

ALGORITMOS: ¿UNA NUEVA HERRAMIENTA CONTRA LA DISCRIMINACIÓN ILEGAL?*

CASS R. SUNSTEIN

Profesor Universitario Robert Walmsley, Harvard University

¿Son parciales los algoritmos? ¿En qué sentido? Estas son grandes preguntas sin una respuesta fácil. A mi modo de ver, la mayoría de argumentos que se oponen al uso de algoritmos se basan generalmente en asunciones mayormente erróneas, si bien es cierto que el tema no es sencillo y requiere de una elaboración más amplia y con matices.

Al promover un debate al respecto, mi objetivo es ofrecer un punto de vista de los algoritmos que se incardina con algunos de los estudios más importantes sobre el uso de algoritmos en el ámbito del derecho y las políticas públicas. Existe un abundante corpus de trabajos teóricos que exploran la idea de que los algoritmos pueden ser tendenciosos, en una u otra dirección. Sin embargo, en el ámbito del derecho y la política, algunas de las investigaciones empíricas más recientes y significativas parecen sugerir una conclusión contraria. Por ejemplo, en el campo jurídico y de la concesión de la libertad bajo fianza, los resultados apuntan a que un algoritmo diseñado para predecir el riesgo de fuga de un preso resulta estadísticamente mucho más certero que un juez humano, debido en parte, a que este último pone un énfasis excesivo en la naturaleza particular del delito cometido.

Esto me lleva a realizar dos afirmaciones. La primera, y la más sencilla, es que los algoritmos pueden contrarrestar los efectos perniciosos de los prejuicios cognitivos que en algunos casos pueden tener una fuerte influencia en aquellas personas cuyo trabajo es, precisamente, mostrarse imparciales. Muchas cuestiones sociales presentan problemas de predicción, y allí donde los sesgos cognitivos pueden hacer descarrilar a las personas, los algoritmos pueden ser de gran ayuda.

La segunda afirmación sobre la que incidiré es que los algoritmos pueden diseñarse con el objetivo expreso de evitar la discriminación racial (o de otro tipo) en sus formas ilegales, y para dirimir cuestiones difíciles sobre cómo conjugar posibles valores sociales en conflicto. Muchos de los temores acerca de la tendenciosidad

algorítmica se apoyan a menudo en la discriminación por razones de raza o sexo. Quisiera puntualizar aquí que la palabra “discriminación” puede interpretarse de maneras distintas. Y debería ser sencillo garantizar que los algoritmos no discriminan en el sentido manifiesto que tiene dicha noción, por lo menos, en la ley estadounidense. En la actualidad, no resulta sencillo gestionar aquellas formas de desigualdad más extendidas, incluida la discriminación racial. Como veremos, los algoritmos pueden contribuir a una transparencia mayor en determinados procesos compensatorios complejos.

El uso de algoritmos viene a menudo motivado por la constatación de las limitaciones de la intuición humana. Tanto en el sector público como en el sector privado, a menudo se le pide a la gente que haga predicciones en condiciones de incertidumbre, y sus intuiciones pueden llevarles fácilmente a cometer errores. Se requiere mucho trabajo para alcanzar las prevenciones necesarias.

Naturalmente, el algoritmo puede conllevar problemas graves si en su diseño utiliza un factor que es en cierto modo consecuencia de una discriminación previa. Por ejemplo, una pobre valoración de crédito o unos antecedentes penales problemáticos pueden ser un artefacto discriminatorio para un ser humano, antes incluso de pedirle al algoritmo que haga su trabajo predictivo. En estos casos existe el riesgo de que los algoritmos contribuyan a perpetuar la discriminación y extender su alcance al utilizar factores que, siendo genuinamente predictivos, son productos de un tratamiento desigual. Esto podría convertir a la discriminación en una especie de profecía autocumplida.

Nada puede garantizar que los algoritmos eviten los sesgos cognitivos, pero se pueden diseñar para que los pongan de manifiesto

Nada puede garantizar, por supuesto, que los algoritmos eviten los sesgos cognitivos. Pero sí es posible diseñarlos para que los pongan de manifiesto. Y debe considerarse también que pueden contribuir a mejorar las decisiones humanas, precisamente, a causa de su imparcialidad. Esto es simplemente una nueva instancia del viejo principio según el cual la predicción estadística resulta a veces más certera que la predicción clínica.

Una de las virtudes más llamativas de los algoritmos es que hacen visibles los equilibrios compensatorios que operan en el resultado de un proceso, de una forma meridianamente clara y con una transparencia sin precedentes: obligan a la gente a pronunciarse previamente sobre los equilibrios compensatorios entre diversos objetivos que pueden ser cruciales y, quizá, también al mismo tiempo incompatibles.

* Esta pieza es un extracto de un artículo más extenso publicado por el mismo autor y que lleva por título “Algorithms, Correcting Biase”, publicado en *Forthcoming Social Research* en diciembre del 2018.

