rápida y precisa de vincular a los jóvenes de alto riesgo con los servicios y las oportunidades que necesitan para tener éxito. La sustitución de los procesos manuales por el análisis puede ayudar a todas las partes del sistema: puede reorientar a los jóvenes para que no tomen caminos que resultan ser peores para ellos y más caros para el estado (IBM, 2011a).

El Departamento de Servicios para Personas sin Hogar de la Ciudad de Nueva York está trabajando con la Fundación SumAll para analizar en detalle los datos de registro de entrada a los refugios. Si se cuenta con información de que el domicilio reciente de una familia se ha convertido en un punto crítico para un gran número de problemas —tales como el desalojo—, la municipalidad puede centrar sus recursos limitados en esa dirección y reducir el riesgo futuro de que las familias tengan que entrar al sistema de falta de vivienda. Al volver su atención hacia las causas de la falta de vivienda en lugar de simplemente reaccionar a sus consecuencias, la ciudad puede avanzar hacia mejores resultados para padres e hijos (Mascarenhas, 2014). Esto se traduce en menores costos para la ciudad y servicios de mejor calidad y más definidos para las poblaciones en riesgo.

Recaudación de impuestos

Los organismos tributarios han comenzado a utilizar el análisis predictivo para hacer un primer análisis de sus propios registros, marcando los casos probables de impuestos impagos o de prácticas contables engañosas. Sobre la base de esas marcas, pueden priorizarse determinados casos para auditorías y exámenes jurídicos (SPSS, 2010). El análisis predictivo también puede utilizarse para complementar

las técnicas de predicción económica existentes, permitiendo que los organismos comprendan mejor los efectos que las nuevas políticas pueden tener sobre los ingresos fiscales futuros (Desigan, 2011).

Por ejemplo, el estado de Nueva York utiliza el análisis predictivo para detectar de forma preventiva devoluciones de impuestos cuestionables en su flujo de trabajo de recaudación de ingresos, al tiempo que se ahorra el costo de tratar esas devoluciones problemáticas más tarde. Desde su implantación, el programa ha ayudado a que se recuperen US\$400 millones que se estaban perdiendo anualmente por devoluciones ilícitas, además de US\$100 millones derivados de una mayor disponibilidad de ingresos. La dirección del departamento está conforme con el hecho de que la inversión se realizó en un mejor clima económico, cuando abundaban los fondos y se podía tomar ese tipo de medidas preventivas. Los esfuerzos están dando sus frutos (IBM, 2011c).

Gestión de desastres

Cuando se produce un desastre, un rápido acceso a predicciones sobre el impacto y estimaciones sobre los recursos necesarios pueden suponer la diferencia entre la vida y la muerte. Las técnicas predictivas permiten analizar —a través de la información que se recibe— las acciones que podrían necesitarse en un tiempo menor que el que le llevaría a una persona darse cuenta de lo que está ocurriendo, realizar un cálculo mental de todas las variables y tomar la decisión consciente de intervenir.

Estados Unidos sigue desarrollando su Marco Nacional de Respuesta (NRF) y el Sistema Nacional de Gestión de Incidentes (NIMS), ambos destinados a gestionar información analítica y obtener información necesaria en tiempo real para quienes están en condiciones de reducir los riesgos de desastres (UICDS, sin fecha). Estas dos estructuras nacionales y de respaldo, que forman parte de la planificación de la seguridad interna del país, están destinadas a aumentar la capacidad de los organismos de gestión de desastres de todo el país —tanto grandes como pequeños—³ con el propósito de responder a los desastres de manera efectiva. La disponibilidad de modelos meteorológicos que siguen mejorando contribuye a tales esfuerzos: ahora los científicos de datos pueden predecir con mayor precisión hacia dónde podría avanzar el mal tiempo, y, en combinación con los datos de SIG, la información puede ser de utilidad en todas las etapas de un desastre. Estos modelos ayudan a predecir el impacto antes de las tormentas y permiten la planificación de la evacuación, o se utilizan durante un desastre para que las agencias locales sepan dónde se necesita ayuda. Su capacidad de planificar con anticipación para dirigir los esfuerzos de asistencia durante el siniestro y la recuperación posterior al siniestro servirán para ahorrar dinero y brindarles mayor seguridad a los ciudadanos estadounidenses.

Salud pública

Cuando Estados Unidos observó el crecimiento de la pandemia de gripe H1N1 en 2009, los Centros para el Control y la Prevención de Enfermedades y los Institutos Nacionales de Salud recurrieron al análisis predictivo para guiar sus actividades y las advertencias de salud sobre la base de la

³ Por ejemplo, la Alianza de Planificación y Capacitación para Incidentes Críticos en la Zona de Operación de Los Ángeles (<http://www.catastrophicplanning.org/alliance.html>) y la Administración de Emergencias del Condado Cowell (<http://www.cowleycounty.org/wp-content/uploads/ESF-5-Emergency-Management.pdf>).

propagación prevista y la fuerza del virus (Desigan, 2011). Herramientas como Flu Trends (evolución de la gripe) de Google y Mapmyhealth pueden añadir nuevas fuentes de datos para ayudar a centrarse en los lugares donde podrían estar propagándose enfermedades como la gripe, mucho antes de que empiecen a repuntar las mediciones tradicionales, como las visitas a las urgencias hospitalarias o las pruebas de laboratorio. Flu Trends de Google recoge y analiza geográficamente las búsquedas de los usuarios sobre recursos relacionados con la gripe, mientras que Mapmyhealth y SickWeather detectan tuits y mensajes de Facebook en tiempo real de personas que indican sentirse enfermas con gripe (Konkel, 2013). Estos pasos pueden ayudar a salvar vidas. Hasta el momento en que un paciente acude al hospital, podría perderse la oportunidad de evitar nuevos brotes de la enfermedad.

Ciertos usos experimentales de datos procedentes de las redes sociales en masa pueden enriquecer la capacidad de predecir y realizar un seguimiento de muchos problemas de salud pública. Por ejemplo, investigadores de la Universidad de Rochester han demostrado que los tuits geolocalizados pueden utilizarse para controlar intoxicaciones alimentarias en restaurantes neoyorquinos, organizando la prioridad de la ardua tarea de los inspectores de alimentos de la ciudad mediante el descubrimiento de datos que les permiten focalizarse en los restaurantes con picos en los “informes” de las redes sociales (Kautz, 2013). No obstante, aún no se ha lanzado la aplicación y sigue siendo un producto de investigación, en gran parte debido a intereses arraigados. Dado que la aplicación utiliza datos que provienen del público, grupos como la Asociación de Restaurantes del Estado de Nueva York se han opuesto a su uso. Les preocupa que los competidores puedan manipular los datos de origen para “encasillar a los restaurantes” (Taney, 2013). Los investigadores están trabajando para mejorar los algoritmos con el fin de hacerlos inmunes a tal manipulación, pero su falta de aplicación apunta a dos cuestiones fundamentales en relación con el análisis predictivo: i) si se utilizan datos que provienen del público, algo que es común, se deben tomar medidas para garantizar que no hayan sido manipulados, y ii) incluso así, puede que ciertos grupos e intereses políticos protejan ferozmente el *statu quo*.

Las técnicas analíticas han permitido a investigadores de Microsoft trabajar hacia la identificación de mujeres en riesgo de depresión posparto grave mediante los cambios en los patrones de sus hábitos en línea y el contenido de sus publicaciones (Kautz, 2013). Otros investigadores de Microsoft han demostrado que pueden descubrir interacciones de medicamentos problemáticas que no se conocían anteriormente a través de un análisis de búsquedas anónimas de usuarios en la web (Greenemeier, 2013). Será fundamental pensar en las implicaciones de estas metodologías respecto de la privacidad y el consumo.

El enfoque predictivo también se utiliza con fuentes de datos médicos más tradicionales. IBM y el Instituto de Tecnología de la Universidad de Ontario están empleando mediciones biomédicas de bebés prematuros controlados a fin de predecir infecciones potencialmente mortales casi un día antes de que puedan observarse de otra forma (IBM, 2010). Ese único día puede darles a los médicos una ventaja esencial para iniciar el tratamiento y prevenir que las infecciones evolucionen hacia complicaciones más graves.

El sistema de distribución ordinaria de los servicios de asistencia sanitaria también sufre cada vez más debido al envejecimiento mundial de la población, un hecho que consume más recursos.

La Ley de Protección al Paciente y Cuidado de Salud Asequible prevé incentivos financieros para los hospitales que reduzcan sus tasas de reingresos de pacientes hospitalizados en los 30 días anteriores, pues esas visitas repetidas tienen un costo anual de US\$30.000 millones (Health Affairs, 2013). Para encontrar maneras de reducir esos costos, el Heritage Health Prize —un concurso organizado por Heritage Provider Network— utiliza datos sobre las solicitudes de reembolsos de gastos médicos para predecir qué pacientes tienen más probabilidades de ser reingresados. Los pacientes reciben una atención especial tras su alta hospitalaria y en el tratamiento posterior a la hospitalización, con el objetivo de reducir o evitar el reingreso. Del mismo modo, el King's Fund británico intenta crear un sistema de software para identificar a los pacientes que tienen un alto riesgo de reingreso hospitalario, de forma tal que los hospitales locales de atención primaria puedan intervenir y reducir el riesgo de que los pacientes reingresen (The Kings Fund, 2014).

El análisis predictivo y el aprendizaje de máquinas también pueden utilizarse para complementar los estudios metodológicos médicos estándar. Utilizando los datos de las solicitudes de reembolso de gastos médicos y farmacéuticos, los científicos de datos pueden comenzar a probar indicadores respecto de múltiples combinaciones de terapias que pueden vincularse con los resultados deseados. Este complejo enfoque —que pocas mentes humanas serían capaces de concebir— ayudará a los proveedores a desarrollar soluciones sanitarias a gran escala, por ejemplo para la diabetes de tipo 2. Hacer frente a estos problemas masivos tendrá un impacto importante en los presupuestos de salud pública; se espera en los próximos años un crecimiento significativo del costo que le genera a la sociedad la diabetes de tipo 2 (Maguire y Dhar, 2013). Este mismo enfoque predictivo impulsado por los datos puede utilizarse para que la prestación de servicios de salud gire en torno a la prevención y no sólo al simple reembolso de visitas y servicios, lo que implica dedicarle atención y recursos a la salud pública (el resultado deseado) en vez de simplemente analizar los resultados de los hospitales públicos y otras instituciones.

En Estados Unidos, los sistemas de protección sanitaria del país para ancianos y pobres —Medicare y Medicaid— han adoptado tecnologías de análisis predictivo con el fin de detectar el fraude. La enorme cantidad de datos recogidos por estos programas se podrían utilizar para comprender mejor los patrones de las actividades de todos los usuarios, no sólo de los implicados en el fraude. En otras palabras, las características agregadas y las acciones de los beneficiarios podrían utilizarse para adaptar la elegibilidad y los planes de cobertura de forma predecible. Como resultado, la población podrá recibir una cobertura más eficiente, que les ofrezca un servicio de salud asequible a más segmentos de la población y construya un mejor nivel de salud para la ciudadanía en general.

Educación

Una importante aplicación del análisis predictivo en la educación que ha surgido recientemente es detectar a estudiantes en riesgo de abandonar la escuela. Las instituciones educativas pueden proporcionar apoyo a esos estudiantes incluso antes de que se den cuenta de la posibilidad de riesgo (Luan, 2004). El apoyo ha sido posible gracias a las mejoras en la capacidad de las instituciones de realizar un seguimiento de los alumnos de una institución a otra, ya que los estudiantes en riesgo tienden a pasar de una escuela a otra. Antes, las barreras legales hacían difícil ese seguimiento

intercentros. Hoy en día, el National Student Clearing House permite establecer correspondencias entre las universidades y los colegios comunitarios, lo que permite una comprensión más completa de las características y los comportamientos de los estudiantes. Cada vez existe mayor cantidad de datos exhaustivos que pueden utilizarse como indicadores; los sistemas pueden incorporar datos sobre calificaciones, sobre ayuda financiera y sobre los estudiantes a fin de construir un modelo predictivo que permita priorizar las intervenciones y proveer recursos adicionales para los estudiantes de alto riesgo (Barber y Sharkey, 2012). Aunque se trata de iniciativas que están comenzando, es evidente que el análisis predictivo ocupará un lugar importante en la educación, en particular para garantizar que los programas públicos de ayuda y apoyo a los estudiantes desfavorecidos sean lo más eficaces posible y generen los índices de graduación más altos que se puedan esperar. Los programas están en condiciones de lograrlo, ayudando a los estudiantes a medida que cambian sus circunstancias en lugar de esperar a que surjan problemas para tomar medidas; entonces, podría ser demasiado tarde.

Conclusiones

Un gobierno de mejor calidad, más rápido y más económico está a nuestro alcance, tanto en los países en desarrollo como en los países desarrollados. Un número cada vez mayor de gobiernos nacionales y subnacionales están aprovechando herramientas de código abierto para poner en marcha usos de datos que darán como resultado gobiernos más eficaces y transparentes. De hecho, la alta penetración de los teléfonos inteligentes en los países en desarrollo puede ser una ventaja adquirida para los actores de la administración pública: los empleados públicos que prestan servicios pueden informar a bajo costo lo que están haciendo a las bases de datos centrales —gracias a las tecnologías de GPS—, y quienes dispongan de dispositivos portátiles podrán visualizar las iniciativas de datos abiertos inmediatamente.

La tecnología ha proporcionado un espectacular conjunto de herramientas que pueden ayudar a los empleados públicos a resolver problemas y producir mejores resultados con la misma cantidad de esfuerzo, pero aplicadas de forma más inteligente. Si nuestro objetivo es aumentar la correlación entre el esfuerzo y los resultados, el análisis predictivo puede mejorar la relación de forma notable. Disponemos de herramientas de computación en la nube, minería de datos y análisis de sentimiento, una conectividad casi universal gracias a los teléfonos inteligentes y otros dispositivos móviles en manos de los trabajadores y los ciudadanos, y herramientas de procesamiento desconocidas o inaccesibles tan solo unos pocos años atrás.

En el camino lo único que se interpone somos nosotros y la forma en que gestionemos el gobierno. Si se ponen en marcha las funciones directivas ejecutivas, las herramientas que abarcan a múltiples organizaciones, la transparencia y los datos abiertos, y nos concentramos en los resultados —y no en las actividades— podremos efectivamente lograr un gobierno más reactivo. Los datos impulsarán mejores servicios, formas más efectivas de gestión y evaluación de los empleados y, mediante la transparencia y las redes sociales, una participación ciudadana más extendida.

Pasos hacia el éxito: por dónde empezar

Un Director de Datos (DD) que llegue a la Alcaldía y quiera utilizar el análisis predictivo tendrá que tener en cuenta diversas cuestiones: la disponibilidad y naturaleza de los datos públicos, las habilidades de los empleados, la influencia política de la oficina y la naturaleza de los problemas que pueden y deben abordarse. A continuación se indican una serie de preguntas básicas que deberá responder cualquier Director de Información o Director de Datos que desee instaurar una actividad de análisis predictivo.

Para los propósitos de este debate, la persona a cargo del programa de análisis predictivo no es directamente responsable de las exigentes y prolongadas tareas relativas a la protección de datos y a las necesidades diarias de tecnología de información de su ciudad o estado (por ejemplo, velar por que funcionen los servidores de correo electrónico u ocuparse de las solicitudes de ayuda técnica de los trabajadores). El DD que tenemos en mente se ocupará de la gestión de las bases de datos, los datos abiertos, el análisis avanzado, la inteligencia empresarial y el almacenamiento de datos.

¿Cuál es la infraestructura de datos existente? La primera capa de cualquier capacidad de análisis predictivo son los datos, que pueden estar (y probablemente estén) distribuidos por diversas bases de datos controladas por diferentes organismos municipales o estatales. Es posible que tanto organismos como ciudades y estados correspondientes aún no estén recopilando datos que serán fundamentales, como la información que fluye desde los dispositivos GPS en autobuses y trenes.

El DD debe elaborar un inventario completo de las bases de datos existentes y de la descripción de los campos de esas bases de datos, a lo que suele llamarse “documentación”. Si la documentación no existiera, el DD velará por su creación. A menudo, diversos individuos aislados dentro de los organismos mantienen bases de datos especializadas y sistemas específicos de la empresa. Esas personas, que conocen lo que hay en el sistema y su utilidad, no han tenido incentivo real alguno para ofrecerle esa información a cualquier otra persona. La investigación inicial del DD será necesaria para conocer los datos o el sistema disponible para la operación.

La ciudad de Chicago ha lanzado recientemente un “diccionario de datos”⁴ que describe los campos contenidos en todas sus bases de datos, el primero de ese tipo. Aunque Chicago no da a conocer todos sus datos al público, está catalogando la totalidad de los distintos tipos de datos, lo que les permite al personal de la alcaldía, los residentes y demás interesados conocer los recursos de datos de la ciudad, cómo se puede acceder a ellos y en qué formatos. Otras ciudades están adoptando diccionarios de datos interoperables similares. Si el DD ha agregado (o encargado) desde el inicio documentación sobre los datos en uso, algún día será posible participar en iniciativas de análisis de datos entre ciudades, lo cual será una herramienta muy potente.

⁴ Véase <http://datadictionary.cityofchicago.org/>.

En este punto también sería útil entender las diferencias entre los datos que está recopilando la oficina gubernamental del DD y los datos recogidos por otras ciudades o estados. En definitiva, el DD puede ayudar al gobierno a aumentar sus esfuerzos de recopilación de datos; para ese paso será esencial contar con un inventario completo y un análisis de deficiencias a mano.

Además de averiguar quién se ocupa de cada tipo de datos recopilados, el DD tendrá que determinar qué requisitos de transparencia, calidad y retención aplicará a los datos con que vaya a trabajar. También será útil una estrecha y cordial coordinación con el departamento jurídico, y con el oficial de seguridad de las TI encargado de tales responsabilidades.

Por último, esta etapa de inventario tendrá que incluir una evaluación sobre la cantidad de datos (y qué tipo de datos) que está bajo el control físico directo del cargo de DD. Diferentes sistemas de datos logran diferentes equilibrios entre centralización y descentralización: el Centro de Administración y Desempeño del estado de Indiana, descrito anteriormente, fue relativamente fácil de crear porque el director de información del estado era quien albergaba los datos de todos los departamentos del estado.

En cambio, la actividad de análisis de datos de Chicago se ha construido a través de bases de datos distribuidas (es decir, muchos conjuntos de datos alojados en diversos departamentos de la ciudad y no en el Departamento Central de Innovación y Tecnología) y ha trabajado con procesos automatizados (de extraer, transformar y cargar, denominados procesos ETL) en las bases de datos distribuidas. Con el tiempo, Chicago ha convencido a los organismos de que otorguen acceso a sus datos en pos del análisis, demostrándoles que es posible trabajar de forma mucho más rápida y eficiente si los alojan en servidores centralizados de avanzada.

En general, las bases de datos transaccionales existentes (sobre delincuencia, servicios municipales, permisos) no deben usarse para el análisis predictivo; están ocupadas prestándoles servicios a la ciudad o al estado y no se debe interferir. En su lugar, el trabajo del DD es construir un motor o capa de análisis que se alimenta constante y automáticamente con datos actualizados extraídos de esas bases de datos transaccionales.

Conclusión: Cuanto más sepa el DD sobre los recursos de datos de su ciudad, más eficaz será.

¿Cuáles son las necesidades analíticas en cada caso? No todos los análisis son predictivos. A veces se necesita simplemente una evaluación: ¿qué porcentaje de graduados está produciendo un sistema escolar en comparación con otros equivalentes? ¿Ha sido un sistema de cámaras de la

calle manipulado por un proveedor sospechoso para poner más multas de tráfico de las que debería? ¿Cuánto dinero ha gastado la alcaldía en un determinado sector transversal?

Si el análisis predictivo es ya una necesidad estratégica del municipio o el estado, el DD tendrá que fijar los principios de su departamento: ¿dará prioridad al análisis que tiene un impacto operacional inmediato (que dé lugar, por ejemplo, a que alguien en un camión con luz amarilla intermitente sea enviado a una determinada calle para hacer una reparación concreta), o al que resulta útil para la planificación económica a largo plazo (un plan de veinte años para un parque)? La ciudad de Chicago definió como prioridad el impacto operacional inmediato, construyendo casos de uso para el análisis predictivo en estrecha coordinación con los departamentos municipales. En Chicago, los arquitectos también están utilizando el análisis predictivo a partir de datos históricos ambientales y de transporte para planificar la construcción de una nueva sección de la ciudad; ese proyecto de construcción se extenderá al menos durante una década. La mayor parte de los DD municipales o estatales han de operar con recursos financieros limitados y fijar prioridades, mediante la elaboración de una norma que asista en la aceptación o el rechazo de los proyectos propuestos, lo que contribuirá con la asignación de fondos que de por sí escasean.

Las promesas exageradas en torno a los beneficios del análisis predictivo —que crean grandes expectativas para los responsables gubernamentales, y los hacen pensar que los datos serán como una bola de cristal que les permitirá ver el futuro de su ciudad o estado— pueden resultar destructivas. En este caso, los directores generales serán propensos a plantear preguntas vagas, tales como “¿cómo podemos mejorar la vida de los ciudadanos?”. Ningún DD puede responder a esa pregunta, porque su respuesta requiere de una enorme cantidad de opciones y supuestos de política (¿qué significa “mejorar”?) que son inherentemente subjetivos. El DD que empiece por pequeños problemas, cuya solución será una mejora pequeña pero visible en servicios municipales, probablemente tendrá más éxito. En particular, dado que toda la idea del análisis predictivo es nueva, el progreso marchará sobre ruedas cuando los resultados sean visibles: es difícil para las personas imaginar la utilidad de alguna cosa si no la pueden ver. Centrarse en éxitos tempranos y visibles estimulará partidarios y entusiasmo para llevar a cabo proyectos más grandes e incorporar más personal.

He aquí un ejemplo de un problema muy simple y pequeño: los sistemas de pago a veces generan errores, y la suma de errores muy pequeños puede suponer un monto considerable. Si el conjunto de datos que contiene información sobre los pagos cumple con los requisitos estadísticos (entre ellos, si es lo suficientemente grande y tiene valores que están ampliamente distribuidos), ejecutar un simple algoritmo de “distribución esperada del primer dígito” en los datos revelará si existen pagos con valores atípicos que deban examinarse.

Otro ejemplo de una clase de problema que podría abordarse con un análisis visible y directo sería preguntar: “¿dónde se están dando realmente los problemas de la ciudad (delincuencia, basura), y cuáles han sido los factores impulsores de esos problemas?”. La mayoría de las ciudades tienen una licencia de ArcGIS de Esri, un software que facilita las representaciones espaciales de los datos. Diversas bases de datos transaccionales (las de delincuencia, datos de llamadas al 311, datos de inspección de edificios, etc.) pueden cargarse en ese software, y luego se pueden

comparar y analizar pequeñas áreas, lo que permite revelar indicadores que guarden una posible relación con la variable dependiente de interés.

Algunos DD deciden comenzar con la creación de mediciones de rendimiento de los departamentos municipales y luego utilizan el análisis para evaluar si se cumple con las mediciones. Otros consideran que se trata de una distracción: el alcance de las preguntas que pueden responderse mediante indicadores clave de rendimiento se verá limitado por los datos que ya están siendo recogidos por el organismo en cuestión (“no tenemos datos sobre eso, así que no podemos medir nuestra actuación allí”), haciendo que el proceso sea circular.

Algunos DD quieren trabajar principalmente en la investigación inductiva, revisando enormes cantidades de datos dispares para encontrar patrones inesperados y luego explorarlos (“¿qué hay en nuestros datos del 311 y del 911 que sea interesante e inesperado?”). Otros se centran más en un trabajo deductivo, en busca de problemas discretos, encontrando factores impulsores y creando soluciones de prueba para mejorar el *statu quo*. Independientemente del esquema que el DD decida priorizar, debe estar al día y estrechamente alineado con las prioridades estratégicas del jefe del Ejecutivo.

Conclusión: El éxito del análisis predictivo conllevará una mayor demanda de proyectos de análisis de datos. El DD debe tener claras sus prioridades, y refinarlas a la luz de lo que vaya aprendiendo con el tiempo.

36

¿Quiénes apoyan el proyecto? La aceptación del jefe del Ejecutivo será esencial para que cualquier actividad de análisis predictivo sea exitosa. Chicago, una vez más, es un ejemplo de ciudad donde el líder que entiende de datos —el Alcalde Rahm Emanuel— les ha otorgado a sus empleados espacio para llevar a cabo una amplia labor de análisis predictivo. El personal de análisis del Alcalde Emanuel, a su vez, ha procurado trabajar en estrecha colaboración con los departamentos municipales a fin de determinar cuáles son los problemas clave de funcionamiento que quieren resolver (incluido el despliegue de recursos de la ciudad).

Contar con apoyo dentro de los organismos, es decir, con personas que conocen los datos que su departamento controla y muestran entusiasmo respecto del poder del gobierno basado en datos, es extremadamente valioso, pero no suficiente sin el respaldo de una autoridad política del más alto nivel de gobierno. Un análisis de datos eficaz requiere de cobertura política, lo que se traduce en mayores recursos y credibilidad asegurada, dos atributos útiles cuando se trata de persuadir a los diferentes integrantes de cualquier gobierno de que se sumen al trabajo colaborativo. Un DD competente tendrá que ser a la vez experto en análisis y eficaz en mantener relaciones laborales fluidas con las burocracias rígidas.

Un antiguo DD señaló que habría deseado saber más sobre política y juegos de poder antes de asumir su cargo. No había entendido la importancia de la diferencia entre los funcionarios de carrera y los designados por políticos, y no estaba acostumbrado a tener que “contar una historia” sobre la importancia del análisis. Se trataba de un experto en datos, impaciente con la lentitud de las funciones públicas, que nada sabía sobre normas de contratación, palabras de moda o demandas de los grupos de presión. En sus palabras: “lo de los datos es fácil; pero todo el mundo te pondrá trabas”.

Conclusión: Es importante que el papel del DD sea visto como un respaldo directo a los objetivos generales del municipio (o del estado) y que el jefe del Ejecutivo le dé su apoyo indiscutido al DD.

¿Qué habilidades tienen los empleados? Un inventario de las bases de datos será tan importante como un inventario de las personas. Identificar y ampliar los talentos será esencial para cualquier tarea de análisis predictivo exitosa.

El DD necesitará administradores de bases de datos que sean curiosos, enérgicos y expertos en análisis estadístico, y que logren hacer cosas. Aunque sus experiencias se hayan limitado al trabajo con software privado (Oracle, por ejemplo), pueden recibir nueva capacitación que les permita trabajar con paquetes de análisis de código abierto, y probablemente sentirán la satisfacción de aprender cosas nuevas.

El DD también necesitará un administrador de ETL a tiempo completo, alguien cuya tarea consista en construir y mantener procesos de ETL que extraigan información estable a partir de bases de datos existentes y la carguen automáticamente en el motor de análisis predictivo. Este es el meollo de la tarea del análisis predictivo; no hay sustituto para los procesos de ETL, y crear y ampliar esa tarea resulta laborioso.

Será esencial contar con gerentes de proyectos especializados en trabajar con departamentos municipales, tanto por sus capacidades políticas como por su experiencia específica; puede tratarse de personas menos técnicas con experiencia en cuestiones de política que son fundamentales para el jefe del Ejecutivo (transporte, clima, etc.), con experiencia en la gestión de procesos entre organismos y con excelentes competencias comunicativas. Tales gerentes pueden estar en contacto constante con los departamentos, comprenderán los problemas que enfrentan los departamentos y tendrán la oportunidad de sugerir la forma de trabajar con el DD que permita facilitar la tarea del organismo.

También es importante contar con personas versátiles, flexibles, que tengan algunos conocimientos técnicos (y, lo que es más importante, experiencia en visualización) y se centren en el

servicio al cliente. Será esencial un trabajo de desarrollo sutil a fin de que la historia que acompaña cualquier experimento de análisis predictivo sea accesible a personas ajenas al tema, y la utilización de herramientas web simples y fáciles de usar para que los departamentos presenten sus propios relatos, lo que ayudará a sumarlos al programa de análisis predictivo del DD. Estos generalistas pueden servir a los empleados técnicos más especializados que trabajan en el análisis o en las bases de datos y dar apoyo en el desarrollo de presentaciones e interfaces web.

También son fundamentales los científicos de datos que se sientan cómodos con las estadísticas, incluidos los elementos necesarios de cualquier declaración estadística válida, y que estén especializados en el trabajo con datos. Deben tener experiencia en pronósticos y predicciones, saber diseñar y analizar una experimentación (test A/B y otras técnicas) que valide los tratamientos diferenciales, ser creativos con soluciones de diseño, muestreo y medición de las pruebas en escenarios no convencionales, ser capaces de escribir guiones y tener experiencia en R u otros paquetes de análisis estadístico. A menudo, será necesario asociarse con las universidades locales para llevar a esas personas al gobierno local o estatal.

Si el DD elige crear un portal de datos abiertos, será muy valioso contar con alguien que tenga conocimiento sobre datos abiertos y pueda aportar ideas acerca de los atributos de las bases de datos que se estén ofreciendo a través del portal. Los datos abiertos pueden ser un motor importante para las actividades de análisis predictivo; los mismos comandos utilizados para automatizar la disponibilidad de datos actualizados a los fines del análisis serán necesarios para poblar un portal de datos, lo que significa que puede usarse el código de programación para propósitos dobles. Los profesionales han observado que los portales de datos abiertos son a menudo incluso más útiles dentro del gobierno que fuera de la alcaldía, pues un organismo puede utilizar el portal para consultar los datos que otros organismos han hecho públicos sin tener que usar el teléfono. Un DD con autoridad tanto en lo que respecta a los datos abiertos como al análisis avanzado puede hacer que las dos funciones se complementen mutuamente y que las inversiones en personal y recursos sean útiles para ambos.

Conclusión: El DD tendrá que combinar la comprensión de los procesos de gobierno y la deservoltura técnica. Poner en marcha un equipo con experiencia en estas áreas y fomentar la comunicación permanente entre los miembros de ese equipo debe ser una prioridad.

¿Con qué proveedores debe trabajar? Los DD gubernamentales son un importante objeto de atención de los proveedores, y cualquier DD tendrá que ocuparse de los procesos de contratación existentes. El sector del análisis predictivo está en auge, y los proveedores saben que son

pocos los empleados gubernamentales veteranos que tienen experiencia en ese campo; a menudo, los empleados de TI actuales están acostumbrados a llevar a cabo procesos de adquisición de TI independientes. Sin embargo, en la actualidad, un DD bien informado tiene poder: los proveedores suelen compartir información con los actores gubernamentales sobre sus aplicaciones en el sector privado, debido a que el gobierno local no está compitiendo con ellos y porque quieren hacer la venta. Por lo tanto, se deben formular preguntas.

Por otra parte, un DD ahora puede recurrir a muchas herramientas de código abierto gratuitas que pueden brindar un intenso apoyo para el análisis. Aún con recursos limitados y diferentes bases de datos transaccionales existentes, un DD puede utilizar múltiples computadoras y servidores de punta, un software de ETL de código abierto (Pentaho) para absorber los datos distribuidos, un software de código abierto que permita el procesamiento distribuido de grandes bases de datos a través de grupos de servidores básicos (Hados), plataformas de análisis de código abierto o capas (PostgreSQL o MongoDB) para lidiar con esos datos, un programa estadístico de código abierto (R) para elaborar cuadros de regresión, un lenguaje de programación de código abierto (Python) para ejecutar procesos con los datos, y un paquete de visualización de bajo costo (Tableau); todo con miras a obtener resultados útiles.

Los empleados que entienden las estadísticas (por ejemplo, ¿presentan los datos una distribución normal? ¿Hay suficientes datos? ¿Cuándo se permiten las regresiones?) pueden aprender a llevar a cabo estas funciones y hablar de los resultados con claridad y conocimiento de causa. Afortunadamente para los empleados del gobierno, la tecnología de la información se está convirtiendo en consumo generalizado; por lo tanto, depender menos de los proveedores privados no sólo permite generarle ahorros considerables al municipio o al estado, sino también liberar recursos para invertir en las personas en el largo plazo. Por otra parte, a medida que más ciudades y estados utilicen estas plataformas y creen paquetes reutilizables, la disponibilidad de proyectos de análisis relativamente económicos será cada vez mayor.

Los DD han de ser cuidadosos acerca de las decisiones de “hacer o comprar”. Existen numerosos productos con una multiplicidad de funciones de compañías como Oracle, IBM y SAP que pueden ser útiles para algunas funciones, pero no necesarios para las demás. Por ejemplo, las plataformas privadas en la nube pueden resultar atractivas para un DD porque resuelven diversos problemas a la vez. (Después de todo, otra persona ha comprado la licencia de hardware y el software utilizado por la nube, y los empleados contratados por la empresa de la nube no son partidas en el presupuesto del DD.) Si un DD trabaja intensamente con datos las 24 horas del día, el alquiler de una instalación en la nube podría convertirse rápidamente en un gasto prohibitivo; cada vez que el DD hace una pregunta, se incurre en otro gasto.

Por otro lado, el alquiler de recursos de la nube puede ser útil cuando el DD necesita tener acceso a capacidades de procesamiento ampliado por un breve período de tiempo. Es posible que el DD desee utilizar una gran cantidad de potencia de computación para llevar a cabo una tarea en particular, y luego llevar los resultados de vuelta a su propio sistema; en este contexto, el alquiler puede tener sentido. Del mismo modo, los servicios de Oracle pueden ser extremadamente útiles, pero la acumulación de gastos es brusca.

No se intenta sugerir que los DD deben operar sin el apoyo de los proveedores. Sin embargo, deberían tener cuidado de no comprometerse con paquetes comerciales de arquitectura muy sofisticada, sobre todo en el nuevo contexto, donde hay una enorme cantidad de alternativas disponibles.

Conclusión: Un DD exitoso formulará preguntas inteligentes a los proveedores y estudiará sus opciones cuidadosamente.

¿Cómo se demostrará la eficacia de los proyectos de datos? Un DD siempre estará en busca de nuevas oportunidades para integrar y aprovechar la información de la ciudad que ya se haya recopilado o que se vaya a recopilar bajo su dirección. Una evaluación constante y visible del impacto de esos proyectos ayudará a reforzar el apoyo a las actividades y a sumar otros empleados a la causa. En las experiencias de análisis predictivo que prosperan se suele comprobar valor, lo que permite que sus partidarios puedan observar la utilidad de los sistemas en la práctica.

40

Conclusión: La evaluación de los proyectos de análisis de datos debe incorporarse en los proyectos desde el inicio. Un DD exitoso buscará victorias tempranas.

¿Cuán útiles son los colegas del DD? De momento, muchos DD municipales y estatales no saben qué hacen sus colegas de otras ciudades. Esto significa que no se están transfiriendo las técnicas maestras y que los pioneros de hoy pueden estar sintiéndose indebidamente aislados y sitiados. Al igual que el éxito de la función del análisis predictivo se basa en la colaboración entre muchas instancias de la administración pública, las funciones de un DD se enriquecerán al colaborar con otras jurisdicciones; eventualmente, las iniciativas analíticas entre jurisdicciones se convertirán en una rutina. Durante la creación de una estrategia de datos, la contratación de empleados, la búsqueda de partidarios del proyecto, el establecimiento de prioridades y la evaluación de los éxitos personales, el DD debe acercarse a sus compañeros para aprender qué cosas funcionan (y cuáles no) en otros lugares.

Conclusión: Actualmente, un DD encargado de llevar a cabo proyectos de análisis predictivo no tiene que trabajar solo. Debería aprovechar las redes de otros DD.

- Barber, R. y M. Sharkey. 2012. "Course Correction: Using Analytics to Predict Course Success." En: S. B. Shum, D. Gasevic y R. Ferguson (Eds.), *In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*. Nueva York: ACM.
- CBS Los Ángeles. 2012. "LAPD the Largest Agency to Embrace 'Predictive Policing'." Los Ángeles: The Associated Press. Extraído de: <http://losangeles.cbslocal.com/2012/07/01/lapd-the-largest-agency-to-embrace-predictive-policing/>.
- CGI Group. 2013. "Predictive Analytics: The Rise and Value of Predictive Analytics in Enterprise Decision Making." Extraído de: <http://www.cgi.com/sites/default/files/white-papers/Predictive-analytics-white-paper.pdf>.
- Desigan, S. 2011. "Predictive Analytics: A Case for Nonprofits and Government Contracting Agencies." AnalyticBridge. Extraído de <http://www.analyticbridge.com/profiles/blogs/predictive-analytics-a-case-for-nonprofits-and-government>.
- Fitz-Enz, J. 2009. "Predicting People: From Metrics to Analytics." *Employer Relations Today* 36, 1–11.
- Gerber, M. S. 2014. "Predicting Crime Using Twitter and Kernel Density Estimation." *Decision Support Systems* 61. Extraído de: <http://www.sciencedirect.com/science/article/pii/S0167923614000268>.
- Goldsmith, S. 2014a. "Digital Transformation: Wiring the Responsive City." Civic Report 87. New York: Manhattan Institute for Policy Research. Extraído de: http://www.manhattan-institute.org/html/cr_87.htm.
- . 2014b. "Big Data Gives a Boost to Health and Human Services." Data-Smart City Solutions. Extraído de: <http://datasmart.ash.harvard.edu/news/article/big-data-gives-a-boost-to-health-and-human-services-380>.
- Greenemeier, L. 2013. "Your Smartphone just Diagnosed you with Postpartum Depression." *Scientific American*. Extraído de: <http://blogs.scientificamerican.com/observations/2013/05/03/your-smartphone-just-diagnosed-you-with-postpartum-depression/>.
- Health Affairs. 2013. "Medicare Hospital Readmissions Reduction Program." Extraído de: http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=102.
- Heaton, B. 2012. "Los Angeles County Uses Analytics to Stop Child-care Fraud." Extraído de: <http://www.govtech.com/health/Los-Angeles-County-Uses-Analytics-to-Stop-Child-Care-Fraud.html>.
- Howard, A. 2013. "On the Power and Perils of 'Preemptive Government'." O'Reilly Radar. Extraído de: <http://radar.oreilly.com/2013/02/preemptive-government-predictive-data.html>.
- Hughes, J. 2014. "Indiana Uses Data Analytics to Lower Infant Mortality, Child Fatality." Government Technology. Extraído de: <http://www.govtech.com/health/Indiana-Uses-Data-Analytics-to-Lower-Infant-Mortality-Child-Fatality.html>.
- Human Capital Media. 2014. "Predictive Hiring: Find Candidates who will Succeed in your Organization." Extraído de: <http://www.slideshare.net/humancapitalmedia/predictive-hiring>.
- IBM. 2010. "University of Ontario Institute of Technology: Leveraging Key Data to Provide Proactive Patient Care." Extraído de: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SNDE_OD_OD_USEN&htmlfid=ODC03157U-SEN&attachment=ODC03157U-SEN.PDF.
- . 2011a. "Simple Solutions for Local Government in an Era of Austerity." Extraído de: <http://public.dhe.ibm.com/common/ssi/ecm/en/yts03031gben/YTS03031GBEN.PDF>.
- . 2011b. "City of Lancaster Takes a Predictive Approach to Policing." Extraído de: <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=AB&htmlfid=YTC03339U-SEN>.

- . 2011c. "New York State Tax: How Predictive Modeling Improves Tax Revenues and Citizen Equity." Extraído de: http://www.ibm.com/smarterplanet/us/en/leadership/nystax/assets/pdf/0623-NYS-Tax_Paper.pdf.
- Kautz, H. 2013. "There's a Fly in my Tweets." *The New York Times*. Extraído de: <http://www.nytimes.com/2013/06/23/opinion/sunday/theres-a-fly-in-my-tweets.html>.
- Konkel, F. 2013. "Predictive Analytics Allows Feds to Track Outbreaks in Real Time." *Federal Computer Weekly*. Extraído de: <http://fcw.com/articles/2013/01/25/flu-social-media.aspx>.
- LaVall, S., E. Lesser, R. Shockley, M. S. Hopkins y N. Kruschwitz. 2011. "Big Data, Analytics and the Path from Insights to Value." *MIT Sloan Management Review* 52 (2).
- Luan, J. 2004. "Data Mining Applications in Higher Education." SPSS. Extraído de: http://www.spss.ch/upload/1122641492_Data%20mining%20applications%20in%20higher%20education.pdf.
- Maguire, J. y V. Dhar. 2013. "Comparative Effectiveness for Oral Anti-diabetic Treatments among Newly Diagnosed Type 2 Diabetics: Data-driven Predictive Analytics in Healthcare." *Health Systems* 2: 73–92.
- Mascarenhas, R. 2014. "Illuminating Housing Challenges with Data." *Data Smart City Solutions*. Extraído de: <http://datasmart.ash.harvard.edu/news/article/illuminating-housing-challenges-with-data-389>.
- Oficina del Comisionado de Protección de la Privacidad de Canadá. 2012. "The Age of Predictive Analytics: From Patterns to Predictions." *Privacy Research Papers*. Quebec, Canadá: Oficina del Comisionado de Protección de la Privacidad de Canadá. Extraído de: http://www.priv.gc.ca/information/research-recherche/2012/pa_201208_e.asp.
- Pearsall, B. 2010. "Predictive Policing: The Future of Law Enforcement"? *NIJ Journal* 266. Extraído de: <http://www.nij.gov/journals/266/Pages/predictive.aspx>.
- Rain Gain. 2012. "Work Package 2: Fine-scale Rainfall Data Acquisition and Prediction." Extraído de: <http://www.raingain.eu/en/fine-scale-rainfall-data-acquisition-and-prediction>.
- Shueh, J. 2014. "Predictive Analytics aboard the London Underground." *Government Technology*. Extraído de: <http://www.govtech.com/transportation/Predictive-Analytics-Aboard-the-London-Underground.html>.
- SPSS. 2010. "Making Critical Connections: Predictive Analytics in Government." Chicago, IL: SPSS Inc. Extraído de: <ftp://ftp.boulder.ibm.com/software/data/sw-library/spss/IMW14284USEN-00.pdf>.
- Stroud, M. 2014. "The Minority Report: Chicago's New Police Computer Predicts Crimes, but is it Racist"? *The Verge*. Extraído de: <http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist>.
- Taney, P. 2013. "New Food Illness Tracking App Getting Mixed Reviews." *WHEC-TV NBC*. Extraído de: http://www.clipsyndicate.com/video/play/4225546/new_food_illness_tracking_app_getting_mixed_reviews?wpid=5435.
- The Kings Fund. 2014. "Predicting and Reducing Re-admission to Hospital." Extraído de: <http://www.kingsfund.org.uk/projects/predicting-and-reducing-re-admission-hospital>.
- UICDS (Unified Incident Command and Decision Support). Sin fecha. "Unified Incident Command and Decision Support (UICDS): A Department of Homeland Security Initiative in Information Sharing." McLean, VA: UICDS.
- Van Barneveld, A., K. E. Arnold y J. P. Campbell. 2012. "Analytics in Higher Education: Establishing a Common Language." *Educause Learning Initiative. ELI Paper 1*. Extraído de: <http://net.educause.edu/ir/library/pdf/ELI3026.pdf>.
- Ward, D. L., R. Tripp y B. Maki. 2012. *Positioned: Strategic Workforce Planning that Gets the Right Person in the Right Job*. Saranac Lake, NY: Amacom.
- Web Builders. 2011. "Predictive Analytics for Federal Government." Extraído de: http://www.informationbuilders.com/pdf/factsheets/FS_Solution_Rstat_GovFed_2011.pdf



BID

Banco Interamericano
de Desarrollo

2016