

BIG DATA, INTELIGENCIA ANALÍTICA Y LA BOLSA DE VALORES¹



HÉCTOR EDUARDO MENDÍVIL RIVAS

— Profesor de Gerencia de Proyectos
Facultad de Ingeniería, UPC

El análisis de datos como fuente de información es una ventaja competitiva.

RESUMEN

La toma de decisiones en las ruedas de bolsa requiere de datos e información, así como de reglas de inferencia a emplear por los agentes económicos. Con las reglas y la información se podrá establecer valores para las variables que le den soporte al análisis de las posibilidades de compra o venta que enfrenta el agente económico en las ruedas de bolsa. Para obtener ventajas en la indicada toma de decisiones dependeremos de la calidad y cantidad de reglas, de la cantidad y calidad de los datos e información, así como de la velocidad de los cálculos. Es así que el uso de *big data*, concepto relacionado con la gestión de diversas fuentes de datos, ofrecería una ventaja en las ruedas de bolsa.

Palabras clave: Big data, Gestión de la información, Gestión de activos financieros

ABSTRACT

Decision-making during trading sessions requires data and information, as well as rules of inference for the use of economic agents. Rules and information make it possible to establish values for the variables that support the analysis of the trading possibilities, faced by economic agents in the stock exchange. In order to gain an advantage in the aforementioned decision-making process, we depend on the quality and quantity of rules, data, and information, as well as the speed of the calculations.

Thus, the use of big data, a concept related to the management of various data sources, would offer an advantage in the stock exchange trading sessions.

Key words: Big data, Information management, Financial asset management

En la actualidad, los tratados de libre comercio (TLC) destacan como los instrumentos de integración y cooperación de más rápido desarrollo, especialmente los promovidos por los Estados Unidos de América con En el

presente documento analizaremos la importancia de los datos y su velocidad de procesamiento para la toma de decisiones. Para ello, se desarrollará el concepto de *dato* y el de *información*, el primero como generador del segundo. Luego de ello, revisaremos la forma en que el manejo de los datos y la información asociada influyen en el comportamiento de los agentes tomadores de decisiones en la bolsa de valores, creando tendencias en el tiempo. A partir de lo anterior, añadiremos al análisis el papel de las fuentes de datos y el significado de *big data*. Se verá que las fuentes de datos son fundamentales para generar información.

Incluiremos el concepto de *inteligencia analítica* como un elemento que da soporte a la toma de decisiones. Finalmente, presentaremos conclusiones a partir de lo expuesto.

LOS DATOS Y LA INFORMACIÓN

Un dato es la unidad mínima de información, y por sí mismo no tiene significado. La información es el resultado del procesamiento de datos, que incluye el empleo de modelos y de algoritmos, y su análisis permite incrementar el conocimiento sobre un tema en particular; en nuestro caso, la bolsa de valores. Asimismo, podemos decir que la información tiene un factor de incertidumbre, que implica la inclusión de variables aleatorias. Por ejemplo, el proceso de estimación del valor de un ADR² será estocástico. Es decir:

$$ADR_p = f(x_1, x_2 \dots) + \mu \quad (1)$$

Donde ADR_p es el precio del ADR, $f(x_1, x_2 \dots)$. Será la función de las variables x que determinan el precio más una variable aleatoria μ que recoge la incertidumbre. A partir de ello, la función de distribución de probabilidades especificará la probabilidad de determinación del precio del ADR; es decir, la variable aleatoria μ estará referida al set de información que no se conoce, pero cuyo comportamiento estadístico se debería determinar. Las variables $x_1, x_2 \dots$ especifican la información conocida, mientras que la variable aleatoria μ corresponde a la información que no se conoce.

1. UPC Review of Global Management, Volumen 3, Número 1, junio 2017

2. *American Depositary Receipt* es un título físico que respalda el depósito en un banco estadounidense de las acciones de compañías que fueron constituidas en el extranjero

Es importante resaltar que los valores de las variables exógenas deberán corresponder a un mismo corte de tiempo; por ejemplo, todas deberían ser mensuales o trimestrales, para que puedan ser comparables. Según la

LOS MERCADOS Y LA INFORMACIÓN

teoría económica, los mercados determinan los precios a partir de la interacción entre la oferta y la demanda. Los agentes económicos toman decisiones afectando los precios o las cantidades transadas. Dichas decisiones parten de un conjunto de datos y de información que los agentes económicos poseen de su entorno y del mercado. La bolsa de valores es un ejemplo claro de ello.

Según Covington (2016), los datos son muy importantes para la toma de decisiones en los negocios: en este sector, los datos de mercado y de productos brindan una ventaja sobre los competidores, y eso hace que estos datos valgan su peso en oro ("in the world of business, market and product data give a business the edge over competitors, and that makes such data worth its weight in gold" (p. 12).

Los agentes económicos poseen información o la crean desde la recopilación de datos. Con la información acumulada o generada, estos agentes definen sus decisiones de compra o venta según sea el caso, ajustando sus precios o tomándolos como aceptantes, de acuerdo con el tipo de mercado. Diremos que las decisiones se pueden tomar sobre las cantidades ofertadas o demandadas, o sobre los precios.

Las reacciones del mercado tienen un rezago; es decir, la ocurrencia de los hechos en la economía generará efectos en el mercado en un delta de tiempo siguiente. Estos rezagos serán diferentes para cada mercado, y dentro de él, para cada agente. El corte de tiempo ya no será estándar para todos: para algunos tomará menos tiempo y para otros más, lo que estará determinado por la velocidad del acopio de la información, su evaluación y la reacción de cada agente. No todos ellos tienen acceso a la información a la misma velocidad, y su capacidad de reacción tampoco será la misma. Asimismo, el set de información con que cada uno cuenta será diferente. Incluso, algunos poseen poca información, por lo que su comportamiento es menos autónomo y actúan más como seguidores.

Podemos decir que la información poseída por cada agente económico estará conformada por la común (la que la gran mayoría tiene) y la particular (aquella generada por cada agente según su acceso a datos particulares). Así, tenemos lo siguiente:

$$SI = Inf - común + Inf - específica \quad (2)$$

Donde *SI* representa el set de información de un agente, *Inf-común* corresponde a aquella a la que todos pueden acceder e *Inf-específica* es la que el agente podrá generar según los datos que obtenga o a los que tenga acceso. Por ejemplo, la información de los estados financieros de todas las empresas que ofertan en la bolsa de valores es de dominio público. Cualquier tomador de decisiones que navegue en internet podrá analizar los estados financieros

de las compañías que juegan en la bolsa de valores. En este caso, las reacciones sobre el mercado se toman sobre la información obtenida de dicha fuente. Cabe resaltar que el movimiento en el precio es paulatino, de acuerdo con la información que se va distribuyendo entre los agentes. Un ejemplo de ello es la acción de Graña y Montero, que mostró en abril un rendimiento errático con tendencia a la baja con el pasar de los días, pues, con cada día, la velocidad de ajuste se puede incrementar. En cuanto a los bancos, ellos no solo tienen información: también la producen. A partir de diversos datos de la economía, pueden generar estimaciones de variables que les permiten obtener un mejor ajuste en sus decisiones, por ejemplo, en la compra de dólares. La decisión de compra de dólares por parte de los bancos será más temprana que la de las personas de a pie, quienes se comportarán como seguidores.

Mientras más formal sea el agente o pertenezca al sector de la economía más moderno, la proporción de información específica que se pueda generar será mayor que la del promedio de agentes económicos. Por lo tanto, sus reacciones ante el mercado tomarán mayor velocidad que el promedio.

Por lo anterior, podemos afirmar que hay dos elementos esenciales en la toma de decisiones: por un lado, el acceso a la información y los datos, y por otro, la capacidad de procesamiento. En este segundo caso, el acceso a la tecnología y a los modelos estadísticos será fundamental. Cabe resaltar que hoy ya existen equipos que pueden procesar todo directamente de la memoria, sin ingresar a sus discos, con lo que logran que los tiempos de respuesta sean muy cortos.

BIG DATA Y LAS FUENTES DE DATOS

Los modelos de análisis de datos trabajan con variables, las cuales deben poseer las observaciones ordenadas temporal o transversalmente. De ser temporales, han de tener un corte de tiempo uniforme, pues todos sabemos que no es posible comparar datos mensuales de una variable con datos trimestrales de otra. Estadísticamente hablando, si los datos no presentan dicha característica, no se podrán procesar. Para obtener las observaciones de las variables se debe haber capturado y procesado datos de manera que tengan las características mencionadas, algo que realizan organismos tanto públicos como privados, como el Banco Central de Reserva, ministerios como el de la Producción o el de Economía y Finanzas, la Bolsa de Valores de Lima y la Superintendencia de Banca, Seguros y AFP, entre otros. Sin embargo, debemos reconocer que hay otras fuentes de datos no estructuradas, como las noticias o las redes sociales.

Hoy es común que los presidentes de diversos Estados anuncien u opinen, por ejemplo, a través de Twitter. ¿Será posible que en dichas fuentes se filtre algún dato que pueda afectar las decisiones de los agentes económicos? La respuesta es obvia: sí es posible. En el caso de Graña

y Montero que se tomó como ejemplo, la información sobre la corrupción no estaba estructurada. Los primeros indicios surgieron de diversas opiniones, muchas de ellas periodísticas. Ello motiva la siguiente pregunta: ¿qué podemos hacer con esas fuentes de datos? Debemos reconocer que las mismas nos enfrentan a tres problemas tecnológicos: el primero de ellos se refiere a la lectura de contenidos de diferentes formatos; el segundo, a los algoritmos que permitan interpretar los contenidos leídos por los sistemas; y, finalmente, el tercero es la gran cantidad de datos y de información que está circulando en internet.

Los problemas tecnológicos mencionados determinarán un elemento adicional que discriminará a los agentes económicos, dado que el soporte de *hardware* y *software* es altamente costoso. Los que cuenten con capacidad de procesamiento de datos más moderna y amplia tendrán ventaja sobre los que no la posean.

Los algoritmos para el análisis de datos son conocidos hoy como inteligencia analítica. Ya lo dijeron Davenport y Harris (2007) en su libro *Competing on Analytics*: "Por inteligencia analítica entendemos el uso exhaustivo de datos, análisis estadísticos y cuantitativos, modelos explicativos y de predicción y dirección y gestión basados en hechos como motor de la toma de decisiones y de las acciones emprendidas" (p. 32).

Sobre la base de lo anterior, cabe esta pregunta: ¿la inteligencia analítica se puede aplicar a la toma de decisiones en la bolsa de valores? La respuesta es sí. Las decisiones en la bolsa requieren de datos, así como del análisis de los mismos, en un contexto estadístico. Por ejemplo, las tendencias de los precios en valores o acciones se podrían determinar a partir de datos pasados y de modelos estocásticos. Incluso, podríamos hablar de predicción de valores y de comportamientos temporales identificando tendencias. Todo ello se puede enriquecer con un análisis de correlación entre diversas variables, no solo determinando causalidad

Para que la inteligencia analítica pueda operar se requiere de datos y de información como insumo primordial. Aquí es cuando introducimos el concepto de *big data*. Según el Informe Ejecutivo del IBM Institute for Business Value (Schroek et al., 2012):

Mucha de la confusión que existe en torno al concepto de *big data* comienza con la propia definición. Para comprender cuál es la definición que los encuestados dan a este término les pedimos que eligieran dos características de *big data*. No hay una determinada característica que predomine sobre el resto, sino que los encuestados mostraron división en sus opiniones acerca de si la mejor manera de describir el concepto de *big data* pasaba por los volúmenes de datos en crecimiento de hoy en día, los nuevos tipos de datos y análisis o los requisitos emergentes de un análisis de la información en tiempo aún más real (p. 3).

De lo anterior, podemos resaltar tres elementos fundamentales:

- Gran cantidad de datos
- Diversidad de tipos de datos y por ende de sus formatos
- Análisis de la información en tiempo real

El primer elemento no solo alude al volumen, sino también a la aparición de nuevas fuentes cada día. Ello implica que los agentes que juegan en las ruedas de bolsa deberían tener un soporte que les permita acceder e identificar nuevas fuentes de datos.

En cuanto al segundo elemento, la diversidad de tipos de datos y de sus formatos hace referencia a que las variadas fuentes expresan sus datos e información (contenidos) en múltiples formatos. Para que los sistemas puedan conectarse a las fuentes, deberá conocerse cada formato, incluyendo los nuevos, de modo que pueda obtenerse la información.

Finalmente, el análisis de la información en tiempo real se relaciona con el espacio de tiempo que hay entre la captura del dato y la obtención de la información. Vale decir, se toman decisiones sobre lo que está ocurriendo en el momento.

Los tres elementos indicados requerirán soporte de *hardware* de alto rendimiento. Por ejemplo, capacidad de memoria, capacidad de almacenamiento, velocidad de proceso. A todo ello se suma una arquitectura que permita recuperar el servicio ante posibles caídas. Las tomas de decisiones en las bolsas no paran. Los agentes ingresan a las ruedas en las bolsas de todo el mundo, por lo que el sistema nunca debería dejar de operar.

CONCLUSIONES

Tras la propuesta presentada, podemos afirmar que los tomadores de decisiones en las ruedas de bolsa tendrán mejores posibilidades de maximizar su ganancia según el set de información disponible. No obstante, los diversos agentes de la economía compiten entre sí en la toma de decisiones, por lo que la velocidad en la captura de datos y la generación de información que dé soporte a la toma de decisiones se vuelve crucial.

La competencia en la toma de decisiones se está manifestando en la cantidad de datos que se capturen, su procesamiento y la reducción del tiempo en que la información resultante se pone a disposición del tomador de decisiones.

Todos los agentes poseen los mismos algoritmos, pero no la misma capacidad de procesamiento de datos ni de acceso a las mismas fuentes. Por ello, los agentes que participan en las ruedas de bolsa se distinguirán unos de otros por las fuentes de información que utilicen y por su soporte tecnológico, pues este permite un tiempo de respuesta muy corto, lo que se conoce, de modo técnico, como *en tiempo real*³.

3. Se dice que se ha obtenido el tiempo real cuando el resultado del procesamiento de un servidor se origina desde lo que está ocurriendo en ese mismo momento; es decir, los tiempos podrían estar expresados en segundos o en unidades incluso más cortas.

También es posible aseverar que los tiempos en que se accede a la información son muy importantes. Puede ocurrir que la posibilidad de acceso a la información sea factible para todos los agentes involucrados; sin embargo, tendrá ventaja el que logre obtenerla primero. En otros términos, los agentes se distinguen por el acceso a las fuentes y la identificación oportuna de las mismas, así como por lograr el acceso a los datos en el momento adecuado.

Las fuentes de información, además, son dinámicas. Vale decir, las mismas cambian en cuanto a su procedencia y sus contenidos. Por ello, es de suma importancia poseer estrategias de selección y de validación de las fuentes. Hoy aparecen en internet muchas direcciones web, por lo que se debe tener muy claro cuáles presentan contenidos válidos y cuáles no lo hacen.

Por ejemplo, el análisis de los mensajes de Twitter en tiempo real podrá incrementar el set de información de los agentes tomadores de decisiones de manera sustancial, mejorando su certeza.

Finalmente, la inteligencia analítica y *big data* están avanzando a tal velocidad que muy pronto tendremos sistemas que tomen las decisiones de compra de valores, y que no sean solo proveedores de información para los agentes.

REFERENCIAS

- Covington, D. (2016). *Analytics: Data Science, Data Analysis and Predictive Analytics for Business* (5.a ed.). CreateSpace Independent Publishing Platform.
- Davenport, T. & Harris, J. (2007). *Competing on Analytics*. Harvard Business Review Press.
- Schroek, M., Shockley, R., Smart, J., Romero-Morales, D. & Tufano, P. (2012). *Analytics: el uso de big data en el mundo real* [Informe ejecutivo]. IBM Institute for Business Value y Escuela de Negocios Saïd de la Universidad de Oxford.