

Analítica del aprendizaje y Big Data: heurísticas y marcos interpretativos

Daniel Domínguez
Facultad de Educación, UNED
ddominguez@edu.uned.es

José Francisco Álvarez
Facultad de Filosofía, UNED
jalvarez@fsof.uned.es

Inés Gil-Jaurena
Facultad de Educación, UNED
inesgj@edu.uned.es

Big Data and Learning Analytics: Heuristics and Interpretive Frameworks

RESUMEN: La capacidad de acceder directamente a informaciones referidas a todo tipo de prácticas sociales mediadas digitalmente y la correspondiente acumulación masiva de los datos, ha situado a la evaluación de los fenómenos sociales en un nuevo terreno que pone en cuestión los modelos analíticos convencionales. La educación es un campo propicio para preguntarse sobre esos movimientos, evaluar su relevancia epistémica y precisar en qué consiste la radicalidad del cambio producido por una nueva capacidad tecnológica. Este artículo contribuye a explicar el impacto de ese nuevo escenario en el terreno de la evaluación del aprendizaje a partir de *big data*, dando cuenta del cambio en la estructura de las categorías empleadas y desarrollando una nueva aproximación a la analítica del aprendizaje basada en heurísticas.

PALABRAS-CLAVE: *Big data*, aprendizaje, heurística, frameworks, racionalidad acotada, cognición distribuida, economía conductual

ABSTRACT: As a result of the capability to directly access information on all types of digitally mediated social practices and the corresponding massive accumulation of data, evaluation of social phenomena has taken a new direction that challenges conventional analytical models. Education is a suitable field for reflecting about these approaches, for analyzing the epistemic relevance of the new methods of data-driven assessment and for exploring the changes that arise from the new technological capabilities. This paper studies the impact of the new scenario in the field of learning analytics from big data, reflecting upon the change in the structure of the categories used in the evaluation of learning, as well as developing a detailed explanation of a new approach to learning analytics based on heuristics.

KEYWORDS: Big data, learning, heuristic, frameworks, bounded rationality, distributed cognition, behavioral economics

1. Introducción

El fenómeno complejo que identificamos como *big data* tiene sus raíces en las nuevas capacidades de gestionar grandes conjuntos de datos digitales, que se basan en una potencia de almacenamiento y procesamiento habitualmente superior a la de un solo ordenador. Una vez generadas las condiciones para procesar la abundante información procedente de todo tipo de prácticas sociales, la puesta en marcha de estudios que se derivan de aplicar el *big data* en la vida cotidiana requiere comprender este fenómeno como algo más que un artefacto tecnológico.

En el caso de las aproximaciones críticas e historicistas se ha puesto el foco en la necesidad de dotar a la gestión de datos masivos de un contexto que ayude a interpretar su impacto en la sociedad (Gitelman, 2013; Crawford, Gray & Miltner, 2014; Boellstorff & Maurer, 2015). Por ejemplo, algunos autores asocian el relato sobre el big data y su impacto en la investigación con un supuesto fin de la teoría y el correspondiente surgimiento de la ciencia libre de hipótesis.



Según este enfoque, en la medida en que se puede acceder a todos los datos que informan sobre un fenómeno, sería posible prescindir de los procedimientos inferenciales —inductivos o deductivos— a la hora de construir ciencia de calidad (Anderson, 2008; Graham, 2012).

Otras interpretaciones consideran que el fenómeno del *big data* es más bien una moda, otra más en el marco de la llamada “narrativa de Silicon Valley” (Weller, 2015). En las últimas décadas la industria digital ha sido capaz de introducir en el discurso social los argumentos en favor de sus intereses y este sería un nuevo caso de éxito. El discurso de Silicon Valley plantea que la aplicación de una determinada tecnología —considerada disruptiva— en un sector concreto altera el funcionamiento normal de ese sector beneficiando con ello al productor de la tecnología responsable del cambio, ya que en adelante la evolución de dicho sector dependerá de los avances en la industria tecnológica. El *big data* se explicaría entonces como la aparición de un nuevo servicio adecuado y dispuesto para la comercialización, cuyo desarrollo responde principalmente a los intereses de la industria digital y no a los objetivos de los diversos sectores de la sociedad donde se aplica la gestión de los datos.

La capacidad de acceder directamente a gran cantidad de datos procedentes de la actividad social genera tensiones en la definición de los fines, en los procedimientos y, por extensión, en la propia consideración de la ciencia social basada en esos datos. Por ejemplo, al mismo tiempo que aparecen nuevos métodos y disciplinas, se abre una reflexión sobre el cambio que supone acceder a la información del comportamiento de las personas directamente, sin filtros previos, o al menos sin el tipo de filtros conceptuales utilizados previamente. Asimismo, también se aprecia que las condiciones epistémicas en las que se apoya la regulación ética de la investigación han cambiado radicalmente a la luz de la disponibilidad de enormes cantidades de datos y, no menos importante, debido a los métodos utilizados para su utilización y gestión (Crawford, 2016; Farrow, 2016; Metcalf, Keller & boyd, 2016).

La educación es un ámbito propicio para experimentar los relatos sobre el *big data* y analizar el alcance de los nuevos enfoques vinculados al análisis de datos masivos. El intercambio y la adquisición de conocimientos en espacios digitales es creciente. Y con ello se abre la posibilidad a utilizar la explosión masiva de información que se genera en esos espacios para mejorar el aprendizaje y, en general, para apoyar la investigación básica sobre los procesos de enseñanza. Métodos como la minería de

datos y las técnicas de análisis del aprendizaje basadas en esos datos, han proliferado especialmente a partir de la expansión de los llamados cursos abiertos masivos *online* —MOOC, por sus siglas en inglés—. La academia, al menos en algunos de sus rincones más dinámicos, se ha abierto a las aportaciones del análisis de datos masivos en el estudio de las experiencias de aprendizaje digital, dando cabida a nuevas disciplinas como el *learning analytics* (Siemens *et al.*, 2011; Buckingham & Ferguson, 2012) o el *data science*, que consisten en una combinación de las ciencias de la computación, las matemáticas y la estadística aplicada (Gitelman, 2013; Kitchin, 2014). Asimismo, la investigación educativa también ha incorporado de manera habitual técnicas como el modelado de estudiantes, los sistemas de recomendación, la minería de datos o el análisis de la traza de los procesos educativos (Breslow *et al.*, 2013; Nesterko *et al.*, 2013; Thille *et al.*, 2014).

En este texto se aborda la naturaleza de las tensiones que se están manifestando en la práctica de la ciencia basada en datos y, específicamente, algunos de los cambios que debido a la aplicación del *big data* se producen en la investigación de los fenómenos educativos mediados digitalmente. A partir de esa aproximación se analiza cómo determinados supuestos sobre la conducta racional de los agentes se ven superados por la posibilidad de proceder a un acercamiento a la conducta que tenga en cuenta la capacidad de obtener y utilizar heurísticas; que se perciben y se pueden poner en práctica a partir de la disponibilidad de gran cantidad de datos muy básicos procedentes de la actividad de los individuos. Se ponen así en discusión muchos modelos teóricos de comportamiento que se alejaban de manera excesiva de las conductas registradas.

El punto de partida es un conjunto de investigaciones que cuestionan la consideración de los usuarios como agentes productores de datos —y especialmente la relación que tienen los usuarios con la superabundancia de datos derivados de sus propias prácticas—, así como las limitaciones que presentan los actuales modelos analíticos, toda vez que la información masiva parece haber superado la narrativa y la inteligibilidad hermenéutica que se apoyaba en el análisis de datos realizado a partir del tratamiento estadístico (Couldry, 2014; Mosco, 2014).

2. *Big data*, mediación digital y analítica del aprendizaje

Al igual que ocurre en otros muchos sectores que trabajan con *big data*, uno de los elementos más controvertidos de la gestión de datos en el ámbito educativo es precisamente todo lo relacionado con su recopilación, análisis y las diversas formas de utilización, incluida su monetización. La mayor parte de los educadores y estudiantes no es consciente de la cantidad de clasificación algorítmica directa que puede producirse en los procesos educativos mediados digitalmente. El análisis y la clasificación de los estudiantes comienza mucho antes de su acceso a la universidad, o de que los padres elijan un centro educativo. A partir del código postal de un candidato es posible establecer su probabilidad de éxito en los estudios. Los sistemas de recomendación sugieren los siguientes cursos que vendrían bien al estudiante. Una vez en el curso, el uso de las herramientas digitales produce modelos predictivos que en general suelen resultar adecuados. Se sugieren recursos de ayuda y apoyo para los estudiantes en situación de riesgo potencial. Y todo esto sucede detrás de la pantalla, en el sistema de gestión del curso, donde los algoritmos, adecuadamente estructurados y diseñados a partir del conjunto de valores que el administrador ha asociado a los procesos de organización y selección de datos, procesan el flujo de entrada y salida de la información tratando de condicionar directa o indirectamente una amplia gama de las prácticas de los estudiantes.

Los algoritmos actúan sobre la mayor parte de las prácticas educativas cuando estas se dan en escenarios mediados digitalmente y sometidos al control de actores principales. Su programación incorpora enfoques pedagógicos más o menos rudimentarios que determinan aspectos básicos que aparentemente no tienen ese mismo nivel de condicionante valorativo, pero que resultan decisivos como sucede con el diseño de las funciones de apoyo o las actividades de estudio. Algunos agentes educativos están comenzando a utilizar los datos masivos procedentes de la comunidad para codificar y diseñar los modelos de la educación de los próximos años. Las investigaciones se realizan principalmente desde la llamada analítica de aprendizaje —*learning analytics*—, una disciplina que crece sometida a una permanente redefinición de sus bases éticas y epistémicas (Borgman, 2015; Metcalf, Crawford & Keller, 2016; Tijerina, 2016). La aplicación de esta disciplina, el diseño de los estudios y los resultados obtenidos a través de esos métodos están generando una serie de tensiones, que han derivado en dos grandes narrativas.

De una parte, una posición maximalista respecto a las posibilidades abiertas por el *big data*, que conduce al intento de medir y contabilizar cualquier actividad conectada directa o indirectamente con el aprendizaje. Es la posición principal defendida por la industria de la tecnología educativa —e incluso por algunos gobiernos (Kamenetz, 2015)—, que insiste en las nuevas capacidades de gestión que surgen cuando es posible disponer masivamente de datos directos de los estudiantes. Su principal argumento es que el análisis de los datos obtenidos en situaciones de aprendizaje resulta clave para desbloquear la “caja negra” de la mente humana, mejorar con ello los procesos educativos haciendo que el aprendizaje sea más eficiente y, a partir de esos datos y su análisis, permitir tomar decisiones mejor informadas (Cuban, 2015; Westervelt, 2015). Se trata de un enfoque que asume un modelo determinista de la acción humana que se sustenta en la supuesta posibilidad de que los agentes puedan optimizar una serie de variables predefinidas que condicionan su conducta. Desde luego, en paralelo a esta actitud maximalista se genera una fuerte preocupación por los posibles abusos de una gestión educativa basada en datos, puesto que su recolección es demasiado amplia, pretende recoger todo tipo de datos y puede suponer una violación de la privacidad. Sin olvidar que el simple almacenamiento de datos conlleva riesgos de seguridad, al tiempo que su análisis y uso propicia la consolidación de un entorno de control y vigilancia algorítmica (Watters, 2015).

De otra parte, en abierto contraste con los modelos tradicionales de la teoría económica, que sugieren la optimización de una función de utilidad ajustada a las propias preferencias, se proponen otras aproximaciones menos exigentes que tienen en cuenta las prácticas efectivas de los sujetos y apuestan por esquemas valorativos explícitos. Estos enfoques tratan de satisfacer conjuntamente algunas variables buscando un equilibrio —aunque sea inestable— que permita avanzar en la acción con menos costes informativos y en una senda de eficiencia reversible y corregible con sujetos activos (Echeverría & Álvarez, 2008). En esta línea aparecen las alternativas que plantean avanzar en los estudios de la analítica de aprendizaje con el fin de comprender el comportamiento de los estudiantes en espacios digitales. Se parte de la asunción de que los grandes conjuntos de datos, ya sea inherentemente o en virtud de su tamaño, no poseen respuestas directas a las preguntas más interesantes. Por ejemplo, en el caso de los MOOC hay pocos estudios que hagan afirmaciones sólidas sobre el aprendizaje de los estudiantes y menos aún sobre los pequeños elementos didácticos que generan mejores aprendizajes. En palabras de

Justin Reich (2015, 34), “tenemos terabytes de datos sobre dónde han hecho *click* los estudiantes y muy poca comprensión de lo que ha cambiado en sus cabezas”.

3. Deslizamiento de las categorías analíticas

La posibilidad de acceder a bases de datos que contienen abundante información sobre el comportamiento de los usuarios en plataformas digitales de aprendizaje abre nuevas líneas a la investigación educativa. Las técnicas de análisis de *big data* proporcionan a los investigadores la información necesaria para monitorizar la totalidad de los comportamientos que son posibles en un sistema digital. Esta circunstancia amplía el rango clásico de la investigación educativa, incorporando de forma creciente las visiones procedentes de otras disciplinas como la economía o la antropología, y desde luego el enfoque propio de la informática.

Desde la aparición de la llamada computación personal, la informática se ha convertido en una disciplina de marcado carácter aplicado que trata de plantear soluciones ante situaciones sociales de todo tipo. Los requisitos de la programación ocupan un lugar central en la mayoría de estudios que requieren del procesamiento de grandes volúmenes de datos. Su prevalencia llega al punto de condicionar los métodos y la forma de plantear cualquier investigación basada en datos. Esta situación, unida a algunos otros factores, hace que abunde el uso de metáforas procedentes de las ciencias de la computación aplicadas para la interpretación de fenómenos propios de las disciplinas de carácter social¹. Se trata de un giro que está afectando especialmente a las dimensiones analíticas que sirven de base a la investigación educativa en contextos digitales.

En el campo de la teoría de la argumentación se denomina argumento resbaladizo a las instancias de un esquema argumental en el que una parte afirma que un resultado particular se da probablemente —o incluso inevitablemente— a partir de la aceptación de un argumento previo. Se apoya en un abuso de la analogía del tenor siguiente: las diferencias son pequeñas, por lo tanto no importan las diferencias. Como ha señalado Carlos Pereda (1994, 274), “ello es falso, diferencias poco importantes son, pese a todo, diferencias y además, acumuladas, a menudo suelen conformar diferencias importantes”. El argumento resbaladizo trata de justificar un posible efecto dominó: propone un pequeño primer paso y mediante una cadena

argumentativa conduce a una serie de eventos relacionados que culminan con un efecto significativo, por lo general negativo. También se asemeja al pequeño empujón dado a un objeto en el borde de una pendiente y que hace que se deslice hasta la parte inferior.

Uno de los efectos negativos del uso falaz de esta figura retórica reside en la linealidad acumulativa que lleva a reformular el contexto argumental y que puede desvirtuar el sentido inicial de la discusión alrededor de un tema (Voloikh, 2003). La introducción en el análisis de fenómenos sociales de determinados significantes a través de metáforas y marcos analíticos procedentes de la informática, así como el énfasis en los modelos deterministas y optimizadores basados en construcciones apriorísticas sobre la eficacia de la educación y la calidad de la instrucción algorítmica —referidos en el apartado anterior—, puede estar propiciando un efecto que equivale en sus consecuencias al de los argumentos resbaladizos. Una de esas consecuencias sería el desplazamiento de las categorías y dimensiones analíticas en la investigación sobre el aprendizaje en espacios digitales. Por ejemplo, se procede a identificar de un modo directo el aprendizaje con los comportamientos que suceden durante la adquisición de nuevos conocimientos, de tal modo que un registro de esos comportamientos —como el acceso a un foro digital, la consulta de un video, la descarga de un material de aprendizaje, etc.— se propone como la medida para valorar el aprendizaje en su conjunto, obviando que este es un fenómeno mucho más complejo de evaluar (Gil-Jaurena, 2013). En este caso, el comportamiento —en el sentido de hacer un *clic* en un entorno digital— se convierte en la dimensión analítica, deslizando al aprendizaje como la dimensión realmente relevante que debería de ser objeto de análisis.

Consideremos cuatro estudios llevados a cabo en las plataformas Udacity, Khan Academy, Google Course Builder y EdX (Collins, 2013; Murphy *et al.*, 2014; Reich *et al.*, 2014; Wilkowski, Deutsch & Russell, 2014). Cada estudio abordó una correlación entre las medidas de éxito —por ejemplo, las puntuaciones en los test de evaluación o la tasa de finalización de un curso— y las medidas de la actividad de los estudiantes. Los cuatro estudios trataron de registrar los datos de la actividad de un modo parecido, reduciendo la enorme cantidad de datos disponibles a una simple variable aplicable a una persona: el número de problemas intentados (Udacity), los minutos de permanencia en el sitio (Khan Academy), la tasa de finalización semanal (Google), o el número de clics por estudiante registrados en una sesión (EdX). A través de com-

paraciones simples o de regresiones, los cuatro llegaron a la conclusión de que existe una correlación positiva entre la actividad de los estudiantes y el éxito. Un resultado altamente probable para el que seguramente no haría falta tanta parafernalia.

Ese tipo de estudios ponen de manifiesto la dificultad que tienen los análisis basados en datos masivos para informar acerca de la complejidad de la actividad del estudiante. No es necesario analizar miles de registros de las entradas a un curso digital para demostrar que el esfuerzo del estudiante se correlaciona con el rendimiento. Lo que no es tan claro es el vínculo causal entre hacer, hacer más y hacerlo mejor (Reich, 2015). Y sin embargo, más allá de exhortar a los estudiantes a ser más activos en un curso digital, no se están planteando nuevas formas de re-diseñar las secuencias de aprendizaje a partir de los datos obtenidos. De manera que la próxima generación de estudios en este terreno necesita adoptar una gama más amplia de diseños de investigación, que presten una mayor atención a los factores causales y funcionales que promueven el aprendizaje del estudiante. Para lo que será necesario crear nuevas herramientas que ayuden a organizar y seleccionar la gran cantidad de datos que es posible analizar mediante el registro de la actividad digital de los estudiantes.

4. Racionalidad acotada, atajos cognitivos y educación

La evolución de las ciencias del comportamiento está abriendo nuevos escenarios para la investigación sobre el análisis de las acciones sociales cotidianas. En el campo educativo actualmente existe una amplia gama de enfoques interpretativos que se definen a partir de las prácticas sociales de los estudiantes, y cuya consideración permite ampliar el rango de la investigación tratando de evitar el determinismo en el establecimiento de las dimensiones analíticas. Es el caso de las teorías de la racionalidad, que se alejan de modelos formales optimizadores de la acción e intentan explicar el comportamiento efectivo de las personas durante la toma de decisiones relativas a la ejecución de acciones (Álvarez, 2002). Entre otros temas plantean que, en los escenarios de práctica real, a diferencia de los modelos de laboratorio, se presentan múltiples variables que afectan a las acciones a realizar y, en esos casos, las personas aplican mecanismos de decisión rápida y poco costosos infor-

mativamente hablando, denominados heurísticos, con los que tratan de adaptarse y explotar en su beneficio los requerimientos del ambiente (March, 2002; Evans, 2006; Gigerenzer & Gaissmaier, 2011).

La caracterización de los heurísticos que se considera en este texto es la propia del modelo de la racionalidad acotada (Kahneman, 2003 & 2012; Klaes & Sent, 2005; Echeverría & Álvarez, 2008) y que también se corresponde con las propuestas desarrolladas en el campo de la ciencia cognitiva (Hutchins, 1995). De acuerdo con Gerd Gigerenzer & Reinhard Selten (2001), para la racionalidad acotada las acciones son racionales a tenor del modo en el que el individuo es capaz de sacar un buen partido a la estructura de acción en la que se encuentra, teniendo en cuenta las restricciones contextuales e informativas en las que se produce la acción. Por su parte, los heurísticos son "atajos" cognitivos que permiten a los individuos realizar evaluaciones de una situación sobre la base de una o varias reglas básicas, evitando así los costes relacionados con una exploración exhaustiva de un conjunto amplio y complejo de opciones y, a su vez, admitiendo las circunstancias que surgen en el contexto. La ventaja de la racionalidad acotada y los heurísticos es que ambas teorías explican sencillamente cómo las personas pueden buscar las opciones adecuadas de forma rápida, fructífera y en función del ambiente. Cuando un sujeto utiliza un heurístico evita el esfuerzo de buscar, conocer y valorar todas las opciones posibles en un contexto en el que éstas no están a nuestra disposición con facilidad (Robles, 2007).

En educación los heurísticos han sido utilizados frecuentemente como recurso metodológico en disciplinas que demandan una alta capacidad cognitiva y meta-cognitiva (p.e. aplicado al caso de las matemáticas: Chávez, 2007; Hoon, Kor & Singh, 2013). A su vez las teorías del aprendizaje también recurren habitualmente a modelos y *frameworks* con fines analíticos, como una forma de operativizar las variables relacionadas con el comportamiento de los sujetos y explicar su funcionamiento en situaciones concretas. Por ejemplo, en el campo específico del aprendizaje conectado en espacios digitales, el uso de heurísticos y *frameworks* se ha empleado en numerosas ocasiones para analizar los procesos relacionados con el comportamiento de los estudiantes en situaciones mediadas tecnológicamente (Conole, Galley & Culver, 2010; Domínguez & Trillo, 2014; Wang, Chen & Anderson, 2014).

De acuerdo con la teoría de la racionalidad acotada, un uso inteligente de los heurísticos también podría aprovecharse para adaptar la estructura del entorno y explotar sus recursos de manera que se propicien determinados *outputs*. La aplicación de esa teoría ha dado resultados contrastados en ámbitos como la economía (Gilovich, Griffin & Kahneman, 2002) y la salud (Thaler & Sunstein, 2008). Nuestra aproximación toma en consideración esos avances y plantea su aplicación en el ámbito de la analítica del aprendizaje, utilizando los heurísticos como *frameworks* capaces de orientar tanto la metodología como la evaluación de las prácticas educativas.

Los "atajos" mentales vinculados a heurísticos generalmente se centran en un aspecto de un problema complejo y hacen caso omiso de los demás. La existencia de reglas heurísticas hace que no sean imprescindibles modelos complejos del mundo para tomar buenas decisiones. Esas reglas funcionan bien en la mayoría de las circunstancias, pero también pueden conducir a aparentes desviaciones sistemáticas de la lógica, la probabilidad o la teoría de la elección racional. Los errores resultantes se denominan "sesgos" cognitivos y pueden ser incentivados de manera exógena situando al sujeto en situaciones donde la información disponible es limitada. Por ello las aportaciones más destacadas de los heurísticos al campo de la analítica del aprendizaje se sitúan principalmente en dos órdenes. Por un lado, en su capacidad para poner en relación las prácticas de los estudiantes con una serie de marcos interpretativos que facilitan la explicación del rendimiento a partir de esas prácticas. Y por otro, en el establecimiento de un conjunto de elementos de juicio que permiten valorar cuánto de adecuadas son las conclusiones obtenidas tras el análisis y de acuerdo con el nivel de satisfacción considerado adecuado, teniendo en cuenta la información disponible y los límites en otros indicadores establecidos previamente.

5. Heurísticas en la analítica de aprendizaje

El caso más extendido de uso de heurísticos con fines de evaluación es el de la usabilidad de sitios web. Se podría decir que la evaluación web basada en heurísticos consiste en el análisis de la interfaz de un portal de acuerdo con un conjunto de criterios o principios de usabilidad. El método heurístico más destacado en la ingeniería de la usabilidad web es el desarrollado por Nielsen (1989) y Molich (*et al.*, 1998), que consiste en analizar la conformidad de la interfaz con unos principios de

uso reconocidos —heurísticos— mediante la inspección de varios evaluadores expertos. La aplicación del método se basa en validar 10 reglas heurísticas, que fueron generadas a partir del análisis factorial de 249 problemas de usabilidad detectados en las prácticas de una muestra de sujetos navegando en sitios web. Este método de evaluación fue desarrollado como una forma rápida y económica de probar interfaces. Su objetivo es encontrar los problemas de usabilidad en el diseño de la interfaz para que estos puedan ser atendidos como parte de un proceso de diseño iterativo (Nielsen & Mack, 1994). Este tipo de evaluación es, por tanto, un método discontinuo pero ampliamente aceptado para diagnosticar problemas potenciales de usabilidad en las interfaces web.

La proyección de estos métodos en la evaluación de aprendizajes basada en datos masivos resulta de utilidad para identificar las mejores prácticas que suceden en un contexto determinado, así como para detectar los elementos específicos que no cumplen con las reglas heurísticas, aportando elementos sobre la forma de corregirlos. También permite definir a partir del contraste con la práctica de los estudiantes el marco general para modelar los comportamientos esperados y ayudar a fijar criterios de satisfacción que cumplan con ciertas formulaciones acordadas previamente (Dzemyda & Sakalauskas, 2011; Reis & McNeill, 2014).

Respecto de la primera aportación, el análisis de datos masivos puede ser visto como un alejamiento de lo social y una aproximación hacia lo relacional. Hay una serie de diferencias fundamentales entre el “antiguo” modelo estadístico, que se apoyaba en la noción de muestra representativa y las correspondientes inferencias, y los nuevos modelos basados en la minería de datos. El más importante de ellos es la divergencia con respecto a los principios de la representación. Cuando la información de la audiencia se produce a través de la captura de datos procedentes de toda la población, se pierde la idea de que los datos se corresponden con el nivel social a través de su representación (Andrejevic, 2013). En esta situación post-referencial, la correlación se convierte en el foco principal y la dimensión explicativa de las estadísticas de representación se vuelve menos importante que el establecimiento de correlaciones entre los patrones de comportamiento, ya sean probables o prefijados como en el caso de las heurísticas. El “quién”, que es socialmente explicable, es menos importante que el “cómo” ha tenido lugar ese comportamiento que ahora es algorítmicamente predecible. Una de las consecuencias de ese enfoque es que, desde el plano de la práctica, si la predicción

del comportamiento grupal es posible, parece anularse la necesidad por parte del evaluador de conocer el dato desagregado a nivel de individuo. Ahora la preocupación fundamental es la detección frecuente de sociogramas o psicogramas que coincidan con las consultas particulares. Sobre este uso en el sector del análisis de usuarios en plataformas digitales, Karlsson (2014; en Bolin & Schwarz, 2015) cita a Jonas Wallentin, responsable de ventas de Google Suecia, que dice: “[En Google] sólo vemos los datos agregados. No a nivel individual, ya que no queremos arruinar la confianza que nuestros usuarios tienen en Google”.

Respecto de la capacidad de modelar el comportamiento del usuario, la principal aportación de la evaluación heurística estaría en el proceso de toma de decisiones en situaciones de complejidad, como las que se producen cuando se dispone de un número abundante de datos. Como se ha dicho, en la teoría de la racionalidad acotada los heurísticos actúan como los mecanismos o las herramientas empleadas por los individuos para adaptarse a los requerimientos del ambiente (Chater *et al.*, 2003; March, 2002). De manera que su aplicación en situaciones de análisis de aprendizaje hace que sea posible elegir grupos de indicadores y ponderar las dimensiones que deben tomarse en consideración.

Ayudar a fijar criterios de satisfacción adecuados, vinculados a indicadores clave de rendimiento (KPIs, por sus siglas en inglés) es quizás la principal aportación de los modelos de toma de decisiones. Normalmente los indicadores se establecen de acuerdo con los objetivos estratégicos en el marco de un determinado sistema. En el caso de un curso pueden ser, por ejemplo, el rendimiento visto en términos de resultados en pruebas estandarizadas o el absentismo durante el ciclo de estudio. Cuando los datos que saturan esos indicadores son escasos, las correlaciones suelen ser simples y, como se indicaba anteriormente, sencillas de interpretar e incluso de predecir. Sin embargo, la principal característica de la gestión de datos masivos es la complejidad. Lo que demanda un profundo conocimiento de la situación y de las personas involucradas en la producción de la información, una idea clara de lo que se quiere medir y claridad acerca de las relaciones entre las categorías.

Cada vez es más frecuente que los sistemas de evaluación basados en datos masivos recurran a KPIs propuestos a partir de supuestos que son evidentes, lo que está dando lugar a efectos negativos como es sustituir la búsqueda de marcos interpretativos adecuados donde poner en contexto los indicadores y los niveles de

satisfacción deseables. Los heurísticos y la racionalidad acotada aportan un enfoque diferencial en ese tipo de circunstancias, como es la introducción del concepto de “resultado satisfactorio” —y por lo tanto “medio satisfactorio” para proponer los datos a evaluar— a tenor de las características del contexto; y también optan por procesos flexibles y adaptativos en la búsqueda de los medios necesarios para el logro de los objetivos. La aplicación de este marco puede reflejarse por ejemplo en la forma de construir los cuadros de mando habituales en las aplicaciones de análisis de datos masivos, que tradicionalmente están asociados a categorías/variables e indicadores de rendimiento en esa variable (Buescher *et al.*, 2012). En última instancia, se trata de redefinir las cuestiones que sirven para estructurar el proceso analítico de los datos masivos: frente a la pregunta de qué KPI utilizar, subyace la alternativa de cuestionarse acerca del uso que se dará a ese KPI, y si tal uso responde a un interés de control de proceso o de conocimiento de la realidad a partir de datos procedentes de la práctica.

6. Discusión y conclusiones

Alrededor del *big data* surgen a menudo interpretaciones y malentendidos que afectan a los principios generales y a los procedimientos relacionados con la obtención y la gestión de los datos (Bolin & Schwarz, 2015). En lo relativo a la evaluación de procesos sociales —y educativos—, basados en datos masivos, algunas de esas dificultades pueden superarse mediante el recurso a ciertos mecanismos prácticos, teniendo en cuenta determinadas propiedades del llamado conocimiento heurístico. En el campo específico de la analítica del aprendizaje, las heurísticas se consideran recomendaciones generales basadas en la evidencia práctica, lo que contrasta con los conjuntos de reglas fijas —algoritmos— que no pueden variar, aunque en la mayoría de las ocasiones garantizan la respuesta correcta (Winker & Gilli, 2004). Los heurísticos no son construcciones *a priori*, sino que su identificación emerge de las prácticas, a menudo cotidianas, que realizan los individuos en diferentes situaciones. En el caso de los espacios mediados digitalmente, esas prácticas proceden de situaciones tales como identificar dónde escribir para hacer una búsqueda desde un navegador o editar un documento de trabajo colaborativo *online*. Las heurísticas pueden ayudar en la comprensión de esas prácticas cuando los responsables de evaluar el conocimiento acceden a ellas a través de conjuntos de datos masivos.

Pero el modelo de los heurísticos está sometido a discusión en lo relativo a sus aplicaciones. La definición de heurísticos no es un proceso deductivo, sino inductivo. El tener que indagar en los mecanismos psicológicos de las acciones racionales del individuo dificulta la tarea. Es por ello que todavía hoy la racionalidad acotada no es un corpus completo y coherente sino un "conjunto robusto de microteorías" (Robles, 2007). Si bien la inclusión de técnicas experimentales en la racionalidad acotada está suponiendo una evolución importante (Gigerenzer, 2015), no obstante, se debe ahondar más en el ámbito empírico.

Precisamente esa discusión ha llevado a acotar en este artículo las referencias a la acción de las heurísticas en la analítica del aprendizaje, limitando su alcance a los ámbitos donde existen evidencias más claras. Por ejemplo, se ha evitado plantear tipos de usos igualmente interesantes pero un tanto alejados, como puede ser el llamado análisis predictivo de comportamientos, un campo en el que las heurísticas también proporcionan marcos adecuados de análisis de datos basados en la satisfacción de criterios procedentes de experiencias demostradas (Lam, 2014).

Referencias

- Álvarez, J.F. (2002): El tejido de la racionalidad acotada y expresiva. En: Michael B. Wrigley (ed.), *Dialogue, Language, Rationality. A Festschrift for Marcelo Dascal* (pp. 11-29), Brasil, CLE/UNICAMP, <http://periodicos.sbu.unicamp.br/ojs/index.php/manuscrito/article/view/8644471/11893>
- Anderson, C. (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *WIRED*, 23 Junio, <http://www.wired.com/2008/06/pb-theory/>
- Andrejevic, M. (2013): *Infoglut. How too Much Information is Changing the Way We Think and Know*, New York, NY, Routledge.
- Boellstorff, T., & Maurer, B. (eds.) (2015): *Data, Now Bigger and Better!* Chicago, Prickly Paradigm Press.
- Bolin, G., & Schwarz, J.A. (2015): Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society*, 2(2). doi: 10.1177/2053951715608406
- Borgman, C.L. (2015): *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge, MA: MIT Press.
- Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A. D., & Seaton, D.T. (2013): Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8(1), pp. 13-25.
- Buckingham, S., & Ferguson, R. (2012): Social Learning Analytics. *Educational Technology and Society*, 15(3), pp. 3-26.
- Buescher, C., Hauck, E., Schilberg, D., & Jeschke, S. (2012): Key Performance Indicators for the Impact of Cognitive Assembly Planning on Ramp-Up Process. *Advances in Decision Sciences*, vol. 2012. doi: 10.1155/2012/798286

- Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003): Fast, Frugal, and Rational: How Rational Norms Explain Behavior. *Organizational Behavior and Human Decision Processes*, 90, pp. 63-86.
- Chavez, J.A. (2007): Enlivening problems with heuristics through learning activities and problem solving. *Learning Science and Mathematics*, 2, pp. 1-8.
- Cioffi-Revilla, C. (2014): *Introduction to Computational Social Science: Principles and Applications*, Berlin: Springer-Verlag.
- Collins, E.D. (2013): *SJSU plus augmented online learning environment: Pilot project report*. The Research & Planning Group for California Community Colleges, Sacramento, CA, http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE%20Report%20Final%20Version_Jan%202014.pdf
- Conole, G., Galley, R., & Culver, J. (2010): Frameworks for understanding the nature of interactions, networking, and community in a social networking site for academic practice. *The International Review of Research in Open and Distance Learning*, 12(3), <http://www.irrodl.org/index.php/irrodl/article/view/914/1666>
- Couldry, N. (2014): *Time and digital media: Time-deepening, time-deficits and narrative configuration*. Draft paper para Time, memory and representation group, Södertörn University, 14 Marzo.
- Crawford, K., Gray, ML, & Miltner, K. (2014): Critiquing Big Data: Politics, Ethics, Epistemology. *International Journal of Communication*, 8, pp. 1663-1672, <http://ijoc.org/index.php/ijoc/article/view/2167/1164>
- Crawford, K. (2016): Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology & Human Values*, 41(1), pp. 77-92. doi: 10.1177/0162243915589635
- Cuban, L. (2015): *Data-Driven Teaching Practices: Rhetoric and Reality*, 6 Octubre, <https://larrycuban.wordpress.com/2015/10/06/another-look-at-data-driven-teaching-practices/>
- Domínguez, D., & Trillo, P. (2014): Learning Competences in Open Mobile Environments: A Comparative Analysis Between Formal and Non-Formal Spaces. *Open Praxis*, 6(3), pp. 235-244. doi: <http://dx.doi.org/10.5944/openpraxis.6.3.131>
- Dzemyda, G., & Sakalauskas, L. (2011): Large-Scale Data Analysis Using Heuristic Methods. *Informatica*, 22(1), pp. 1-10.
- Echeverría, J., & Álvarez, J.F. (2008): Bounded Rationality in Social Sciences. En: Agazzi, E., Echeverría, J., & Rodríguez, A.G. (eds.), *Epistemology and the Social: A Feedback Loop* (pp. 173-191), Amsterdam/New York, Rodopi.
- Evans, J. (2006): The heuristic-analytical theory of reasoning: extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), pp. 378-395.
- Farrow, R. (2016): A Framework for the Ethics of Open Education. *Open Praxis*, 8(2), pp. 93-109. doi: <http://dx.doi.org/10.5944/openpraxis.8.2.291>
- Gigerenzer, G., & Gaissmaier, W. (2011): Heuristic Decision Making. *Annual Review of Psychology*, 62, pp. 451-482.
- Gigerenzer, G., & Selten, R. (Eds.) (2001): *Bounded Rationality. The Adaptive Toolbox*, Cambridge, MA., The MIT Press.
- Gigerenzer, G. (2015): *Simply rational: Decision making in the real world*, New York, Oxford University Press.
- Gil-Jaurena, I. (2013): Learning assessment in open, distance and flexible education. *Open Praxis*, 5(3), pp. 195-196, <http://dx.doi.org/10.5944/openpraxis.5.3.82>
- Gilovich, T., Griffin, D., & Kahneman, D. (eds.) (2002): *Heuristics and biases: The psychology of intuitive judgment*, New York, Cambridge University Press.
- Gitelman, L. (ed.) (2013): *Raw Data Is an Oxymoron*, Cambridge, MA., MIT Press.

- Graham, M. (2012): Big data and the end of theory? *The Guardian*, 9 Marzo, <http://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>
- Hoon, T.S., Kor, L.K., & Singh, P. (2013): Learning Mathematics Using Heuristic Approach. *Procedia - Social and Behavioral Sciences*, 90, pp. 862-869.
- Hutchins, E. (1995): *Cognition in the Wild*, Cambridge, MA., The MIT Press.
- Kahneman, D. (2003): A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, pp. 697-720.
- Kahneman, D. (2012): *Thinking Fast and Slow*, London, Penguin Group.
- Kamenetz, A. (2015): *The Test: Why Our Schools Are Obsessed with Standardized Testing-But You Don't Have to Be*, New York, PublicAffairs.
- Karlsson, M. (2014): Vi träffar Jonas Wallentin som är operativ chef på Google och diskuterar Analys. *Dagensanalys.se*, 9 Junio.
- Kitchin, R. (2014): Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), pp. 1-12.
- Klaes, M., & Sent, E-M. (2005): A conceptual history of the emergence of bounded rationality. *History of Political Economy*, 37(1), pp. 27-59.
- Lam, D. W. (2014): *A Survey of Predictive Analytics in Data Mining with Big Data*. Edmonton, Alberta, Canada, Athabasca University, https://www.academia.edu/8825157/A_Survey_of_Predictive_Analytics_in_Data_Mining_with_Big_Data
- March. B. (2002): Heuristics as Social Tools. *News Ideas in Psychology*, 20, pp. 49-57.
- Metcalf, J., Crawford, K., & Keller, E.F. (2015): *Pedagogical Approaches to Data Ethics*. Council for Big Data, Ethics, and Society report, <http://bdes.datasociety.net/council-output/pedagogical-approaches-to-data-ethics-2/>
- Metcalf, J., Keller, E.F., & boyd, d. (2016): *Perspectives on Big Data, Ethics, and Society*. Council for Big Data, Ethics, and Society, <http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/>
- Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D., & Kirakowski, J. (1998): Comparative evaluation of usability tests. *Proceedings of the Usability Professionals Association 1998 Conference* (189-200), Chicago, IL., UPA.
- Mosco, V. (2014): *To the Cloud: Big Data in a Turbulent World*. Boulder, CO, Paradigm Publishers.
- Murphy, R., Gallagher, L., Krumm, A., Mislavy, J., & Hafter, A. (2014): *Research on the Use of Khan Academy in Schools (Research Brief)*. Menlo Park, CA., SRI International, https://www.sri.com/sites/default/files/publications/2014-03-07_implementation_briefing.pdf
- Nesterko, S., Dostenko, S., Hu, Q., Seaton, D., Reich, J., Chuang, I., Ho, A. (2013): Evaluating geographic data in MOOCs. Presentado en el *Neural Information Processing Systems Workshop on Data Driven Education*, Lake Tahoe, NV, <http://nesterko.com/files/papers/nips2013-nesterko.pdf>
- Nielsen, J. (1989): *Coordinating user interfaces for consistency*, Boston, MA., Academic Press.
- Nielsen, J., & Mack, R.L. (1994): *Usability Inspection Methods*, New York, NY: John Wiley & Sons.
- Pereda, C. (1994): *Vértigos argumentales*, Barcelona y México, Anthropos y UAM-Iztapalapa.
- Reich, J., Emanuel, J., Nesterko, S., Seaton, D.T., Mullaney, T., Waldo, J., Chuang, I., & Ho, A.D. (2014): HeroesX: The Ancient Greek Hero: Spring 2013 Course Report. *HarvardX Working Paper Series*, 3, pp. 1-19, <https://dash.harvard.edu/handle/1/11988100>
- Reich, J. (2015): Rebooting MOOC Research. Improve assessment, data sharing, and experimental design. *Science*, 347(6217), pp. 34-35. doi: 10.1126/science.1261627
- Reis, C.A., & McNeill, F. (2014): *Heuristics in Analytics: A Practical Perspective of What Influences Our Analytical World*, SAS Institute Inc., Cary, North Carolina.

- Robles, J.M. (2007): Bounded rationality: a realistic approach to decision making. *Theoria*, 16(1), pp. 41-58.
- Siemens, G, Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R.S.J.d. (2011): *Open Learning Analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. Concept paper, http://www.elearnspace.org/blog/wp-content/uploads/2016/02/ProposalLearningAnalyticsModel_SoLAR.pdf
- Thaler, R., & Sunstein, C. (2008): *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT, Yale University Press.
- Tille, C., Schneider, E., Kizilcec, R.F., Piech, C., Halawa, S.A., & Greene, D.K. (2014): The future of data-enriched assessment. *Research & Practice in Assessment*, 9(2), pp. 5-16.
- Tijerina, B. (2016): Campus Support Systems for Technical Researchers Navigating Big Data Ethics. *EDUCAUSE review*, Julio/Agosto, pp. 48-49, <http://er.educause.edu/articles/2016/6/campus-support-systems-for-technical-researchers-navigating-big-data-ethics>
- Volokh, E. (2003): The Mechanisms of the Slippery Slope. *Harvard Law Review*, 116(4), 1026–1137. doi: 10.2307/1342743
- Wang, Z., Chen, L., & Anderson, T. (2014): A framework for interaction and cognitive engagement in connectivist learning contexts. *The International Review of Research in Open and Distance Learning*, 15(2), <http://www.irrodl.org/index.php/irrodl/article/view/1709/2838>
- Watters, A. (2015): Top Ed-Tech Trends of 2015: The Compulsion for Data. *Hackeducation*, 16 Diciembre, <http://hackeducation.com/2015/12/16/trends-data>
- Weller, M. (2015): MOOCs and the Silicon Valley Narrative. *Journal of Interactive Media in Education*, 2015(1), pp. 1–7. doi: <http://doi.org/10.5334/jime.am>
- Westervelt, E. (2015): Meet The Mind-Reading Robo Tutor In The Sky. *nprEd*, 13 Octubre, <http://www.npr.org/sections/ed/2015/10/13/437265231/meet-the-mind-reading-robo-tutor-in-the-sky>
- Wilkowski, J., Deutsch, A., & Russell, D. M. (2014): Student skin and goal achievement in the mapping with Google MOOC. Proceedings of the *first ACM Conference on Learning@Scale*, pp. 3–10, ACM, New York.
- Winker, P., & Gilli, M. (2004): Application of optimization heuristics to estimation and modeling problems. *Computational Statistics and Data Analysis*, 47(2), pp. 211–223.

Notas

1. En realidad, se trata de un proceso de influencia mutua donde la ciencia computacional se nutre de metáforas sociales para devolver de nuevo esas metáforas a su ámbito de procedencia cargadas ahora de nuevos significados (Cioffi-Revilla, 2014).