

# “Científico de datos”, la profesión del presente

José Antonio Álvarez Jareño, Vicente Coll-Serrano

*Métodos de información*, ISSN 2173-1241, Vol. 9, n. 16, 2018, p. 113-129

Todos los clientes no son igual de rentables para una empresa. Si ordenásemos los clientes en función del beneficio que generan para la empresa, el 20% de los más rentables aportan el 80% del beneficio. Al mismo tiempo, el quintil menos rentable no solo no aporta ningún beneficio, sino que incluso destruye valor. Identificar a los clientes que destruyen el valor e invitarles a que se busquen otro proveedor es el trabajo de muchos gerentes. Este problema puede ser resuelto por unos profesionales que han llegado para revolucionar la toma de decisiones en la empresa y en otros muchos campos, los “científicos de datos”. Los especialistas en Big Data analizan los grandes datos que posee una empresa y los usan como indicadores fiables para proponer medidas correctoras y para ayudar en la toma de decisiones sobre el rumbo que tiene que seguir un determinado negocio. Los especialistas en Big Data abarcan un amplio espectro de roles que van desde los centros de investigación, las grandes empresas, hasta las administraciones públicas. Su principal virtud será entender qué son y qué representan esos grandes conjuntos de datos. Para esta labor no se dispone de ningún procedimiento o técnica, y los profesionales se enfrentan a los datos con su intuición y su experiencia. Sin embargo, sí se dispone de una importante caja de herramientas para analizar los datos y buscar correlaciones o patrones de comportamiento, que llevará a hacer predicciones. En una primera fase, se pretende obtener un modelo que describa los datos lo mejor posible sin que haya sobreajuste. Después se pasa a una segunda fase donde se predice el comportamiento del subconjunto de comprobación, y por último, es conocer la tecnología disponible y saber cómo utilizarla. Se han creado diferentes ecosistemas de programación que permiten manejar grandes bases de datos y realizar cálculos a gran velocidad, como el Hadoop. La mayoría de la metodología estadística que se utiliza en ciencia de datos se descubrió hace 50 años o más, pero muchas de ellas solo se han podido poner en práctica cuando se ha dispuesto de la potencia de cálculo adecuada. No es exacto afirmar que el “científico de datos” es una nueva profesión, más bien es la democratización de la profesión de científico o investigador. En definitiva, es un nuevo nombre para un trabajo muy antiguo.

Resumen realizado por José María Amate Sánchez