
EL ANALISTA DE INTELIGENCIA COMPETITIVA EN LA ERA DE LA AUTOMATIZACIÓN

ENRIC ESCORSA O'CALLAGHAN

IALE Tecnología

La actividad principal de los analistas de inteligencia en las organizaciones ha consistido típicamente en llevar a cabo estrategias de búsqueda en diferentes fuentes, obtener datos, procesarlos y analizarlos con la ayuda de ciertas herramientas para, finalmente, comunicar eficientemente los resultados de este análisis a aquéllos que puedan tomar decisiones en su organización.

Sin embargo, en un contexto como el actual, caracterizado por el crecimiento exponencial de la información y la multiplicidad de formatos, canales y estructuras, junto con la creciente capacidad para procesar esta información (poderosas herramientas de minería de datos y textos, cada vez mayor capacidad de almacenamiento e infraestructuras distribuidas) el rol analista de inteligencia está siendo redefinido y se está tomando cada vez más complejo y poliédrico, ya que para abordar estas cuestiones, el analista hoy día debe ser capaz de abarcar tareas adicionales tales como la construcción de modelos analíticos y predictivos (algoritmos de Aprendizaje Automático) y la gestión de volúmenes crecientes de datos de naturaleza radicalmente distinta y en flujo permanentemente (incluyendo datos de sensores procedentes de dispositivos conectados, datos financieros, datos genómicos, datos geográficos, datos multimedia, datos de las redes sociales, etc.). Esta situación desafía los procesos analíticos tradicionales, normalmente restringidos a sólo unas pocas fuentes -científicas y técnicas- y al uso de métricas (cienciometría) que eran sólo válidas en el contexto de aquel ámbito restringido.

El nuevo perfil de los analistas de inteligencia exige funciones adicionales; el analista hoy está pasando de ser un buscador eventual y un usuario de software a convertirse en un explorador permanente (*scout*), un agregador e integrador (*data wrangler*) -tareas éstas que suelen ocupar claramente el mayor porcentaje de horas de dedicación de todo el proceso analítico- y un modelador y dotador de sentido (*sense maker*) -¡todo a la vez!- y por tanto requiere cada vez más de experiencia en habilidades de programación -también llamadas informalmente habilidades de «hacker»- que básicamente le permitan recuperar datos e informaciones de distinto tipo, de distintas fuentes y por vías distintas- no sólo mediante programación de ecuaciones de búsqueda en bases de datos, sino también mediante procesos de extracción y estructuración de datos de la web (conjunto de técnicas conocidas como *webscraping*) o mediante acceso a datos de proveedores a través de APIs (*Application Programming Interfaces*) o mediante habilitar opciones de parametrizado y filtrado del flujo constante de información de las redes sociales, etc.- para luego poder analizar, integrar y modelar toda esta información a fin de responder a una multiplicidad de propósitos.

Así, un requerimiento específico en una organización actual puede involucrar por ejemplo, el tener que establecer un canal de acceso y comunicación con la API de un servicio externo, o configurar un flujo de datos o *pipeline* a medida canalizado en un cuadro de mandos o *dashboard* que permita accionar los datos u otro tipo de plataforma híbrida (o *mashup*) que nos permita aglutinar fuentes varias tales como bases de datos distribuidas en red y alimentarnos de sus datos a tiempo real. Estos entornos complejos actuales en cuanto a flujos de datos, implican nuevos retos y a la vez nuevas oportunidades para las organizaciones de crear productos novedosos de información integrada capaces de aportar valor accionable.

Pero para ello se requiere cada vez más, de conocimientos sólidos en estadística; Hal Varian, chief economist de Google, predijo en 2009 que los estadísticos serían...«la profesión más sexy de la próxima década!». Definitivamente esta disciplina transversal se ha convertido en una habilidad crucial, ya que las organizaciones deben poder sacar conclusiones de manera eficiente, precisa y certera de los datos que están tratando. En el proceso de análisis, particularmente en los procesos de modelado de algoritmos de aprendizaje automático, necesitamos asegurarnos de que somos capaces de controlar el sobreajuste (*overfitting*) al entrenar nuestros datos para que sigan siendo válidos cuando extendamos su aplicación a datos reales, también necesitamos ser capaces de detectar valores aislados o atípicos (*outliers*) que pueden afectar notablemente nuestra visión del conjunto que estamos analizando y necesitamos ser capaces de distinguir la correlación de la causalidad (tarea no baladí), para finalmente poder interpretar correctamente el sesgo y entender y validar los resultados obtenidos.

Por otra parte, en un contexto de trabajo colaborativo y sistemas inteligentes en red (acceso abierto, innovación abierta y organizaciones difusas), se requiere de flujos de trabajo flexibles y reproducibles de investigación; es decir, los procesos analíticos pueden configurarse en base a un determinado conjunto de datos estudiados y luego aplicarse a otros conjuntos de datos, manteniendo una trazabilidad con respecto al proceso analítico que se ha llevado a cabo. El análisis puede entonces ser compartido y reutilizado y se pueden construir nuevos modelos encima de los ya existentes a través de procesos incrementales de adquisición de conocimiento. En una encuesta reciente a investigadores realizada por Wiley, aproximadamente el 70% indicaron que habían compartido los datos de sus investigaciones, representando un incremento del 17% con respecto al año anterior. Se observa, pues, una tendencia a acompañar los resultados de las investigaciones con los datos analizados. El concepto de reproducibilidad toma cada vez mayor relevancia. Organizaciones privadas –varias grandes compañías tecnológicas– abrazan ecosistemas abiertos en sus modelos de negocio como una forma de incentivar la innovación. En estos entornos, se utilizan cada vez más los lenguajes de programación estadística de código abierto (tales como por ejemplo Python o R, por citar

sólo algunos de los más populares) que incorporan una gran cantidad de bibliotecas y paquetes o funciones para abordar problemas específicos y están fuertemente respaldados por comunidades de apoyo.

Finalmente, quizás la mayor amenaza –para todas las industrias, de hecho– proviene de los avances de la Inteligencia Artificial (AI). Los sistemas autónomos son ya capaces de tomar decisiones basadas en datos a través de algoritmos de testeo iterativo para encontrar las mejores soluciones (Aprendizaje Profundo o «*Deep Learning*») e incluso entregar una narrativa automática a sus resultados (Generación de Lenguaje Natural).

¿Qué posición deberá ocupar entonces el analista de inteligencia en el futuro para seguir aportando valor añadido a la organización?

HABILIDADES NECESARIAS PARA EL ANALISTA DE INTELIGENCIA DE HOY ↓

Algunos de los aspectos que se han mencionado con respecto al rol cambiante del analista de inteligencia aparecen representados en el Diagrama de Venn de la Ciencia de Datos o *Data Science* (Conway, Drew) que ilustra las actividades de los hoy llamados «*Data Scientists*». (Ver Figura 1)

FIGURA 1
DIAGRAMA DE VENN DE LA CIENCIA DE DATOS



Fuente: Drew Conway

Una parte fundamental de la actividad del analista de inteligencia tiene que ver con lo que entendemos por experiencia substancial en un dominio específico, esto es, una comprensión de la realidad en la que existen los problemas o cuestiones a resolver, de los factores críticos del negocio.

Pero adicionalmente al conocimiento de un sector, de una industria, de un mercado o de una área de experiencia determinada (sea ésta la biotecnología, la

FIGURA 2
EJEMPLOS DE ALGORITMOS DE APRENDIZAJE SUPERVISADO

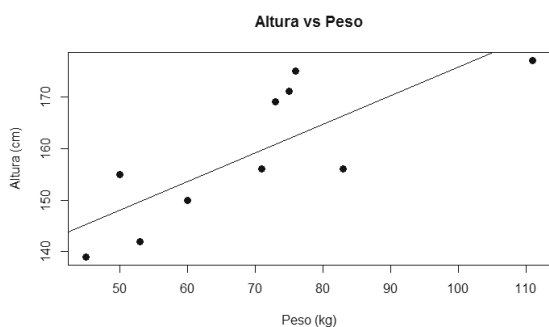


Figura 2a: Problema de Regresión, Regresión lineal.
Fuente: R

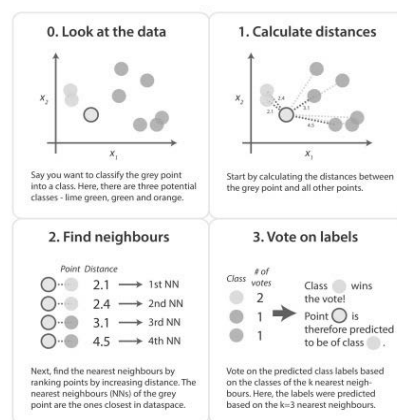


Figura 2b: Problema de Clasificación.
Algoritmo kNN (Nearest Neighbour)
Fuente: Cambridgecoding¹²

Fuente: Elaboración propia

salud, las finanzas, el comercio o cualquier otra), para poder afrontar hoy sus desafíos, se requiere adicionalmente de un conocimiento de métricas y procesos estadísticos, que nos permita trazar las informaciones desde el muestreo o captura de datos a partir de observaciones de la realidad, hasta la obtención de conocimiento sobre ésta.

Si además le añadimos a ello habilidades informáticas o de programación, seremos capaces de ir un paso más allá y generar, testear y probar algunos modelos, desarrollar algoritmos de aprendizaje automático para aprender de los datos y proyectar o hacer previsiones a partir de ellos.

AUGE DE LA MINERÍA DE DATOS Y EL APRENDIZAJE AUTOMÁTICO ↓

Hemos introducido la Minería de Datos y el Aprendizaje Automático, pero ¿a qué nos referimos exactamente con estos conceptos?

Minería de datos (en inglés *Data mining*) es el proceso de extracción de información implícita y potencialmente útil en grandes volúmenes de datos. Consiste básicamente en métricas y procesos estadísticos que podemos usar para detectar patrones y asociaciones en nuestros datos. Es decir, para describirlos y caracterizarlos, a menudo por medio de distintos tipos de visualizaciones, con el fin de encontrar información interesante (*insights*).

Pero no solamente podemos realizar un análisis descriptivo sino que también podemos realizar un análisis predictivo; esto es, podemos construir modelos analíticos que nos permitan predecir a partir de los datos y automatizar estos modelos; decimos entonces que podemos aprender de los datos y es por ello que llama-

mos a estos procesos Aprendizaje Automático (en inglés *Machine Learning*).

Yo puedo, por ejemplo, clasificar un conjunto de datos en clases o categorías y luego, aplicar el mismo modelo de clasificación a nuevos datos. Este tipo de procesos analíticos son los que puedo automatizar por medio del Aprendizaje automático.

Dentro de los procesos de Aprendizaje Automático distinguimos tres tipos principales: Aprendizaje Supervisado, Aprendizaje No supervisado y Aprendizaje reforzado.

El Aprendizaje Supervisado (*Supervised Learning*) es el que parte de unos datos previamente entrenados (clasificados o etiquetados). Consiste básicamente en encontrar una respuesta (o variable dependiente) que puede predecirse a partir de una serie de atributos (o variables independientes). Por ejemplo, si tenemos un set de pacientes o síntomas etiquetados como «sanos» o «enfermos»; o un set de mails etiquetados como «ham» o «spam», etc. podemos, en base a estos atributos, clasificar nuevos casos -nuevos pacientes, o nuevos mails que nos lleguen- de forma automática. Dentro del aprendizaje supervisado nos encontramos básicamente con dos tipos de problemas (ver Figura 2): problemas de regresión, cuando la respuesta es continua (por ejemplo, para saber cuánto mide alguien en función de cuánto pesa) y de clasificación, cuando la respuesta es categórica (clases), por ejemplo, para clasificar tipos de productos según su forma, tamaño o color.

A lo largo de los años los matemáticos y estadistas han propuesto –y siguen proponiendo hoy activamente- varios modelos o algoritmos para optimizar este tipo de funciones. Algunos ejemplos de algoritmos de aprendizaje supervisado son: regresión, árboles de decisión, Random Forest, kNN, regresión logística, etc.

FIGURA 3
EJEMPLOS DE ALGORITMOS DE APRENDIZAJE NO SUPERVISADO

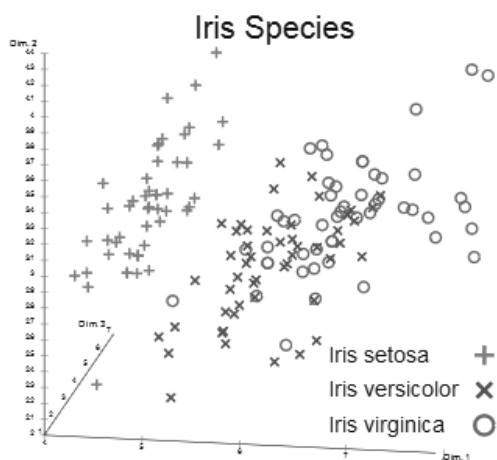


Figura 3a: Ejemplo de clusterización de grupos de flores del género Iris en base al largo y anchura de sus sépalos, mediante el método de clusterización K-means.
Fuente: wikimedia

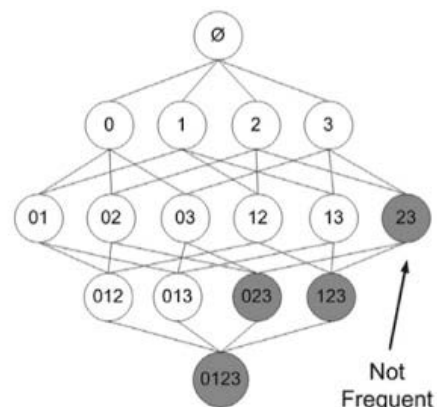
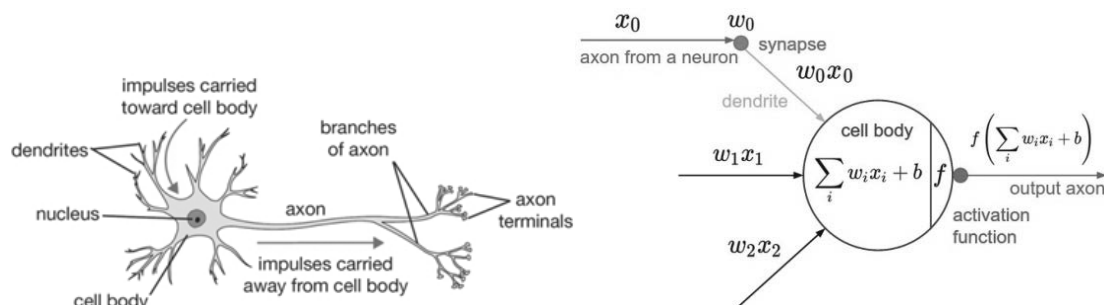


Figura 3b. Ejemplo clásico de descubrimiento de asociaciones en el comportamiento de usuarios mediante Algoritmo A Priori: Compra de Pañales y Cerveza los Jueves en el Supermercado.
La figura representa todas las combinaciones posibles de 4 productos (0, 1, 2 y 3) que pueden ser comprados.
Fuente: AImotion¹³

Fuente: Elaboración propia

FIGURA 4
REPRESENTACIÓN DE LOS ALGORITMOS DE REDES NEURONALES (DERECHA) INSPIRADOS EN EL FUNCIONAMIENTO TAL Y COMO LO CONOCEMOS DE LA SINAPSIS EN LAS NEURONAS BIOLÓGICAS (IZQUIERDA)



Fuente: Universidad de Stanford (curso Cs231)¹⁵

El Aprendizaje No Supervisado (*Unsupervised Learning*) es otro tipo de aprendizaje que usamos, en este caso, cuando no tenemos una respuesta o variable a predecir o estimar. No partimos de ningún set entrenado, ni de ninguna clasificación previa. Lo usamos para clusterizar una población en diferentes grupos. Por ejemplo, para segmentar grupos de clientes o para detectar tópicos subyacentes o clústeres temáticos en un conjunto de documentos.

Algunos ejemplos de Algoritmos de Aprendizaje no supervisado son: K-means, Apriori algorithm, Clusterización jerárquica, etc. (ver Figura 3)

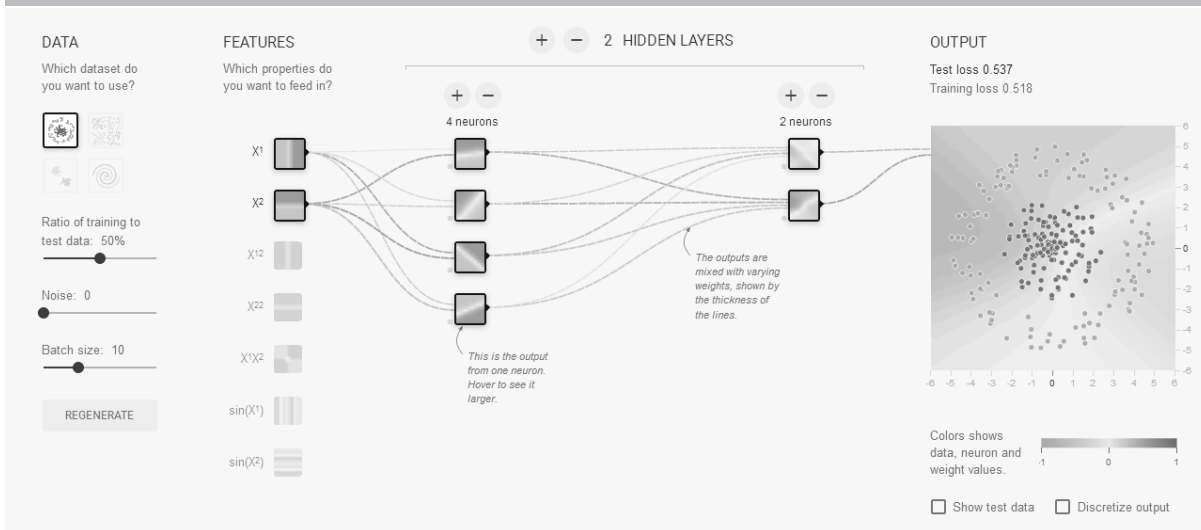
Finalmente, en el El Aprendizaje Reforzado (*Reinforced Learning*), se modelan automatismos que se exponen

a un entorno dónde pueden autoentrenarse continuamente mediante prueba y error. El sistema aprende de la experiencia pasada e intenta capturar el mejor conocimiento para tomar decisiones precisas.

Como ejemplos de algoritmos de Aprendizaje Reforzado destacan los Procesos de decisión de Markov y yendo un paso más allá en cuenta a complejidad y alcance podemos situar las Redes Neuronales (ver Figura 4) o el Aprendizaje Profundo (*Deep Learning*)¹⁴.

A grandes rasgos, las redes neuronales y el *deep learning* son técnicas empleadas para construir programas que aprenden de datos basados vagamente en cómo creemos que funciona nuestro cerebro. Primero, se crea una colección de «neuronas» de software

FIGURA 5
TENSORFLOW DE GOOGLE ES OPENSOURCE Y PERMITE EXPERIMENTAR CON REDES NEURONALES



Fuente: <http://playground.tensorflow.org>

FIGURA 6
MODELOS DE VISIÓN POR COMPUTACIÓN PARA DETECCIÓN E IDENTIFICACIÓN AUTOMÁTICA DE OBJETOS EN IMÁGENES. TECNOLOGÍA OPEN SOURCE DE GOOGLE (TENSORFLOW)



Fuente: Google research¹⁷

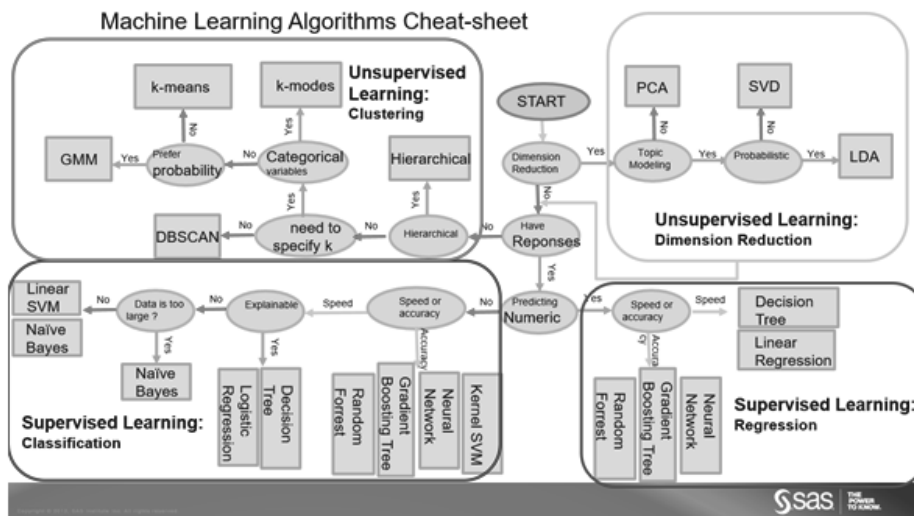
conectadas juntas, de modo a que puedan mandarse mensajes las unas a las otras. Luego se insta a la red a resolver un problema, que intenta resolver una y otra vez, cada vez estrechando aquellas conexiones que llevan a acertar y disminuyendo aquellas que llevan a fallar.

Encontramos aplicaciones de las redes neuronales por ejemplo en visión artificial y en coches autónomos, entre muchas otras, cada día más. Este ámbito de investigación, se encuentra hoy en pleno apogeo, y concentra una actividad muy intensiva en cuanto a nuevos desarrollos e implementaciones.

Muchas de las grandes empresas tecnológicas (Google, Facebook, Microsoft, Amazon, IBM, Baidu, etc.) han desarrollado entornos y librerías y las han puesto Open Source a disposición de los usuarios para que puedan experimentar. (ver Figura 5)

Algunos ejemplos de ellas son: Torch y Pytorch (Facebook); TensorFlow¹⁶ (Google); CNTK (deep learning) i DMTK (machine learning) de Microsoft; Deepmask (deep learning para visión artificial de Facebook); Caffe2 (deep learning para visión artificial de Yangqing Jia, Facebook); DSSINE (deep learning) de Amazon; Mahout (Machine learning) y BigDL (Deep Learn-

FIGURA 7
CLASIFICACIÓN DE LOS PRINCIPALES ALGORITMOS DE APRENDIZAJE AUTOMÁTICO



Fuente: Hui Lui, SAS¹⁹

FIGURA 8
TIPOS DE RAZONAMIENTO

TIPOS DE RAZONAMIENTO...	
<p>RAZONAMIENTO DEDUCTIVO:</p> <ul style="list-style-type: none"> • Comúnmente asociado con la "lógica formal". • Implica razonar desde premisas, conocidas o asumidas como ciertas, hasta una conclusión certera. • Sus conclusiones son certeras e inevitables. 	<p>RAZONAMIENTO INDUCTIVO:</p> <ul style="list-style-type: none"> • Comúnmente conocido como "lógica informal". • Implica trazar inferencias inciertas, basadas en razonamiento probabilístico. • Sus conclusiones son probables, razonables, plausibles y creíbles.
... Y SU ROL EN LA PRÁCTICA DE LA CIENCIA DE DATOS	
<p>RAZONAMIENTO DEDUCTIVO:</p> <ul style="list-style-type: none"> • Formular hipótesis sobre relaciones y modelos subyacentes. • Desarrollar experimentos con los datos para probar hipótesis y modelos. 	<p>RAZONAMIENTO INDUCTIVO:</p> <ul style="list-style-type: none"> • Análisis exploratorio de datos para descubrir o perfeccionar hipótesis. • Descubrir nuevas relaciones, percepciones y pautas analíticas de los datos.

TIPOS DE RAZONAMIENTO Y SU ROL EN EL TRATAMIENTO DE LA CIENCIA DE DATOS

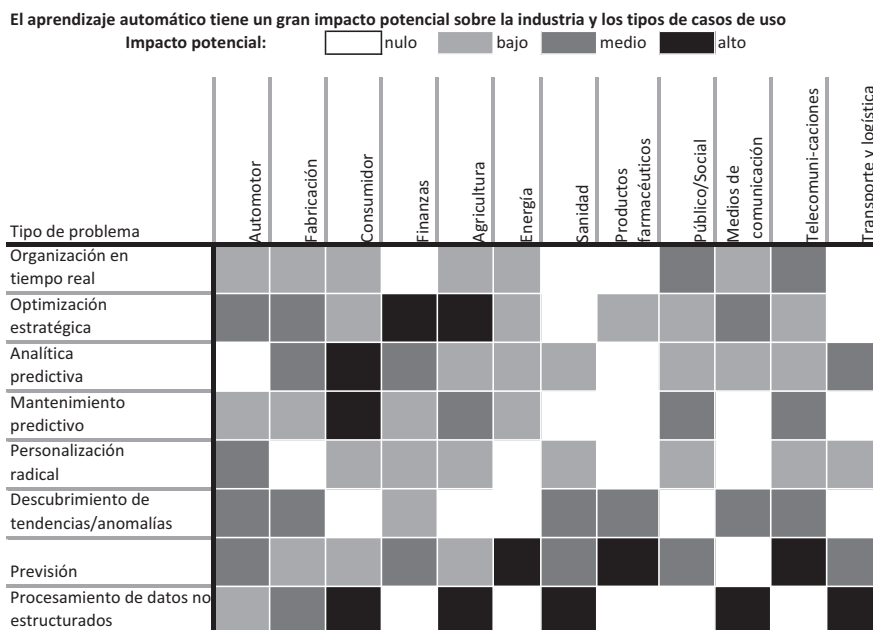
Fuente: Booz Allen Hamilton

ing) de Apache; SystemML by IBM (también Watson y Bluemix); PaddlePaddle de Baidu (Andrew Ng); DL4J o MxNet (de Pedro Domingos, Universidad de Washington/Amazon) y Keras (Python API para Tensorflow desarrollada por el ingeniero de Google François Chollet); (ver Figura 6).

En definitiva, un analista de inteligencia actual que trabaja con grandes volúmenes de datos, debe hoy conocer los principales algoritmos de aprendizaje

automático y entender cuando es más conveniente y eficiente aplicar cada uno de ellos y debe estar al día de los avances en las posibilidades y opciones de implementación de estos algoritmos que se encuentran disponibles (nuevas herramientas, entornos, librerías). Para ello es cada vez más necesario un conocimiento de lenguajes de programación y de entornos de desarrollo -tales como por ejemplo la librería *Scikit Learn*¹⁸ en el lenguaje de programación Python o los paquetes CARET, e1071, random-

FIGURA 9
IMPACTO POTENCIAL DEL APRENDIZAJE AUTOMÁTICO (ML) EN VARIAS INDUSTRIAS



Fuente: McKinsey Global Institute (MGI), December 2016 «The age of analytics: competing in a data-driven world»²³

Forest, XGBoost, entre otros, para el lenguaje R, etc.- que nos permiten cargar los datos que queremos analizar y acomodar distintos modelos y algoritmos para entrenar y modelar los datos y generar resultados (ver Figura 7).

EL ANALISTA DE INTELIGENCIA SE REPOSICIONA

Pero el analista de inteligencia no es únicamente un programador informático. Su función en última instancia es impulsar los procesos de adquisición de conocimiento y mejorar de la toma de decisiones en su organización. Los distintos métodos analíticos y algoritmos de aprendizaje son herramientas para apoyar a ese fin. El perfil del analista, volviendo al diagrama de Venn (ver Figura 1), es amplio y multidisciplinar.

Consideremos ahora los razonamientos implícitos en los métodos de investigación reflejados en ese esquema general: fijémonos primero en el método científico tradicional; el método científico típicamente involucra llevar a cabo un experimento en el cual una parte de la realidad se toma, se observa de cerca y es manipulada y diseccionada para luego obtener algunos datos que se anotan o registran; estos datos son revisados por el/la investigador/a que elabora una hipótesis, que luego intenta validar. Por tanto, hay un razonamiento deductivo a partir de una idea previa o teoría, esto es, tiene lugar un análisis de datos confirmatorio.

Contrariamente, en los procesos de ciencia de datos actuales basados en grandes volúmenes de datos y en los procesos de aprendizaje automático, la

realidad es percibida en conjunto, de manera comprensiva y orgánica, tomando datos secundarios que están ya disponibles. A partir de estos datos, el/la científico/a actual lleva a cabo análisis exploratorios y predictivos; esto es, hay un razonamiento inductivo; de este modo, en base a estos datos, se crean nuevas ideas o teorías (ver Figura 8).

En este contexto, un aspecto clave a considerar es que ambos procesos de aprendizaje (razonamiento deductivo e inductivo) son de hecho, complementarios entre sí y deben, por tanto, ir siempre juntos. Ninguno de los dos basta por sí mismo e incluso podemos decir que es peligroso remplazar el uno por el otro (Kuonen, D.)²⁰; la realidad, afortunadamente, siempre escapa tanto a reducciones como a generalizaciones; no podemos asegurar que nuestros experimentos sean universales sólo porque fueron validados eventualmente en una ocasión, ni tampoco podemos afirmar, del mismo modo, que los datos hablen por sí mismos (si los torturas suficientemente, tal y como decía Ronald H. Coase²¹). Es otra falacia afirmar que todos los datos tienen el mismo valor o calidad. De hecho, no es que los datos puedan estar a veces sesgados, sino que los datos –todos los datos– son por definición sesgados, en tanto que sistemas informativos artificiales construidos por humanos.

De ahí la importancia de una figura profesional dedicada a su gestión, capaz de entender su naturaleza, valorar su origen, los procesos de adquisición empleados, su calidad y su contexto y proponer o derivar usos posibles en función de motivaciones o problemas organizacionales o industriales específicos.

En la actualidad, numerosas industrias están implementando algoritmos de aprendizaje automático y profundo para distintas aplicaciones. Algunas de las cuales están generando ya un alto impacto (ver Figura 9) y dado el altísimo dinamismo actual en los avances tecnológicos (cada día aparecen nuevas publicaciones sobre aplicaciones insólitas en campos como la salud, las telecomunicaciones, la logística, la energía, la agricultura, las finanzas, etc.) se prevén prontas disrupciones de negocio en cada ámbito específico²².

Pero paralelamente al alto dinamismo en esta área de conocimiento, empiezan también a aflorar algunas preocupaciones en relación a si la ciencia puede/debe avanzar en base únicamente a conclusiones derivadas por máquinas y automatismos resultantes de modelos que en muchos casos son «cajas cerradas» o, en todo caso, difíciles de entender a ojos de la mayoría de los humanos²⁴.

Aparecen algunas voces críticas con respecto al modo en que ciertos automatismos ya implantados empiezan a tomar decisiones que afectan a las empresas y, a veces, incluso a nuestra cultura y nuestras vidas (por ejemplo decisiones sesgadas en base a raza, ubicación o género) y en contrapartida se empieza a hacer énfasis en los aspectos éticos de la ciencia de datos y la determinación de los modos de tomar decisiones responsables, justas y éticas al usar el Aprendizaje Automático y Inteligencia Artificial (Katharine Jarmul)²⁵

Es precisamente aquí dónde el rol del analista de inteligencia va recuperando su posición estratégica y su valor: cómo el puente entre las necesidades humanas y/o de la organización y los algoritmos que pueden aumentar sus capacidades; cómo asegurador del diálogo entre el razonamiento inductivo y el deductivo, cómo aquella persona que apoya en la definición en primer lugar de las políticas estratégicas y prioridades, motivaciones y estrategias, apalancando –en base a ellas– los datos necesarios, contrastando su veracidad y la utilidad de los modelos, contrastando el valor de las informaciones y discerniéndolas del ruido informativo y proponiendo modos creativos de identificar problemas y de desplegar narrativas compartidas con sentido.

¡Un futuro brillante, altamente motivador y lleno de retos aguarda al analista de inteligencia!

NOTAS ↴

- [1] Booz Allen Hamilton. «The Field Guide to Data Science» 2015. Web Accessed 1 February 2017. SSRN: https://www.boozallen.com/content/dam/boozallen_site/sig/pdf/publications/2015-field-guide-to-data-science-160211215115.pdf
- [2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. Hung Byers. «Big data: The next frontier for innovation, competition, and productivity» Report McKinsey Global Institute, 2011.
- [3] CSC Leading Edge Forum (LEF). «DATA rEVOLUTION». 2013. URL: <http://goo.gl/qlh6Ld>

- [4] Hal Varian «the sexy job in the next ten years will be statisticians» <http://online.library.wiley.com/doi/10.1111/j.1740-9713.2011.00476.x/pdf>
- [5] Wiley Open Science Researcher Survey 2016 https://figshare.com/articles/Wiley_Open_Science_Researcher_Survey_2016/4748332
- [6] Lenguajes y herramientas populares para el análisis de datos <http://r4stats.com/articles/popularity/>
- [7] Hadley Wickham and Garrett Grolemund «R for Data Science» O'Reilly Media. ISBN-13: 978-1491910399
- [8] Perera R, Nand P (2017). «Recent Advances in Natural Language Generation: A Survey and Classification of the Empirical Literature». *Computing and Informatics*. 36 (1): 1–32.
- [9] Conway, Drew. «The Data Science Venn Diagram.» March 2013.
- [10] Web. Accessed 15 October 2013. SSRN: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- [11] Top 10 Data Mining Algorithms <https://www.devteam.space/blog/top-10-data-mining-algorithms/>
- [12] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009) «The Elements of Statistical Learning: Data Mining, Inference and Prediction» Disponible en: <https://statweb.stanford.edu>
- [13] <http://blog.cambridgecoding.com/2016/01/16/machine-learning-under-the-hood-writing-your-own-k-nearest-neighbor-algorithm/>
- [14] <http://aimotion.blogspot.com.es/2013/01/machine-learning-and-data-mining.html>
- [15] Ver Neural Networks and Deep Learning de Michael Nielsen y Deep Learning de Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
- [16] Universidad de Stanford. Curso Cs231 <http://cs231n.github.io/neural-networks-1/>
- [17] TensorFlow de Google <http://playground.tensorflow.org>
<https://research.googleblog.com/2017/06/supercharge-your-computer-vision-models.html?m=1>
- [18] <http://scikit-learn.org/stable/>
- [19] <http://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use>
- [20] Kuonen, Diego Keynote NITS 2017, Brussels, Belgium March 14, 2017 available at <https://www.slideshare.net/kuonen/big-data-data-science-machine-intelligence-and-learning-demystification-trends-challenges-and-opportunities-for-official-statistics-in-the-context-of-datainformed-policy-making>
- [21] Gordon Tullock, «A Comment on Daniel Klein's 'A Plea to Economists Who Favor Liberty'», *Eastern Economic Journal*, Spring 2001
- [22] «The age of analytics : competing in a data-driven world» du cabinet de conseil américain McKinsey, datée de décembre 2016. <http://www.journaldunet.com/solutions/reseau-social-d-entreprise/1191979-machine-learning-12-secteurs/>
- [23] <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world>
- [24] David Weinberger «Alien Knowledge. When Machines justify knowledge» 18th April 2017. Accessed: <https://backchannel.com/our-machines-now-have-knowledge-well-never-understand-857a479dccc0e>
- [25] PyData, Anmsterdam 2017 <https://www.youtube.com/watch?v=hDgXlUM3Rmw>