

Transcripción del vídeo

Aplicaciones de la ciencia de datos

(Rafael B. Da Silva): Creo que una de las nuevas aplicaciones de la ciencia de datos es en el campo de la medicina, por ejemplo, en el suministro de medicamentos o el tratamiento del cáncer.

(Diana Zárate-Díaz): Creo que una muy interesante es cómo las empresas pueden utilizar la información que están recopilando de sus clientes para desarrollar nuevos productos que respondan a sus necesidades.

(Luis O. Martins): Una nueva aplicación de la ciencia de datos fue la gran cantidad de noticias que generó Pokémon Go, donde utilizaron Ingress, los datos de la aplicación Ingress, la última aplicación de la misma empresa, para elegir las ubicaciones de los Pokémons y los gimnasios Pokémons. Al utilizar los datos de la última aplicación, aprendían de sus errores.

(Mandeep Kaur): El buscador de Google es una aplicación de ciencia de datos. Utilizamos el buscador de Google siempre que queremos buscar algo. Creo que Google es lo que es gracias a la ciencia de datos.

(Stephen Sherman): La realidad aumentada es mi nueva implementación favorita de la ciencia de datos. Creo que no podemos ver ninguna nueva tecnología y no encontrar la ciencia de datos en ella. Pero la realidad aumentada es la que más me impresiona, la capacidad de andar y ver cosas en las paredes o alrededor nuestro que no están realmente ahí. Los Pokémons fueron el principio.

(Murtaza Haider): Lo que ha pasado es que ahora que están disponibles las herramientas y los conjuntos de datos, las personas los aplican sin demasiada diligencia. Creo que uno de los casos más extraños que leí en la prensa es la historia de un padre que fue a los almacenes Target en EE.UU. para quejarse de que le estaban enviando correos a su hija adolescente sobre pañales y leche de fórmula para bebés. Estaba enfadado y dijo "por qué quieren que mi hija adolescente tenga un hijo". Estaba molesto por ese correo o la campaña de publicidad, y ellos obviamente se disculparon. Pero dos semanas más tarde el padre tuvo que volver y pedir perdón, explicando que no sabía que su hija estaba embarazada. La pregunta es cómo se enteró Target antes que el padre. Lo que ocurrió es que ellos observan el comportamiento de compras de las personas, por lo que si se compra algún tipo de suplemento o vitaminas, asumen que está en el primer trimestre del embarazo. De esta forma, saben qué productos enviarles, al suponer que la persona que compró esos suplementos estaba embarazada.

Es una buena historia sobre la ciencia de datos y cómo puede pronosticar y predecir comportamientos de clientes, incluso antes de que lo sepa la familia. Lo que me parece preocupante y extraño, por muchos motivos, en primer lugar es que para cada predicción correcta, tenemos cientos de predicciones incorrectas. Es lo que llamamos los falsos positivos, pero ningún experto en datos los publicita, solo promocionamos las buenas predicciones, pero los cientos de veces que fallamos, no se cuentan. En segundo lugar, creo que es un abuso de los datos, básicamente no dan mucha información, solo se ha encontrado una correlación, pero alguien podría estar comprando el mismo material para otra persona, por lo que la probabilidad de equivocarse y obtener un falso positivo es mucho más alta.

Me parece extraño y creo que da una impresión falsa de nuestra capacidad de predecir el futuro. La realidad sobre la ciencia de datos y lo más importante que deben saber los compañeros expertos en datos es que todas las predicciones son incorrectas. Son útiles, pero son incorrectas. No deberíamos estar seguros de que ahora podemos hacer análisis predictivos, de que podemos resolver todos los problemas. Creo que un buen ejemplo es el buscador de Google. Google publicó un documento

afirmando que podían predecir las epidemias de gripe antes que el Centro de control de enfermedades. Lo que hicieron es consultar qué personas hacían búsquedas de síntomas de la gripe en Google. Google vio las búsquedas de síntomas de gripe antes que nadie y pudieron predecirlas. La verdad es que estas búsquedas son buenas y se correlacionan con algunos resultados, pero no necesariamente todas las veces. Cuando Google lo anunció, tuvo mucha repercusión, todo el mundo dijo: es una nueva era para el análisis predictivo... sin embargo, varios años después, se dieron cuenta de que Google empezó a predecir falsos positivos, a predecir cosas que no ocurrían o a realizar predicciones que no eran tan precisas por distintos motivos. Probablemente cambiaron sus algoritmos y los conjuntos de datos ya no se correlacionaban con los resultados.

La lección que debemos aprender aquí es que hay que evitar lo que se denomina la arrogancia de los datos. No podemos creer demasiado en los modelos porque nos pueden engañar. La ciencia de datos tiene un gran potencial de generar cambios en partes del mundo... en partes de nuestra sociedad que han estado marginadas durante años. Podemos ver grandes ejemplos de aplicación de la ciencia de datos en países en desarrollo, donde se proporciona ayuda humanitaria, comida y otras ayudas a las personas en lugares donde no llegaban antes. La razón por la que ocurre ahora es gracias a la mayor disponibilidad de los datos, los modelos y la analítica para identificar dónde hay una mayor necesidad. La capacidad de diseñar y realizar experimentos permite ver si se deben proporcionar pequeños préstamos de microcréditos a hogares muy pobres en países del mundo en vías de desarrollo, y comprobar cómo permiten salir de la pobreza a hogares individuales, así como mejorar colectivamente el bienestar económico de la comunidad local con inyecciones de efectivo o crédito muy pequeñas.

Estos experimentos se realizan en todo el mundo y son el resultado directo de nuestra capacidad de analizar datos y diseñar experimentos, y de dedicar enormes esfuerzos para ofrecer ayuda humanitaria con créditos que dan oportunidades a aquellos que han estado marginados en el pasado, para que se puedan unir al resto del mundo en niveles de prosperidad, salud y felicidad.