

# De la búsqueda de personas diligentes a la valoración de opciones: cómo el big data y el machine learning están conformando el sector financiero

Carlo Sala  
Profesor del Departamento de Economía,  
Finanzas y Contabilidad de Esade

Nos encontramos en el umbral de la cuarta revolución industrial, que, como ocurrió con las tres anteriores, tiene el potencial de cambiar radicalmente nuestra forma de vida. Como también sucedió en las tres últimas revoluciones industriales, en la venidera hay un elemento único que es el principal impulsor del cambio. Piensa en internet, que fue el impulsor fundamental de la tercera revolución industrial. Considera cómo, con anterioridad, el petróleo permitió que una parte del mundo entrara en la era del consumo de masas entre finales del siglo xix y principios del xx. Antes incluso, el vapor puso en marcha la difusión de la máquina automática y empezó a sustituir a los trabajadores humanos en los lugares de trabajo durante el siglo xviii. ¿Cuál es el impulsor de esta cuarta revolución industrial? Los datos o, para ser más precisos, el big data.

Pero, ¿cómo de 'grande' es el big data? Si bien no hay una respuesta definitiva a esta difícil pregunta, un análisis del Foro Económico Mundial muestra que, a finales de 2020, la cantidad de datos creada por los principales actores del sector (redes sociales, instituciones financieras, plataformas comerciales, medios de comunicación, por citar algunos) era de 44 zettabytes. Un zettabyte tiene 21 ceros, lo que equivale a 1.000 bytes elevado a 7. Un estudio de IBM calculó que en 2018 el mundo produjo 2,5 trillones de bytes cada día (2.500 seguido de 15 ceros).

Una de las frases más habituales en los últimos años es que los datos son el nuevo petróleo, pero ¿por qué? En primer lugar, al igual que el petróleo, los datos son el principal impulsor de una revolución industrial. En segundo lugar, también como el petróleo, los datos hay

Los datos, como el petróleo, hay que refinarlos, pero son mucho más fáciles de obtener, más baratos de almacenar, no son un producto escaso y tienen múltiples usos y facetas.

que refinarlos. De hecho, el valor intrínseco de un byte de datos es escaso, y es necesario manipularlo de forma adecuada para que sea utilizable. Pero los datos tal vez sean mejores que el petróleo. Son mucho más fáciles de obtener, más baratos de almacenar, no son un producto escaso y tienen múltiples usos y facetas.

Esos datos no se habrían creado ni habrían sido distribuidos por todo el mundo sin la existencia de una robusta potencia de computación y una forma barata de almacenarlos. Además de los datos, la potencia de computación y la capacidad de almacenamiento están creciendo exponencialmente con los años. Piensa que la potencia de la CPU (siglas en inglés de unidad central de procesamiento) y la capacidad de almacenamiento del ordenador de navegación del Apollo (el software instalado en cada módulo de mando del Apollo y en su módulo lunar para controlar las primeras naves espaciales que aterrizaron en la Luna) y de la consola

En casi todos los análisis financieros, la información tiene una importancia clave. En el sector de las finanzas, el big data tendrá uno de los impactos más significativos.

Nintendo son comparables. El crecimiento exponencial de la potencia computacional hizo que algunos ordenadores fueran superiores a un cerebro de ratón (en algunas tareas) y nos ha traído a la época actual de los superordenadores. En mayo de 2021, el ordenador más potente del mundo es el japonés FUGAKU, con un rendimiento de 442 PFLOPS (operaciones de coma flotante por segundo).

Cabe señalar que tanto la potencia computacional como el big data son una consecuencia directa de la tercera revolución industrial. Conocida también como la primera revolución de la información, la tercera revolución industrial hizo que la información estuviera disponible para todo el mundo. Si antes de internet los datos eran escasos, ahora estamos desbordados por una cantidad de datos infinita. Hemos entrado, por tanto, en la era del big data, en la que todos intentan obtener un beneficio de esos datos.

Vale la pena preguntarse cómo está utilizando esos datos el sector financiero. No hace falta decir que en casi todos los análisis financieros la información tiene una importancia clave. Por ejemplo, la idea de eficiencia del mercado, uno de los principios centrales de la disciplina de la economía financiera, afirma que un mercado se considera eficiente si sus precios reflejan plenamente todos los datos disponibles. La disponibilidad de los datos puede, por tanto, crear una importante diferencia en términos de eficiencia.



En la era del big data, algo que hasta ahora solo se consideraba desde un punto de vista teórico está volviéndose (mucho más) asequible empíricamente. Así, no es sorprendente ver que, entre todos los sectores, en el de las finanzas el big data tendrá uno de los impactos más significativos. Aunque el sector financiero lo utiliza para infinidad de aplicaciones, en este artículo me centraré principalmente en algunas aplicaciones de la banca y algunos ejercicios de fijación de precios. En el proceso, también presentaré otra herramienta revolucionaria de los últimos años: el machine learning.

En las finanzas, una de las características más importantes y, al mismo tiempo, más difíciles de cuantificar es el perfil de preferencia de riesgo de una persona. Piensa, por ejemplo, en un asesor financiero que tiene que optimizar una cartera para un cliente. En el planteamiento habitual, el asesor financiero le pregunta al cliente cuánto dinero está dispuesto a invertir, su ventana temporal para la inversión y sus preferencias de riesgo. Con esos parámetros, al asesor le resulta sencillo elaborar una estrategia de inversión. Por desgracia, aunque la cantidad de dinero a invertir y el plazo son fáciles de determinar, un inversor normal suele tener dificultades para definir sus preferencias de riesgo. No es fácil agregar distintas características personales de carácter psicológico, financiero, económico y social, y cuantificar las preferencias de riesgo en una sola cifra, ni siquiera para personas consideradas expertas en finanzas.

Dada esta compleja mezcla de elementos, ¿por qué no utilizar las redes sociales para inferir la actitud de una persona frente al riesgo? Por ejemplo, una persona a la que le encantan los deportes extremos y los practica podría ser más 'amante del riesgo' que una persona que va al aeropuerto con varias horas de antelación para asegurarse de que no pierde el vuelo.

No es de extrañar que un estudio de Wu Youyou, Michal Kosinski y David Stillwell (2015) demuestre que Facebook te conoce mejor que tu familia, tus amigos, tus compañeros de trabajo y solo un poco peor que tu cónyuge. Mediante la aplicación de diferentes modelos econométricos, los autores demuestran que la fiabilidad del juicio de los humanos sobre la personalidad es, de media, peor que la de los ordenadores. Para calcularlo, parten de los cinco grandes rasgos de personalidad (apertura, amabilidad, extraversión, responsabilidad y neurosis), un modelo de la personalidad ampliamente aceptado y usado en la psicología académica. A partir de más de 70.000 encuestas, los autores comparan los resultados de los juicios humanos con los obtenidos de los ordenadores, que se basan en las preferencias de Facebook de los usuarios, expresadas mediante sus 'me gusta'. La fiabilidad del ordenador respecto a la media de los cinco rasgos recibió una puntuación de 0,56, justo por debajo de la del cónyuge (0,58), pero muy por encima de la precisión media de los humanos (0,49) o de la de los compañeros de trabajo (0,27).

En las finanzas, una de las características más importantes y, al mismo tiempo, más difíciles de cuantificar es el perfil de preferencia de riesgo de una persona. ¿Por qué no utilizar las redes sociales para inferir la actitud de una persona frente al riesgo?

Este estudio (entre otros) muestra que la explotación del enorme mundo del big data puede predecir los rasgos de la personalidad humana sin la intervención de las habilidades sociocognitivas humanas. Si te preguntas cómo podrían ayudar esos hallazgos a la industria financiera, ¿y si te dijera que si trabajas en el departamento de riesgo crediticio de un banco, tienes que buscar acreedores diligentes? En igualdad de condiciones, una persona diligente suele mostrar una probabilidad mucho menor de impago que una que no lo es.

En esta misma línea, Oded Netzer, Alain Lemaire y Michal Herzenstein (2016) utilizan la minería de textos y diferentes técnicas estadísticas para evaluar 18.000 solicitudes de préstamos de Prosper.com, una plataforma online de micromecenazgo, para mostrar qué palabras son más o menos representativas de una persona con una buena o mala solvencia. Los autores intentan responder a un problema muy delicado. Imagina de nuevo que eres un gestor de riesgo crediticio, o un simple oficinista que trabaja en un banco, y necesitas saber si debes prestarle o denegarle dinero a un cliente. ¿Qué palabras te gustaría escuchar de tu cliente? ¿Cuáles preferirías no oír? Resulta que palabras como 'dios', 'gracias', 'hospital', 'prometo', 'después de impuestos' y 'pagaré' las suelen decir personas con un mal historial crediticio, mientras palabras como 'graduado', 'tipo de interés más bajo' o 'pago mínimo' identifican a personas con un buen historial crediticio.

La principal ventaja frente a la programación genérica, que requiere algunos inputs y un algoritmo de programación formal para producir un output, es que las técnicas de machine learning 'aprenden' a programar de forma empírica mediante el uso de datos existentes.

Tanto el estudio de Youyou, Kosinski y Stillwell (2015) con el análisis de Facebook, como el de Netzer, Lemaire y Herzenstein (2016) que investiga las mejores palabras para determinar la solvencia de la gente, se basan en el uso de técnicas de machine learning. ¿Pero qué es exactamente el machine learning? Formalmente, es la combinación de diferentes herramientas estadísticas y econométricas y de una cantidad de datos significativa que permite a los ordenadores aprender de los datos sin que se les haya programado explícitamente para hacerlo.

La principal ventaja frente a la programación genérica, que requiere algunos inputs y un algoritmo de programación formal para producir un output, es que las técnicas de machine learning 'aprenden' a programar de forma empírica mediante el uso de datos existentes. Una vez que la máquina ha aprendido empíricamente el patrón que hay en los datos (la fase de entrenamiento), puede producir nuevos outputs a partir de nuevos inputs (la fase de predicción).

Hay que destacar que el machine learning es una subdivisión del campo mucho más amplio de la inteligencia artificial. Entre las diferentes técnicas de machine learning, se está convirtiendo en un enfoque especialmente exitoso el uso de redes neuronales. Teóricamente, las principales razones de la popularidad de las redes neuronales están relacionadas con el teorema de aproximación universal de Hornik et al. (1989). Según este teorema, (bajo condiciones específicas) si se le enseña adecuadamente con suficientes datos, una red neuronal siempre es capaz de aproximar una función continua cualquiera. A partir de esta sólida base teórica, las redes neuronales pueden aproximar cualquier función mediante la detección empírica de patrones a partir de grandes conjuntos de datos. De hecho, este hallazgo tiene consecuencias convincentes, en particular para problemas que implican funciones muy complejas.

Al trabajar con grandes grupos de datos (y ser, por tanto, una técnica basada por completo en datos), tanto el aumento constante de la potencia de computación como la disponibilidad de datos clasificados y no clasificados han sido elementos clave para explicar los impresionantes resultados logrados por las redes neuronales en una amplia gama de tareas, como la clasificación de imágenes. Una vez más, una consecuencia de la tercera revolución industrial presente en la cuarta revolución industrial.

Cuando la máquina aprende cómo funcionan los datos, puede repetirse el mismo proceso para crear nuevos outputs y hacer predicciones utilizando datos nuevos.

2

La principal idea detrás de las redes neuronales es imitar el comportamiento del cerebro humano, en particular las actividades de las neuronas del cerebro, y crear una formulación matemática de ese cerebro para extraer información de los datos de entrada. Dejando a un lado los tecnicismos, la réplica del cerebro humano a través de una red neuronal empieza por alimentarla con algunos inputs (1, 2, n) y por la producción de uno o más outputs mediante diferentes reelaboraciones entre las capas ocultas. Los resultados de cada nodo en las capas ocultas y de salida, las mencionadas reelaboraciones, no son más que combinaciones lineales de las capas previas que imitan las actividades de las neuronas en el cerebro.

Desde los inputs hasta los outputs, esta operación se repite muchas veces para entrenar al modelo, es decir, para dejar que la máquina detecte el patrón del conjunto de datos. Luego, cuando la máquina aprende cómo funcionan los datos, puede repetirse el mismo proceso para crear nuevos outputs y hacer predicciones utilizando datos nuevos.

Las técnicas de machine learning suponen una gran ventaja siempre que (1) los inputs y los outputs tengan una relación no lineal, y (2) no exista una solución de fórmula cerrada (es decir, una fórmula única que represente la resolución de nuestro problema).

3

La abundancia de datos y la naturaleza del problema han estimulado inherentemente la investigación sobre la aplicación de técnicas de machine learning para valorar opciones.

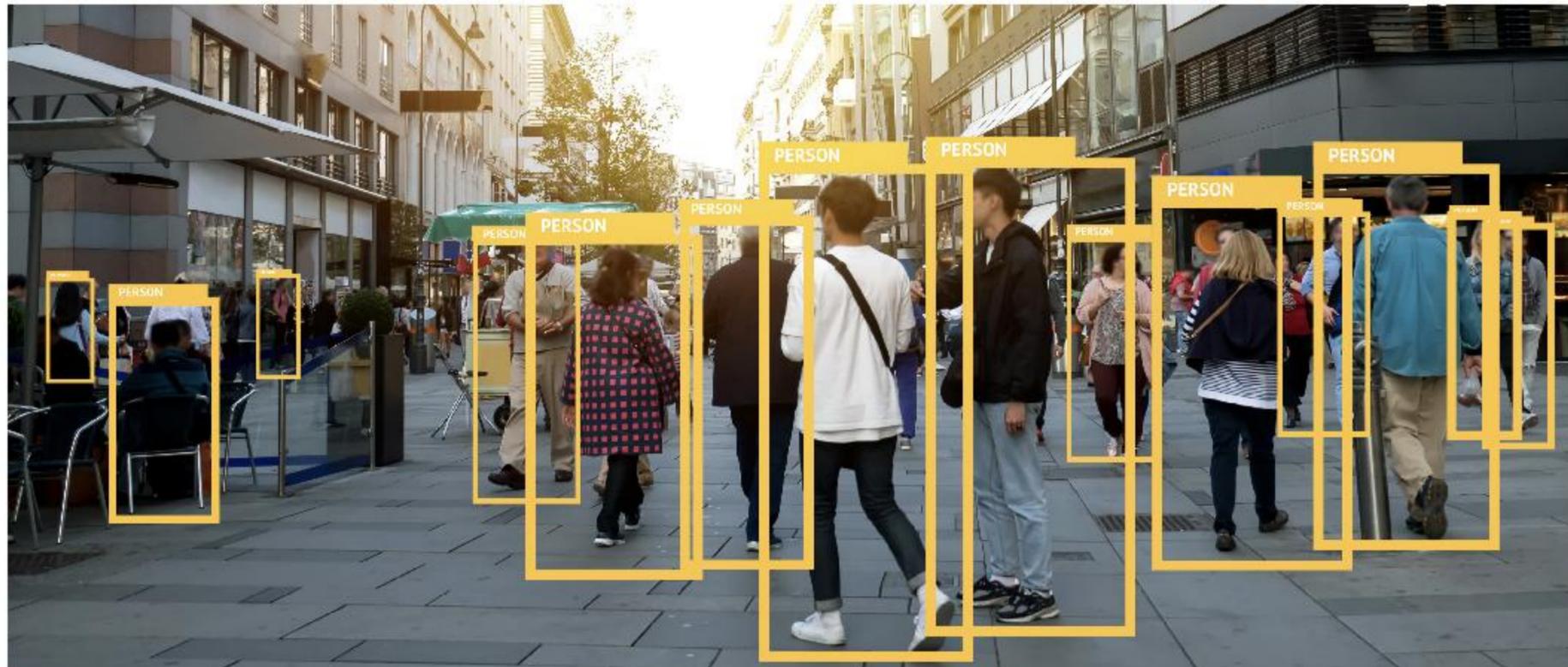
La ausencia de una solución de fórmula cerrada y la no linealidad de los inputs son exactamente dos problemas a los que tiene que enfrentarse cualquier persona que quiera valorar el precio de una opción. Las opciones europeas y estadounidenses son instrumentos financieros que dan el derecho –pero no la obligación– a comprar (call) o vender (put) un determinado activo (llamado el subyacente) a un precio predeterminado (el precio de ejercicio), en una fecha predeterminada (estilo europeo) o incluso antes (estilo estadounidense). Las opciones son, por tanto, una apuesta sobre el valor futuro del activo subyacente. Si el precio del activo subyacente es mayor (menor) que el precio predeterminado, eso hace que tenga sentido ejercer el derecho incluido en la opción de compra (de venta). De hecho, el valor futuro se determina, entre otros factores, por la volatilidad, el precio de ejercicio y cuánto tiempo tiene el inversor para ejercer su derecho. Por tanto, es necesario encontrar una función que conecte todas esas variables de una manera coherente y económicamente aceptable para valorar una opción.

La resolución de este problema tan complejo se expuso teóricamente en la famosa investigación de Fisher Black y Myron Scholes que, al mismo tiempo que Robert Merton, propusieron el modelo Black-Scholes-Merton en 1973. Este modelo valió a los tres profesores el Premio Nobel en 1997, debido a su enorme papel en el desarrollo de las finanzas cuantitativas, en particular del mercado de opciones. A pesar de este impacto enorme y la elegancia de su solución (que ya no estaba relacionada con la imposibilidad de estimar las preferencias de riesgos de los inversores, sino con la posibilidad de estimar la volatilidad subyacente), el modelo Black-Scholes-Merton no puede reproducir algunos hallazgos empíricos importantes del mercado de opciones.

Dado el momento en que se produjo su innovación, debería excusarse a los autores por no haber proporcionado una respuesta basada en datos al problema de la valoración de opciones. De hecho, hay que señalar que el Chicago Board Options Exchange abrió en el mismo momento que se publicó el artículo (1973), lo que hacía imposible que Black, Scholes y Merton dispusieran de series temporales de datos para probar su análisis. Además, entonces, las pruebas empíricas no eran tan fáciles como ahora: IBM introdujo su primer PC en 1981, y el disquete apenas tenía dos años de vida.

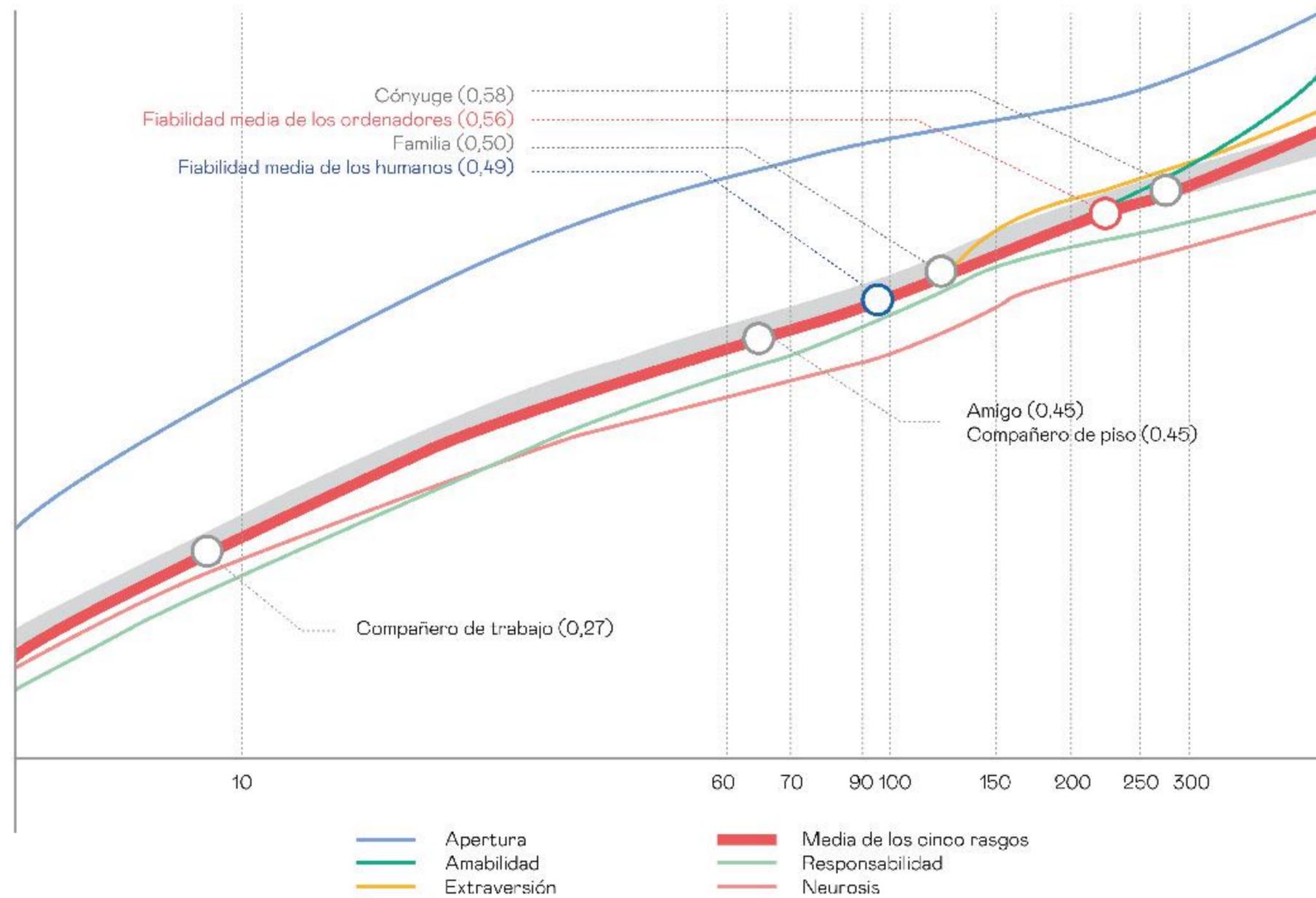
Hoy la situación es muy diferente, tanto en lo referente a la disponibilidad de datos como a la potencia computacional. La Option Clearing Corporation (OCC) ha documentado que en los últimos tres años en Estados Unidos se intercambiaron opciones por valor de más de 3 trillones de dólares. En la actualidad, contamos con técnicas econométricas y la potencia computacional necesarias para explotarlos. Esta abundancia de datos y la naturaleza del problema han estimulado inherentemente la investigación sobre la aplicación de técnicas de machine learning para valorar opciones (véase, entre otros muchos, Cazenave, Sala y Sohm-Queron, 2020, y sus referencias).

El mundo ha cambiado mucho desde 1973, cuando se publicó el artículo fundamental de Black, Scholes y Merton. Si bien ellos proporcionaron la base teórica para la valoración de opciones, la llegada del big data y algunas tecnologías nuevas (machine learning) podría suponer una valiosa adición al conjunto de herramientas del que disponen los analistas de finanzas cuantitativas, y convertirse, con el tiempo, en alternativas válidas a los métodos clásicos y tradicionales en muchas áreas de aplicación ●



Número de 'me gusta' en Facebook

1



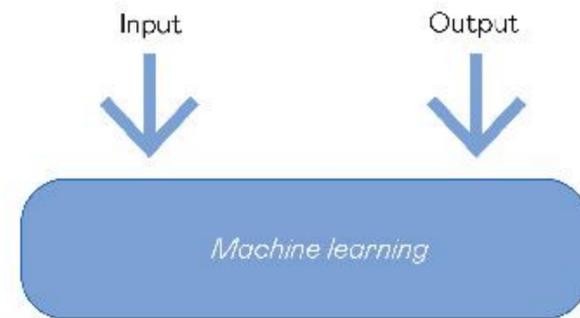
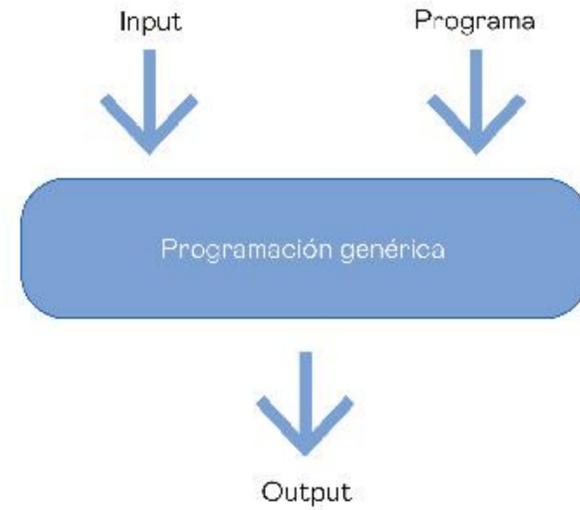
Fuente: Youyou, Kosinski y Stillwell (2015)

El estudio de Youyou, Kosinski y Stillwell muestra que la explotación del enorme mundo del big data puede predecir los rasgos de la personalidad humana sin la intervención de las habilidades sociocognitivas humanas.

[← Volver al texto](#)

Diferencia entre programación genérica y machine learning

2

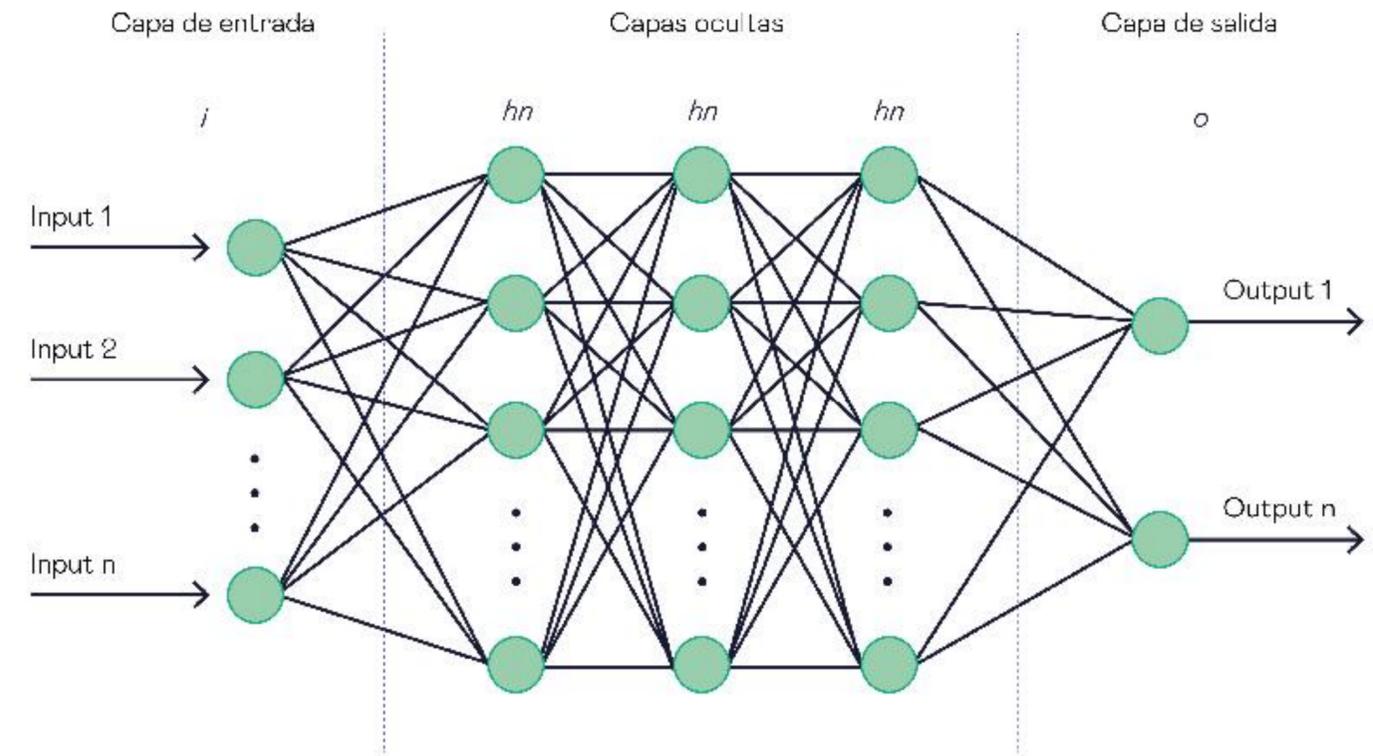


Nuevo input → Programa aprendido → Nuevo output

Fuente: Elaboración propia

Esquema de funcionamiento de las redes neuronales

3



Fuente: Elaboración propia