

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CONTADURÍA PÚBLICA Y ADMINISTRACIÓN



TESIS

**EFFECTOS DE BIG DATA Y DATA SCIENCE EN EL DESEMPEÑO FINANCIERO
DE LAS EMPRESAS QUE COTIZAN EN LAS BOLSAS DE VALORES DE LOS
ESTADOS UNIDOS**

PRESENTA

IVÁN VENZOR CÁRDENAS

**COMO REQUISITO PARA OBTENER EL GRADO ACADÉMICO DE DOCTOR
EN CONTADURÍA**

ENERO, 2019

DECLARACIÓN DE AUTENTICIDAD

Declaro solemnemente que el presente documento es fruto de mi propio trabajo y, hasta donde estoy enterado, no contiene material previamente publicado o escrito por otra persona, excepto aquellos materiales o ideas que por ser de otras personas les he dado el debido reconocimiento y los he citado debidamente en la bibliografía y referencias.

Declaro además que tampoco contiene material que haya sido aceptado para el otorgamiento de cualquier otro grado o diploma de alguna universidad o institución.

Para colaborar lo anterior, el presente trabajo fue analizado utilizando la herramienta *Turnitin* para la detección de plagio. Las únicas observaciones que tuvo la herramienta fueron los nombres de las empresas listadas en la sección de apéndices, el resto del documento estuvo libre de cualquier comentario.

Nombre y firma: Iván Venzor Cárdenas

DEDICATORIA

Para Mariel e Ivanna.

AGRADECIMIENTOS

Al Doctor Klender por su guía y apoyo, al Doctor Cruz y al Doctor Wong por ser parte de mi comité doctoral durante estos años.

Al Doctor Klender, al Doctor Cruz, al Doctor Wong, al Doctor Treviño y a la Doctora Araiza por sus revisiones y comentarios para lograr la versión final de la tesis.

A todos los doctores que directa o indirectamente hicieron este trabajo posible.

ABREVIATURAS y TERMINOS TECNICOS

Artificial Intelligence: inteligencia artificial, rama de las Ciencias Computacionales que busca automatizar tareas intelectuales que normalmente son desempeñadas por humanos.

Big Data: se refiere a los datos, que son muy grandes, que se generan muy rápido o que no son estructurados, así como también a las técnicas creadas para manipularlos ya que no pueden ser manejados eficientemente por bases de datos relacionales.

CAPEX: capital expenditures (inversiones en bienes de capitales). Son inversiones de capital que crean beneficios.

Data-driven decision making: se refiere a basar la toma de decisiones en datos en lugar de confiar únicamente en la intuición.

Data mining: minería de datos. Proceso que intenta descubrir patrones en grandes volúmenes de datos a través de machine learning, estadística y técnicas de bases de datos.

Data Science: consiste en la extracción generalizable de información a partir de datos y se caracteriza por su capacidad de predicción y por la generación de modelos a partir de muy pocas consideraciones iniciales.

Data Scientist: científico de datos, especialista que combina capacidades analíticas, conocimientos avanzados de estadística, machine learning y otros tópicos técnicos, con conocimientos profundos de negocio.

Deep Learning: rama de Machine Learning, se distingue por múltiples capas sucesivas de redes neuronales donde ocurre el aprendizaje.

EBITDA: Earnings before interests, taxes, depreciations and amortizations. Es el resultado empresarial antes de intereses, impuestos, depreciaciones y amortizaciones productivas.

EDGAR: Electronic Data Gathering, Analysis, and Retrieval. Plataforma de la SEC que contiene los estados financieros de las empresas que cotizan en las bolsas de valores de los Estados Unidos.

Hadoop: popular plataforma de Big Data.

Internet of Things (IoT): múltiples dispositivos y sensores están comunicados en tiempo real proporcionando información relevante.

IT: information technologies (tecnologías de la información).

IT business value: generación de valor de las tecnologías de la información en los negocios.

IS: information systems (sistemas de la información). A menudo tecnologías de la información y sistemas de la información se usan como sinónimos.

Machine learning: aprendizaje automático o aprendizaje de máquina, rama de la inteligencia artificial.

R&D: research and development (investigación y desarrollo).

Scopus: base de datos de artículos científicos.

SEC: Securities and Exchange Commission de Estados Unidos, equivalente a la Comisión Bancaria y de Valores en México.

TI: tecnologías de la información.

10-K: informe anual requerido por la SEC que brinda un resumen completo del desempeño financiero de una compañía.

TABLA DE CONTENIDO

DECLARACIÓN DE AUTENTICIDAD	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
ABREVIATURAS y TERMINOS TECNICOS	iv
TABLA DE CONTENIDO	v
INDICE DE TABLAS.....	vii
INDICE DE FIGURAS.....	viii
1. NATURALEZA Y DIMENSIÓN DEL ESTUDIO	1
1.1. <i>Introducción.....</i>	1
1.2. <i>Planteamiento del problema y pregunta de investigación</i>	9
1.3. <i>Importancia y justificación del estudio</i>	12
1.4. <i>Hipótesis de investigación</i>	14
1.5. <i>Objetivos de la investigación</i>	15
1.6. <i>Estructura y contenido.....</i>	15
2. BIG DATA Y DATA SCIENCE EN LOS NEGOCIOS	17
2.1. <i>Big Data y Data Science.....</i>	17
2.2. <i>Artificial Intelligence, Machine Learning y Deep Learning</i>	24
2.3. <i>Estatus del uso de Big Data y Data Science</i>	26
2.4. <i>Los sistemas de información y generación de valor</i>	30
3. LA Q DE TOBIN Y LOS INTANGIBLES	38
3.1. <i>Métricas contables y la generación de valor</i>	38
3.2. <i>Definición de la Q de Tobin.....</i>	39
3.3. <i>Método de cálculo de la Q de Tobin</i>	40
3.4. <i>Aproximación de Chung y Pruitt.....</i>	43
4. MODELO Y METODOLOGÍA.....	44
4.1. <i>Modelo propuesto</i>	45
4.2. <i>Diseño de la investigación</i>	47
4.3. <i>Población, marco muestral y muestra</i>	47
4.4. <i>Métodos de análisis.....</i>	48
5. RESULTADOS	50

5.1.	<i>Análisis de los datos</i>	50
5.2.	<i>Resultados del modelo</i>	54
5.3.	<i>Análisis por elementos clave</i>	56
5.4.	<i>Prueba de correlación espuria</i>	57
6.	CONCLUSIONES	59
6.1.	<i>Consideraciones y recomendaciones finales</i>	59
6.2.	<i>Líneas sugeridas de investigación</i>	63
	REFERENCIAS BIBLIOGRÁFICAS	65
	APENDICES	73

INDICE DE TABLAS

TABLA 1. <i>BÚSQUEDA RELEVANTE DE INFORMACIÓN AL PRIMER SEMESTRE DE 2015 EN SCOPUS</i>	11
TABLA 2. <i>OCURRENCIA DE LAS PALABRAS CLAVE</i>	51
TABLA 3. <i>ESTADÍSTICA DESCRIPTIVA</i>	52
TABLA 4. <i>COEFICIENTES DE LA REGRESIÓN CUANTÍLICA</i>	54
TABLA 5. <i>CA A LO LARGO DEL TIEMPO</i>	56
TABLA 6. <i>ANÁLISIS INDIVIDUAL DE LAS PALABRAS CLAVE</i>	57
TABLA 7. <i>LISTADO DE EMPRESAS ANALIZADAS</i>	73

INDICE DE FIGURAS

FIGURA 1. <i>LÍNEA DE TIEMPO DE LOS ANTECEDENTES HISTÓRICOS</i>	4
FIGURA 2. <i>CRECIMIENTO DEL UNIVERSO DIGITAL</i>	5
FIGURA 3. <i>POPULARIDAD DE BIG DATA</i>	6
FIGURA 4. <i>POPULARIDAD DE DATA SCIENTIST</i>	7
FIGURA 5. <i>POPULARIDAD ERP</i>	8
FIGURA 6. <i>MAPA CONCEPTUAL DEL PROBLEMA BAJO ESTUDIO</i>	10
FIGURA 7. <i>ARTIFICIAL INTELLIGENCE, MACHINE LEARNING Y DEEP LEARNING</i>	24
FIGURA 8. <i>MODELO DE DEHNING Y RICHARDSON</i>	32
FIGURA 9. <i>MODELO ORIENTADO A LA PRODUCCIÓN</i>	33
FIGURA 10. <i>MODELO ORIENTADO A LOS PROCESOS</i>	34
FIGURA 11. <i>MODELO BASADO EN RECURSOS</i>	35
FIGURA 12. <i>MODELO DE SCHRYEN</i>	36

1. NATURALEZA Y DIMENSIÓN DEL ESTUDIO

En este capítulo se fijan las bases del presente trabajo de investigación, ya que contiene los antecedentes históricos, el planteamiento del problema, la pregunta de investigación, la hipótesis y los objetivos.

1.1. Introducción

El modelo relacional de base de datos fue propuesto académicamente por Codd (1970), como respuesta a las limitaciones de las bases de datos¹ y de las aplicaciones de aquella época. En 1979 Oracle lanzó al mercado el primer sistema gestor de base de datos relacional comercial², el cual ganó amplia aceptación y la empresa se convirtió en la compañía líder del mercado de bases de datos³. Los beneficios y popularidad del modelo relacional, propiciaron que las aplicaciones empresariales fueran ahora desarrolladas bajo este paradigma. Por ejemplo, en 1992 SAP presentó su ERP (*Enterprise Resource Planning*) R/3⁴, que hacía uso exclusivo de bases de datos relacionales⁵.

Paulatinamente las aplicaciones incrementaron su complejidad, lo que propició que las organizaciones tuvieran un aumento en la cantidad de datos que almacenaban, por lo que fue necesario introducir nuevos elementos que permitieran facilitar su análisis. De esta manera, surge el concepto de *Data Warehouse* (almacén de datos), que consiste en

¹ Una base de datos es fundamentalmente un sistema computarizado que almacena y proporciona acceso a la información (Silberschatz, Stonebraker y Ullman, 1996).

² Para más información, ver Oracle (2007).

³ En el segmento de bases de datos relacionales Oracle es el líder con una participación de mercado del 48.3% (Graham, et al., 2013).

⁴ Para más información, visitar: <http://www.sap.com/corporate-en/about/our-company/history/1982-1991.html>

⁵ En el segmento de ERPs, SAP es el líder con una participación de mercado del 24% (Pang, et al., 2014).

copiar los datos transaccionales, estructurándolos de forma especial para facilitar su análisis (Kimball, 1996). Al respecto, en 1997 SAP lanzó al mercado el producto BIW (*Business Warehouse Information System*), enfocado al reporte, análisis y almacenamiento masivo de datos⁶.

Dado el continuo incremento en la generación de los datos, en 1998 Silicon Graphics (SGI) comenzó a utilizar el término *Big Data*. Por su parte, Weiss y Indurkha (1998) hicieron referencia al nuevo concepto de *Big Data* en el contexto del análisis de grandes *data warehouses* mediante minería de datos (*data mining*)⁷. Sin embargo, al popularizarse el internet y surgir empresas digitales, no solo aumentó el volumen de datos, sino también su variedad y su velocidad de generación. Para Laney (2001), esas tres características son elementos esenciales del paradigma *Big Data*.

Las tecnologías que componen *Big Data* surgieron a partir de los trabajos de Google en algoritmos de procesamiento distribuido en 2004⁸, ya que los elementos tecnológicos disponibles en aquel momento (tales como los *data warehouses* relacionales) no tenían las capacidades requeridas para el procesamiento de conjuntos de datos muy grandes. Al respecto, Jacobs (2009) afirma que el *data warehouse* es una solución lenta y limitada, y sostiene que para los conjuntos de datos realmente grandes es necesario abandonar el modelo relacional puro, ya que *Big Data* son aquellos datos que no pueden ser correctamente manejados por las capacidades de las bases de datos relacionales⁹.

⁶ Para más información, visitar:
<http://wiki.scn.sap.com/wiki/display/BI/Overview+of+SAP+History+and+BW-BI+Evolution>

⁷ Para más información, ver Diebold (2012).

⁸ Para más información, ver Tambe (2014).

⁹ En 2013 las bases de datos relacionales representaban todavía el 90.8% del ranking de popularidad de bases de datos. Para más información, visitar: http://db-engines.com/en/blog_post/23

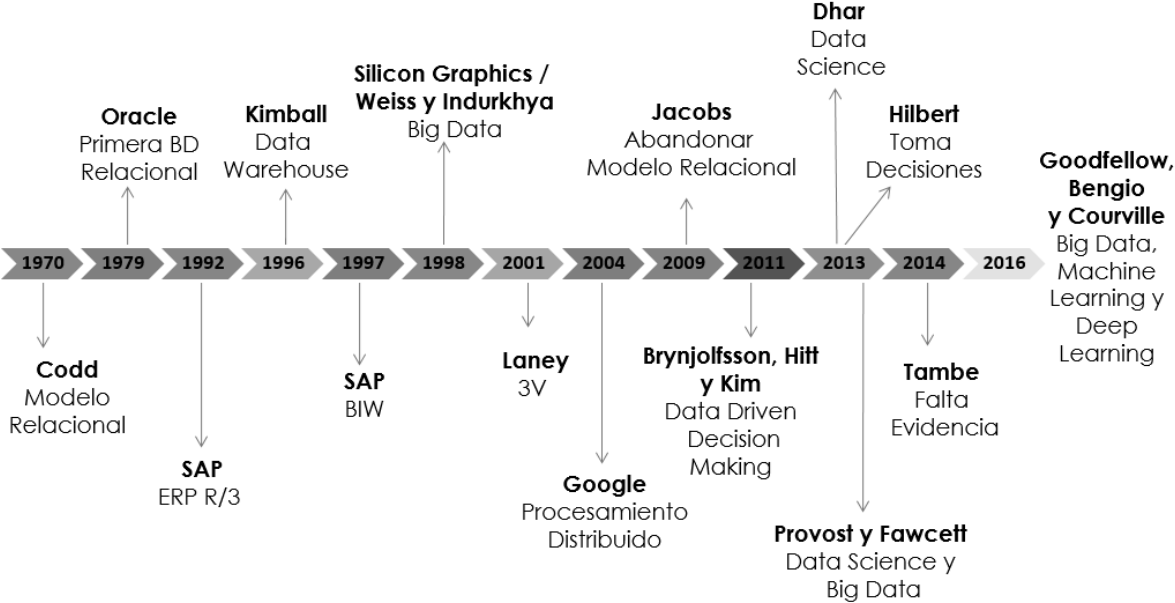
En el ámbito empresarial y financiero, Brynjolfsson, Hitt y Kim (2011), estudiaron los efectos de las tecnologías de la información y las prácticas organizacionales en empresas públicas de los Estados Unidos. Definieron *data-driven decision making* (toma de decisiones basadas en datos), como la práctica de basar la toma de decisiones en el análisis de los datos a través del uso de las tecnologías de la información. Encontraron que una desviación estándar mayor con respecto a la media en el uso de *data-driven decision making* está asociada con un incremento de 4-6% en las ventas de las empresas y dicha relación no parece ser provocada por causalidad inversa. El estudio anterior es importante porque el concepto de *Big Data* está íntimamente relacionado con *Data Science* (ciencia de datos) y *data-driven decision making* (Provost y Fawcett, 2013). *Data Science* es la extracción generalizable de conocimiento a partir de los datos (Dhar, 2013). De esta manera, *Big Data* y *Data Science* buscan mejorar la toma de decisiones en las empresas. Esto es, la característica esencial del paradigma *Big Data* es el análisis de los datos para una mejor toma de decisiones (Hilbert, 2016). Por ejemplo, en la industria financiera se utiliza *Data Science* para determinar en tiempo real las puntuaciones crediticias, así como la detección de fraudes en el momento en que los eventos suceden¹⁰.

Sin embargo, para Tambe (2014) la evidencia empírica de los beneficios de estas nuevas tecnologías de datos ha sido restringida a casos específicos de ciertas empresas. Tambe argumenta que hay un debate activo sobre si las tecnologías de *Big Data* han conducido a beneficios económicos en las empresas y sostiene que hay una necesidad de estudios que proporcionen evidencias a gran escala del impacto de estas inversiones en el desempeño de las organizaciones.

¹⁰ Para más información, ver Provost y Fawcett (2013).

La Figura 1 resume los principales antecedentes históricos que enmarcan el surgimiento y desarrollo de las tecnologías de *Big Data* y *Data Science*, desde la aparición del modelo relacional de base de datos, hasta el surgimiento de conceptos recientes como *Deep Learning*, *Machine Learning* y el resurgimiento de la Inteligencia Artificial.

Figura 1. Línea de tiempo de los antecedentes históricos

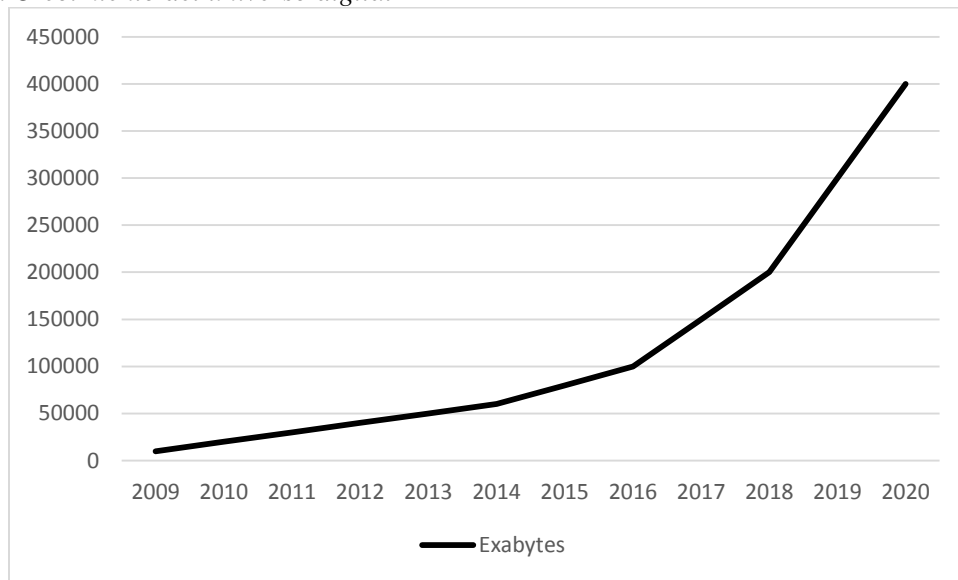


Fuente: Elaboración propia.

Después de establecer los principales antecedentes históricos, procedemos a examinar la popularidad de estos conceptos y sus aplicaciones. En la Figura 2 se observa que la cantidad de información en el mundo (universo digital) doblará su tamaño cada dos años y continuará esa tendencia al menos hasta el 2020 (IDC 2012; 2014). El crecimiento de los datos fue iniciado en el paso de la era analógica a la digital: generación de contenidos de alta definición, videos, *streaming*, redes sociales, sensores, *weblogs* y *webclicks*, sensores, dispositivos móviles, dispositivos biomédicos y, en general, de toda

la comunicación máquina a máquina mediante el Internet de las Cosas (*Internet of Things*).

Figura 2. *Crecimiento del universo digital*



Fuente: Elaboración propia con datos de IDC (2012).

El volumen de la información que está representado en la gráfica anterior está expresado en *exabytes*. Según IDC (2007), un *exabyte* está conformado por 1000 *petabytes*, un *petabyte* por 1000 *terabytes* y un *terabyte* por 1000 *gigabytes*. El *gigabyte* es probablemente la medida de almacenamiento más conocida hoy en día, ya que las capacidades de las memorias *usb*, los celulares y las tabletas se miden en *gigabytes* (por ejemplo, una memoria *usb* de 8 *gigabytes*).

Para tener contexto del tamaño del universo digital, en 2013 estaría contenido en la memoria de una pila de *ipads* extendida por dos tercios de la distancia de la tierra a la luna. Dicho de otra manera, hay aproximadamente la misma cantidad de bits en el universo digital que estrellas en el universo físico. Sin embargo, lo más relevante de los

estudios de IDC es que en 2012 solo el 0.5% del total de la información generada es analizada. Esto representa un área de oportunidad significativa. Tal como lo mencionan Mayer-Schönberger y Cukier (2013), debido a que la cantidad de información seguirá su ritmo de crecimiento en los próximos años, aquellas organizaciones que analicen y utilicen a su favor los datos, tendrán beneficios y ventajas competitivas sobre aquellas que no lo hagan. Esta es la promesa de *Big Data*, *Data Science* y otros términos y tecnologías afines.

Big Data y *Data Science* han despertado un creciente interés desde 2011 al día de hoy. Mediante *Google Trends* podemos ver la evolución en la popularidad de estos conceptos. Como comparativa también se incluye el *ERP*¹¹, el cual es un concepto maduro, conocido y bien establecido. Podemos visualizar esta comparativa en las Figuras 3, 4 y 5. El eje de las ordenadas (eje Y) representa el índice de popularidad de búsqueda proporcionado por Google.

Figura 3. Popularidad de Big Data



Fuente: Elaboración propia con datos de Google Trends.

¹¹ De acuerdo a Hillegrberg y Kumar (2000), los sistemas ERP (*Enterprise Resource Planning*) integran la información y los procesos dentro y a través de las áreas funcionales de una organización.

Para entender mejor el gráfico anterior, citamos a Google¹²:

Los números reflejan el interés de búsqueda en relación con el valor máximo de un gráfico en una región y un periodo determinados. Un valor de 100 indica la popularidad máxima de un término, mientras que 50 y 0 indican que un término es la mitad de popular en relación con el valor máximo o que no había suficientes datos del término, respectivamente.

Figura 4. Popularidad de Data Scientist

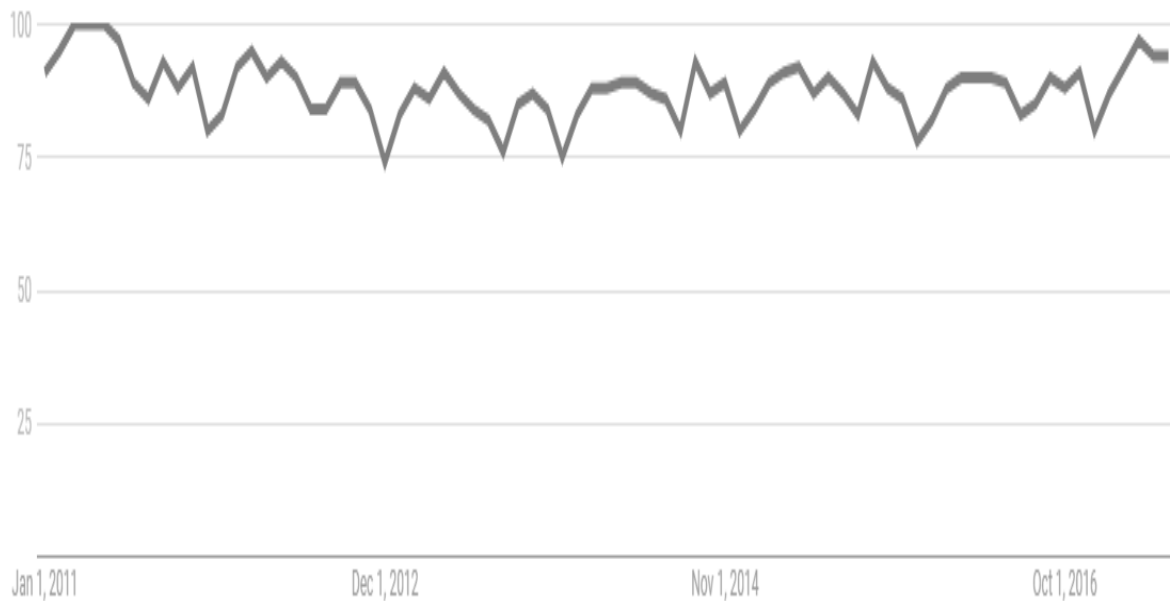


Fuente: Elaboración propia con datos de Google Trends.

Semejante a *Big Data*, *Data Science* también ha tenido un aumento vertiginoso en cuanto a su popularidad ya que ambos conceptos han incrementado notablemente su relevancia a partir del año 2011. Eso contrasta con la evolución en el interés de otros conceptos más maduros y establecidos, como en el caso del *ERP* (Enterprise Resource Planning).

¹² <https://support.google.com/trends/answer/4355164?hl=es&rd=1>

Figura 5. Popularidad *ERP*



Fuente: Elaboración propia con datos de Google Trends.

Podemos mencionar que, aunque *ERP* es un concepto muy arraigado y su uso es prácticamente indispensable en todas las empresas, su popularidad es estable a través de los años. Al comparar las figuras 3, 4 y 5, podemos ver que los conceptos de *Big Data* y *Data Science* han despertado un interés significativo en los últimos años. Sin embargo, a pesar del incremento de la popularidad de los conceptos de *Big Data* y *Data Science*, Andlinger¹³ encontró que en 2013 las bases de datos relacionales aún dominan el mercado de base de datos con una popularidad del 90.8%.

¹³ Para más información, visitar: http://db-engines.com/en/blog_post/23

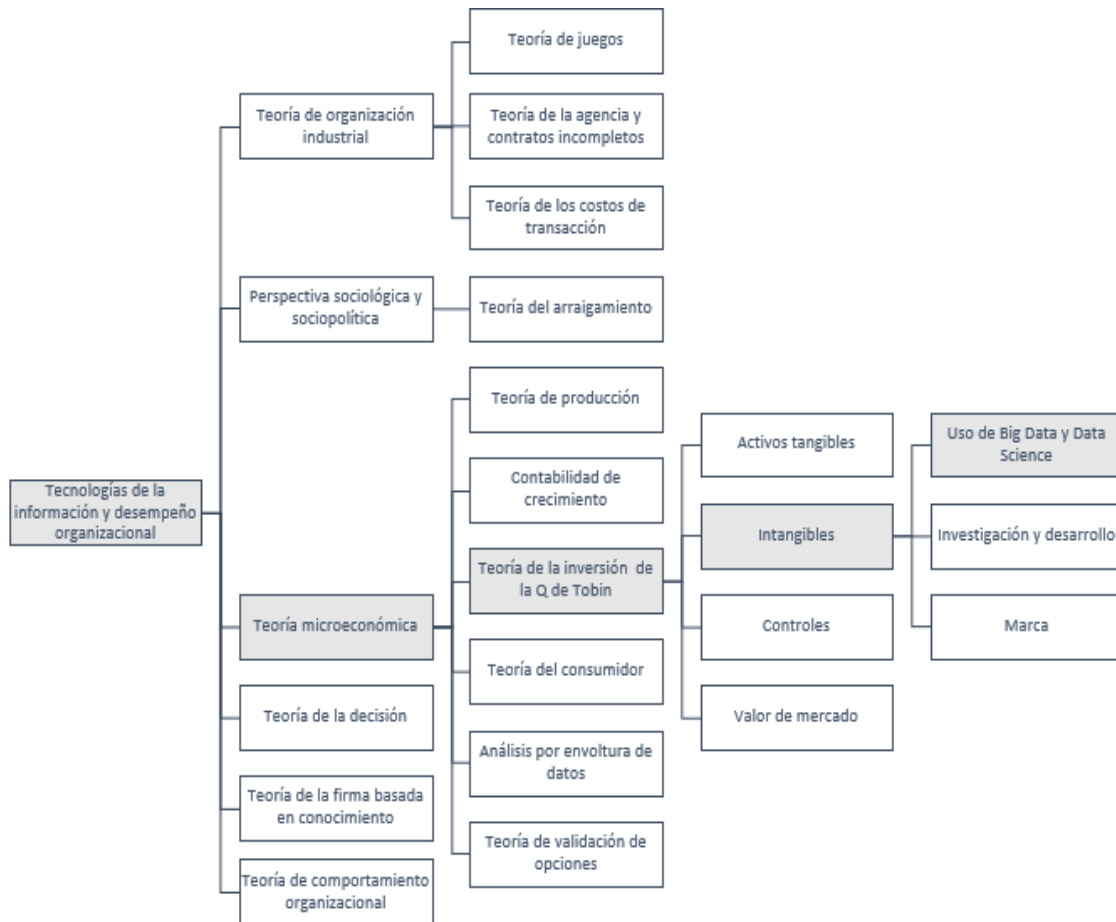
1.2. Planteamiento del problema y pregunta de investigación

Como se describió en los antecedentes, en los últimos años se ha presentado un aumento significativo en la cantidad de la información digital generada. Empero, resulta relevante y contrastante que tan solo una fracción muy pequeña de todos esos datos es analizada. *Big Data* y *Data Science* son herramientas y conceptos tecnológicos que han surgido recientemente como respuesta a las limitantes de las tecnologías tradicionales para la manipulación de estos datos.

La mayoría de la literatura académica aborda los temas de *Big Data* y *Data Science* desde el punto de vista de su definición conceptual, su popularidad y el impacto que estas tecnologías pueden tener para los individuos, las empresas y la sociedad en su conjunto. Sin embargo, aparte de casos de éxito en el ámbito empresarial encontrados en la literatura no académica, Tambe (2014) argumenta que faltan estudios formales que analicen las implicaciones de su uso en las compañías. Por lo tanto, la presente investigación pretende estudiar los efectos financieros de *Big Data* y *Data Science* en aquellas compañías que los utilizan, tratando con ello de responder a la falta de estudios anteriormente señalado. En la Figura 6 se expone el mapa conceptual realizado a través de la revisión previa de la literatura.

En la literatura se encontró que el efecto de las tecnologías de la información en el desempeño de las compañías ha sido analizado desde múltiples puntos de vista. Por ejemplo, desde la perspectiva sociológica y sociopolítica, teoría de organización industrial, teoría microeconómica, teoría de la decisión, teoría de la firma basada en conocimiento y teoría de comportamiento organizacional.

Figura 6. Mapa conceptual del problema bajo estudio



Fuente: Elaboración propia a partir de Melville, et al. (2004), Schryen (2013), Brynjolfsson, et al. (2011) y Saunders y Tambe (2014).

Abordaremos el estudio de los efectos de *Big Data* y *Data Science* en el desempeño financiero de las compañías a partir de la teoría microeconómica, específicamente mediante el valor de mercado de las empresas que cotizan en bolsa y la Q de Tobin (la razón entre el valor de mercado de una compañía y el valor de reposición, o contable, de los activos de la compañía). Las razones para utilizar estas métricas y el detalle de la Q de Tobin es explicado en el Capítulo 3.

Con el objetivo de determinar los trabajos existentes en esta área de investigación, se realizó una búsqueda en *Scopus*¹⁴, el cual es una base de datos de artículos académicos de Elsevier. El primer corte corresponde a la búsqueda de las palabras clave que aparezcan en el título, *abstract* o *keywords*, sin filtros adicionales. El segundo corte filtra únicamente artículos. El tercer corte filtra los documentos relativos a las ciencias sociales, negocios, administración, contabilidad, ciencias de la decisión, economía, econometría o finanzas. El cuarto corte es la intersección de los dos conjuntos, en este caso la intersección de los conjuntos del primer corte de ambas variables. Esto es, para nuestra primera variable se utilizan en la búsqueda de Scopus los términos de “*big data*” o “*data science*”. Para la segunda variable se utilizan los términos “*tobin*”, “*market value*”, “*IT value*” o “*IS value*”. El resultado de esta búsqueda se muestra en la Tabla 1.

Tabla 1. *Búsqueda relevante de información al primer semestre de 2015 en Scopus*

Palabra Clave	Corte 1	Corte 2	Corte 3	Corte 4
Variable de Interés: <i>big data</i> o <i>data Science</i>	9,194 ¹⁵	2,767 ¹⁶	700 ¹⁷	NA
Variable Y: <i>tobin</i> , <i>market value</i> , <i>IT value</i> o <i>IS value</i>	9,056 ¹⁸	6,352 ¹⁹	3,713 ²⁰	2 ²¹

Fuente: Elaboración propia con información de Scopus (www.scopus.com)

¹⁴ <http://www.scopus.com/>

¹⁵ (TITLE-ABS-KEY ("big data") OR TITLE-ABS-KEY ("data science"))

¹⁶ (TITLE-ABS-KEY ("big data") OR TITLE-ABS-KEY ("data science")) AND (LIMIT-TO (DOCTYPE , "ar"))

¹⁷ (TITLE-ABS-KEY ("big data") OR TITLE-ABS-KEY ("data science")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "ECON"))

¹⁸ (TITLE-ABS-KEY ("tobin") OR TITLE-ABS-KEY ("market value") OR TITLE-ABS-KEY ("IT value") OR TITLE-ABS-KEY ("IS value"))

¹⁹ (TITLE-ABS-KEY ("tobin") OR TITLE-ABS-KEY ("market value") OR TITLE-ABS-KEY ("IT value") OR TITLE-ABS-KEY ("IS value")) AND (LIMIT-TO (DOCTYPE , "ar"))

²⁰ (TITLE-ABS-KEY ("tobin") OR TITLE-ABS-KEY ("market value") OR TITLE-ABS-KEY ("IT value") OR TITLE-ABS-KEY ("IS value")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "DECI"))

²¹ (TITLE-ABS-KEY ("tobin") OR TITLE-ABS-KEY ("market value") OR TITLE-ABS-KEY ("IT value") OR TITLE-ABS-KEY ("IS value")) AND (TITLE-ABS-KEY ("big data") OR TITLE-ABS-KEY ("data science"))

En el último corte donde se relacionan ambas variables, solo hay dos artículos que cumplen con las condiciones de búsqueda, los cuales fueron revisados para determinar su pertinencia hacia el tema de estudio. Esto es, aunque ciertos aspectos de las tecnologías de la información ya han sido estudiados, no se encontraron artículos académicos que aborden los efectos de *Big Data* y *Data Science* en el desempeño financiero de las empresas.

A partir de la revisión de los antecedentes y del planteamiento del problema, concebimos la siguiente pregunta central de investigación:

¿Cuál es el impacto que tiene el uso de *Big Data* y *Data Science* en el valor de mercado de las empresas que cotizan en las Bolsas de Valores de los Estados Unidos expresado mediante la Q de Tobin?

El resto de la tesis estará enfocado a proveer lo necesario para lograr responder a la pregunta de investigación que hemos planteado.

1.3. Importancia y justificación del estudio

Mediante un estudio de las prácticas organizacionales y del uso de las tecnologías de información de las compañías que cotizan en las Bolsas de Valores en los Estados Unidos, Brynjolfsson, et al. (2011), encontraron que una desviación estándar mayor respecto a la media en el uso de la toma de decisiones basadas en datos, está asociada con un incremento de 4-6% en el desempeño de las compañías. Esto es, recopilaron

información interna de las empresas mediante encuestas a altos ejecutivos y también utilizaron información financiera pública, para determinar la contribución marginal de la variable de interés. Posteriormente, Provost y Fawcett (2013) afirmaron que la toma de decisiones basadas en datos está íntimamente relacionada con *Big Data* y *Data Science*, mientras que Tambe (2014) argumenta que existe una necesidad de estudios con evidencia a gran escala del impacto de estas inversiones en el desempeño de las organizaciones.

De esta manera, esta investigación pretende responder a la falta de estudios señalada por de Tambe (2014), al medir los efectos de *Big Data* y *Data Science* en el desempeño financiero de las empresas que cotizan en las Bolsas de Valores de los Estados Unidos, mediante el uso de información financiera pública. Aunque el estudio aborda conceptos tecnológicos, el enfoque está dirigido hacia los efectos financieros derivados de ellos.

Esta investigación es relevante por las siguientes razones:

- 1) Concluir si *Big Data* y *Data Science* ofrecen ventajas competitivas a las empresas que los utilizan.
- 2) Al establecer el efecto esperado de estas tecnologías, los tomadores de decisiones en las organizaciones pueden emitir mejores juicios relativos a las inversiones en estos conceptos o a sus posibles beneficios al implementarlos.
- 3) Por último, se responde a la falta de evidencia señalada por Tambe (2014).

1.4. Hipótesis de investigación

Estudios anteriores han investigado acerca de *Big Data* y *Data Science* (Provost y Fawcett, 2013) o de la Q de Tobin en el contexto de las tecnologías de la información y el valor de mercado (Brynjolfsson et. al., 2011; Saunders y Tambe, 2014), pero no de ambos temas como parte de una pregunta o hipótesis conjunta de investigación. El presente estudio pretende precisamente investigar ambos conceptos integrados.

Por lo anterior y, siguiendo con lo expuesto en la pregunta central de investigación, formulamos las siguientes hipótesis:

H₀: No existe una relación significativa entre el valor de mercado de una empresa expresado mediante la Q de Tobin y el uso de *Big Data* y *Data Science*.

H_a: Existe una relación significativa entre el valor de mercado de una empresa expresado mediante la Q de Tobin y el uso de *Big Data* y *Data Science*.

La variable dependiente es la Q de Tobin, mientras que nuestra variable de interés es dicotómica y se construye a partir de la ocurrencia de las palabras clave relacionadas a *Big Data* y *Data Science* contenidas en los estados financieros de las empresas. La construcción de la variable de interés se describe en el Capítulo 4. Se incluyen, además, variables de control como la razón de los activos intangibles y el total de activos, el total de activos, la razón de los pasivos y el total de activos, la razón de inversiones de capital (CAPEX) y total de activos y, por último, la razón de los beneficios antes de intereses, impuestos, depreciaciones y amortizaciones (EBITDA) y el total de activos.

1.5. Objetivos de la investigación

A partir del planteamiento del problema, de la pregunta central de investigación y de las hipótesis planteadas, se desprenden los siguientes objetivos.

Objetivo general: establecer el impacto del uso de las tecnologías de la información basadas en *Big Data* y *Data Science* en el valor de mercado, expresado mediante la Q de Tobin, de las empresas públicas de los Estados Unidos.

Objetivos específicos:

- 1) Documentar el grado en el que se utilizan de *Big Data* y *Data Science* en las empresas que cotizan en las Bolsas de Valores de los Estados Unidos.
- 2) Derivar el efecto específico de *Big Data* y *Data Science* en las empresas de acuerdo a su tamaño.
- 3) Especificar la duración del impacto de *Big Data* y *Data Science* en el tiempo.
- 4) Contrastar el efecto de los principales componentes de *Big Data* y *Data Science* en la Q de Tobin.

1.6. Estructura y contenido

La tesis tiene la siguiente estructura. El Capítulo 1 es introducción, donde se plantearon los antecedentes, el planteamiento del problema y la pregunta central de investigación, la importancia y justificación del estudio, la hipótesis de investigación, los objetivos y finalmente una descripción de la estructura de los demás capítulos. El Capítulo 2, *Big*

Data y Data Science en los negocios, contiene la narrativa de la popularización de estos conceptos y otros conceptos afines, tales como *Artificial Intelligence* y *Machine Learning*, así como el estatus el uso de estos conceptos en la industria. En el Capítulo 2 adicionalmente se detalla como los sistemas de información generan valor en las compañías. El Capítulo 3, la Q de Tobin y los intangibles, contiene un recuento de las métricas contables, la definición de la Q de Tobin, su método de cálculo y una aproximación que se va a utilizar para su cálculo. El Capítulo 4, modelo y metodología, contiene el modelo propuesto, así como la metodología utilizada. El Capítulo 5, resultados, expone los resultados obtenidos. El Capítulo 6, conclusiones, contiene las conclusiones obtenidas en esta investigación, así como líneas futuras sugeridas de investigación. Por último, hay dos apartados adicionales que contienen los apéndices y las referencias bibliográficas utilizadas.

2. BIG DATA Y DATA SCIENCE EN LOS NEGOCIOS

En este capítulo se revisa la literatura de *Big Data* y *Data Science* en el contexto de los negocios y empresas.

2.1. Big Data y Data Science

Es difícil hoy en día leer una publicación popular y no encontrar referencias a *Big Data*, *Data Science*, *Analytics* o *Data-Driven Decision Making*. Tal como lo mencionan Provost y Fawcett (2013) y Agarwal y Dhar (2014), estos conceptos están tan relacionados unos con otros y son tan relativamente recientes, que las fronteras y definiciones de cada uno de ellos no están bien establecidas o es común que se utilice indistintamente cualquiera de los términos como si fueran sinónimos.

Antes de continuar, hay que recapitular cómo estos conceptos llegaron a ser tan populares. En 2003 se publicó el libro *Moneyball*²², que fue la base para la película estelarizada por Brad Pitt en 2011. *Moneyball* describe como un equipo de beisbol de la liga profesional de los Estados Unidos, utilizó el poder de modelos analíticos para poder competir con otros equipos con presupuestos mucho más altos. En 2006, Davenport publica en *Harvard Business Review* el artículo *Competing on Analytics*, que fue el más popular de ese año²³ y derivó en un libro en 2007²⁴. Otro hito fue el trabajo de Manyika et al. (2011) de McKinsey, acerca de *Big Data* y sus beneficios en las empresas. En 2012, Davenport y Patil publicaron en *Harvard Business Review* el artículo *Data Scientist: The Sexiest Job of the 21st Century*²⁵. Mientras que en 2015, el gobierno de los Estados

²² Para más información, ver: Lewis, M. (2003).

²³ Para más información, visitar: <http://www.informationweek.com/software/information-management/analytics-at-work-qanda-with-tom-davenport/d/d-id/1085869?>

²⁴ Para más información, ver: Davenport, T. H. y Harris, J. G. (2007).

²⁵ Para más información, ver: Davenport T.H. y Patil D.J. (2012).

Unidos nombró a Patil como el primer *Chief Data Scientist*²⁶. Sin duda, los eventos anteriormente descritos fueron importantes para ayudar a popularizar los conceptos que a continuación serán analizados.

Resulta oportuno preguntarse, ¿qué es en realidad *Big Data* y por qué se llama así? Como se ha discutido anteriormente, es innegable que nos encontramos en el periodo más frenético de generación de información digital de la historia humana, por lo que los términos de “big data” o “grandes datos” parecen hacer sentido. Sin embargo, coincidimos con Davenport (2014) cuando afirma que el término *Big Data* es un nombre muy limitado para un fenómeno mucho más amplio. Lo que hoy son datos muy grandes el día de mañana no lo serán, debido a la evolución constante del poder de cómputo. Por ejemplo, hasta hace unos años, las computadoras personales tenían lectores de discos flexibles de 3.5 pulgadas, que podían almacenar un poco más de 1 *mega byte* de información. Hoy en día una memoria USB “pequeña” de 1 *gigabyte*, tiene 1000 veces más capacidad de almacenamiento que un disco flexible de 3.5 pulgadas. Así que lo que al día de hoy es relativamente grande, el día de mañana seguramente no lo será. Por ello, el nombre y el uso del término *Big Data* es un tanto desafortunado.

Para Jacobs (2009) y Dumbill (2013), *Big Data* es un término que engloba a todos aquellos datos que no pueden ser correctamente manejados mediante las bases de datos relacionales. El tamaño es solo uno de los factores. Si bien es cierto que la explosión de datos ha generado en técnicas novedosas para su manejo, hay otros elementos que tampoco son correctamente manejados por las bases de datos tradicionales. Laney (2001) fue el primero en establecer las tres ‘V’ de *Big Data*:

²⁶ Para más información, visitar: <https://www.whitehouse.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist>

- 1) Volumen: esta característica es la más conocida ya que es la relativa al tamaño de los datos. Como se ha visto anteriormente, el tamaño y el volumen en general de información ha aumentado considerablemente.

- 2) Velocidad: se refiere a la tasa en la que los datos son generados. No necesariamente los datos son grandes, pero son generados y deben procesarse en tiempo real. Por ejemplo, los datos producidos por sensores pueden generar decenas o cientos de observaciones por segundo, que quizá en su conjunto no sean muy grandes, pero deben ser procesadas y analizadas en tiempo real.

- 3) Variedad: se refiere a la falta de estructura de las nuevas fuentes de datos que no fácilmente es representada mediante filas y columnas, o visto de otra manera, que no puede ser almacenada y procesada eficientemente en formatos tabulares semejantes al excel. Por ejemplo: textos y códigos de las páginas web, videos, imágenes, audios, geocalizadores, *tweets*, mensajes de *facebook*, etc. Para Davenport (2014), la falta de estructura de los datos es el aspecto más complicado de *Big Data*.

Kitchin (2014) detalla las características de los datos de *Big Data*. Sugiere que son enormes en volumen (medidos en *petabytes*), tienen alta velocidad (creados en tiempo real), variables (pueden ser estructurados o no estructurados y comúnmente referenciados en forma temporal o espacial), exhaustivos (intenta capturar todos los individuos o elementos de un sistema o cuando menos capturar muestras grandes de poblaciones), relacionables por naturaleza (se pueden unir con otros conjuntos de datos),

flexibles y escalables (puede incorporar nuevos elementos y expandirse rápidamente de tamaño).

Los orígenes de las tecnologías generadas para manejar *Big Data* se remontan a los trabajos en procesamiento distribuido en Google en 2004 (Tambe; 2014). No es el objetivo de este trabajo entrar al detalle de la parte técnica de los sistemas de información, puesto que es un doctorado en Contaduría, pero las tecnologías *Big Data* permiten a las empresas extraer información a partir de datos muy grandes, casi en tiempo real. Por ejemplo, Morgan Stanley ha usado *Big Data* para determinar en tiempo real como los eventos del mercado financiero afecta la actividad de su sitio web al examinar los *logs*²⁷, actividad que en el pasado podía tomar meses (Groenfeldt; 2012).

Una de las herramientas más utilizadas en el paradigma *Big Data* es *Hadoop*, el cual nace en Yahoo a partir de las ideas de los trabajos de Ghemawat et al. (2003), así como Dean y Ghemawat (2004). El objetivo original de esta herramienta era ayudar a Yahoo a mejorar el procesamiento de la indexación de las páginas web para con ello mejorar el servicio de búsqueda hacia los usuarios. *Hadoop* posibilitó el procesamiento eficiente de grandes volúmenes de datos mediante la repartición de tareas en múltiples servidores y posteriormente agregar los resultados. Este paradigma de divide y vencerás, es aplicable a múltiples casos de uso. Actualmente *Hadoop* es la herramienta de *Big Data* más conocida y en torno a ella gira todo un ecosistema de otras herramientas. Según lo encontrado por Zion Market Research, el mercado global de *Hadoop* en 2016 fue valorado aproximadamente en 7 mil de millones de dólares y esperan que siga creciendo hasta llegar a 87 mil de millones de dólares en 2022.

²⁷ Los *logs* son archivos de registro de las aplicaciones y sistemas de cómputo donde se registran eventos específicos. Por ejemplo, cada *click* que hacemos en una página web se almacena, junto con otra información, en los archivos de *logs* del portal que estamos visitando.

Sin embargo, en este punto donde es necesario establecer las diferencias entre los conceptos. Tal como mencionan Provost y Fawcett (2013), estrictamente hablando *Big Data* se encarga del manejo, almacenamiento y procesamiento de los datos, mientras que *Data Science* se encarga de la extracción de conocimiento. Esto es, *Data Science* puede utilizar *Big Data* como una entrada de datos. *Big Data* también puede utilizarse para temas de procesamiento de datos y transacciones donde implícitamente no haya una extracción de conocimiento en dichos procesos, sino que únicamente se trate de una mejora operativa o de procesamiento.

Se establece entonces la siguiente relación:

Big Data -> Data Science -> Data Driven Decision Making

Entonces, habiendo definido que *Big Data* es un medio y no un fin por si mismo, examinaremos ahora lo relativo a *Data Science*. Para Dhar (2013), *Data Science* es el estudio de la extracción generalizable de información a partir de datos. Quizá el concepto más relacionado a *Data Science* es *Data Mining* (minería de datos), que nace en la década de los 90, el cual utiliza técnicas de *Machine Learning* con fines de predicción y explicación. Estos métodos detectan relaciones en los datos sin tener que establecer previamente supuestos acerca de las relaciones entre variables, tales como linealidad o parámetros de distribución. Lo anterior es muy relevante porque en lugar de recopilar datos para probar alguna teoría, ahora a partir de conjuntos grandes de datos estos métodos pueden producir modelos predictivos de forma automática, lo que para Popper (1963), es un criterio epistemológico clave para determinar que tan seriamente debemos tomar una teoría o un nuevo conocimiento.

Esta característica de predicción dio lugar a un debate académico muy interesante iniciado por la publicación de un artículo en una revista de divulgación. Anderson²⁸, expuso en 2008 que, dada esa capacidad de crear modelos predictivos, el método científico es ahora obsoleto. Anderson argumenta que realmente no es relevante saber por qué las personas hacen lo que hacen a través de modelos imperfectos, sino lo relevante es tener la capacidad de predecir lo que hacen las personas, aunque no entendamos del todo la razón. Por ejemplo, Google, Amazon o Netflix pueden predecir los gustos, hábitos o compras de las personas, sin entender realmente la causa raíz. Para Anderson la correlación es más importante que la causalidad y la ciencia puede avanzar sin modelos coherentes, teorías unificadas o ninguna explicación mecanicista en absoluto. Para Kitchin y Lauriault (2015) este punto de vista es un nuevo tipo de empirismo donde los datos hablan por si mismos libres de teorías. Por supuesto, semejante postura desencadenó múltiples respuestas, por ejemplo, Boyd y Crawford (2012) y González-Bailón (2013).

No es el objetivo del presente trabajo discutir la postura epistemológica de unos y otros, pero coincidimos con Agarwal y Dhar (2014) cuando afirman que la predicción como base para la construcción de teorías es muy relevante en un mundo donde los patrones emergen antes de comprender las razones por las cuales lo hacen. Por ejemplo, un estudio que busque predecir en lugar de explicar, puede revelar asociaciones entre variables que formen la base para el desarrollo de una teoría que posteriormente pueda ser objeto de pruebas exhaustivas. Shmueli (2010) proporciona un resumen donde el modelaje predictivo puede ser útil en el mundo científico.

En este mismo sentido Han et. al (2011), exponen que para grandes cantidades de datos se han optimizado y adaptado técnicas estadísticas como regresiones, construcción

²⁸ Para más información, visitar: http://www.edge.org/3rd_culture/anderson08/anderson08_index.html

de modelos, visualización de datos y mapas, así como el uso de nuevas técnicas de *Machine Learning* y visualización analítica que pueden detectar, clasificar y segmentar patrones, relaciones, asociaciones y comportamiento entre variables y producir modelos predictivos, de simulación y de optimización. Siegel (2013) explica que el proceso de construir modelos empieza con un modelo sencillo que se modifica aplicando reglas de aprendizaje, como si fueran mutaciones genéticas, hasta que evoluciona en un modelo robusto. Con *Machine Learning*, cientos de modelos utilizando diversas técnicas estadísticas como regresión, redes neuronales, vecino más cercano, análisis de factores y árboles de decisión, pueden ser aplicados a un conjunto de datos para determinar cuál de ellos explica el fenómeno de mejor manera o se desempeña mejor.

Retomando a Provost y Fawcett (2013), *Data Science* es mucho más que algoritmos de *data mining*. Los problemas de negocio se ven a través de una perspectiva de datos, donde los principios fundamentales del análisis causal deben ser entendidos. La estadística es elemental para *Data Science*, así como los métodos para la visualización de datos.

Concordamos con Hilbert (2016) cuando argumenta que la característica principal de estos nuevos paradigmas consiste en que el tratamiento analítico de los datos es sistemáticamente colocado al frente de la toma de decisiones. El fin último de *Data Science* consiste en mejorar la toma de decisiones.

En este sentido, Brynjolfsson et al. (2011), definieron *data-driven decision making* como la práctica de basar las decisiones en el análisis de datos apoyándose en las tecnologías de la información. Demostraron que aquellas empresas que más utilizan los datos son más productivas. También encontraron que *data-driven decision making* está

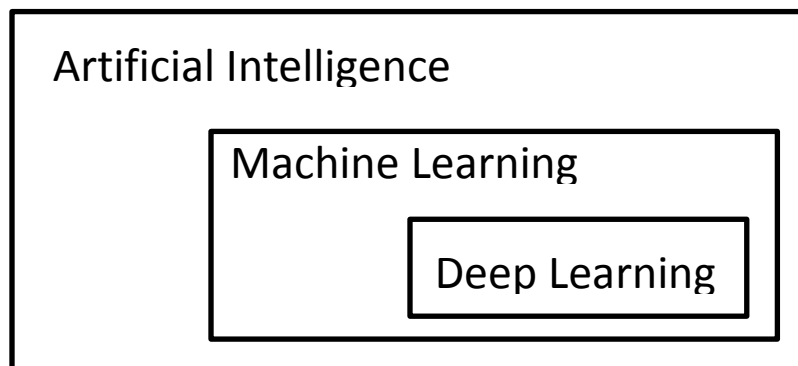
correlacionada con un mayor rendimiento en activos, retorno de capital, utilización de activos y valor de mercado, y dicha relación parece ser causal.

Por último, cabe aclarar que, aunque en el presente trabajo se ha definido lo que es *Big Data* y *Data Science*, en la mayoría de las publicaciones no académicas, con frecuencia se utiliza *Big Data* para referirse también a *Data Science*.

2.2. Artificial Intelligence, Machine Learning y Deep Learning

Existen otros términos que están muy ligados a *Big Data* y *Data Science*: *Machine Learning* y *Deep Learning*. *Machine Learning* pertenece al campo de la Inteligencia Artificial (*Artificial Intelligence*), mientras que a su vez *Deep Learning* es un subcampo de *Machine Learning* (Goodfellow, Bengio y Courville; 2016). Podemos ver esta relación en la Figura 7.

Figura 7. *Artificial Intelligence, Machine Learning y Deep Learning*



Fuente: Elaboración propia a partir de Chollet (2017)

Para Chollet (2017), el campo de la Inteligencia Artificial inicia en 1950 y busca automatizar tareas intelectuales que normalmente son desempeñadas por humanos. Inteligencia Artificial engloba a *Machine Learning* y *Deep Learning*, pero además contiene

otros campos que no tienen que ver con aprendizaje. Por ejemplo, cuando Deep Blue derrotó al campeón mundial de ajedrez Gary Kasparov en 1997 no tuvo intrínsecamente un proceso de aprendizaje, ya que solo realizó búsquedas en el espacio de soluciones establecido por el conjunto de reglas formales que describen completamente al ajedrez. A este tipo de Inteligencia Artificial se le llama simbólica.

Precisando a Goodfellow, Bengio y Courville (2016), en los primeros días de la Inteligencia Artificial se resolvieron problemas que eran intelectualmente difíciles para seres humanos, pero relativamente directos para las computadoras: procesos que pueden ser completamente descritos por una serie de reglas matemáticas formales, como el ajedrez. Sin embargo, el reto para la Inteligencia Artificial resultó ser aquellos problemas que son fáciles para las personas pero que son difíciles de describir formalmente, como por ejemplo reconocer rostros en imágenes, reconocer palabras habladas, traducir lenguajes, etc. Estos problemas que para nosotros son intuitivos, apenas a partir del año 2010, pudieron ser manejados correctamente por las computadoras mediante el aprendizaje a través de la experiencia en términos de una jerarquía de conceptos, donde cada concepto se define en términos de su relación con otros conceptos más sencillos.

Aunque el campo de *Machine Learning* empezó a florecer en la década de los noventa, fue hasta la llegada de *Big Data* que hubo información suficiente para poder entrenar correctamente a los algoritmos (Chollet; 2017). Así nace el campo conocido como *Deep Learning* y actualmente es el más popular y exitoso dentro de la Inteligencia Artificial.

Algunos de los principales logros que ha tenido *Deep Learning* son:

- Clasificación de imágenes a un nivel casi humano.
- Reconocimiento del habla a un nivel casi humano.

- Transcribir escritura manual a un nivel casi humano.

Además de su relación con *Big Data*, estos campos están íntimamente ligados con *Data Science*. *Machine Learning* es parte central de *Data Science* y la define como una inducción estadística enfocada a generar modelos predictivos robustos (Dhar; 2013). Estos métodos detectan la estructura en los datos de manera relativamente sencilla sin tener que hacer supuestos importantes respecto a linealidad, monotonía, parámetros o distribuciones.

2.3. Estatus del uso de Big Data y Data Science

Se analizaron diversos estudios no académicos, principalmente provenientes de firmas de consultoría de tecnologías de la información. Para Wagner et al. (2014) de *The Boston Consulting Group*, los líderes en el uso de *Big Data* generan 12% más ingresos que aquellos que no lo usan. Dado la habilidad de *Big Data* de generar nuevos productos, mercados y flujos de ingresos, se esperaba que *Big Data* estuviera situada dentro de los programas de innovación de las empresas. Sin embargo, 75% de los encuestados mencionaron que en sus empresas *Big Data* no forma parte de los programas de innovación. La industria de software es la única donde más del 50% considera que *Big Data* tendrá un impacto significativo en innovación. Además de las industrias de tecnología, la industria de las Aseguradoras es la que más importancia le da a *Big Data*.

General Electric y Accenture (2015), sostienen que en *Big Data* los datos generados por equipos industriales tienen más valor potencial al negocio, en comparación a otros tipos de datos asociados a la web social, el internet del consumidor y otras fuentes. Bajo esta premisa, *Big Data* tiene especial impacto en los sectores de aeronáutica, gas y petróleo, transportación, generación de energía y distribución,

manufactura y minería. De las industrias mencionadas, más del 80% mencionó que las iniciativas de analítica mediante *Big Data* se encuentran entre las tres prioridades más importantes en esa industria. Especialmente en aeronáutica, el 61% menciona que la analítica es su prioridad más importante. El 73% de las compañías ya invierten más del 20% de su presupuesto de tecnologías de la información en *Big Data*. El 84% opina que la analítica basada en *Big Data* tiene el potencial para cambiar el panorama competitivo en la industria en un año o menos. El 89% opina que aquellos que no inviertan en *Big Data* en el próximo año pueden perder participación de mercado. De hecho, la segunda consecuencia más mencionada sobre el fallo en la implementación de *Big Data*, es que los inversionistas perderán la confianza en la habilidad de la compañía para crecer.

Accenture (2014) encontró que las empresas grandes, con ingresos mayores a mil millones de dólares, opinan que *Big Data* es importante y para un 67% de ellas es extremadamente importante. El 89% opina que *Big Data* revolucionará los negocios como lo hizo el Internet, mientras que el 85% espera que cambie dramáticamente la manera en la que hacen negocios. El 79% indica que aquellos que no utilicen *Big Data* perderán ventaja competitiva. Aunque el 83% ya han comenzado con algún proyecto de *Big Data*, el 91% incrementará sus capacidades de *Data Science* en el corto plazo, mientras que el 69% de las empresas grandes lo hará en un año o menos.

Para Manyika, et al. (2011), *McKinsey Global Institute*, los sectores de cómputo, productos electrónicos e informáticos son los mejor posicionados para beneficiarse del uso de *Big Data*. Los sectores financieros, de seguros y de gobierno pueden beneficiarse ampliamente del uso de *Big Data* si las barreras a su uso son superadas.

McKinsey Center for Business Technology (2012), expone que, en salud, servicios de gobierno, *retail* y manufactura, *Big Data* puede mejorar la productividad entre un 0.5 y un 1% anualmente.

El *Centre for Economics and Business Research* (2012), realizó un estudio en el Reino Unido y encontraron que en 2011 la adopción de *Big Data* era de 34%. Las industrias que más lo utilizaban eran la banca de inversión, aseguramiento, telecomunicaciones y manufactura. Estimaban que en 2015 la adopción sea de un 52%.

En un estudio realizado por *Tata Consultancy Services* (2013), se expone que más del 50% de las empresas poseen iniciativas de *Big Data*. Sin embargo, hay una polarización en cuanto al gasto. Las que más lo hacen son las de telecomunicaciones, de viajes (incluye hotelería y aviación), alta tecnología y los sectores bancarios y financieros. Las que menos gastan son ciencias de la salud, *retail* y las de energía. Solo un tercio de las compañías de los Estados Unidos no tienen alguna iniciativa de *Big Data* y el 80% de las compañías reporta mejoras en la toma de decisiones.

Datameer (2014), plantea que las industrias de servicios financieros, tecnología, comunicaciones y *retail*, son las que más usan *Big Data* y que el 56% de las empresas en norteamérica ya han invertido o están invirtiendo en *Big Data*.

Tech Pro Research (2015), indica que el 30% de las empresas ha iniciado al menos un proyecto de *Big Data*. Los sectores líderes son telecomunicaciones, construcción e ingeniería, gobierno, tecnología de la información, así como finanzas, la banca y las aseguradoras.

Por último, Devlin et al. (2012) de *Enterprise Management Associates*, encontraron que el 36% de los encuestados ya se encuentra operando con algún proyecto de *Big Data*

y que las compañías con más de 5000 empleados son las que más adoptan la analítica basada en *Big Data*. Las industrias de medios de información, *retail*, industriales, finanzas y manufactura son las que cuentan con más proyectos en operación.

Los estudios provenientes de la industria muestran resultados interesantes, ya que fueron elaborados por distintas compañías en distintos años y muestran información que difiere entre los diferentes estudios, sobre todo en relación al porcentaje de empresas que realmente usa *Big Data*. Sin embargo, hay elementos donde si hay cierta concordancia. En primer lugar, todos indican un grado importante de uso e interés en las tecnologías *Big Data*. Todos los estudios también mostraron beneficios cualitativos, aunque muy pocos establecieron beneficios cuantitativos. Ciertos sectores fueron mencionados con frecuencia como aquellos con mayor grado de uso: telecomunicaciones, banca, finanzas, aseguradoras, *retail* y las de tecnología o tecnologías de la información. También, algunos estudios indican que aquellas empresas consideradas grandes ya sea en ventas o en empleados parecen tener más disposición hacia el uso de *Big Data*.

En este capítulo se revisó lo relativo a los conceptos de *Big Data* y *Data Science*. Establecimos que *Big Data* se refiere tanto a los datos, que son muy grandes, que se generan muy rápido o que no son estructurados, así como también a las técnicas creadas para manipularlos, ya que no pueden ser manejados correctamente por bases de datos relacionales. *Data Science* consiste en la extracción generalizable de información a partir de datos y se caracteriza por su capacidad de predicción y por la generación de modelos a partir de muy pocas consideraciones iniciales. También se indicó que usualmente el término *Big Data* se utiliza como sinónimo o para englobar a otros conceptos como *Data Science*, Analítica y *Data-driven decision making*, ya que la finalidad de estas tecnologías

es la mejora en la toma de decisiones. Se describió además la relación de *Big Data* y *Data Science* con *Machine Learning* y *Deep Learning*, los cuales son campos de la Inteligencia Artificial. Por supuesto, ya que los conceptos anteriores están inmersos en los sistemas de información, pueden ser considerados como intangibles. Ambos tópicos serán discutidos en las siguientes secciones.

2.4. Los sistemas de información y generación de valor

Los sistemas de información se definen como el conjunto de la infraestructura, la organización, el personal y los componentes para la recolección, procesamiento, transmisión, visualización, diseminación y disposición de la información (ATIS Telecom Glossary; 2007). A su vez, Melville (2004) define que la generación de valor de las tecnologías de la información (*IT business value*) se refiere a los impactos en el desempeño organizacional de las tecnologías de la información, incluyendo mejoras a la productividad, mejoras a la rentabilidad, reducción de costos, ventajas competitivas, reducción de inventario y otras mediciones de desempeño.

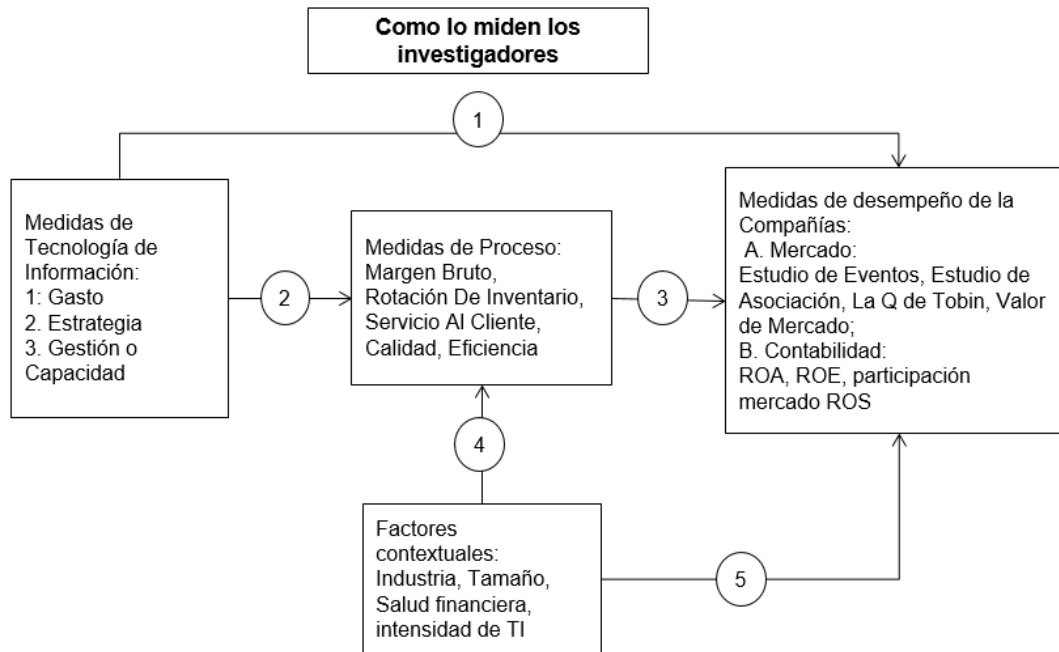
Schryen (2013) analiza más de 300 artículos académicos relacionados al valor de los sistemas de información en las empresas y resume los paradigmas teóricos que se han utilizado:

- 1) Microeconomía
 - a. Teoría de producción
 - b. Contabilidad de crecimiento
 - c. Teoría del consumidor

- d. Análisis por envoltura de datos
 - e. Teoría de la inversión Q de Tobin
 - f. Teoría de valoración de opciones
- 2) Teoría de la organización industrial
- a. Teoría de juegos
 - b. Teoría de la agencia y contratos incompletos
 - c. Teoría de los costos de transacción
- 3) Perspectivas sociológicas y sociopolíticas
- a. Teoría del arraigamiento
- 4) Teoría del comportamiento organizacional
- 5) Teoría del enfoque de recursos
- 6) Teoría de la decisión

Hay cuatro modelos que se han postulado para la generación de valor de las tecnologías de la información en las organizaciones, los cuales pueden observarse en las Figuras 8, 9, 10 y 11.

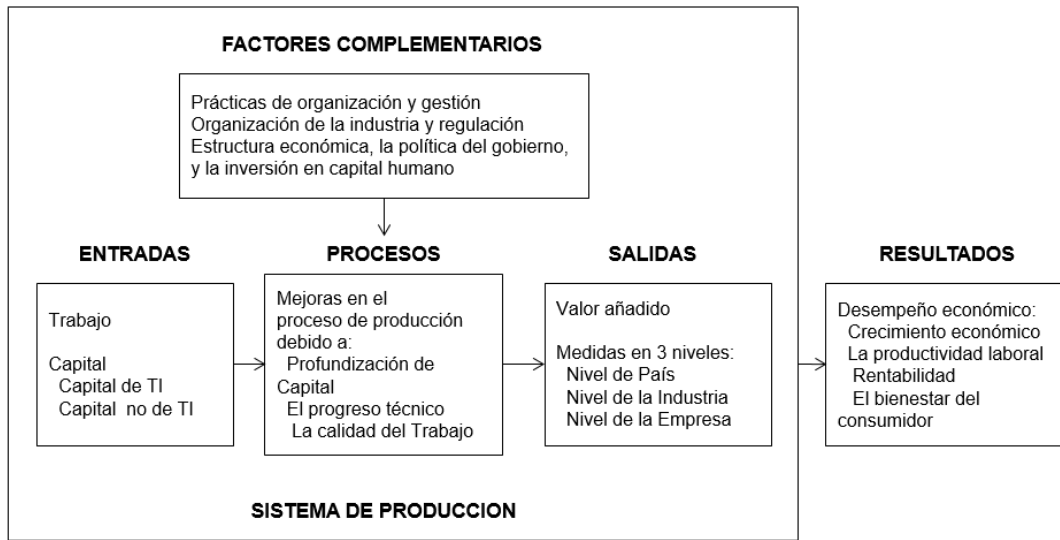
Figura 8. *Modelo de Dehning y Richardson*



Fuente: Elaboración propia a partir de Dehning y Richardson (2002).

En la Figura 8 se expone un modelo mediante el cuál las tecnologías de la información pueden verse como un medio para mejorar temas de gasto, de estrategia y de gestión o capacidad. Por consiguiente, algunos elementos de los procesos que pueden medirse son los márgenes brutos, rotación de inventario, servicio al cliente, calidad y eficiencia. A nivel empresa, se utilizan diversas métricas para determinar la contribución de las tecnologías de la información. En el valor de mercado, se usan estudio de eventos, de asociación, la Q de Tobin. En métricas contables se utilizan, por ejemplo, el retorno de activos y el retorno de capital.

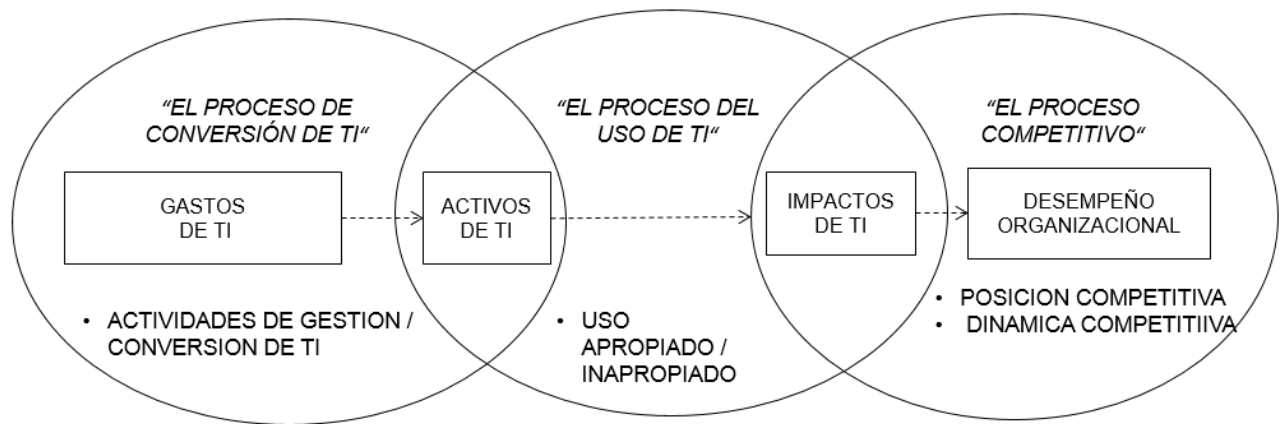
Figura 9. *Modelo orientado a la producción*



Fuente: Elaboración propia a partir de Dedrick et al. (2003)

El modelo orientado a la producción de la Figura 9, sugiere que las tecnologías de la información coadyuvan a la mejora de resultados económicos, de productividad, de rentabilidad e incluso del bienestar del consumidor. Sin embargo, este modelo no ayuda a determinar el efecto de las tecnologías de la información por sí mismas.

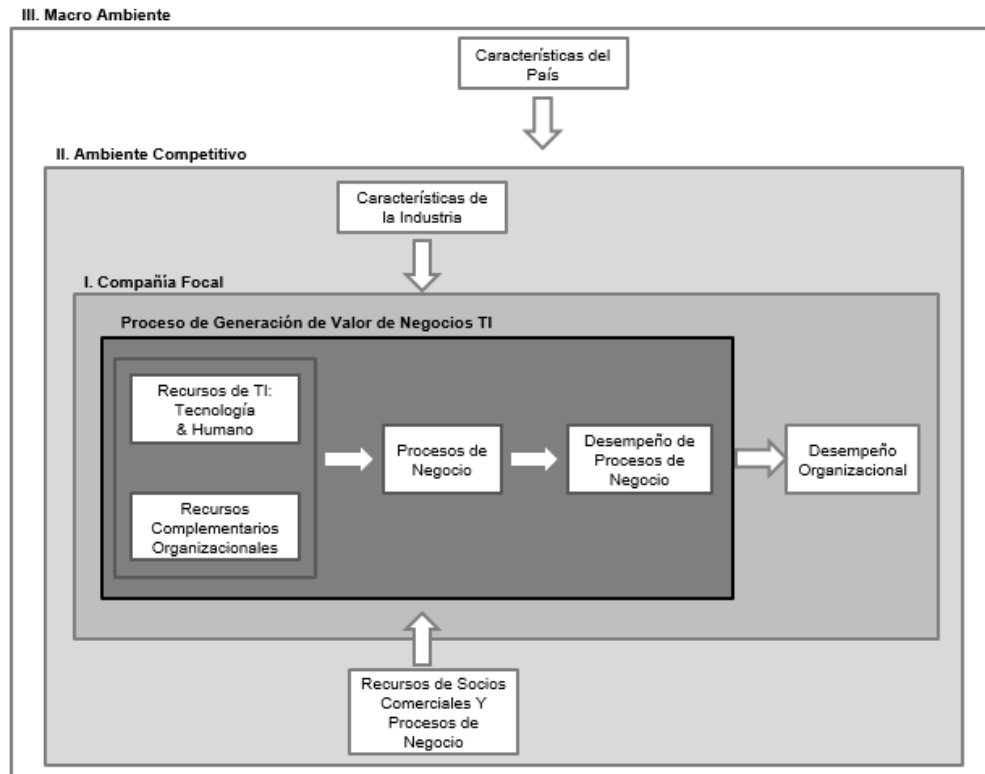
Figura 10. *Modelo orientado a los procesos*



Fuente: Elaboración propia a partir de Soh y Markus (1995)

El modelo orientado a procesos de la Figura 10 ayuda a visualizar como los gastos o las inversiones en tecnologías de la información afectan, luego de una serie de pasos, al desempeño organizacional. Sin embargo, este modelo no explica cómo y en qué medida los sistemas de información afectan el desempeño organizacional. Mientras que, el modelo basado en recursos de la Figura 11 indica como los recursos de la tecnología de la información impactan los procesos de negocio y por consiguiente al desempeño organizacional. Sin embargo, no se especifica cómo puede ser medido o cuantificado.

Figura 11. *Modelo basado en recursos*



Fuente: Elaboración propia a partir de Melville et al. (2004)

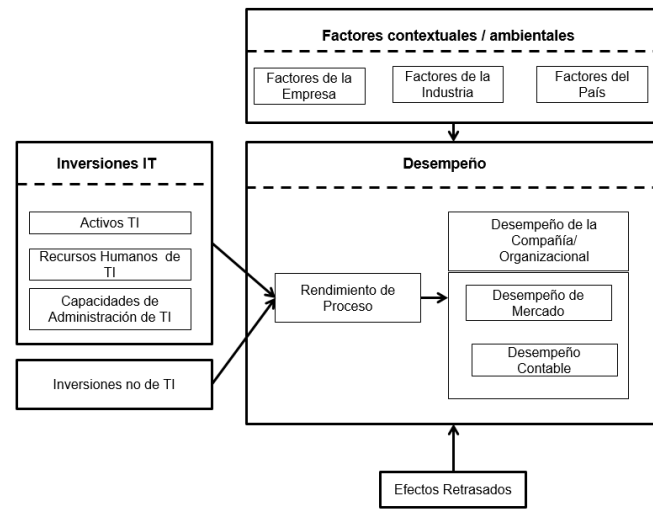
Para Schryen (2013), los modelos anteriores muestran consenso en tres puntos principales:

- 1) El impacto de las inversiones en las tecnologías de la información puede tratarse a través de numerosas medidas de desempeño. Estas últimas pueden ser divididas en medidas de mercado y financieras-contables.
- 2) El impacto de las inversiones en tecnologías de la información en el desempeño de las compañías es afectado por factores ambientales y contextuales de la empresa, de la industria y del país.

- 3) Las inversiones en tecnologías de la información y sus activos resultantes pueden ocurrir en forma de gastos (hardware, software), recursos humanos técnicos y habilitadores de administración.

La Figura 12 contiene el modelo propuesto por Schryen (2013) que intenta unificar los modelos de Dehning y Richardson, el modelo orientado a la producción, el modelo orientado a los procesos y el modelo orientado a los recursos.

Figura 12. *Modelo de Schryen*



Fuente: Elaboración propia a partir de Schryen (2013).

El modelo anterior es muy útil, porque especifica los tipos de recursos involucrados en las inversiones de los sistemas de información, pero también clarifica que hay otras inversiones que no son de TI que también afectan los procesos de la compañía que a su vez afectan el desempeño organizacional, medido a través de métricas de mercado y contables. Este efecto está también influido por factores específicos de la empresa, de la industria y del país. Por último, los efectos de las inversiones en TI suelen ser retardados, esto es pudieran más de un año en manifestarse.

Debido a que *Big Data* y *Data Science* son inversiones tecnológicas o que son acompañadas de elementos tecnológicos, los modelos anteriormente propuestos son aplicables para explicar la generación de valor de estos conceptos. La teoría de la información propuesta por Blackwell (1953) y el enfoque de procesamiento de información propuesto por Galbraith (1974), sugieren que entre más precisa y correcta sea la información, debería de facilitar su uso en el proceso de toma de decisiones y, por lo tanto, ayudar a la empresa a tener un mejor desempeño.

Blackwell (1953) describe matemáticamente cuándo un conjunto de información imperfecta es mejor que otro para un tomador de decisiones que intenta escoger la acción que produzca los mejores resultados. Un tomador de decisiones racional que actúa con mejor información debería alcanzar mayores beneficios esperados. En este sentido, mejor información siempre mejora el desempeño.

Por su parte, Galbraith (1974) expone que la realización de tareas complejas requiere una cantidad mayor de información a ser procesada, por lo que las organizaciones deben ser diseñadas para facilitar este procesamiento de información. Aquellas tecnologías que habiliten una mayor recopilación de información o que faciliten una mejor distribución de información dentro de una organización deberían minimizar costos y mejorar el desempeño.

De esta manera, como se expuso anteriormente, la función primordial de estas tecnologías consiste en el procesamiento de mayores o más variadas fuentes de información de una forma eficiente, para posteriormente extraer conocimiento y mejorar la toma de decisiones en las organizaciones. Dichas características encajan con lo mencionado por Blackwell y Galbraith. Además, al ser elementos de sistemas de información, los modelos de generación de valor son aplicables a estos conceptos.

3. LA Q DE TOBIN Y LOS INTANGIBLES

En este capítulo se expone por qué las métricas contables algunas veces no son la mejor opción para medir el desempeño organizacional, lo que nos lleva al uso de la Q de Tobin. Se analiza el método de cálculo de la Q de Tobin, así como una aproximación muy útil que simplifica enormemente su cálculo y la información requerida para obtener el indicador.

3.1. Métricas contables y la generación de valor

En el trabajo de Brynjolfsson, et al. (2011), encontramos que las métricas contables tales como el retorno de activos, el retorno de capital y el retorno en ventas, tienen algunas debilidades al capturar el desempeño de las empresas:

- 1) Usualmente solo reflejan información pasada y no miran al futuro.
- 2) No han sido ajustadas por riesgos.
- 3) Son distorsionadas por efectos temporales de desequilibrio, leyes fiscales y normas contables.
- 4) No capturan el valor de activos intangibles.
- 5) No son sensibles a efectos retardados que podrían ser importantes para el estudio del valor generado por los sistemas de información.

Bajo este contexto, las métricas financieras y de mercado pueden ser alternativas viables al uso de métricas contables. Retomando a Brynjolfsson, et al. (2011), la Q de Tobin ha sido utilizada para medir el desempeño del negocio, al tomar en cuenta los activos intangibles, activos tecnológicos y el valor de la marca. El valor de mercado ha sido utilizado para estimar el valor de intangibles asociados a los activos de TI o en el desempeño organizacional. Bharadwaj et al. (1999) y Brynjolfsson et al. (2002) fueron los primeros trabajos al respecto.

3.2. Definición de la Q de Tobin

El concepto de la Q de Tobin fue propuesto por Tobin (1969), quien posteriormente en 1981 recibió el Premio Nobel en Economía. La Q de Tobin se define como la razón entre el valor de mercado de una compañía y el valor de reposición, o contable, de los activos de la compañía.

$$Q = \frac{VM}{ATAN} \quad (1)$$

Donde VM es el valor de capitalización de mercado de la compañía, mientras que la $ATAN$ es el valor de reemplazo de sus activos tangibles.

Hall (2000; 2001) y Brynjolfsson, et al. (2002) exponen que el valor de mercado de una empresa debe ser igual a la suma de sus activos tangibles e intangibles.

$$VM = AI + ATAN \quad (2)$$

Donde AI es el valor de los activos intangibles de la compañía.

Si dividimos (2), por *los activos tangibles*, tenemos a la Q de Tobin expresada mediante la razón de activos intangibles sobre activos tangibles.

$$Q = \frac{AI}{ATAN} + 1 \quad (3)$$

Desde un punto de vista empírico, la Q de Tobin es un *proxy* para los activos intangibles de las empresas como resultado del tratamiento contable de los intangibles (Lev; 2001), Esto es, los activos tangibles son capitalizados (reconocidos como activos en el balance), mientras que los intangibles son expresados como gastos, semejante a

salarios, rentas o intereses. Como resultado, el valor en libros de los activos no refleja los intangibles, pero el valor de mercado si lo hace. Bajo este sentido, Lindenberg y Ross (1981), encontraron que la Q de Tobin era anormalmente alta en empresas que tienen mucha investigación y desarrollo, o publicidad, ambos elementos asociados a activos intangibles.

Villalonga (2004) lista algunos estudios que han usado la Q de Tobin para medir activos intangibles específicos al tomar el valor predicho en una regresión de la Q de Tobin sobre datos contables o datos provenientes de encuestas del activo intangible de interés. En el ámbito de las tecnologías de la información, algunos investigadores han utilizado este método para determinar el valor intangible de los sistemas de información en las empresas. Los primeros trabajos al respecto fueron los de Bharadwaj et al. (1999) y Brynjolfsson et al. (2002).

3.3. Método de cálculo de la Q de Tobin

El método siguiente para calcular la Q de Tobin, fue propuesto por Lindenberg y Ross (1981), modificado posteriormente por Smirlock, et al. (1984) y utilizado subsecuentemente por Lang y Stulz (1994), McGahan (1999), Villalonga (2004), Saunders y Tambe (2014), entre otros.

$$Q_{LR} = \frac{PAC * NACC + VAP + DCP + DLPA - VACC}{VLTA - VLCP + CSA} \quad (4)$$

Donde Q_{LR} representa el valor de la Q de Tobin calculado a través del método de Lindenberg y Ross. En el numerador, PAC es el precio de las acciones comunes al final del año fiscal, NACC es el número de las acciones comunes en circulación, VAP es el

valor de las acciones preferentes, DLPA es el valor de la deuda a largo plazo ajustado por la edad de la compañía, DCP es el valor en libros de la deuda a corto plazo y VACC es el valor neto de los activos de corto plazo. En el denominador, VLTA es el valor en libros del total de activos, VLCP es el valor en libros del capital social y el CSA es el capital social ajustado de acuerdo a la inflación.

Una modificación del método anterior es expuesta por Villalonga (2004). El denominador es construido por tres partes. La primera es el costo de reemplazo de propiedad, planta y equipo, el cual se asume que es igual a su valor en libros en 1970 o bien cuando su información comenzó a ser listada en bases de datos como Compustat o Bloomberg. El valor de reemplazo del año anterior es reducido por un 5% de depreciación y luego se ajusta por cambios en precios de acuerdo al producto interno bruto para activos fijos no residenciales. El cambio en el valor en libros del valor bruto de propiedad, planta y equipo durante el año, es agregado al valor de reemplazo del año anterior. El parámetro de cambio tecnológico se asume que es cero.

La segunda parte del denominador es el costo de reemplazo de los inventarios en 1970 o bien cuando su información comenzó a ser listada en bases de datos como Compustat o Bloomberg. El costo de reemplazo de los inventarios es igual a su valor en libros. Posteriormente se toma en cuenta la inflación anual dependiendo del método de valuación de inventarios y el factor de deflación para inventarios. Si no hubiera claridad respecto al método de valuación, se asume que el costo de reemplazo del inventario es igual a su valor en libros. Finalmente, la tercera parte del denominador incluye todos los demás activos que se encuentran en el balance, que se asumen que están listados a su valor de reemplazo. Esto equivale a restar de los activos totales el valor en libros de propiedad, planta y equipo, así como de los inventarios.

Otro elemento de la ecuación son los activos de investigación y desarrollo (*research and development; R&D*), los cuales se calculan al tomar el gasto nominal reportado por las empresas y se convierten a flujos constantes de dólares. Para el modificador de precio se usan las series de precios publicadas en el Bureau of Economic Analysis.²⁹ De acuerdo a Hall et al. (2005) se utiliza un 15% de tasa de depreciación y, como valor inicial, el primer año disponible del gasto de *R&D* o bien el de 1959. Las compañías tienen que reportar este gasto si excede el 1% de las ventas (Zhao; 2002). Si no hay información, se asume que es cero.

Semejante a *R&D*, para los activos de publicidad se convierte el gasto nominal reportado por las empresas a flujos constantes de dólares. Se comienza con el primer año disponible donde la empresa reporte el gasto en publicidad, o bien el del año 1959. Se utilizaron tres series de precios para formar el modificador para publicidad³⁰ y se asume una depreciación anual de 45% de acuerdo a Villalonga (2004). Las compañías están obligadas a reportar estos gastos si son materiales. Si no se reporta, se asume que es cero.

La rotación de inventario se define como la razón del costo de los bienes vendidos y los inventarios como están registrados en el balance. Se eliminan los percentiles 1 y 99 para eliminar extremos. A continuación, se calcula el logaritmo natural de la razón de las vueltas de inventario y la vuelta de inventarios promedio de la industria.

²⁹ Aggregate Input Price Indexes for R&D Investment, 1959-2007. Disponible en <http://www.bea.gov/national/newinnovation.htm>

³⁰ Para 1970-1976 y 1977-1994 usaron información de http://www.bea.gov/industry/gdpbyind_data.htm, mientras que para 1995-2012 usaron información de <http://www.bls.gov/ppi/>.

3.4. Aproximación de Chung y Pruitt

Como se expuso en el apartado pasado, el cálculo de la Q de Tobin con el método de Lindenberg y Ross, requiere un número considerable de información financiera como punto de partida que después es ajustada por factores que causan divergencias entre los valores de mercado y los contables.

Chung y Pruitt (1994) propusieron un método que utiliza menos información financiera y requiere mínimos ajustes. Este método no pretende dar un valor exacto de la Q de Tobin, sino que funciona como una aproximación. Según los autores, el valor del coeficiente de determinación R^2 en las regresiones nunca bajó de 0.966, lo que indica que el 96.6% de la variabilidad total del valor calculado por el método de Lindenberg y Ross puede ser explicado por la aproximación de Chung y Pruitt. La aproximación es la siguiente:

$$Q_{CP} = \frac{PAC * NACC + VAP + DLP + DCP}{ATAN} \quad (5)$$

Donde en el numerador PAC es el precio de las acciones comunes al final del año fiscal, NACC es el número de las acciones comunes en circulación, VAP es el valor de las acciones preferentes, DLP es el valor en libras de la deuda a largo plazo, DCP es la deuda a corto plazo. En el denominador, ATAN significa valor en libras del total de activos tangibles. Todos los términos necesarios en este método están contenidos en los informes financieros básicos de las compañías. DaDalt et al. (2003) argumenta que este método de aproximación es preferible excepto en casos donde la precisión de la Q de Tobin sea de vital importancia.

4. MODELO Y METODOLOGÍA

A partir de lo expuesto en capítulos anteriores, tenemos que el uso de *Big Data* y *Data Science* posibilitan una mejor toma de decisiones en las empresas y también pueden generar beneficios operativos. Además, existe evidencia en la literatura no académica acerca del uso de *Big Data* y *Data Science* en las compañías. Sin embargo, en la literatura no académica se listan beneficios generalmente cualitativos del uso de *Big Data* o *Data Science*, ya que la mayoría son estudios que se limitan a encuestas de opinión. La literatura académica posibilita el estudio de *Big Data* y *Data Science* mediante la microeconomía a través de la Q de Tobin, al estudiar dichos efectos como un intangible. Por último, *Big Data* y *Data Science* están estrechamente relacionados con *Machine Learning* y *Deep Learning*, los cuales son campos de la Inteligencia Artificial.

Basados en literatura previa, Saunders y Tambe (2015) utilizaron la siguiente ecuación de estimación para determinar impacto de las prácticas basadas en datos en las empresas:

$$Q_{it} = \alpha + \beta_1 \frac{I_{1it}}{A_{it}} + \beta_2 \frac{I_{2it}}{A_{it}} + \dots + \beta_n \frac{I_{nit}}{A_{it}} + \delta PalabrasClave_{it} + Controles_{it} + u_{it} \quad (6)$$

Donde I_{nit} es el activo intangible número n de la empresa i en el año t , *PalabrasClave* captura la variable de interés, los *Controles* consisten en las siguientes razones financieras: total de pasivos entre activo total, *CAPEX* entre activo total, *EBITDA* entre activo total y logaritmo natural de activo total. Estos controles son incluidos para tomar en cuenta oportunidades previas de rentabilidad y crecimiento. Por último, u_{it} corresponde al error no observado.

4.1. Modelo propuesto

A partir de la ecuación (4) proponemos el siguiente modelo, tomando en cuenta nuestra variable de interés del uso de *Big Data* y *Data Science*, así como las variables de control.

$$Q_{it} = \beta_0 + \beta_1 \frac{AI_{it}}{TA_{it}} + \beta_2 \ln(TA_{it}) + \beta_3 CA_{it} + \beta_4 \frac{P_{it}}{TA_{it}} + \beta_5 \frac{CAPEX_{it}}{TA_{it}} + \beta_6 \frac{EBITDA_{it}}{TA_{it}} + e_{it} \quad (7)$$

donde:

Q_{it} es la Q de Tobin de la compañía i en el año t , ésta es la variable dependiente y utilizaremos la aproximación de Chung y Pruitt (1994), ecuación (5), para su cálculo.

CA es nuestra variable de interés, la cual es dicotómica y se construye a partir de la ocurrencia de las palabras clave relativas a *Big Data* y *Data Science* en los estados financieros de las empresas. Su construcción será especificada en breve.

AI corresponde al total de los activos intangibles de la empresa, de acuerdo a (6). Estudios anteriores como Bharadwaj et al. (1999), Brynjolfsson et al. (2002), Bardhan et al. (2013) y Saunders y Tambe (2015) han utilizado los conceptos de investigación y desarrollo, gastos en tecnologías de la información o publicidad como elementos de los activos intangibles.

El logaritmo natural de los activos totales, **TA**, se utiliza como variable de control respecto al tamaño de la empresa. Bharadwaj et al. (1999) utiliza las ventas como variable de control, mientras que Brynjolfsson et al. (2002) y Saunders y Tambe (2015) utilizan activos.

CAPEX, inversiones de capital, es otra variable de control, utilizada también por Bharadwaj et al. (1999) y Saunders y Tambe (2015).

EBITDA, beneficios antes de intereses, impuestos, depreciaciones y amortizaciones es otra variable de control, utilizada también por Saunders y Tambe (2015).

P corresponde a los pasivos de la empresa, utilizada también por Saunders y Tambe (2015).

B₀ es el intercepto de la regresión.

B₁ es el coeficiente de la razón de los activos intangibles y el total de activos.

B₂ es el coeficiente del logaritmo natural del total de activos.

B₃ es el coeficiente de la variable de interés dicotómica del uso de *Big Data* y *Data Science*.

B₄ es el coeficiente de la razón de los pasivos y el total de activos.

B₅ es el coeficiente de la razón de las inversiones de capital (CAPEX) y el total de activos.

B₆ es el coeficiente de la razón del EBITDA y el total de activos.

e es el error.

Para el cálculo de nuestra variable de interés, nos basamos en Archak et al. (2011) y Ghose et al. (2012). Utilizando el análisis de textos financieros, se analizarán los reportes anuales de las compañías mediante la técnica de bolsa de palabras, es decir, un conteo de palabras clave. En este caso, el listado de palabras clave que se buscará en los estados financieros, provienen de la literatura consultada y expuesta en los capítulos anteriores: *big data*, *data scien*³¹, *artificial intelligence*, *machine learning*, *deep learning* y *hadoop*. Al contabilizar la ocurrencia de estas palabras en el reporte anual, se genera una variable dicotómica llamada Constructo Ajustado (CA), la cual es la variable que mide el

³¹ Se utilizará “data scien” para englobar tanto a Data Science como a Data Scientist.

uso de *Big Data* y *Data Science* en las empresas. Esta variable toma el valor de 1 si al menos hay alguna ocurrencia de cualquiera de las palabras clave, y tomará el valor de 0 si no hay ocurrencias.

Para el cálculo de la Q de Tobin, nos basamos en la aproximación de Chung y Pruitt (1994), ecuación (5). Esto es, sumamos la capitalización de mercado, el valor de las acciones preferentes y el total de la deuda, para después dividirlo entre el total de activos tangibles.

4.2. Diseño de la investigación

A partir de Hernández et al. (2010), el diseño de la presente investigación es cuantitativa, no experimental y longitudinal de panel. Es cuantitativa porque nos basamos en métodos estadísticos para probar una hipótesis. Es no experimental ya que no estamos manipulando variables sino únicamente observando datos de hechos ya ocurridos. Los diseños panel de investigación son de tipo no experimental y longitudinal donde se recolectan datos a través del tiempo en ciertos periodos para los mismos participantes. Es decir, las mismas entidades son medidas o observadas en todo momento para hacer inferencias respecto a los cambios, sus determinantes y sus consecuencias.

4.3. Población, marco muestral y muestra

Mediante la base de datos *Economática* (2016), se recopila información financiera de 1260 empresas públicas de los Estados Unidos en el período comprendido entre diciembre de 2011 y diciembre de 2015. Del total de las compañías, dos de ellas cotizan en la bolsa de valores AMEX (*American Stock Exchange*), 362 en NASDAQ (*National*

Association of Securities Dealers Automated Quotation) y 896 en NYSE (*New York Stock Exchange*).

Además de la información proveniente de Economatica, para dichas empresas se analizarán los formatos 10-K (reportes anuales) que son requeridos por la SEC (*Securities and Exchange Commission*), a través de la plataforma EDGAR (*Electronic Data Gathering, Analysis, and Retrieval*)³² *Full-Text Search*. Se analizará la ocurrencia de las palabras clave relacionadas a *Big Data* y *Data Science*.

Se divide la recolección de información en dos partes. La primera, consiste en la recopilación de datos financieros a través de Economatica (2016), lo cual posibilitará el cálculo la variable dependiente, la Q de Tobin, para cada empresa, además del cálculo de las variables independientes de control. La segunda parte consiste en los análisis de los reportes anuales de las empresas para la identificación de palabras claves asociadas a la variable de *Big Data* y *Data Science*. En ambos casos, se analizarán los últimos 4 años, debido a que este periodo de tiempo es el máximo disponible para la búsqueda de texto en la plataforma EDGAR.

4.4. Métodos de análisis

Múltiples estudios anteriores han utilizado regresiones lineales, particularmente por el método de mínimos cuadrados ordinarios, para determinar el efecto de las variables independientes sobre la Q de Tobin en datos panel, por ejemplo, Lang y Stulz (1994), Bharadwaj et al. (1999), Brynjolfsson et al. (2002), Bardhan et al. (2013), Saunders y Tambe (2015). Sin embargo, dicho método asume que los datos presentan una

³² Para más información, visitar:
https://searchwww.sec.gov/EDGARFSCClient/jsp/EDGAR_MainAccess.jsp

distribución normal. Otros autores, como Coad y Rao (2006), han decidido utilizar otros métodos, tales como la regresión cuantílica, debido a que habitualmente la Q de Tobin no presenta una distribución normal.

La regresión cuantílica fue propuesta por Koenker y Basset (1978) y se caracteriza por que puede ser utilizada cuando no se cumplen los supuestos de normalidad y homocedasticidad. Adicionalmente, se utiliza este tipo de regresión cuando se requiere analizar la relación de variables más allá de una tendencia central y explorar varias zonas a lo largo de toda la distribución. Eso es, a diferencia de la regresión lineal por mínimos cuadrados ordinarios, podemos analizar el efecto de la variable de interés en ciertos grupos y no en toda la población. Por ejemplo, si utilizamos la regresión lineal por mínimos cuadrados ordinarios, analizaremos el efecto de *Big Data* y *Data Science* en todas las empresas, independientemente de su tamaño. La regresión cuantílica permite analizar el efecto de *Big Data* y *Data Science* en empresas 'pequeñas', 'medianas' y 'grandes' con respecto a su Q de Tobin. Esto es muy relevante ya que esperamos que el efecto de *Big Data* y *Data Science* varíe con respecto al tamaño de la empresa.

Así pues, utilizar la regresión cuantílica nos brinda dos ventajas. En primer lugar, es robusta a efectos de sesgo en los datos (no normalidad y homocedasticidad) y adicionalmente nos permite estudiar el efecto de la variable de interés en distintos grupos de acuerdo a su Q de Tobin.

5. RESULTADOS

En este capítulo se exponen los resultados obtenidos de la investigación, mediante el uso de tablas que condensan la información relevante.

5.1. Análisis de los datos

De acuerdo a la información financiera contenida en la base de datos de Economatica (2016) se construye la variable dependiente de la Q de Tobin a partir de lo especificado en el capítulo anterior. Además, se incorpora a este conjunto de datos el resultado de la construcción de la variable dicotómica que captura las palabras clave proveniente de la herramienta EDGAR de la SEC, por lo que se obtuvieron 5040 registros en 4 años de informes financieros pertenecientes a 1260 empresas. De acuerdo al modelo propuesto, estudiamos 7 variables, una de ellas dependiente y las otras seis independientes. De esta manera, se obtuvo un conjunto de 35,280 observaciones.

La Tabla 2 contiene la ocurrencia de las palabras clave en las empresas analizadas. Si la palabra clave es contenida en el informe 10-K una o más ocasiones, se contabiliza como una ocurrencia. La variable de interés del uso de *Big Data* y *Data Science* se construye a partir de las palabras clave, esto es, si al menos una palabra clave tiene una ocurrencia, la variable dicotómica toma el valor de uno y toma el valor de cero en caso contrario.

Tabla 2. *Ocurrencia de las palabras clave*

Palabra Clave	Año	Ocurrencias	% Empresas
Big Data	2012	16	1.27
Big Data	2013	34	2.70
Big Data	2014	52	4.13
Big Data	2015	64	5.08
Data Scien	2012	4	0.32
Data Scien	2013	11	0.87
Data Scien	2014	15	1.19
Data Scien	2015	20	1.59
Machine Learning	2012	4	0.32
Machine Learning	2013	7	0.56
Machine Learning	2014	15	1.19
Machine Learning	2015	20	1.59
Artificial Intelligence	2012	4	0.32
Artificial Intelligence	2013	3	0.24
Artificial Intelligence	2014	7	0.56
Artificial Intelligence	2015	11	0.87
Hadoop	2012	5	0.40
Hadoop	2013	9	0.71
Hadoop	2014	10	0.79
Hadoop	2015	9	0.71
Deep Learning	2012	0	0
Deep Learning	2013	0	0
Deep Learning	2014	1	0.08
Deep Learning	2015	2	0.16
CA ⁽¹⁾	2012	33	2.62
CA	2013	64	5.08
CA	2014	100	7.94
CA	2015	126	10

Fuente: Elaboración propia a partir de los Datos de Economatica utilizando Eviews. El periodo de tiempo es de 2012 a 2015 y el número de observaciones es 5040.

⁽¹⁾ CA es un constructo ajustado correspondiente a nuestra variable dicotómica de interés. Es una variable independiente que toma el valor de 1 si en el informe anual aparece alguna de las palabras clave y 0 en caso contrario.

Se encontraron 166 ocurrencias de *Big Data*, 50 ocurrencias de *Data Scien*, 46 ocurrencias de *Machine Learning*, 25 ocurrencias de *Artificial Intelligence*, 33 ocurrencias de *Hadoop* y 3 ocurrencias de *Deep Learning*. De nuestra variable de interés, alguna de las palabras clave aparecieron en el 2.62% de las empresas en 2012, en 2013 fue 5.08%, en 2014 fue 7.94% y en 2015 10%.

Utilizando el software *Eviews* (*Quantitative Micro Software*, 2010), se analiza la distribución de las variables, Tabla 3.

Tabla 3. *Estadística descriptiva*

Estadístico	Q⁽¹⁾	CA⁽²⁾	AI/ TA⁽³⁾	ln(TA)⁽⁴⁾	P/TA⁽⁵⁾	CAPEX/ TA⁽⁶⁾	EBITDA/ TA⁽⁷⁾
Observaciones	3,672	5,040	4,031	4,840	4,555	3,865	4,299
Media	3.35	0.05	0.24	15.72	0.70	0.07	0.10
Mediana	2.67	0	0.18	15.59	0.66	0.05	0.10
Máximo	39.078	1	0.89	21.67	25.33	0.70	1.30
Mínimo	0.30	0	0	8.38	0.00	-0.67	-2.83
Desviación estándar	2.91	0.23	0.23	1.52	0.55	0.10	0.13
Asimetría	3.34	4.16	0.74	0.34	23.62	1.43	-3.92
Curtosis	23.79	18.29	2.51	4.00	953.37	11.34	81.34
Jarque-Bera ⁽⁸⁾	73	64	0.4	0.3	172,000	13	1,110

Fuente: Elaboración propia utilizando Eviews. El periodo de tiempo es de 2012 a 2015.

⁽¹⁾ Q de Tobin, corresponde a la variable dependiente del modelo. Su construcción se detalla en la ecuación (5).

⁽²⁾ CA es un constructo ajustado correspondiente a nuestra variable dicotómica de interés. Es una variable independiente que toma el valor de 1 si en el informe anual aparece alguna de las palabras clave y 0 en caso contrario.

⁽³⁾ Razón de activos intangibles y total de activos. Variable independiente de control.

⁽⁴⁾ Logaritmo natural del total de activos. Variable independiente de control.

⁽⁵⁾ Razón de pasivos y total de activos. Variable independiente de control.

⁽⁶⁾ Razón de inversiones en bienes de capitales y total de activos. Variable independiente de control.

⁽⁷⁾ Razón del resultado empresarial antes de intereses, impuestos, depreciaciones y amortizaciones productivas, y el total de activos. Variable independiente de control.

⁽⁸⁾ Cifras expresadas en miles

Jarque y Bera (1978) propusieron un método para determinar normalidad en los datos, que se basa en la asimetría y la curtosis. Una distribución es simétrica si es igual a la izquierda y a la derecha del punto central; si no es simétrica, tenemos distribuciones con 'colas' hacia la izquierda o derecha. La curtosis indica el grado de presencia de datos en las colas; un valor negativo indica que hay pocos datos en ellas, mientras que un valor positivo indica mayor presencia de datos en esas zonas. Para una distribución normal, la asimetría debe ser cero y la curtosis, tres. De acuerdo a los valores de la Tabla 3, para todas las variables tienen asimetría diferente de cero y, salvo en un caso, fuerte grado de

curtosis, lo que nos indica que tenemos una distribución de datos con colas y muchos valores en ellas.

Aunque con la asimetría y la curtosis podemos inferir que los datos no siguen una distribución normal, el método de Jarque-Bera pone a prueba la hipótesis nula de que los datos siguen una distribución normal, esto es:

H_0 : los datos siguen una distribución normal.

H_a : los datos no siguen una distribución normal.

El estadístico de prueba del método Jarque-Bera se basa en la asimetría y la curtosis y está definido por la siguiente ecuación:

$$JB_{\text{Estadístico de Prueba}} = n \left[\frac{\widehat{\text{Asimetría}}^2}{6} + \frac{(\widehat{\text{Curtosis}}-3)^2}{24} \right] \quad (8)$$

Donde $JB_{\text{Estadístico de Prueba}}$ es el valor de Jarque-Bera que será utilizado para aceptar o rechazar la hipótesis nula. Asimetría y Curtosis son los valores contenidos en la Tabla 3.

El valor del estadístico de prueba del método de Jarque-Bera se contrasta con los valores críticos de la distribución χ^2 (chi-cuadrada) con dos grados de libertad. En este caso, para un valor de significancia $\alpha=0.001$, el valor crítico corresponde a 13.816. Para la prueba de hipótesis, rechazaremos H_0 si $\chi^2_{2,1-\alpha} < JB_{\text{Estadístico de Prueba}}$. Para nuestras variables en todos los casos 13.816 es menor que el estadístico de prueba de Jarque-Bera, por lo que rechazamos H_0 y aceptamos H_a . Esto es, aceptamos la hipótesis alternativa que los datos no siguen una distribución normal.

La prueba Jarque-Bera nos indica que la Q de Tobin no está distribuida de manera normal por lo que los métodos más comunes de regresiones lineales no son directamente

aplicables. Particularmente para la variable dependiente Q de Tobin, la distribución que encontramos es consistente con otros estudios, por ejemplo, Coad y Rao (2006). Se utilizará entonces la regresión cuantílica propuesta por Koenker y Bassett (1978), ya que no requiere supuestos de normalidad y es también robusta a datos atípicos.

5.2. Resultados del modelo

Se realizaron tres regresiones cuantílicas en los percentiles 25, 50 y 75. Los resultados obtenidos están contenidos en la Tabla 4.

Tabla 4. *Coefficientes de la regresión cuantílica*

Percentil	CA ⁽¹⁾	AI/TA ⁽²⁾	ln(TA) ⁽³⁾	P/TA ⁽⁴⁾	CAPEX/TA ⁽⁵⁾	EBITDA/TA ⁽⁶⁾
75 ⁷	0.89 [*] (4.28)	6.10 [*] (30.27)	-0.50 [*] (-17.17)	0.93 [*] (62.80)	4.40 [*] (6.94)	4.62 [*] (11.27)
50 ⁸	0.53 [*] (4.33)	4.94 [*] (40.12)	-0.26 [*] (-14.49)	0.98 [*] (7.29)	1.04 [*] (3.35)	4.86 [*] (9.25)
25 ⁹	0.18 ^{***} (1.68)	3.89 [*] (35.43)	-0.17 [*] (-10.84)	0.92 [*] (11.98)	0.33 (1.46)	3.29 [*] (7.81)

Fuente: Elaboración propia utilizando Eviews. El periodo de tiempo es de 2012 a 2015. El valor entre paréntesis es el estadístico t.

*Significancia 0.99

**Significancia 0.95

***Significancia 0.90

⁽¹⁾ CA es un constructo ajustado correspondiente a nuestra variable dicotómica de interés. Es una variable independiente que toma el valor de 1 si en el informe anual aparece alguna de las palabras clave y 0 en caso contrario.

⁽²⁾ Razón de activos intangibles y total de activos. Variable independiente de control.

⁽³⁾ Logaritmo natural del total de activos. Variable independiente de control.

⁽⁴⁾ Razón de pasivos y total de activos. Variable independiente de control.

⁽⁵⁾ Razón de inversiones en bienes de capitales y total de activos. Variable independiente de control.

⁽⁶⁾ Razón del resultado empresarial antes de intereses, impuestos, depreciaciones y amortizaciones productivas, y el total de activos. Variable independiente de control.

⁽⁷⁾ Pseudo R-cuadrada de 0.23

⁽⁸⁾ Pseudo R-cuadrada de 0.24

⁽⁹⁾ Pseudo R-cuadrada de 0.24

En el percentil 25 podemos ver que para las empresas con Q de Tobin pequeñas las variables que más aportan al modelo son los activos intangibles y el EBITDA. Nuestra

variable de interés CA es positiva y estadísticamente significativa al 90%. En el percentil 50, empresas con Q de Tobin medianas, las variables que más aportan al modelo son los activos intangibles y el EBITDA, pero aún en mayor medida que lo observado en el percentil 25. Nuestra variable de interés CA es tres veces mayor que lo observado en el percentil 25 y estadísticamente significativa al 99%. En el percentil 75, las empresas con Q de Tobin grandes, los activos intangibles y el EBITDA siguen siendo los que más aportan al modelo, pero también los CAPEX se vuelven muy relevantes. Adicional, nuestra variable de interés CA es 75% mayor que lo observado en las empresas medianas y estadísticamente significativa al 99%.

De acuerdo al modelo de Schryen (2013)., los efectos de las tecnologías de la información, pudieran tener un efecto retardado en el tiempo. Por lo anterior, se analizaron dos años de efectos retardados de la variable de interés CA. Para ello se estimó nuevamente la ecuación (7), agregando los efectos mencionados, como se muestra en (8) y (9).

$$Q_{it} = \beta_0 + \beta_1 \frac{AI_{it}}{TA_{it}} + \beta_2 \ln(TA_{it}) + \beta_3 CA_{it-1} + \beta_4 \frac{P_{it}}{TA_{it}} + \beta_5 \frac{CAPEX_{it}}{TA_{it}} + \beta_6 \frac{EBITDA_{it}}{TA_{it}} + e_{it} \quad (9)$$

$$Q_{it} = \beta_0 + \beta_1 \frac{AI_{it}}{TA_{it}} + \beta_2 \ln(TA_{it}) + \beta_3 CA_{it-2} + \beta_4 \frac{P_{it}}{TA_{it}} + \beta_5 \frac{CAPEX_{it}}{TA_{it}} + \beta_6 \frac{EBITDA_{it}}{TA_{it}} + e_{it} \quad (10)$$

En la Tabla 5 se presentan los resultados de los coeficientes estimados (β_3) correspondientes a los modelos (9) y (10).

Tabla 5. *CA a lo largo del tiempo*

Percentil	$CA_t^{(1,2)}$	$CA_{t-1}^{(3)}$	$CA_{t-2}^{(4)}$
75	0.89 [*] (4.28)	0.71 ^{***} (1.77)	0.33 (0.83)
50	0.53 [†] (4.33)	0.45 [†] (3.21)	0.24 (1.22)
25	0.18 ^{***} (1.68)	0.17 (1.21)	0.22 (1.10)

Fuente: Elaboración propia utilizando Eviews. El periodo de tiempo es de 2012 a 2015. El valor entre paréntesis es el estadístico t.

[†] Significancia 0.99

^{**} Significancia 0.95

^{***} Significancia 0.90

⁽¹⁾ CA es un constructo ajustado correspondiente a nuestra variable dicotómica de interés. Es una variable independiente que toma el valor de 1 si en el informe anual aparece alguna de las palabras clave y 0 en caso contrario.

⁽²⁾ Coeficiente del constructo ajustado en el año correspondiente (t).

⁽³⁾ Coeficiente del constructo ajustado retrasado un año (t-1).

⁽⁴⁾ Coeficiente del constructo ajustado retrasado dos años (t-2).

El efecto de la variable de interés en el segundo año para empresas medianas y grandes es significativo con retraso de un año. No se encontró que los efectos retrasados a dos años fueran significativos.

5.3. Análisis por elementos clave

Por último, se estudió el efecto individual de las palabras clave en la Q de Tobin. Como se describió en el capítulo anterior, la variable de interés CA, es una variable dicotómica construida a partir de la ocurrencia de un conjunto de palabras clave. Los resultados de efecto individual de cada palabra clave está contenido en la Tabla 6.

Tabla 6. *Análisis individual de las palabras clave*

Percentil	Hadoop	Big Data	Data Scien ⁽¹⁾	Machine Learning	Artificial Intelligence	Deep Learning
75	2.83** (2.04)	0.30 (1.35)	4.70* (6.74)	1.15* (3.17)	1.60* (3.16)	-0.62*** (-1.94)
50	1.63* (4.40)	0.42* (3.31)	1.59** (2.27)	0.47 (1.35)	1.66** (2.70)	-0.61 (-1.61)
25	1.38** (2.23)	0.17 (1.58)	0.58 (0.78)	-0.14 (-0.48)	0.80* (2.98)	-0.12 (0.41)

Fuente: Elaboración propia utilizando Eviews. El periodo de tiempo es de 2012 a 2015. El valor entre paréntesis es el estadístico t.

* Significancia 0.99

** Significancia 0.95

*** Significancia 0.90

(1) Se utiliza Data Scien para englobar las palabras clave de Data Science y Data Scientist.

En el percentil 75 la palabra clave *Big Data* no es significativa, mientras que *Deep Learning* es negativa. En el percentil 50, la palabra clave *Machine Learning* no es significativa, mientras que *Deep Learning* es negativa. En el percentil 25, las palabras clave de *Data Scien*, *Machine Learning* y *Deep Learning* no son significativas.

5.4. Prueba de correlación espuria

Se realizó una prueba para confirmar que los resultados obtenidos de modelo son confiables y no son debido a alguna correlación espuria o fenómeno aleatorio. Esta prueba consistió en escoger 10 filas aleatorias de las 5140 observaciones, en esas filas se le especificó un 1 a la variable de interés de *Big Data* y *Data Science* y a las demás se les especificó el valor de 0. Se realizaron las regresiones cuantílicas para los percentiles 25, 50 y 75 y en todos los casos el modelo arroja que la variable no es significativa. Esta

evidencia sugiere que el modelo efectivamente captura relaciones reales entre las variables que no son producto de correlaciones espurias o efectos aleatorios.

6. CONCLUSIONES

En este apartado se exponen las conclusiones de la investigación, así como futuras líneas sugeridas de investigación.

6.1. Consideraciones y recomendaciones finales

En este apartado procedemos a enumerar las conclusiones obtenidas en este trabajo, recalcando que los objetivos planteados al inicio de la investigación fueron satisfactoriamente alcanzados.

Primera conclusión:

Nuestra primera conclusión responde a la hipótesis que nos propusimos probar:

H_0 : No existe una relación significativa entre el valor de mercado de una empresa expresado mediante la Q de Tobin y el uso de *Big Data* y *Data Science*.

H_a : Existe una relación significativa entre el valor de mercado de una empresa expresado mediante la Q de Tobin y el uso de *Big Data* y *Data Science*.

Adicionalmente, esta conclusión también responde al objetivo general que establecimos: determinar el impacto sobre el valor de mercado expresado mediante la Q de Tobin del uso de las tecnologías de la información basadas en *Big Data* y *Data Science* en las empresas públicas de los Estados Unidos.

De acuerdo a la información contenida en la Tabla 4, rechazamos la hipótesis nula y aceptamos la hipótesis alternativa. Esto es, tanto para empresas pequeñas, medianas y

grandes, el uso de *Big Data* y *Data Science* mejora los resultados financieros de las empresas. Lo anterior sugiere que los tomadores de decisiones de las empresas obtendrán beneficios al invertir en estas tecnologías.

Segunda conclusión:

Esta conclusión responde al segundo objetivo específico que propusimos: determinar el efecto de *Big Data* y *Data Science* en empresas de acuerdo a su tamaño.

De acuerdo a la Tabla 4, el uso de *Big Data* y *Data Science* es cinco veces mejor en las empresas grandes comparadas con las pequeñas, mientras que en las empresas medianas es tres veces mejor comparadas con las pequeñas. Aunque todas las empresas se benefician del uso de *Big Data* y *Data Science*, al aumentar el tamaño de la empresa se observan mayores beneficios. De acuerdo a Davenport y Dyché (2013) lo anterior puede ser explicado debido a que las empresas grandes han tenido que lidiar con *Big Data* y *Data Science* por más tiempo que empresas de menor tamaño.

Tercera conclusión:

Esta conclusión responde al tercer objetivo específico: determinar la duración del impacto de *Big Data* y *Data Science* en el tiempo.

De acuerdo a la Tabla 5, encontramos que para las empresas medianas y grandes el efecto de *Big Data* y *Data Science* se sigue manifestando hasta un año posterior. En todos los casos no encontramos evidencia de impacto después del segundo año. Esto es consistente con Brynjolfsson, et al. (2011) y Schryen (2013) donde estudian el efecto de

las tecnologías de la información a lo largo del tiempo. Lo anterior señala que el impacto de la adopción de estas tecnologías es de corto y mediano plazo, lo que sugiere que las empresas deben estar en un proceso constante de innovación para la creación de valor.

Cuarta conclusión:

Esta conclusión responde al primer objetivo específico que nos planteamos: cuantificar el uso de *Big Data* y *Data Science* en las empresas que cotizan en las Bolsas de Valores de los Estados Unidos.

De acuerdo a la Tabla 2, cuantificamos el uso de cada una de las palabras clave en el periodo de 2012 a 2015. Encontramos que el uso de *Big Data* y *Data Science* ha pasado de un 2.62% en 2012 a un 10% en 2015, lo que implica un incremento de casi cuatro veces. Lo anterior es consistente con lo encontrado por Dresner Advisory Services (2017), quienes situaron la adopción en 17% en 2015. Por lo tanto, la adopción de estas tecnologías ha crecido año con año, por lo que los datos sugieren que las empresas que no inviertan en estas tecnologías corren el riesgo de estar en desventaja con sus competidores.

Quinta conclusión:

Esta conclusión responde al cuarto y último objetivo específico que nos planteamos: Analizar el efecto de los principales componentes de *Big Data* y *Data Science* en la Q de Tobin.

De acuerdo a la Tabla 6, encontramos que *Hadoop* e Inteligencia Artificial son significativos para las empresas de cualquier tamaño. Esto es, los datos sugieren que las empresas obtienen mejores resultados al utilizar tecnologías específicas para manipulación de datos en conjunto con todos los métodos posibles de análisis de inteligencia artificial, ya sean métodos tradicionales o vanguardistas. Lo anterior es consistente con Provost y Fawcett (2013) que argumentaban que la generación de valor de *Big Data* se encuentra al extraer la información mediante *Data Science*.

Sexta conclusión:

De acuerdo a la Tabla 6, los datos sugieren que invertir en tecnologías extremadamente de vanguardia como *Deep Learning* puede restar valor. Lo anterior puede estar justificado por el dilema entre exploración y explotación de nuevas tecnologías (Belderbos, 2010), donde el efecto de explorar nuevas tecnologías no necesariamente se refleja en mejor desempeño financiero. Los tomadores de decisiones en las empresas deben tener cuidado con las inversiones en tecnologías de vanguardia, ya que por si solas no generan valor para el negocio.

Séptima conclusión:

De acuerdo a la Tabla 6, encontramos que en lo relativo al procesamiento de datos las empresas obtienen mejores resultados si se invierte en tecnologías específicas. Esto es, invertir particularmente en *Hadoop* tiene mejores resultados que referirse a conceptos generales de *Big Data*. Lo anterior es consistente con Tambe (2014), donde encontré beneficios de crecimiento de productividad en las empresas, específicamente con el uso

de *Hadoop*. Por consiguiente, los datos sugieren que el mercado premia mejor aquellas inversiones en tecnologías puntuales en relación a inversiones donde no es claro el producto o la tecnología específica.

6.2. Líneas sugeridas de investigación

Para líneas futuras de investigación se sugiere ampliar el ámbito de estudio, para no sólo incluir empresas públicas de los Estados Unidos, sino también incluir países adicionales. De particular relevancia sería realizar un estudio semejante en México para poder contrastar el uso y los efectos de *Big Data* y *Data Science* en nuestro país, comparado con Estados Unidos. Por último, ampliar el estudio a países latinoamericanos permitiría comprender de mejor manera el uso de estas tecnologías en nuestra región, ya que desde el ámbito académico se han realizado contribuciones relevantes en estos campos (Fernández, 2014).

También se sugiere el estudio del desempeño de empresas por sector económico, para analizar si existen diferencias entre los diferentes giros industriales de las compañías. Lo anterior buscaría ampliar el trabajo realizado por Manyika, et al. (2011), aunque extendiendo su ámbito para incorporar conceptos adicionales a *Big Data*.

Por último, también se sugiere ampliar el ámbito del estudio al contemplar palabras clave adicionales como *Internet of Things* e Industria 4.0 (Lasi, 2014). Lo anterior es particularmente relevante para el estado de Nuevo León, ya que en 2017 el gobierno del

estado propuso la iniciativa Nuevo León 4.0³³ que busca fomentar el desarrollo económico del estado mediante la innovación en tecnologías afines a Industria 4.0.

³³ Para más información, visitar: <https://www.nuevoleon40.org/>

REFERENCIAS BIBLIOGRÁFICAS

- Accenture. (2014). *Big success with Big Data* [Whitepaper]. Disponible en <http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Big-Data-POV.PDF>
- Agarwal, R. y Dhar, V. (2014). Big Data, Data Science, and analytics: the opportunity and challenge for IS research. *Information Systems Research*, 25(3): 443-448.
- Archak, N., Ghose A., y Ipeirotis, P.G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485-1509.
- ATIS. (2007). ATIS Telecom Glossary. *Alliance for telecommunications industry solutions* [Manual]. Disponible en <http://www.atis.org/glossary/>
- Belderbos, R., Faems, D., Leten, B., y Looy, B.V. (2010). Technological activities and their impact on the financial performance of the firm: exploitation and exploration within and between firms. *Journal of Product Innovation Management*, 27(6), 869-882.
- Bardhan, I., Krishnan, V., y Lin, S. (2013). Business value of Information Technology: testing the interaction effect of IT and R&D on Tobin's Q. *Information Systems Research*, 24(4), 1147-1161.
- Bharadwaj, A. S., Bharadwaj, S.G. y Konsynski, B.R. (1999). Information technology effects on firm performance as measured by Tobin's Q. *Management Science*, 45(7), 1008-1024.
- Blackwell, D. (1953). Equivalent comparison of experiments. *The Annals of Mathematical Statistics*, 24(2), 265-272.
- Boyd, D. y Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662-679.
- Brynjolfsson, E., Hitt, L.M., y Kim, H.H. (2011). Strength in numbers: wow does data-driven decisionmaking affect firm performance? *SSRN Electronic Journal*, 1, doi:10.2139/ssrn.1819486.
- Brynjolfsson, E., Hitt, L., y Yang, S. (2002). Intangible assets: computers and organizational capital. *Brookings Papers on Economic Activity*, 2002(1), 137-181.

- Centre for Economics and Business Research. (2012). *Data equity: unlocking the value of Big Data* [Whitepaper]. Disponible en <http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf>
- Chollet, F. (2017). *Deep learning with Python*. Greenwich: Manning Publications Company.
- Chung, K., & Pruitt, S. (1994). A simple approximation of Tobin's Q. *Financial Management*, 23(3), 70-74.
- Coad, A., y Rao, R. (2006). Innovation and market value: a quantile regression analysis. *Economics Bulletin*, 15(13), 1-10.
- Codd, E.F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387.
- DaDalt, P., Donaldson, J. y Garner, J. (2003). Will any Q do? Firm characteristics and divergences in estimates of Tobin's Q. *Journal of Financial Research*, 26(4), 535-551.
- Datameer. (2014). *Big Data: a competitive weapon for the enterprise* [Whitepaper]. Disponible en http://www.datameer.com/pdf/infographics/State_of_the_Industry.pdf
- Davenport, T.H. and Dyché, J., 2013. Big Data in big companies. *International Institute for Analytics*, 3.
- Davenport, T. H. y Harris, J. G. (2007). *Competing on analytics: the new science of winning*. Boston: Harvard Business School Press.
- Davenport, T.H., y Patil, D.J. Patil. (2012). Data Scientist: the sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70-76.
- Davenport, T.H. (2014). *Big Data at work: dispelling the myths, uncovering the opportunities*. Boston: Harvard Business Review Press.
- Dean, J. y Ghemawat, S. (2004). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.
- Dedrick, J., Gurbazani, V. y Kraemer, K.L. Information technology and economic performance: a critical review of the empirical evidence. *ACM Computing Surveys*, 35(1), 1-28.

- Dehning, B. y Richardson, V.J. (2002). Returns on investments in information technology: a research synthesis. *Journal of Information Systems*, 16(1), 7-30.
- Devlin, B., Rogers, S. y Myers, J. (2012). *Big Data comes of age* [Whitepaper]. Disponible en http://www-03.ibm.com/systems/hu/resources/big_data_comes_of_age.pdf
- Dhar, V. (2013). Data Science and prediction. *Communications of the ACM*, 56(12), 64-73.
- Diebold, F.X. (2012). A personal perspective on the origin(s) and development of 'Big Data': the phenomenon, the term, and the discipline. *PIER Working Paper No. 13-003*, doi:10.2139/ssrn.2202843.
- Dresner Advisory Services. (2017). *Big Data Analytics market study* [Whitepaper]. Disponible en <http://www.bigdataanalyticsreport.com/>
- Dumbill, E. (2013). Making sense of Big Data. *Big Data*, 1(1), 1-2.
- Economica. (2016). Pre-saved screens for data analysis of USA Stocks. Recuperado el 7 de noviembre de 2016 de la base datos Economica.
- Fernández, A., Gómez, Á., Lecumberry, F., Pardo, A. y Ramírez, I. (2014). Pattern recognition in Latin America in the Big Data era. *Pattern Recognition*, 48(4), 1185-1196.
- Galbraith, J.R. (1974). Organization design: an information processing view. *Interfaces*, 4(3), 28-36.
- Geyskens, I., Gielens, K., and Dekimpe, M.G. 2002. The market valuation of internet channel additions. *Journal of Marketing*, 66(2), 102-119.
- General Electric y Accenture. (2015). *Industrial Internet insights report for 2015* [Whitepaper]. Disponible en <http://www.accenture.com/sitecollectiondocuments/pdf/accenture-industrial-internet-changing-competitive-landscape-industries.pdf>
- Ghemawat, S., Howard, G., y Shun-Tak, L. (2003). The google file system. *Proceedings of the 17th ACM Symposium on Operating Systems Principles*, 29-43.
- Ghose, A., Ipeirotis, P.G., y Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content. *Marketing Science*, 31(3), 493-520.

- González-Bailón, S. (2013). Social Science in the era of Big Data. Policy & Internet. *Policy & Internet*, 5(2), 147-160.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. Cambridge: MIT Press.
- Graham, C., Dharmasthira, Y., Eschinger, C., Granetto, B.F., Correia, J.M, Wurster, L.F., De Silva, F., Eid, T., Contu, R., Biscotti, F., Pang, C., Horiuchi, H., Coyle, D.M., Sommer, D., Cheung, M., Swinehart, H.H., Sood, B., Raina, A., Nagashima, Y., Arcaris, C., Zhang, J., Warrilow, M. y Motoyoshi, K. (2013). *Market share: all software markets, worldwide, 2012*. ID:G00250553. Recuperado de la base de datos de Gartner.
- Groenfeldt, T. (2012). *Morgan Stanley takes on Big Data with Hadoop* [Whitepaper]. Disponible en <http://www.forbes.com/sites/tomgroenfeldt/2012/05/30/morgan-stanley-takes-on-big-data-with-hadoop/>
- Han, J., Kamber, M. y Pei. (2011). *Data Mining: concepts and techniques*. Waltham: Morgan Kaufmann.
- Hall, R.E. (2000). E-capital: the link between the stock market and the labor market in the 1990s. *Brookings Papers on Economic Activity*, 2000(2), 73-118.
- Hall, R.E. (2001). The stock market and capital accumulation. *The American Economic Review*, 91(5), 1185-1202.
- Hall, B.H., Jaffe, A., y Trajtenberg, M. 2005. Market value and patent citations. *The RAND Journal of Economics*, 36(1), 16-38.
- Hernández, R., Fernández, C., y Baptista, M. (2010). *Metodología de la investigación*. México: McGraw-Hill.
- Hilbert, M. (2016). Big Data for development: a review of Ppromises and challenges. *Development Policy Review*, 34(1), 135-174
- IDC. (2007). *The expanding digital universe* [Whitepaper]. Disponible en <http://www.idema.org/wp-content/downloads/1785.pdf>

- IDC. (2012). *The Digital Universe in 2020: Big Data, bigger digital shadows, and biggest growth in the Far East* [Whitepaper]. Disponible en <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- IDC. (2014). *The digital universe of opportunities* [Whitepaper]. Disponible en <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf>
- Jacobs, A. (2009). The pathologies of Big Data. *Communications of the ACM*, 52(8), 36-44.
- Kimball, R. (1996). *The Data Warehouse toolkit: practical techniques for building dimensional Data Warehouses*. New York: John Wiley & Sons.
- Kitchin, R. (2014). The real-time City? Big Data and smart urbanism. *GeoJournal*, 79(1), 1-14.
- Kitchin, R. y Lauriault, T.P. (2015). Small data in the era of Big Data. *GeoJournal*, 80(4), 463-475.
- Koenker, R. y Bassett, G. (1978). Regression Quantiles. *Econometrica* 46, 33-50.
- Hillegersberg, J.V. y Kumar K. (2000). ERP experience and evolution. *Communications of the ACM*, 43(4): 22-26.
- Laney, D. (2001). 3-D Data management: controlling data volume, velocity and variety. *META group research note*, 6(70), 1.
- Lang, L.H.P. y Stulz, R.M. (1994). Tobin's Q, corporate diversification, and firm performance. *The Journal of Political Economy*, 102(6), 1248-1280.
- Lasi, H., Fettke, P., Kemper, H.G., Feld, T. and Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, 6(4), 239-242.
- Lev, B. (2001). *Intangibles: management, measurement, and reporting*. Washington: Brookings Institution Press.
- Lewis, M. (2003). *Moneyball: the art of winning an unfair game*. New York: W.W. Norton.
- Lindenberg, E.B. y Ross, S.A. (1981). Tobin's Q ratio and industrial organization. *Journal of Business*, 54, 1-32.

- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung-Byers, A. (2011). *Big Data: the next frontier for innovation, competition and productivity* [Whitepaper]. Disponible en "http://www.mckinsey.com/mgi/publications/big_data/index.asp"
- Mayer-Schönberger, V. y Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. London: John Murray.
- McGahan, A.M. 1999. The performance of US corporations: 1981-1994. *The Journal of Industrial Economics*, 47(4), 373-398.
- McKinsey Center for Business Technology. (2012). *Perspectives on digital business* [Whitepaper]. Disponible en "http://www.mckinsey.com/~media/mckinsey/dotcom/client_service/BTO/PDF/MCBT_Compensum_Perspectives_on_Digital_Business.ashx"
- Melville, N., Kraemer, K. y Gurbaxani, V. (2004). Review: information technology and organizational performance: an integrative model of IT business value. *MIS Quarterly* 28(2), 283–322.
- Oracle. (2007). Oracle celebrates 30 years. *Profit*, 12(2), 26-33.
- Pang, C., Dharmasthira, Y., Eschinger, C., Kenneth, F.B. y Motoyoshi, K. Gartner. (2014). *Market share analysis: ERP software, worldwide, 2013* [Whitepaper]. ID:G00263541.
- Popper, K. (1963). *Conjectures and refutations*. Londres: Routledge.
- Provost, F. y Fawcett, T. (2013). Data Science and its relationship to Big Data and data-driven decision-making. *Big Data*, 1(1), 51-59.
- Quantitative Micro Software. (2010). Eviews 7.1 [Computer Software].
- Silberschatz, A., Stonebraker, M. y Ullman, F. (1996). Database research: achievements and opportunities into the 21st century. *ACM Sigmod Record*, 25(1), 52-63.
- Saunders, A. y Tambe, P. (2015). A data-driven digital divide? Evidence from 10-K filings. *SSRN Electronic Journal*, 1, doi:10.2139/ssrn.2537089.
- Siegel, E. (2013). *Predictive analytics*. Hoboken: Wiley.

- Schryen, G. (2013). Revisiting IS business value research: what we already know, what we still need to know, and how we can get there. *European Journal of Information Systems*, 22(2), 139-169
- Smirlock, M., Gilligan, T., y Marshall, W. (1984). Tobin's Q and the structure-performance relationship. *The American Economic Review*, 74(5), 1051-1060.
- Soh, C. y Markus, M.L. (1995). How IT creates business value: a process theory synthesis. *ICIS 1995 Proceedings*, 4, 29-41.
- Tambe, P. (2014). Big Data investment, skills, and firm value. *Management Science*, 60(6), 1452-1469.
- Tata Consultancy Services. (2013). *The emergins big returns of Big Data* [Whitepaper]. Disponible en http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf
- Tech Pro Research. (2015). *The power of IoT and Big Data* [Whitepaper]. Disponible en <http://www.techproresearch.com/downloads/the-power-of-big-data-and-iot/>
- Tobin, J. (1969). A general equilibrium approach to Monetary Theory. *Journal of Money, Credit and Banking*, 1(1), 15-29.
- Villalonga, B. (2004). Intangible resources, Tobin's Q, and sustainability of performance differences. *Journal of Economic Behavior & Organization*, 54(4), 205-230.
- Wagner, K., Taylor, A., Zablit, H. y Foo, E. (2014). The most innovative companies [Whitepaper]. Disponible en https://www.bcgperspectives.com/content/articles/innovation_growth_digital_economy_innovation_in_2014/
- Weiss, S.M. y Indurkha, N. (1998). *Predictive Data Mining: a practical guide*. San Francisco: Morgan Kaufmann Publishers.
- Zhao, R. (2002). Relative value relevance of R&D reporting: an international comparison. *Journal of International Financial Management and Accounting*, 13(2), 153-174.

Zion Market Research. (2017). *Hadoop market by type (software, hardware and services) for BFSI, government sector, IT & ITES, healthcare, telecommunication, retail and others end-uses: global industry perspective, comprehensive analysis, size, share, growth, segment, trends and forecast, 2016 – 2022* [Whitepaper]. Disponible en <https://www.zionmarketresearch.com/report/hadoop-market>

APENDICES

Tabla 7. Listado de empresas analizadas

1 800 Flowers Com Inc	3d Systems Corp
3M Company	Aaron'S Inc
Abbott Laboratories	Abbvie Inc.
Abercrombie & Fitch	Abiomed Inc
Acadia Healthcare Company, Inc.	Acadia Pharmaceuticals Inc
Accenture Plc	Activision Blizzard, Inc.
Acuity Brands Inc	Acxiom Corp
Adobe Systems	Adtran Inc
Advance Auto Parts Inc	Advanced Micro Devices Inc
Advansix Inc.	Aecom
AES Corp	Aetna Inc /Pa/
Affiliated Managers Group, Inc.	Aflac Inc
Agco Corp /De	Agilent Technologies
Agios Pharmaceuticals Inc	Air Lease Corp
Air Products & Chemicals	Aircastle Ltd
Ak Steel Holding Corp	Akamai Technologies Inc
Akorn Inc	Alaska Air Group, Inc.
Albemarle Corp	Alere Inc.
Alexandria Real Estate Equities Inc	Alexion Pharmaceuticals
Align Technology Inc	Alkermes Plc.
Alleghany Corp /De	Allegheny Technologies Incorporated
Allegion Plc	Allergan Plc
Allete Inc	Alliance Data Systems
Alliant Energy Corp	Allied World Assurance Co Holdings, Ag
Allison Transmission Holdings Inc	Allscripts Healthcare Solutions, Inc.
Allstate Corp	Ally Financial Inc.
Alnylam Pharmaceuticals, Inc.	Alphabet Inc.
Alphabet Inc.	Altria Group, Inc.
Amazon Com Inc	AMC Networks Inc.
Amdocs	Amerco /NV/
Ameren Corp	American Airlines Group Inc.
American Axle & Manufacturing Holdings Inc	American Campus Communities Inc
American Capital Agency	American Capital, Ltd
American Eagle Outfitters	American Electric Power Co Inc
American Express Company	American Financial Group Inc
American Homes 4 Rent	American International Group Inc
American National Insurance Co /Tx/	American Tower Corp /Ma/
American Water Works Company, Inc.	Ameriprise Financial Inc

Tabla 8. Listado de empresas analizadas (continuación)

Amerisourcebergen	Ametek Inc/
Amgen Inc.	Amkor Technology, Inc.
AMN Healthcare Services Inc	Amphenol Corp /De/
Amsurg Corp	Amtrust Financial Services, Inc.
Anadarko Petroleum Corp	Analog Devices Inc
Anheuser Busch Inbev	Annaly Capital Management
Ansys Inc	Antero Resources Corp
Anthem, Inc.	Aon Plc
Apache Corp	Apartment Investment & Management Co
Apollo Education Group Inc	Apple Hospitality Reit, Inc.
Apple Inc	Applied Materials
Applied Micro Circuits Corp	Aptargroup Inc
Aqua America Inc	Aramark
Arcbest Corp /De/	Arch Capital Group
Archer Daniels Midland Co	Archrock
Arconic Inc.	Ares Capital Corp
Ariad Pharmaceuticals Inc	Arista Networks, Inc.
Arlington Asset Investment Corp.	Armstrong World Industries Inc
Arris Group	Arrow Electronics Inc
Artisan Partners Asset Management Inc.	Asbury Automotive Group Inc
Ascena Retail Group	Ashland
Aspen Insurance Holdings Ltd	Associated Banc-Corp
Assurant Inc	Assured Guaranty Ltd
Astoria Financial Corp	AT&T Inc.
Athenahealth Inc	Atmos Energy Corporation
Atwood Oceanics Inc	Autodesk
Autoliv Inc	Automatic Data Processing Inc
Autonation, Inc.	Autozone Inc
Avalonbay Communities Inc	Avangrid, Inc.
Avery Dennison Corp	Avis Budget Group, Inc.
Avnet Inc	Avon Products Inc
AVX Corp	Axalta Coating Systems Ltd.
Axis Capital Holdings Ltd	B/E Aerospace Inc
Baker Hughes Inc	Ball Corp
Bancorpsouth Inc	Bank Of America Corp /De/
Bank Of Hawaii Corp	Bank Of New York Mellon Corp
Bankunited, Inc.	Bard C R Inc /Nj/
Barnes & Noble Inc	Baxter International Inc
BB&T Corp	Beazer Homes USA Inc

Tabla 9. Listado de empresas analizadas (continuación)

Bebe Stores, Inc.	Becton Dickinson
Bed Bath & Beyond	Belmond Ltd.
Bemis Co Inc	Berkley W R Corp
Berkshire Hathaway Inc	Berkshire Hathaway Inc
Berry Plastics Group	Best Buy
Big Lots	Bio Rad Laboratories Inc
Biogen Inc.	Biomarin Pharmaceutical Inc
Bioscrip, Inc.	Bio-Techne Corp
Black Box Corp	Black Knight Financial Services, Inc.
Blackrock Inc.	Blue Buffalo Pet Products, Inc.
Bluebird Bio, Inc.	Boeing Co
Bok Financial Corp Et Al	Booz Allen Hamilton Holding Corp
Borgwarner Inc.	Boston Properties Inc
Boston Scientific Corporation	Boyd Gaming Corp
Brandywine Realty Trust	Briggs & Stratton Corp
Brinker International Inc	Brinks Co
Bristol Myers Squibb Co	Brixmor Property Group Inc.
Broadcom Ltd	Broadridge Financial Solutions, Inc.
Brocade Comm Systems	Brookdale Senior Living Inc.
Brooks Automation	Brown & Brown Inc
Brown-Forman	Brown-Forman
Bruker Corp	Brunswick Corp
Bunge Ltd	Burlington Stores
Bwx Technologies, Inc.	C H Robinson Worldwide Inc
Ca, Inc.	Cabelas Inc
Cable One, Inc.	Cabot
Cabot Microelectronics Corp	Cabot Oil & Gas Corp
Caci International Inc /De/	Cadence Design Systems Inc
Calatlantic Group, Inc.	California Resources Corp
Callaway Golf Co	Calpine Corp
Camden Property Trust	Campbell Soup
Capital One Financial Corp	Capitol Federal Fnanical
Carbo Ceramics Inc	Cardinal Health Inc
Care Capital Properties, Inc.	Career Education Corp
Carlisle Companies Inc	Carmax
Carnival	Carpenter Technology Corp
Carters Inc	Caseys General Stores
Caterpillar Inc	Cato Corp
CBL & Associates Properties Inc	CBOE Holdings, Inc.
CBRE Group, Inc.	CBS Corporation

Tabla 10. Listado de empresas analizadas (continuación)

CDK Global, Inc.	CDW Corp
CEB Inc.	Celanese Corporation
Celestica	Celgene Corp /De/
Centene Corp	Centerpoint Energy Inc.
Central European Media Enterprises Ltd.	Century Aluminum Co
Centurylink, Inc	Cerner Corp /Mo/
CF Industries Holdings, Inc.	Charles River Laboratories International Inc
Charter Communications, Inc. /Mo/	Cheesecake Factory Inc
Cheniere Energy Inc	Chesapeake Energy Corporation
Chevron Corp	Chicago Bridge & Iron Co N V
Chicos Fas Inc	Childrens Place
Chimera Investment	Chipotle Mexican Grill Inc
Choice Hotels International Inc /De	Christopher & Banks
Chubb Ltd	Church & Dwight Co Inc /De/
Ciena	Cigna Corporation
Cimarex Energy Co	Cimpress N.V.
Cincinnati Financial Corp	Cinemark Holdings, Inc.
Cintas	Cisco Systems, Inc.
Cit Group Inc	Citigroup Inc
Citizens Financial Group Inc/Ri	Citrix Systems Inc
Clean Harbors Inc	Clear Channel Outdoor Holdings, Inc.
Cliffs Natural Resources Inc.	Clorox Co /De/
CME Group Inc.	CMS Energy Corp
CNA Financial Corp	CNO Financial Group
Coach Inc	Cobalt International Energy, Inc.
Coca Cola Co	Coca Cola European Partners
Cognex Corp	Cognizant Technology Solutions Corp
Colfax Corp	Colgate Palmolive Co
Columbia Property Trust, Inc.	Columbia Sportswear Co
Comcast Corp	Comerica Inc /New/
Commerce Bancshares Inc /Mo/	Commercial Metals Co
Commscope Holding Company, Inc.	Communications Sales & Leasing, Inc.
Community Health Systems Inc	Compass Minerals International Inc
Computer Sciences Corp	Comstock Resources Inc
Conagra	Concho Resources Inc
Conocophillips	Consol Energy Inc
Consolidated Edison Inc	Constellation Brands
Continental Resources, Inc	Convergys Corp.
Cooper Companies	Cooper Tire & Rubber Co
Copart	Core Laboratories N V

Tabla 11. *Listado de empresas analizadas (continuación)*

Corelogic, Inc.	Corning Inc /NY
Corporate Office Properties Trust	Corrections Corp Of America
Costar Group Inc	Costco Wholesale
Coty Inc.	Cousins Properties Inc
Covanta Holding Corp	Cracker Barrel Old Country Store, Inc
Crane Co /De/	Credit Acceptance Corp
Cree Inc	Crocs, Inc.
Crown Castle International Corp	Crown Holdings Inc
CSG Systems International Inc	Csra Inc.
CST Brands, Inc.	CSX
Cubsmart	Cullen/Frost Bankers, Inc.
Cummins Inc	Cumulus Media Inc
CVR Energy Inc	CVS Health Corporation
Cypress Semiconductor Corp /De/	Cyrusone Inc.
Danaher Corp /De/	Darden Restaurants
Davita Inc.	DCT Industrial Trust Inc.
DDR Corp	Dean Foods Co
Deckers Outdoor Corp	Deere & Co
Delphi Automotive Plc	Delta Air Lines
Deluxe Corp	Denbury Resources Inc.
Dentsply Sirona Inc.	Devon Energy Corp/De
Devry Education Group Inc.	Dexcom Inc
Diamond Offshore Drilling Inc	Diamondback Energy, Inc.
Dicks Sporting Goods Inc	Diebold Inc
Digital Realty Trust, Inc.	Dillard
Discover Financial Services	Discovery Communications, Inc.
Discovery Communications, Inc.	Dish Network Corp
Dolby Laboratories, Inc.	Dollar General
Dollar Tree	Dominion Resources Inc /Va/
Dominos Pizza Inc	Domtar Corp
Donaldson	Douglas Emmett Inc
Dover Corp	Dow Chemical Co /De/
Dr Pepper Snapple Group, Inc.	Dril-Quip Inc
DST Systems Inc	DSW
DTE Energy Co	Duke Energy Corporation
Duke Realty Corp	Dun & Bradstreet Corp/Nw
Dunkin' Brands Group, Inc.	Dupont E I De Nemours & Co
E Trade Financial Corp	E.W. Scripps Co
Eagle Materials Inc	East West Bancorp Inc
Eastman Chemical Co	Eaton Corp Plc

Tabla 12. Listado de empresas analizadas (continuación)

Eaton Vance	Ebay Inc
Echostar Corp	Ecolab Inc.
Edison International	Edwards Lifesciences Corp
Electronic Arts Inc.	Emerson Electric Co
Emmis Comm	Empire State Realty Trust, Inc.
Endo International Plc	Endurance Specialty Holdings Ltd.
Energen Corp	Energizer Holdings, Inc.
Enesco Plc	Entercom Communications Corp.
Entergy Corp /De/	Entravision Communications Corp
Envision Healthcare Holdings, Inc.	EOG Resources Inc
Ep Energy Corp	EPR Properties
EQT Corp	Equifax Inc
Equinix Inc	Equity Commonwealth
Equity Lifestyle Properties Inc	Equity One, Inc.
Equity Residential	Erie Indemnity
Essex Property Trust Inc	Estee Lauder Companies Inc
Ethan Allen Interiors Inc	Euronet Worldwide Inc
Everest Re Group Ltd	Eversource Energy
Evine Live	Exco Resources Inc
Exelon Corp	Expedia, Inc.
Expeditors International Of Washington Inc	Express Scripts Holding Co.
Extended Stay America, Inc.	Extra Space Storage Inc.
Extreme Networks Inc	Exxon Mobil Corp
F5 Networks Inc	Facebook Inc
Factset Research Systems Inc	Fair Isaac Corp
Fastenal	Federal Realty Investment Trust
Federal-Mogul Holdings Corporation	Federated Investors Inc /Pa/
Fedex	Ferrellgas Partners L P
Fidelity National Financial, Inc.	Fidelity National Information Services, Inc.
Fifth Third Bancorp	Finish Line
Fireeye, Inc.	First American Financial Corp
First Bancorp /Pr/	First Citizens Bancshares Inc /De/
First Data Corp	First Horizon National Corp
First Industrial Realty Trust Inc	First Midwest Bancorp Inc
First Republic Bank	First Solar, Inc.
Firstenergy Corp	Fiserv Inc
Fitbit Inc	Fleetcor Technologies Inc
Flex Ltd.	Flir Systems Inc
Flowers Foods Inc	Flowserve Corp
Fluor Corp	FMC Corporation

Tabla 13. *Listado de empresas analizadas (continuación)*

FMC Technologies Inc	FNB Corp/Pa/
Foot Locker	Ford Motor Co
Forest City Realty Trust, Inc.	Fortinet Inc
Fortive Corp	Fortune Brands Home & Security, Inc.
Fossil Group, Inc.	Franklin Resources
Frank'S International N.V.	Freds
Freeport-Mcmoran Inc	Frontier Communications Corp
FTI Consulting Inc	Fulton Financial Corp
Gallagher Arthur J & Co	Gamestop
Gaming & Leisure Properties, Inc.	Gap
Garmin Ltd	Gartner Inc
Gatx Corporation	General Cable Corp /De/
General Dynamics Corporation	General Electric Company
General Growth Properties, Inc.	General Mills
General Motors Company	Genesee & Wyoming Inc.
Genpact Ltd	Gentex Corporation
Genuine Parts Co	Genworth Financial Inc
Gilead Sciences Inc	Global Payments
GNC Holdings, Inc.	Godaddy Inc.
Golar Lng Ltd	Goldman Sachs Group Inc
Goodyear Tire & Rubber Co /Oh/	Gopro, Inc.
Graco Inc.	Graham Holdings Co
Grainger W W Inc	Granite Construction Inc
Graphic Packaging Holding Co	Great Plains Energy Inc
Green Dot Corp	Greenhill & Co Inc
Greif	Group 1 Automotive Inc
Groupon, Inc.	Guess
Guidewire Software	Gulfport Energy Corp
H&R Block	Hain Celestial Grp
Half Robert International Inc /De/	Halliburton Company
Hanesbrands Inc.	Hanover Insurance Group, Inc.
Harley Davidson Inc	Harman International Industries Inc /De/
Harris Corp /De/	Harsco Corp
Harte Hanks Inc	Hartford Financial Services Group Inc/De
Hasbro Inc	Hawaiian Electric Industries Inc
HCA Holdings, Inc.	HCP, Inc.
Hd Supply Holdings	Healthcare Realty Trust Inc
Healthcare Trust Of America, Inc.	Heico Corporation
Heico Corporation	Helix Energy Solutions Group Inc
Helmerich & Payne	Henry Jack & Associates Inc

Tabla 14. *Listado de empresas analizadas (continuación)*

Henry Schein Inc	Herbalife Ltd.
Herc Holdings Inc.	Hershey Co
Hess Corp	Hewlett Packard Enterprise
Hexcel Corp /De/	Highwoods Properties Inc.
Hillenbrand, Inc.	Hill-Rom Holdings, Inc.
Hilton Worldwide Holdings Inc.	HNI Corp
Hollyfrontier Corp	Hologic Inc
Home Depot Inc	Honeywell International Inc
Hormel Foods	Horton D R
Hospitality Properties Trust	Host Hotels & Resorts, Inc.
Hovnanian Entrp	Howard Hughes Corp
Hp	Hubbell Incorporated
Humana	Hunt J B Transport Services Inc
Huntington Bancshares Inc/Md	Huntington Ingalls Industries, Inc.
Huntsman Corp	Hyatt Hotels Corp
IAC/Interactivecorp	Icu Medical Inc/De
Idacorp Inc.	Idex Corp /De/
Idexx Laboratories Inc /De	IDT
Illinois Tool Works Inc	Illumina Inc
Imation Corp	Incyte Corp
Ingersoll-Rand Plc	Ingram Micro Inc
Ingredion Inc	Innoviva, Inc.
Inovalon Holdings, Inc.	Insight Enterprises Inc
Integrated Device Technology Inc	Intel Corp
Interactive Brokers	Intercept Pharmaceuticals Inc
Intercontinental Exchange, Inc.	Interdigital, Inc.
International Bancshares Corp	International Business Machines Corp
International Flavors & Fragrances Inc	International Paper Co /New/
Interpublic Group Of Companies, Inc.	Intersil Corp/De
Intl Speedway	Intrepid Potash, Inc.
Intrexon Corp	Intuit Inc
Intuitive Surgical Inc	Invesco Ltd.
Investment Technology Group, Inc.	Ionis Pharmaceuticals Inc
IPG Photonics Corp	Iron Mountain Inc
Itron Inc /Wa/	ITT Inc.
J M Smucker Co	Jabil Circuit Inc
Jacobs Engineering Group	Janus Capital Group Inc
Jazz Pharmaceuticals Plc	Jetblue Airways Corp
Johnson & Johnson	Johnson Controls International
Jones Lang Lasalle Inc	Joy Global

Tabla 15. *Listado de empresas analizadas (continuación)*

Jpmorgan Chase & Co	Juniper Networks Inc
Juno Therapeutics, Inc.	Kansas City Southern
KAR Auction Services, Inc.	Kate Spade & Co
Kb Home	KBR, Inc.
Kellogg Co	Kemper Corporation
Kennametal Inc.	Keycorp /New/
Keysight Technologies	Kilroy Realty Corp
Kimberly Clark Corp	Kimco Realty Corp
Kinder Morgan, Inc.	Kirby Corp
Kla Tencor Corp	Knowles Corp
Kohls	Kosmos Energy Ltd.
Kraft Heinz Co	Kroger
Kronos Worldwide Inc	L 3 Communications Holdings Inc
L Brands	Laboratory Corp Of America Holdings
Lam Research Corp	Lamar Advertising Co/New
Lancaster Colony Corp	Landstar System Inc
Laredo Petroleum, Inc.	Las Vegas Sands Corp
Lattice Semiconductor Corp	Lazard Ltd
La-Z-Boy	Lear Corp
Lee Enterprises	Legg Mason, Inc.
Leggett & Platt Inc	Leidos Holdings, Inc.
Lendingclub Corporation	Lennar Corp /New/
Lennar Corp /New/	Lennox Intl
Leucadia National	Level 3 Communications Inc
Lexmark International Inc /Ky/	Liberty Broadband Corp
Liberty Broadband Corp	Liberty Property Trust
Liberty Siriusxm	Liberty Siriusxm
Liberty Ventures	Life Storage, Inc.
Lifepoint Health, Inc.	Lilly Eli & Co
Lincoln Electric Holdings Inc	Lincoln National Corp
Linear Technology Corp /Ca/	Linkedin Corp
Lions Gate Entertainment Corp /Cn/	Live Nation Entertainment, Inc.
LKQ Corp	Lockheed Martin Corp
Loews Corp	Louisiana-Pacific Corp
Lowes Companies	LPL Financial Holdings Inc.
Lululemon Athletica	Lyondellbasell Industries N.V.
M&T Bank Corp	Macerich Co
Mack Cali Realty Corp	Macquarie Infrastructure Corp
Macys	Madison Square Garden Co
Mallinckrodt	Manhattan Associates Inc

Tabla 16. *Listado de empresas analizadas (continuación)*

Manitowoc Co Inc	Manitowoc Foodservice, Inc.
Mannatech Inc	Manpowergroup Inc.
Marathon Oil	Marathon Petroleum Corp
Markel Corp	Marketaxess Holdings Inc
Marriott International Inc /Md/	Marsh & McLennan Companies, Inc.
Martin Marietta Materials Inc	Marvell Tech Group
Masco Corp /De/	Mastercard Inc
Match Group, Inc.	Matson, Inc.
Mattel Inc /De/	Maxim Integrated Products Inc
MBIA Inc	Mcclatchy Co
Mccormick & Co	Mcdermott International Inc
Mcdonalds Corp	Mckesson Corp
MDC Holdings Inc	MDU Resources Group Inc
Mead Johnson Nutrition Co	Media General Inc
Mednax, Inc.	Medtronic Inc
Merck & Co., Inc.	Mercury General Corp
Meredith Corp	Meritage Homes Corp
Meritor	Metlife Inc
Mettler Toledo International Inc/	MFA Financial, Inc.
MGIC Investment Corp	MGM Resorts International
Michael Kors Holdings Ltd	Michaels Companies
Microchip Technology Inc	Micron Technology Inc
Microsoft Corporation	Mid America Apartment Communities Inc
Middleby Corp	Miller Herman
MKS Instruments Inc	Mohawk Industries Inc
Molson Coors Brewing Co	Mondelez International, Inc.
Moneygram International Inc	Monsanto Co /New/
Monster Beverage Corp	Monster Worldwide, Inc.
Moodys Corp /De/	Morgan Stanley
Morningstar, Inc.	Mosaic Co
Motorola Solutions, Inc.	MRC Global Inc.
MSC Industrial Direct Co Inc	MSCI Inc.
Msg Networks Inc.	Murphy Oil Corp /De
Murphy USA Inc.	Mylan N.V.
Myriad Genetics Inc	Nabors Industries Ltd
Nasdaq, Inc.	National Fuel Gas Co
National Instruments Corp /De/	National Oilwell Varco Inc
National Retail Properties, Inc.	Nationstar Mortgage Holdings Inc.
Navient Corp	Navistar Intl
NCR Corp	Nektar Therapeutics

Tabla 17. Listado de empresas analizadas (continuación)

Nelnet Inc	Netapp
Netflix Inc	Netsuite Inc.
Neurocrine Biosciences Inc	Neustar Inc
New York Community Bancorp Inc	New York Times Co
Newell Brands Inc	Newfield Exploration Co /De/
Newmarket Corp	Newmont Mining Corp /De/
News Corp	News Corp
Nextera Energy Inc	Nielsen Holdings Plc
Nike Inc	Nisource Inc/De
Noble Corporation	Noble Energy Inc
Nordson	Nordstrom
Norfolk Southern Corp	Northern Trust Corp
Northrop Grumman Corp /De/	Northstar Asset Management Group Inc.
Northstar Realty Finance Corp.	Norwegian Cruise Line Holdings Ltd.
Now Inc.	NRG Energy, Inc.
Nu Skin Enterprises Inc	Nuance Comm
Nucor Corp	Nutri System Inc /De/
Nvidia Corp	NVR Inc
O Reilly Automotive Inc	Oasis Petroleum Inc.
Occidental Petroleum Corp /De/	Oceaneering International Inc
Ocwen Financial Corporation	Office Depot Inc
OGE Energy Corp.	Oil States International, Inc
Old Dominion Freight Line Inc/Va	Old National Banc
Old Republic International Corporation	Omega Healthcare
Omnicom Group Inc.	On Semiconductor Corporation
Onebeacon Insurance Group, Ltd.	Onemain Holdings, Inc.
Oneok Inc /New/	Opko Health, Inc.
Oracle Corp	Orbital ATK
Ormat Technologies, Inc.	Oshkosh Truck
Outfront Media Inc.	Owens Corning
Owens Illinois Inc /De/	Paccar Inc
Packaging Corp Of America	Pacwest Bancorp
Palo Alto Networks Inc	Pandora Media, Inc.
Panera Bread Company	Paramount Group, Inc.
Park National Corp /Oh/	Parker Hannifin Corp
Parsley Energy, Inc.	Patterson Cos
Patterson UTI Energy Inc	Paychex
Paypal Holdings, Inc.	PBF Energy Inc.
PDL Biopharma, Inc.	Penn National Gaming Inc
Penney (Jc)	Penske Automotive Group, Inc.

Tabla 18. *Listado de empresas analizadas (continuación)*

Pentair Plc	People'S United Financial, Inc.
Pepsico Inc	Perkinelmer Inc
Perrigo Co Plc	Pfizer Inc
PG&E Corp	Philip Morris International Inc.
Phillips 66	Photronics
Piedmont Office Realty Trust, Inc.	Pier 1 Imports
Pilgrims Pride Corp	Pinnacle Foods Inc.
Pinnacle West Capital Corp	Pioneer Natural Resources Co
Pitney Bowes Inc /De/	Plantronics Inc /Ca/
Platform Specialty Products Corp	Plexus Corp
PNC Financial Services Group, Inc.	PNM Resources Inc
Polaris Industries Inc/Mn	Pool Corp
Popular Inc	Post Holdings, Inc.
Post Properties Inc	Potash Corp Of Saskatchewan Inc
Potlatch Corp	PPG Industries Inc
PPL Corp	Praxair Inc
Premier, Inc.	Price T Rowe Group Inc
Priceline Group Inc.	Pricesmart Inc
Principal Financial Group Inc	Proassurance Corp
Procter & Gamble Co	Progressive Corp/Oh/
Prologis, Inc.	Prudential Financial Inc
PTC Inc.	Public Service Enterprise Group Inc
Public Storage	Pultegroup Inc/Mi/
Puma Biotechnology, Inc.	PVH Corp
QEP Resources, Inc.	Qiagen NV
Qorvo, Inc.	Qualcomm Inc/De
Quanta Services Inc	Quest Diagnostics Inc
Quintiles IMS Holdings, Inc.	Rackspace Hosting, Inc.
Radian Group Inc.	Radio One, Inc.
Ralph Lauren Corp	Rambus Inc
Range Resources Corp	Raymond James Finl
Rayonier Inc.	Raytheon Co/
Realnetworks Inc	Realogy Holdings Corp.
Realty Income Corp	Red Hat
Regal Beloit Corp	Regal Entertainment Group
Regency Centers Corp	Regeneron Pharmaceuticals Inc
Regions Financial Corp	Regis Corp
Reinsurance Group Of America Inc	Reliance Steel & Aluminum Co
Renaissancere Holdings Ltd.	Rent A Center Inc De
Republic Services, Inc.	Resmed Inc

Tabla 19. Listado de empresas analizadas (continuación)

Retail Properties Of America, Inc.	Reynolds American Inc
Rice Energy Inc.	Rite Aid
Rockwell Automation Inc.	Rockwell Collins Inc
Rollins Inc	Roper Technologies Inc
Ross Stores	Rowan Companies Plc
Royal Caribbean Cruises Ltd	Royal Gold Inc
RPC Inc	RPM International
Rr Donnelley & Sons Co	Ruby Tuesday
Ryder System Inc	S&P Global Inc.
Sabre Corp	Salesforce Com Inc
Sally Beauty Holdings, Inc.	Sanmina Corp
Santander Consumer USA Holdings Inc.	SBA Communications Corp
Scana Corporation	Schlumberger Limited/NV
Schnitzer Steel Industries Inc	Scholastic
Schwab Charles Corp	Scientific Games Corp
Scotts	Scripps Networks Interactive, Inc.
Seacor Holdings Inc /New/	Seadrill Limited
Seagate Technology Plc	Sealed Air Corp/De
Sears Hldgs	Seattle Genetics Inc /Wa
Seaworld Entertainment, Inc.	SEI Investments Company
Sempra Energy	Semtech
Senior Housing Properties Trust	Service Corporation International
Servicemaster Global Holdings Inc	Servicenow, Inc.
Sherwin Williams Co	Shire Plc
Signature Bank	Silgan Holdings Inc
Silicon Laboratories Inc	Simon Property Group Inc /De/
Sirius XM Holdings Inc.	Six Flags Entertainment Corp
Skechers USA Inc	Skywest Inc
Skyworks Solutions	SI Green Realty Corp
SLM Corporation	SM Energy Co
Smith A O Corp	Snap-On Inc
Sohu Com Inc	Solarcity Corp
Sonic Automotive Inc	Sonic Corp.
Sonoco Products Co	Sonus Networks Inc
Sothebys	Southern Co
Southern Copper Corp/	Southwest Airlines Co
Southwestern Energy Co	Spectra Energy Corp.
Spectrum Brands Holdings, Inc.	Spirit Aerosystems Holdings, Inc.
Spirit Airlines, Inc.	Spirit Realty Capital, Inc.
Splunk	Sprint Corp

Tabla 20. Listado de empresas analizadas (continuación)

Sprouts Farmers Market, Inc.	SPX Corp
Square, Inc.	Ss&C Technologies Holdings Inc
St Joe Co	St Jude Medical Inc
Stanley Black & Decker, Inc.	Staples Inc
Starbucks Corp	Starwood Property Trust
Starz	State Street Corp
Steel Dynamics Inc	Steelcase
Stericycle Inc	Steris Plc
Store Capital Corp	Strayer Education Inc
Stryker Corp	Suburban Propane
Sun Communities	Sunpower Corp
Suntrust Banks Inc	Superior Energy Services Inc
Supervalu Inc	SVB Financial Group
Symantec Corp	Synchrony Financial
Synopsys	Synovus Financial Corp
Sysco Corp	Tableau Software Inc
Tahoe Resources	Take Two Interactive Software Inc
Tanger Factory Outlet Centers Inc	Targa Resources Corp.
Target Co	Taubman Centers Inc.
Taylor Morrison Home Corp	TCF Financial
TD Ameritrade Holding Corp	Te Connectivity Ltd.
Tech Data	Tegna Inc
Teleflex Incorporated	Telephone & Data Systems Inc /De/
Tempur Sealy International, Inc.	Tenet Healthcare Corp
Teradata	Teradyne, Inc
Terex Corp	Terraform Power
Tesla Motors Inc	Tesoro Corp /New/
Tetra Technologies Inc	Texas Instruments Inc
Textron Inc	TFS Financial
Thermo Fisher Scientific Inc.	Thomson Reuters
Thor Industries	Tidewater Inc
Tiffany	Time Warner Inc.
Timken Co.	Tivo Corporation
TJX	T-Mobile US, Inc.
Toll Brothers	Tootsie Roll Industries Inc
Torchmark Corp	Toro
Total System Services Inc	Tractor Supply Co /De/
Transdigm Group Inc	Transocean Ltd.
Transunion	Travelers Companies, Inc.
Treehouse Foods, Inc.	Tribune Media Company

Tabla 21. *Listado de empresas analizadas (continuación)*

Trimble Inc.	Trinity Industries Inc
Tripadvisor, Inc.	Triumph Group Inc
Trustmark Corp	Tuesday Morning Corp/De
Tupperware Brands Corp	Twenty-First Century Fox, Inc.
Twenty-First Century Fox, Inc.	Twitter, Inc.
Two Harbors Investment Corp.	Tyler Technologies Inc
Tyson Foods	UDR, Inc.
UGI	Ulta Salon Cosmetics
Ultimate Software Group Inc	Under Armour
Under Armour	Union Pacific Corp
Unisys Corp	Unit Corp
United Bankshares Inc/Wv	United Continental Holdings
United Parcel Service Inc	United Rentals Inc /De
United States Cellular Corporation	United States Steel Corp
United Technologies Corp /De/	United Therapeutics Corp
Unitedhealth Group Inc	Universal Health Services Inc
Unum Group	Urban Outfitters
US Bancorp \De\	US Foods Holding Corp.
USG Corp	Utstarcom
V F Corp	Vail Resorts
Valero Energy Corp/Tx	Valhi Inc /De/
Validus Holdings Ltd	Valley National Bancorp
Valmont Industries Inc	Valspar
Vantiv, Inc.	Varian Medical Sys
Vca Inc	Vectren Corp
Veeva Systems	Ventas Inc
Vereit, Inc.	Verifone Holdings
Verisign Inc/Ca	Verisk Analytics, Inc.
Verizon Communications Inc	Versum Materials
Vertex Pharmaceuticals Inc / Ma	Viacom Inc.
Viacom Inc.	Viad Corp
Viavi Solutions Inc.	Visa Inc.
Vishay Intertechnology Inc	Vista Outdoor Inc.
Visteon Corp	Vmware, Inc.
Vornado Realty Trust	Voya Financial, Inc.
Vulcan Materials Co	Vwr Corporation
W R Grace & Co	W&T Offshore Inc
W. P. Carey Inc.	Wabco Holdings Inc.
Waddell & Reed Financial Inc	Wal Mart Stores
Walgreens Boots Alliance, Inc.	Walt Disney

Tabla 22. *Listado de empresas analizadas (continuación)*

Washington Federal	Washington Prime Group Inc.
Waste Connections	Waste Management Inc
Waters Corp /De/	Watsco Inc
Weatherford International Plc	Webmd Health Corp.
Webster Financial Corporation	Wec Energy Group, Inc.
Weight Watchers International Inc	Weingarten Realty Investors /Tx/
Weis Markets Inc	Wellcare Health Plans, Inc.
Wells Fargo & Company/Mn	Welltower Inc. /De/
Wendy'S Co	Werner Enterprises Inc
Wesco International Inc.	West Pharmaceutical Services Inc
Westamerica Bancorporation	Westar Energy Inc /Ks
Western Alliance Bancorporation	Western Digital Corp
Western Refining, Inc.	Western Union Co
Westinghouse Air Brake Technologies Corp	Westlake Chemical Corp
Westrock Co	Wex Inc.
Weyerhaeuser Co	WGL Hldg
Whirlpool Corp /De/	White Mountains Insurance Group Ltd
Whitewave Foods Co	Whiting Petroleum Corp
Whole Foods Market Inc	Wiley John & Sons, Inc.
Williams Companies Inc	Williams Sonoma
Willis Towers Watson Plc.	Windstream Holdings, Inc.
Winnebago Industries Inc	Workday
World Fuel Services Corp	Worthington Ind
WPX Energy, Inc.	Wyndham Worldwide Corp
Wynn Resorts Ltd	Xcel Energy Inc
Xerox	Xilinx Inc
XL Group Ltd	Xylem Inc.
Yahoo Inc	Yelp Inc
Yrc Worldwide Inc.	Yum Brands Inc
Zayo Group Holdings, Inc.	Zebra Technologies Corp
Zillow Group, Inc.	Zillow Group, Inc.
Zimmer Biomet Holdings, Inc.	Zions Bancorporation /Ut/
Zoetis Inc.	Zynga Inc

Fuente: base de datos Economatica (2016). Se recopiló información financiera de 1260 empresas públicas de los Estados Unidos en el periodo comprendido entre diciembre de 2011 y diciembre de 2015. Del total de las compañías, dos de ellas cotizan en la bolsa de valores AMEX (American Stock Exchange), 362 en NASDAQ (National Association of Securities Dealers Automated Quotation) y 896 en NYSE (New York Stock Exchange).