

Transcripción del vídeo

Un día en la vida de un experto en datos

(Stephen Sherman): He creado un motor de recomendación previamente como parte de una gran organización y he trabajado con todo tipo de ingenieros para contemplar las distintas partes del problema. Es uno de los trabajos de los que estoy más satisfecho porque, al final, conseguí una solución muy simple y fácil de entender para todos los niveles, desde los ejecutivos a los ingenieros y desarrolladores. En última instancia, fue tan eficaz como una solución realmente compleja, y podían haber dedicado mucho más tiempo.

(Rafael B. Da Silva): Cuando estaba en la universidad, tuvimos un problema cuando intentamos predecir la floración de algas. La floración de algas podía provocar un aumento de la toxicidad del agua, lo que era un problema para la empresa de tratamiento del agua. No podíamos realizar la predicción con nuestros conocimientos de ingeniería química. Por lo tanto, utilizamos redes neuronales artificiales para predecir cuándo se producirían estas floraciones. De esta forma, las empresas de tratamiento del agua pudieron manejar este problema mejor.

(Murtaza Haider): En Toronto, el transporte público está dirigido por la Toronto Transit Commission. También conocida como la TTC. Es uno de los mayores organismos de transporte de la región en Norteamérica. Un día se pusieron en contacto conmigo y me dijeron "Tenemos un problema". "De acuerdo, ¿cuál es el problema?" "Tenemos datos de reclamaciones y nos gustaría analizarlos, para lo que necesitamos su ayuda". "Muy bien, encantado de poder ayudarles". Les pregunté: "¿Cuántas reclamaciones tienen?". "Unas cuantas". "¿Cuántas?". Alrededor de medio millón. "Perfecto, vamos a empezar a trabajar con eso".

Así que empecé a analizar los datos. Básicamente, habían hecho un gran trabajo manteniendo algunos datos en formato tabular y había datos no estructurados. En este caso, los datos tabulares incluían cuándo llegó la reclamación, quién la recibió, qué tipo de reclamación era, si se resolvió y quién era responsable. La parte no estructurada era el intercambio de correos electrónicos y faxes. Imaginemos cómo intentar obtener respuestas de medio millón de intercambios de correos electrónicos.

Lo primero que quise saber es por qué se quejaba la gente y si había algún patrón, por ejemplo, días en los que había más reclamaciones que otros. Consulté los datos y los analicé con distintos formatos, y no podía encontrar el motivo de que se realizan más reclamaciones un determinado día y menos otros. Así continuó durante un mes aproximadamente. Un día, me estaba bajando del autobús en Toronto pensando todavía en esto. Salí sin mirar al suelo y pisé un charco, un charco de agua. Me vi con el pie metido en agua hasta el tobillo, un pie mojado y el otro seco. Me enojó muchísimo. Empecé a andar y entonces caí en la cuenta: "Espera un segundo. Hoy ha llovido inesperadamente y no estaba preparado. Por eso me he mojado, no estaba preparado". ¿Era posible que existiera una relación entre las condiciones meteorológicas extremas y el tipo de reclamaciones que recibía la TTC?

Fui al sitio web de Environment Canada y busqué datos sobre lluvias y precipitaciones, vientos y luz. Ahí encontré algo muy interesante. Los 10 días con más reclamaciones. Los 10 en los que la gente se quejó más fueron los días con peor tiempo. Fue una lluvia inesperada, una caída extrema de las temperaturas, demasiada nieve y un día muy ventoso. Volví a hablar con los ejecutivos de la TTC y les dije "Tengo buenas noticias y malas noticias". Las buenas noticias es que sé por qué se quejaba la gente excesivamente determinados días. Sé cuál es la razón. Las malas noticias es que no podéis hacer nada al respecto.