

Prólogo

En estos últimos años hemos sido testigos del rápido crecimiento en la capacidad de las empresas y las administraciones públicas para explotar los avances recientes en las tecnologías de la información y de las comunicaciones (TIC), de la investigación operativa y de la modelización estadística, para recopilar y procesar datos sociales y de mercado, de sensores y de operaciones, para apoyar sus procesos de toma de decisiones. La captura de datos a través de aplicaciones *online* y móviles produce cantidades ingentes de información potencial que debemos aprehender para entender cómo actuamos, nos sentimos, nos movemos e interactuamos y cómo respondemos frente a las políticas de los gobiernos y las decisiones de las empresas. Los datos alcanzan así cada vez mayor valor para las organizaciones que intentan vislumbrar cómo aprovecharlos para mejorar las relaciones con los ciudadanos o los clientes, personalizar servicios y productos y automatizar todo tipo de procesos. Hasta hace poco, la toma de decisiones y el diseño de políticas en muchos ámbitos se hacía según criterios basados en información cualitativa, cuando no directamente según impresiones subjetivas. La irrupción del *big data* está implicando que, cada vez más, la información proporcionada por los datos constituya la base del análisis y las decisiones, posibilitando

procesos más objetivos y automáticos, con las ventajas y riesgos que ello supone.

En cierta forma, existe algo de exageración sobre el fenómeno que hoy se conoce como big data, motivado en parte por éxitos indudables en aplicaciones industriales como los coches autónomos, la traducción automática o el reconocimiento del habla. Esta exageración se refleja en titulares como “El diluvio de datos vuelve obsoleto el método científico” (Wired, 2008) o “el big data salvará la política” (Technology Review, 2013) y ha llevado a cierta confusión en empresas y administraciones al querer realizar proyectos big data *per se*, cuando realmente no los necesitan, y en confundir big data con una mera tecnología, típicamente asociada con Hadoop o Spark.

En este libro introducimos los conceptos, tecnologías y metodologías básicas del big data y describimos algunas aplicaciones actuales y potenciales para contribuir a un mejor desarrollo de la sociedad. Intentamos desmontar algunas ideas equivocadas y promover una concepción más global de este fenómeno orientada a la obtención de valor de los datos a través de un uso responsable de la tecnología y la ciencia dirigido a apoyar la toma de decisiones basada en evidencia.

Tras una introducción al fenómeno del big data, identificando los pilares estadístico-matemático, informático-tecnológico y aplicado que deben sostener este tipo de proyectos, hacemos una breve incursión en sus tecnologías principales, con la obvia advertencia de la rápida evolución de las mismas. Revisamos, después, el otro pilar fundamental referido a los métodos estadísticos y de aprendizaje automático dentro del paradigma de la ciencia de datos. Los siguientes tres capítulos se refieren a aplicaciones en campos de interés principal como son su uso para la promoción de políticas públicas, en lo que denominaríamos “analítica de políticas”, y en la comunicación política, en particular en referencia al fenómeno de las *fake news*; su empleo en la promoción de la salud, y, en tercer lugar, en su adopción para el desarrollo de sociedades más ciberseguras. Finalmente, hacemos una

referencia a los aspectos éticos y de responsabilidad social del tratamiento de los big data, para concluir con una mirada al futuro. Cada capítulo se completa con una breve reseña bibliográfica con la que el lector podrá ampliar los contenidos.

Comienza la singladura.

DAVID RÍOS INSUA Y DAVID GÓMEZ-ULLATE OTEIZA

Pantín-El Puerto de Santa María