

CATHY O'NEIL



# ARMAS DE DESTRUCCIÓN MATEMÁTICA

CÓMO EL **BIG DATA** AUMENTA LA  
DESIGUALDAD Y AMENAZA LA DEMOCRACIA

Capitán Swing

# ARMAS DE DESTRUCCIÓN MATEMÁTICA

CÓMO EL **BIG DATA** AUMENTA LA  
DESIGUALDAD Y AMENAZA LA DEMOCRACIA

CATHY O'NEIL

Traducción

**Violeta Arranz de la Torre**

*Capitán Swing* 



## **Agradecimientos**

Quiero dar las gracias a mi marido y a mis hijos por su increíble apoyo. Gracias también a John Johnson, Steve Waldman, Maki Inada, Becky Jaffe, Aaron Abrams, Julie Steele, Karen Burnes, Matt LaMantia, Martha Poon, Lisa Radcliffe, Luis Daniel y Melissa Bilski. Y también gracias a las personas sin las que este libro no existiría: Laura Strausfeld, Amanda Cook, Emma Berry, Jordan Ellenberg, Stephen Baker, Jay Mandel, Sam Kanson-Benanav y Ernie Davis.

*Este libro está  
dedicado a todos los  
desamparados*

## Introducción

Cuando era pequeña, solía quedarme mirando el tráfico a través de la ventanilla del coche estudiando los números de las matrículas de los vehículos que pasaban. Me gustaba reducir cada matrícula a sus elementos básicos —los números primos que la componen—.  $45 = 3 \times 3 \times 5$ . Esto se denomina factorización y era mi pasatiempo de investigación favorito. Yo era una empollona de matemáticas en ciernes y sentía una fascinación particular por los números primos.

Mi amor por las matemáticas acabó convirtiéndose en una auténtica pasión. Fui a un campamento de matemáticas cuando tenía catorce años y volví a casa agarrando con fuerza un cubo de Rubik contra mi pecho. Las matemáticas me ofrecían un refugio ordenado frente al desorden del mundo real. Avanzaban a grandes pasos y su ámbito de conocimiento se ampliaba inexorablemente, prueba tras prueba. Y yo podía contribuir a su desarrollo. Me especialicé en matemáticas en la facultad y decidí estudiar un doctorado. Mi tesis versaba sobre la teoría de los números algebraicos, un campo cuyas raíces estaban precisamente en esa factorización que me gustaba hacer cuando era niña. Con el tiempo, conseguí un trabajo de

profesora con opción a un contrato fijo en la universidad de Barnard College, que compartía con la Universidad de Columbia el departamento de Matemáticas.

Al cabo de un tiempo, decidí cambiar radicalmente mi carrera profesional. Dejé mi puesto y empecé a trabajar como analista cuantitativa (*quant* en inglés) para D. E. Shaw, un destacado fondo de cobertura. Al dejar el mundo académico por las finanzas, llevé las matemáticas de la teoría abstracta a la práctica. Las operaciones que hacíamos con números se traducían en billones de dólares que pasaban de una cuenta a otra. Al principio estaba entusiasmada y asombrada con la idea de trabajar en este nuevo laboratorio, la economía mundial, pero en otoño de 2008, cuando llevaba poco más de un año en ese mundo, todo se desplomó.

La crisis financiera dejó bien claro que las matemáticas, que una vez habían sido mi refugio, no solo estaban profundamente involucradas en los problemas del mundo, sino que además agravaban muchos de ellos. La crisis inmobiliaria, la ruina de grandes entidades financieras, el aumento del desempleo: todo esto había sido impulsado e inducido por matemáticos que blandían fórmulas mágicas. Además, gracias a los extraordinarios poderes que tanto amaba, las matemáticas podían combinarse con la tecnología para multiplicar el caos y la desgracia, lo que añadía eficacia y magnitud a unos sistemas que entonces comprendí que eran defectuosos.

Si hubiéramos estado lúcidos, habríamos dado un paso atrás en este punto para analizar cómo habíamos hecho un mal uso de las matemáticas y cómo podíamos evitar una

catástrofe similar en el futuro. Sin embargo, en lugar de eso, justo después de la crisis, las nuevas técnicas matemáticas estaban más de moda que nunca y se extendían a un creciente número de áreas. Funcionaban veinticuatro horas al día procesando petabytes de información, en gran parte datos extraídos de las redes sociales o de páginas web de comercio electrónico. Y en lugar de prestar cada vez más atención a los movimientos de los mercados financieros mundiales, se dedicaban cada vez más a analizar a los seres humanos, a nosotros. Los matemáticos y los especialistas en estadísticas estudiaban nuestros deseos, nuestros movimientos y nuestro poder adquisitivo. Predecían nuestra solvencia y calculaban nuestro potencial como estudiantes, trabajadores, amantes o delincuentes.

Esta era la economía del *big data*, y prometía ganancias espectaculares. Un programa de ordenador era capaz de procesar miles de currículos o solicitudes de préstamos en un par de segundos y clasificarlos en listas bien ordenadas, con los candidatos más prometedores situados en los primeros puestos. Estos programas no solo permitían ahorrar tiempo, sino que además se anunciaban como procesos más justos y objetivos. Al fin y al cabo, eran procesos en los que no había seres humanos, con sus prejuicios, escarbando en montones de papel, sino simplemente máquinas procesando números de manera objetiva. En el año 2010 aproximadamente las matemáticas se habían impuesto como nunca antes en los asuntos humanos, y el público en general recibió el cambio con los brazos abiertos.

Y, sin embargo, yo veía problemas en el horizonte. Estas aplicaciones fundamentadas en las matemáticas que alimentaban la economía de los datos se basaban en decisiones tomadas por seres humanos que no eran infalibles. Seguro que algunas de esas decisiones se tomaban con la mejor de las intenciones, pero muchos de estos modelos programaban los prejuicios, las equivocaciones y los sesgos humanos en unos sistemas informáticos que dirigían cada vez más nuestras vidas. Cuales dioses, estos modelos matemáticos eran opacos y sus mecanismos resultaban invisibles para todos, salvo para los sumos sacerdotes del sector: los matemáticos y los ingenieros informáticos. Sus veredictos, incluso cuando estaban equivocados o eran perjudiciales, eran indiscutibles e inapelables y solían castigar a los pobres y los oprimidos de nuestra sociedad, al tiempo que enriquecían a los ricos.

Se me ocurrió un nombre para este tipo de modelos perniciosos: *armas de destrucción matemática* o ADM. Veamos un ejemplo, en el que iré señalando sus características destructivas poco a poco.

Como ocurre a menudo, este caso empezó con un objetivo loable. En 2007, el nuevo alcalde de Washington D. C., Adrian Fenty, estaba decidido a corregir la situación de las escuelas deficientes de la ciudad. Tenía un gran desafío por delante: en ese momento, apenas uno de cada dos alumnos de instituto llegaba a la graduación después del noveno curso<sup>[1]</sup> (el equivalente a tercero de secundaria) y solo un 8 % de los alumnos de octavo curso<sup>[2]</sup> (el equivalente a segundo de secundaria) tenía un nivel de matemáticas



acorde a su curso. Fenty contrató a la experta en reformas educativas Michelle Rhee para que ocupara un importante puesto de nueva creación: sería nombrada rectora de los centros educativos de primaria y secundaria de Washington.

La teoría generalmente aceptada era que los alumnos no aprendían lo suficiente porque sus profesores no trabajaban bien. De modo que, en 2009, Michelle Rhee puso en marcha un plan para extirpar del sistema a los docentes de bajo rendimiento. Esta era la tendencia generalizada en los distritos escolares con problemas en todo el país y, desde el punto de vista de la ingeniería de sistemas, este razonamiento tiene mucho sentido. Hay que evaluar a los profesores, deshacerse de los peores y colocar a los mejores donde puedan producir el mayor efecto positivo posible. En el lenguaje de los científicos de datos, de este modo «se optimiza» el sistema escolar y presuntamente se garantizan mejores resultados para los alumnos. Exceptuando a los «malos» profesores, ¿quién podría no estar de acuerdo con este razonamiento? Rhee desarrolló una herramienta de evaluación del personal docente[3] a la que llamó IMPACT y, a finales del curso académico 2009-2010, el distrito escolar despidió a todos los docentes cuyas puntuaciones los situaban en el 2 % inferior.[4] A finales del siguiente curso, echaron a otro 5 %, es decir a 206 maestros y profesores.[5]

No parecía que Sarah Wysocki, una maestra de quinto curso[6] (el equivalente a quinto de primaria), tuviera nada por lo que preocuparse. Llevaba trabajando solo dos años en el colegio MacFarland, pero el director del colegio y los

padres de sus alumnos tenían ya una excelente opinión de ella. En una evaluación la elogiaban por lo atenta que era con los niños;[7] en otra se decía que era «una de las mejores maestras con las que he tratado nunca».

No obstante, a finales del curso de 2010-2011, Wysocki sacó una penosa puntuación en su evaluación de IMPACT.[8] El problema fue un nuevo sistema de puntuación llamado modelación de valor añadido, que pretendía medir su eficacia en la enseñanza de competencias lingüísticas y matemáticas. Esa puntuación, generada por un algoritmo, representaba la mitad de su valoración global,[9] y tenía más peso que las valoraciones positivas de los cargos directivos del colegio y de la comunidad. De modo que el distrito escolar no tuvo más remedio que despedirla, junto con otros 205 docentes que habían tenido puntuaciones de IMPACT por debajo del umbral mínimo.

No parecía tratarse de una caza de brujas ni de un ajuste de cuentas. De hecho, el enfoque del distrito escolar tenía lógica. Al fin y al cabo, los cargos directivos de los centros educativos podían ser amigos de personas que hicieran muy mal su trabajo, o quizá admiraran su estilo o su aparente dedicación. Los malos docentes pueden *parecer* buenos. Por lo tanto, el distrito de Washington, al igual que otros muchos sistemas escolares, decidió que debía minimizar este sesgo humano y prestar más atención a las puntuaciones basadas en resultados irrefutables: las puntuaciones de rendimiento de los alumnos en matemáticas y lectura. Los números hablarían alto y claro, prometieron los funcionarios del distrito escolar. Serían más justos.

Obviamente, Sarah Wysocki pensó que los números eran terriblemente injustos y quiso saber de dónde venían. «No creo que nadie los entendiera», me dijo después. ¿Cómo podía una buena maestra sacar tan malas puntuaciones? ¿Qué medía exactamente el modelo de valor añadido?

Según descubrió Sarah Wysocki, la respuesta era muy complicada. El distrito había contratado a una consultora con sede en Princeton, Mathematica Policy Research, para que creara el sistema de evaluación.[10] Mathematica se enfrentó al reto de medir el progreso educativo de los alumnos del distrito y a continuación calcular qué parte de ese progreso o retroceso podía atribuirse a sus maestros y profesores. Evidentemente, no fue tarea fácil. Los investigadores sabían que muchas variables diferentes, desde el contexto socioeconómico de los alumnos hasta las dificultades de aprendizaje, podían afectar a los resultados de los alumnos. Los algoritmos tenían que tener en cuenta dichas diferencias, lo que explicaba en parte por qué eran tan complejos.

Efectivamente, intentar reducir el comportamiento, el rendimiento y el potencial humanos a algoritmos no es tarea fácil. Para comprender lo que intentaba evaluar Mathematica, imaginemos a una niña de diez años de un barrio pobre del sureste de Washington D. C. Al final de un curso escolar, hace la prueba homologada de quinto curso. Y luego la vida sigue. Puede que haya problemas en su familia o que tengan dificultades económicas. Quizá se muden a una nueva casa o a la niña le preocupen los problemas que tiene su hermano mayor con la ley. Tal vez esté descontenta con su peso, o quizá esté asustada porque

la acosan en el colegio. Pase lo que pase, al final del curso siguiente hace otra prueba homologada, la prueba diseñada para los alumnos de sexto curso.

Al comparar los resultados de ambas pruebas, la puntuación debería mantenerse estable o, con suerte, subir. Mientras que si sus resultados caen en picado, resulta fácil calcular la diferencia entre su rendimiento y el de los alumnos con buenos resultados.

Pero ¿qué parte de esa diferencia es responsabilidad de su maestra? Es difícil saberlo, y los modelos de Mathematica solo pueden comparar unas cuantas cifras. En empresas de *big data* como Google, por el contrario, los investigadores están constantemente haciendo pruebas y controlan miles de variables. Pueden preparar dos versiones de un mismo anuncio —una con las letras en azul y otra en rojo—, presentar cada una de estas versiones a diez millones de personas y hacer un seguimiento para saber cuál de las dos recibe más clics. Utilizan esta retroalimentación para pulir sus algoritmos y ajustar su funcionamiento. Aunque considero que Google presenta muchos problemas, como veremos más tarde, es cierto que en este tipo de pruebas hace un uso efectivo de las estadísticas.

Intentar calcular el impacto que una persona puede tener sobre otra a lo largo de un curso escolar es un proceso mucho más complejo. «Hay tantos factores implicados en la enseñanza y en el aprendizaje que sería muy difícil medirlos todos», decía Sarah.[11] Por otra parte, intentar puntuar la eficacia de un docente analizando los resultados de una prueba de solo veinticinco o treinta alumnos no

tiene solidez estadística y es incluso ridículo. El número de valores es insuficiente si se tiene en cuenta todo lo que puede fallar. De hecho, si quisiéramos analizar a los docentes con el rigor estadístico de un motor de búsqueda, tendríamos que probarlos en miles o incluso millones de alumnos seleccionados al azar. Los estadísticos utilizan grandes cifras para compensar las excepciones y las anomalías (y las ADM, como veremos más adelante, a menudo castigan a personas concretas que resultan *ser* la excepción).

Otro aspecto igualmente importante es que los sistemas estadísticos requieren una retroalimentación, algo que les indique cuándo se están desviando. Los estadísticos utilizan los errores para enseñar a sus modelos y hacerlos más inteligentes. Si Amazon.com empleara una correlación defectuosa y empezara a recomendar libros sobre el cuidado del césped a chicas adolescentes, el número de clics caería en picado y la empresa se aseguraría de corregir el algoritmo hasta que funcionara bien. Sin embargo, si no se tiene en cuenta la retroalimentación, un motor estadístico puede seguir realizando análisis defectuosos y perjudiciales sin aprender nunca de sus errores.

Muchas de las ADM de las que hablaré en este libro, incluido el modelo de valor añadido del distrito escolar de Washington, se comportan así. Definen su propia realidad y la utilizan para justificar sus resultados. Este tipo de modelo se autoperpetúa y es altamente destructivo —y además está muy extendido—.

Cuando el sistema de puntuación de Mathematica etiquetó a Sarah Wysocki y a otros 205 docentes como malos enseñando, el distrito escolar los despidió. Pero ¿cómo puede confirmar el sistema que su análisis fue correcto? No puede. El sistema decidió él solito que estos docentes eran un desastre y los trataron como tal. Doscientos seis «malos» enseñantes menos. Este simple hecho ya parece demostrar lo efectivo que es el modelo de valor añadido: está limpiando el distrito de los maestros de bajo rendimiento. En lugar de buscar la verdad, la puntuación que produce el modelo la personifica.

Este es un ejemplo del bucle de retroalimentación de un ADM. Veremos muchos otros en este libro. Las empresas, por ejemplo, utilizan cada vez más las calificaciones de solvencia crediticia para evaluar a los posibles candidatos. Se basan en la creencia de que las personas que pagan pronto sus facturas tienen más probabilidades de llegar puntualmente a su puesto de trabajo y de cumplir las normas. Aunque en realidad hay muchas personas responsables y buenos trabajadores que tienen mala suerte y cuya calificación crediticia cae en picado. No obstante, esta idea de que una mala calificación crediticia está relacionada con un mal rendimiento en el trabajo hace que las personas que tienen una calificación más baja tengan menos probabilidades de encontrar trabajo. El desempleo los empuja a la pobreza, lo que a su vez empeora aún más sus calificaciones de solvencia, con lo que les resulta aún más difícil encontrar trabajo. Es una espiral que se retroalimenta. Y las empresas nunca descubren cuántos buenos empleados han dejado de contratar por centrarse

únicamente en las calificaciones crediticias. En las ADM, hay muchas premisas perniciosas camufladas bajo las matemáticas y se mantienen sin que nadie las verifique ni las cuestione.

Esto nos lleva a hablar de otra característica común de las ADM: suelen castigar a los pobres. Esto se debe, en parte, a que han sido diseñadas para evaluar grandes cantidades de personas. Están especializadas en trabajar con grandes volúmenes, y son baratas. Eso forma parte de su atractivo. Los ricos, en cambio, reciben a menudo un trato más personal. Un bufete de abogados de clase alta o un exclusivo instituto privado se basarán más en recomendaciones y entrevistas personales durante los procesos de selección que una cadena de comida rápida o un distrito escolar urbano con escasos fondos. Los privilegiados, como veremos una y otra vez, son analizados por personas; las masas, por máquinas.

El hecho de que Sarah Wysocki no consiguiera encontrar a nadie que le pudiera explicar su pésima puntuación también es muy revelador. Los veredictos de las ADM son como los dictados de los dioses de los algoritmos. El modelo en sí es una caja negra, su contenido, un secreto corporativo fieramente guardado. De este modo, las consultoras como Mathematica pueden cobrar más, aunque este secretismo sirve también a otros propósitos: se supone que, si las personas evaluadas no saben cómo se hace la evaluación, es menos probable que intenten engañar al sistema. Y así no les quedará otra alternativa que simplemente trabajar duro, cumplir las normas y rezar por que el modelo registre y valore sus esfuerzos. Por otra

parte, al esconder los detalles de su funcionamiento, también resulta más difícil cuestionar la puntuación resultante o protestar contra ella.

Durante años, los docentes de Washington se quejaron de las puntuaciones arbitrarias y pidieron los detalles de los aspectos que se valoraban. Es un algoritmo, les dijeron, es muy complejo. Esto disuadió a muchos de seguir presionando. Desgraciadamente, hay muchas personas que se sienten intimidadas por las matemáticas. Pero Sarah Bax, una profesora de Matemáticas, siguió presionando al administrador del distrito, Jason Kamras, un antiguo colega suyo, para que le diera los detalles.[12] Después de meses de tira y afloja, Jason Kamras le dijo que esperara un poco, que se publicaría pronto un informe técnico. Sarah Bax le contestó: «¿Cómo puedes justificar el hecho de que estéis evaluando a personas con un método que no sois capaces de explicar?». Pero esa es precisamente la naturaleza de las ADM. El análisis se subcontrata a programadores y estadísticos, y, por lo general, dejan que sean las máquinas las que hablen.

A pesar de todo, Sarah Wysocki sabía que las puntuaciones de las pruebas estandarizadas de sus alumnos tenían un gran peso en la fórmula. Y tenía ciertas sospechas sobre ese punto. Antes de empezar el que sería su último curso en el centro de enseñanza media MacFarland, se había llevado una alegría al descubrir que sus nuevos alumnos de quinto curso habían tenido unas notas sorprendentemente buenas en las pruebas que habían hecho al finalizar el curso anterior. En el centro de primaria Barnard, del que venían muchos de sus alumnos,



el 29 % de los alumnos[13] obtuvo la calificación de «nivel avanzado» en lectura. Era un porcentaje cinco veces mayor que la media del distrito escolar.

Sin embargo, cuando empezaron las clases, Sarah descubrió que a muchos de sus alumnos les costaba leer incluso frases sencillas. Mucho después, los periódicos *Washington Post* y *USA Today* publicaron que se había descubierto un alto número de tachaduras en las pruebas homologadas de 41 centros escolares del distrito, incluido el Barnard.[14] Encontrar un alto número de respuestas corregidas apunta a una mayor probabilidad de que se hayan hecho trampas en el examen. En algunos centros se sospechaba de hasta el 70 % de las clases.

¿Y qué tiene esto que ver con las ADM? Pues un par de cosas. En primer lugar, los algoritmos de evaluación de docentes son una herramienta poderosa para la modificación del comportamiento. Ese es su propósito, y en los centros educativos de Washington sirvieron al mismo tiempo de palo y de zanahoria. Los maestros sabían que si sus alumnos tropezaban en la prueba, sus propios empleos estarían amenazados. Y esto supuso una fuerte motivación para los maestros, que querían asegurarse de que sus alumnos aprobarían, especialmente en un momento en el que la Gran Recesión azotaba el mercado laboral. Al mismo tiempo, si sus alumnos obtenían mejores resultados que otros alumnos del mismo curso, los maestros y los directivos de los centros recibirían primas de hasta 8.000 dólares.[15] Si se combinan estos poderosos incentivos con las pruebas del caso —el alto número de respuestas borradas y las puntuaciones anormalmente altas en las

pruebas—, hay razones para sospechar que los maestros de cuarto curso, cediendo ante el miedo o la codicia, habían corregido los exámenes de sus alumnos.

Esto significa que existe la posibilidad de que los alumnos de quinto curso de Sarah Wysocki empezaran el curso escolar con unas puntuaciones artificialmente infladas. Si esto fue así, los resultados que obtuvieron al final del quinto curso condujeron a la falsa conclusión de que sus competencias habían retrocedido... y de que tenían una mala maestra. Sarah estaba convencida de que este había sido su caso. Esta explicación encajaría con los comentarios de los padres, los compañeros y su director, que decían que era una buena docente. Esto aclararía toda la confusión. Sarah Wysocki tenía argumentos sólidos.

Sin embargo, no es posible apelar la decisión de un ADM. Esto forma parte de su temible poder. No escuchan. Ni se doblegan. Son sordas, y no solo al encanto, las amenazas y las adulaciones, sino también a la lógica, incluso cuando hay buenas razones para cuestionar los datos que alimentan sus conclusiones. Sí, es cierto que si resulta evidente que los sistemas automatizados están metiendo la pata de forma vergonzosa y sistemática, los programadores abrirán los sistemas y retocarán los algoritmos, pero en casi todos los casos los programas emiten veredictos inquebrantables, y a los seres humanos que los utilizan solo les queda encogerse de hombros como si dijeran: «Bueno, ¿y qué podemos hacer?».

Y esa es precisamente la respuesta que Sarah Wysocki recibió finalmente del distrito escolar. Jason Kamras declaró después al *Washington Post* que las tachaduras

eran «sugerentes»[16] y que era posible que hubiera algún error en las puntuaciones de los alumnos de quinto curso de Sarah Wysocki, pero que las pruebas no eran concluyentes. Según Jason Kamras la habían tratado con justicia.

¿Es clara la paradoja? Un algoritmo procesa un montón de estadísticas y produce como resultado una cierta probabilidad de que una persona concreta *pueda* ser un mal empleado, un prestatario de riesgo, un terrorista o un pésimo maestro. Esa probabilidad se condensa en una puntuación, que puede llegar a destrozar la vida de alguien. Y, sin embargo, cuando esa persona decide defenderse, las «sugerentes» pruebas en contra del veredicto son insuficientes para aclarar las cosas. El caso debe quedar blindado. Como veremos a lo largo de este libro, en lo que a las pruebas se refiere, las víctimas humanas de las ADM tienen que responder a un nivel de exigencia mucho mayor que los propios algoritmos.

Después de la terrible sorpresa de su despido, Sarah Wysocki solo estuvo unos cuantos días en paro.[17] Había mucha gente dispuesta a responder por su trabajo como maestra, entre ellas su director, y en seguida la contrataron en un colegio de un próspero distrito al norte de Virginia. La consecuencia final fue que, gracias a un modelo altamente cuestionable, un colegio pobre perdió a una buena maestra, y un colegio rico que no despedía a los docentes por las puntuaciones que obtuvieran sus alumnos ganó una.

\* \* \*

Tras el estallido de la burbuja hipotecaria, fui consciente de la proliferación de ADM en la banca y del peligro que suponían para nuestra economía. A principios de 2011 dejé mi trabajo en el fondo de cobertura. Más tarde, tras añadir a mi currículum las palabras «científica de datos», empecé a trabajar en una *start-up* de comercio electrónico. Desde ese lugar estratégico veía que había legiones de ADM operando en todos los sectores imaginables, muchas de ellas exacerbando la desigualdad y castigando a los pobres. Estaban en el corazón de la atroz economía de los datos.

Quería hacer correr la voz sobre las ADM y empecé a escribir un blog, MathBabe. Con él pretendía movilizar a otros matemáticos contra el uso de estadísticas chapuceras y de modelos sesgados que crean sus propios bucles de retroalimentación perniciosos. Mi blog atrajo en particular a muchas personas especializadas en datos, que me advirtieron de la propagación de ADM a otros sectores. Pero a mediados de 2011, cuando arrancó el movimiento Occupy Wall Street al sur de Manhattan, comprendí que teníamos mucho trabajo que hacer con el gran público. Miles de personas se habían reunido para exigir justicia económica y rendición de cuentas, pero cuando escuchaba las entrevistas que hacían a los miembros del movimiento, a menudo saltaba a la vista que no sabían nada de los aspectos más elementales de las finanzas. Era obvio que no habían leído mi blog (aunque debo añadir que no es necesario entender todos los detalles de un sistema para saber que ha fracasado).

Me di cuenta de que tenía dos opciones: podía criticarlos o unirme a ellos; y decidí unirme a ellos. Poco después

empecé a dirigir las reuniones semanales del Grupo de Banca Alternativa en la Universidad de Columbia, donde debatíamos sobre la reforma financiera. Durante este proceso, descubrí que mis dos aventuras fuera del mundo académico, una en finanzas y la otra en la ciencia de los datos, me habían abierto las puertas a la tecnología y a la cultura que alimentan las ADM.

En la actualidad, los modelos matemáticos mal diseñados microgestionan la economía, desde la publicidad hasta las cárceles. Estas ADM presentan muchas de las características que tiene el modelo de valor añadido que arruinó la carrera de Sarah Wysocki en los centros educativos públicos de Washington. Son opacas, nadie las cuestiona, no dan ningún tipo de explicaciones y operan a tal escala que clasifican, tratan y «optimizan» a millones de personas. Al confundir sus resultados con la realidad sobre el terreno, la mayoría de ellas crean bucles de retroalimentación perniciosos.

No obstante, hay una diferencia importante entre un modelo de valor añadido de un distrito escolar y, por ejemplo, un ADM que busca a posibles clientes para los abusivos préstamos rápidos: las recompensas que obtienen son diferentes. En el caso del distrito escolar, la recompensa es una especie de divisa política, la sensación de que se están arreglando los problemas. Mientras que en el caso de las empresas, la recompensa es la divisa habitual: dinero. En muchas de las empresas que utilizan estos algoritmos canallas, el dinero que ganan a raudales parece demostrar que sus modelos funcionan bien. Si lo consideramos desde su punto de vista, tiene mucho sentido.

Construyen sistemas estadísticos para encontrar clientes o manipular a prestatarios desesperados, y sus crecientes ingresos parecen demostrar que van por el buen camino. El *software* cumple su función. El problema es que los beneficios acaban actuando como un valor sustitutivo de la verdad. Veremos que esta peligrosa confusión se repite una y otra vez.

Esto ocurre porque los científicos de datos se olvidan con demasiada frecuencia de la gente que está al otro lado de la transacción. Sin duda alguna saben que un programa de procesamiento de datos malinterpretará a las personas parte del tiempo, lo que significa que las clasificará en grupos que no les corresponden, o les negará el acceso a un empleo o la oportunidad de comprar la casa de sus sueños. Y, sin embargo, en términos generales, las personas que ejecutan las ADM no se paran a pensar en esos errores. Su retroalimentación es el dinero, que es también su incentivo. Sus sistemas están diseñados para engullir más y más datos y perfeccionar sus análisis con el objetivo de que les llueva más dinero. Los inversores, evidentemente, celebran estos rendimientos y colman a las empresas de ADM con más dinero.

¿Y las víctimas? Bueno, un científico de datos nos diría que ningún sistema estadístico es *perfecto*. Esas personas son daños colaterales. Y, con frecuencia, como en el caso de Sarah Wysocki, son consideradas indignas y prescindibles. Olvidémonos de las víctimas por un momento, nos diría, y pensemos en todas las personas que reciben sugerencias provechosas de los motores de recomendación y que encuentran música que les apasiona en la radio en línea

Pandora, su trabajo ideal en LinkedIn o quizá el amor de su vida en Match.com. Pensemos en la prodigiosa escala de estos modelos e ignoremos las imperfecciones.

El *big data* cuenta con muchos predicadores, pero yo no me cuento entre sus filas. Este libro va claramente en otra dirección, trata sobre el daño que infligen las ADM y las injusticias que perpetúan. Repasaremos ejemplos dañinos que afectan a las personas en momentos cruciales de la vida: empezar la universidad, pedir un préstamo, ser condenado a prisión o encontrar y conservar un trabajo. Todos estos ámbitos de nuestras vidas están cada vez más en manos de unos modelos secretos que blanden castigos arbitrarios.

Bienvenidos al lado oscuro del *big data*.

- 
- [1] Robert Stillwell, *Public School Graduates and Dropouts from the Common Core of Data: School Year 2006-07*, NCES 2010-313 (Washington D. C.: National Center for Education Statistics, Institute of Education Sciences, US Department of Education, 2009), p. 5, <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010313>.
- [2] Jihyun Lee, Wendy S. Grigg y Gloria S. Dion, *The Nation's Report Card Mathematics 2007*, NCES 2007-494 (Washington D. C.: National Center for Education Statistics, Institute of Education Sciences, US Department of Education, 2007), p. 32, <https://nces.ed.gov/nationsreportcard/pdf/main2007/2007494.pdf>.
- [3] Bill Turque, «Rhee Dismisses 241 D.C. Teachers; Union Vows to Contest Firings», *Washington Post*, 24 de julio de 2010, [www.washingtonpost.com/wp-dyn/content/article/2010/07/23/AR2010072303093.html](http://www.washingtonpost.com/wp-dyn/content/article/2010/07/23/AR2010072303093.html).
- [4] Steven Sawchuck, «Rhee to Dismiss Hundreds of Teachers for Poor Performance», *Education Week Blog*, 23 de julio de 2010, [http://blogs.edweek.org/edweek/teacherbeat/2010/07/\\_states\\_and\\_districts\\_across.html](http://blogs.edweek.org/edweek/teacherbeat/2010/07/_states_and_districts_across.html).
- [5] Bill Turque, «206 Low-Performing D.C. Teachers Fired», *Washington Post*, 15 de julio de 2011, [www.washingtonpost.com/local/education/206-low-performing-dc-teachers-fired/2011/07/15/gIQANEj5GI\\_story.html](http://www.washingtonpost.com/local/education/206-low-performing-dc-teachers-fired/2011/07/15/gIQANEj5GI_story.html).

- [6] Bill Turque, «“Creative... Motivating” and Fired», *Washington Post*, 6 de marzo de 2012, [www.washingtonpost.com/local/education/creative—motivating-and-fired/2012/02/04/gIQAwzZpvR\\_story.html](http://www.washingtonpost.com/local/education/creative—motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html).
- [7] *Ibid.*
- [8] *Ibid.*
- [9] *Ibid.*
- [10] *Ibid.*
- [11] Sarah Wysocki, entrevista por correo electrónico llevada a cabo por la autora, 6 de agosto de 2015.
- [12] Guy Brandenburg, «DCPS Administrators Won’t or Can’t Give a DCPS Teacher the IMPACT Value-Added Algorithm», *GFBrandenburg’s Blog*, 27 de febrero de 2011, <https://gfbrandenburg.wordpress.com/2011/02/27/dcps-administrators-wont-or-cant-give-a-dcps-teacher-the-impact-value-added-algorithm/>.
- [13] Turque, «“Creative... Motivating” and Fired».
- [14] Jack Gillum y Marisol Bello, «When Standardized Test Scores Soared in D.C., Were the Gains Real?», *USA Today*, 30 de marzo de 2011, [http://usatoday30.usatoday.com/news/education/2011-03-28-1Aschooltesting28\\_CV\\_N.htm](http://usatoday30.usatoday.com/news/education/2011-03-28-1Aschooltesting28_CV_N.htm).
- [15] *Ibid.*
- [16] Turque, «“Creative... Motivating” and Fired».
- [17] *Ibid.*



## Partes de una bomba

¿Qué es un modelo?

Era una calurosa tarde de agosto de 1946. Lou Boudreau, el jugador y director técnico del equipo de los Cleveland Indians, llevaba un día terrible. En el primer partido de un doble juego, Ted Williams, que jugaba con el adversario, casi había logrado aniquilar a todo su equipo él solito. Ted, quien tal vez fuera el mejor bateador de la historia hasta el momento, había anotado tres jonrones y llevado a ocho jugadores al *home*. Los Indians acabaron perdiendo 11 a 10.

Lou Boudreau tenía que hacer algo. De modo que, cuando Ted Williams apareció por primera vez durante el segundo partido, los jugadores de los Indians cambiaron de posición. Lou, el campocorto, corrió hasta donde solía colocarse el jugador de segunda base y este se retiró al campo exterior derecho. El jugador de tercera base se fue hacia la izquierda, al agujero del campocorto. Resultaba obvio que Lou Boudreau, quizá por pura desesperación,<sup>[18]</sup> estaba cambiando toda la orientación de su defensa para intentar poner a Ted William *out* cada vez que bateara.

En otras palabras, Lou Boudreau estaba pensando como un científico de datos. Había analizado los datos primarios, en general mediante observación: Ted Williams *normalmente* bateaba la pelota al campo derecho. Así que Lou se adaptó a eso. Y funcionó. Los jardineros consiguieron interceptar más veloces líneas de Williams que antes (aunque no pudieron hacer nada contra los jonrones que volaban sobre sus cabezas).

En cualquier partido de béisbol de una liga importante en la actualidad, los defensores tratan a casi todos los jugadores como Lou Boudreau trató a Ted Williams. Mientras que Lou simplemente observó dónde solía batear la pelota Ted Williams, los directores técnicos de hoy en día saben exactamente dónde ha bateado la pelota cada jugador durante la última semana, en el último mes, a lo largo de su carrera, contra lanzadores zurdos, cuando lleva dos *strikes*, etc. Utilizan estos datos históricos para analizar cada situación y calcular el posicionamiento de los jugadores que está asociado a la mayor probabilidad de éxito, lo que en ocasiones implica colocar a los jugadores en posiciones alejadas.

Cambiar a los defensores de posición es solo una parte de una pregunta mucho más amplia: ¿qué medidas pueden tomar los equipos de béisbol para maximizar la probabilidad de ganar? En su búsqueda de respuestas, los estadísticos del béisbol han desmenuzado cada una de las variables que han podido cuantificar y le han adjudicado un valor. ¿Qué diferencia de valor hay entre un doble y un sencillo? ¿Cuándo merece la pena dar un toque de bola

para que un jugador pase de primera a segunda base?  
¿Tiene algún sentido hacerlo?

Las respuestas a todas estas preguntas se mezclan y combinan en los modelos matemáticos de este deporte. Estos modelos constituyen universos paralelos del mundo del béisbol, y cada uno de ellos es un complejo tapiz de probabilidades. Incluyen todas las relaciones medibles entre los distintos componentes del béisbol, desde las bases por bolas hasta los jonrones, pasando por los propios jugadores. La finalidad del modelo es ejecutar diferentes escenarios en cada coyuntura para dar con las combinaciones óptimas. Si los Yankees ponen a un lanzador diestro para enfrentarse a Mike Trout, el bateador de los Angels que batea con más fuerza, en lugar de dejar al lanzador que tenían, ¿tendrán más probabilidades de eliminarlo? ¿Y cómo afectará eso a sus probabilidades de ganar el partido?

El béisbol es el entorno ideal para probar los modelos matemáticos predictivos. Tal y como explicó Michael Lewis en su *best seller* de 2003, *Moneyball*,<sup>[19]</sup> este deporte ha atraído a empollones obsesionados con los datos a lo largo de toda su historia. Hace décadas, los aficionados leían con atención las estadísticas que aparecían en el reverso de las cartas de béisbol para analizar los patrones de jonrón de Carl Yastrzemski o comparar los *strikes* totales de Roger Clemens y Dwight Gooden. Sin embargo, desde la década de 1980, algunos estadísticos profesionales empezaron a investigar lo que significaban realmente todas estas cifras, junto con una avalancha de nuevos datos: cómo estas cifras

se traducían en victorias y cómo los directivos podían maximizar el éxito con una pequeña inversión.

«*Moneyball*» es ahora sinónimo de cualquier enfoque estadístico aplicado a terrenos que han estado tradicionalmente dominados por el instinto. El béisbol es un caso práctico sin efectos perniciosos, y nos servirá de ejemplo positivo con el que comparar los modelos tóxicos o ADM que están aflorando en tantísimas áreas de nuestra vida. Los modelos del béisbol son justos, en parte, porque son transparentes. Todo el mundo tiene acceso a las estadísticas y todos entienden más o menos bien cómo interpretarlas. Es cierto que el modelo de un equipo puede conceder más valor a los bateadores que batean jonrones, mientras que tal vez otro reduzca un poco su valor, porque los bateadores suelen hacer muchos *strikeouts*. En cualquier caso, el número de jonrones y *strikeouts* está ahí y todo el mundo tiene acceso a esas cifras.

El béisbol tiene también rigor estadístico. Sus gurús cuentan con un inmenso conjunto de datos, la inmensa mayoría de los cuales están directamente relacionados con la actuación de los jugadores en los partidos. Además, sus datos son muy relevantes para los resultados que intentan predecir. Esto puede parecer obvio, pero, como veremos en este libro, es muy habitual que los que construyen ADM a menudo no dispongan de datos relativos a los comportamientos que más les interesan, por lo que los reemplazan por datos sustitutivos o *proxies*. Establecen correlaciones estadísticas entre el código postal de una persona o sus patrones de uso del lenguaje y su potencial para devolver un préstamo o realizar un trabajo. Estas

correlaciones son discriminatorias y algunas de ellas incluso ilegales. Los modelos del béisbol, en su mayoría, no emplean datos sustitutivos porque utilizan datos relevantes como bolas, *strikes* y *hits*.

Y, lo más importante, estos datos se acumulan de manera constante, gracias a las nuevas estadísticas que generan una media de doce o trece partidos diarios durante la temporada, que va de abril a octubre. Los estadísticos pueden comparar los resultados de estos partidos con las predicciones de sus modelos e identificar en qué se han equivocado. Quizá predijeron que un relevista zurdo perdería muchos *hits* frente a bateadores diestros, y, sin embargo, arrasa. En ese caso, el equipo de estadística tiene que retocar el modelo e investigar por qué se equivocaron. ¿Afectó la nueva bola de tornillo del lanzador a sus estadísticas? ¿Lanza mejor por la noche? Cuando descubren algo nuevo, lo introducen en el modelo para perfeccionarlo. Así es como funcionan los modelos fiables. Mantienen un constante ir y venir con los elementos del mundo real que intentan comprender o predecir. Y cuando las condiciones cambian, el modelo debe cambiar también.

En este punto podríamos preguntarnos cómo se nos habrá podido ocurrir comparar el modelo del béisbol, que incluye miles de variables cambiantes, con el modelo utilizado para evaluar a los docentes de los centros educativos de Washington D. C. En el primer caso se modelan todos y cada uno de los elementos del deporte hasta el más mínimo detalle y se incorpora información actualizada de forma continua. Mientras que el otro modelo, rodeado de misterio, parece basarse en gran

medida en la comparación de los resultados de un puñado de pruebas de un curso al siguiente. ¿Podemos considerarlo realmente un modelo?

Y la respuesta es sí. Un modelo, al fin y al cabo, no es más que una representación abstracta de ciertos procesos, ya se trate de un partido de béisbol, de la cadena de suministro de una petrolera, de las acciones de un Gobierno extranjero o de la asistencia del público a un cine. Independientemente de que se ejecute en un programa de ordenador o en nuestra cabeza, el modelo coge lo que sabemos y lo utiliza para predecir respuestas en distintas situaciones. Todos nosotros tenemos miles de modelos en la cabeza. Nos dicen qué esperar y nos guían al tomar decisiones.

Veamos un ejemplo de un modelo informal que yo empleo a diario. Tengo tres hijos, y soy yo la que cocina en casa — mi marido, bendito sea, ni siquiera se acuerda de echar sal al agua cuando cuece pasta—. Cada noche, cuando me pongo a cocinar para toda la familia, de manera interna e intuitiva, modelo el apetito que tendrá cada uno de los miembros de mi familia. Sé que uno de mis hijos se comerá solo la pasta (con queso parmesano rallado) y que a otro le encanta el pollo (pero odia las hamburguesas). Pero también tengo que tener en cuenta que el apetito de una persona varía de un día a otro, por lo que si se produce un cambio, la nueva situación podría pillar a mi modelo por sorpresa. Siempre hay una inevitable cantidad de incertidumbre.

Los datos de entrada de mi modelo interno para cocinar son la información que tengo sobre mi familia, los