



UNIVERSIDAD DE SEVILLA  
TESIS DOCTORAL

---

# Factores que afectan a la adopción del Big Data como herramienta de Marketing en las empresas

---

Programa de Doctorado en Gestión Estratégica y Negocios Internacionales

Doctorando:

Juan Pedro Cabrera Sánchez

Director:

Dr. Ángel Francisco Villarejo Ramos



ADMINISTRACIÓN DE  
EMPRESAS Y MARKETING

Sevilla, 2019

## Índice

<b>CAPÍTULO 0 .....</b>	<b>5</b>
<b>Introducción y objetivos .....</b>	<b>5</b>
Introducción .....	6
Objetivos de la tesis doctoral .....	9
Estructura de la tesis doctoral .....	10
Referencias .....	12
<b>CAPÍTULO 1 .....</b>	<b>15</b>
<b>Big Data y Adopción de Tecnología .....</b>	<b>15</b>
Big Data Analytics (BDA) .....	16
Machine Learning o Aprendizaje Automático .....	20
Aprendizaje Supervisado .....	21
Aprendizaje No Supervisado .....	23
Aprendizaje Semi-Supervisado .....	24
Aprendizaje Por Refuerzo .....	24
Aprendizaje Profundo ( <i>Deep Learning</i> ) .....	25
Minería de Datos .....	28
Métodos estadísticos .....	29
Aplicaciones al Marketing .....	30
Software Asociado .....	31
Adopción de Tecnología .....	32
Teoría de la acción razonada (TRA) .....	33
Teoría del comportamiento planeado (TPB) .....	33
Modelo de aceptación de la tecnología (TAM) .....	34
TAM2 .....	35
TAM3 .....	36
Modelo del uso de ordenadores personales (MPCU) .....	37
Teoría cognitiva social .....	37
Teoría unificada de la aceptación y uso de la tecnología (UTAUT) .....	38
UTAUT2 .....	40
Nuevas variables .....	41
Riesgo percibido .....	41

Resistencia al uso .....	42
Coste de oportunidad.....	43
Modelos propuestos .....	44
Referencias.....	47
<b>CAPÍTULO 2 .....</b>	<b>59</b>
<b>Extendiendo el modelo UTAUT para evaluar los factores que afectan la adopción del Big Data en empresas españolas .....</b>	<b>59</b>
Resumen.....	60
Abstract .....	60
<b>CAPÍTULO 3 .....</b>	<b>61</b>
<b>Factors affecting the adoption of Big Data Analytics in companies .....</b>	<b>61</b>
Abstract .....	62
Resumo.....	62
Resumen.....	63
<b>CAPÍTULO 4 .....</b>	<b>65</b>
<b>Acceptance and use of Big Data Techniques in services companies .....</b>	<b>65</b>
Abstract .....	66
<b>CAPÍTULO 5 .....</b>	<b>67</b>
<b>Conclusiones, implicaciones, limitaciones y futuras líneas de investigación.....</b>	<b>67</b>
Conclusiones.....	68
Implicaciones teóricas .....	72
Implicaciones prácticas .....	74
Limitaciones de la investigación.....	75
Futuras líneas de investigación .....	75
Referencias.....	78



# CAPÍTULO 0

## Introducción y objetivos

## Introducción

Hace tiempo ya que oímos hablar del término Big Data y aún hoy parece que es algo ajeno a nosotros y percibido como algo del futuro. Nada más lejos de la realidad, los datos proporcionados por el [INE<sup>1</sup>](#) en su encuesta de uso TIC y Comercio Electrónico en las empresas 2017-2018 nos mostraban que un 11,24% de las empresas censadas usaban técnicas de Big Data para explotar sus datos, y el porcentaje aumentaba, considerablemente, hasta el 31,01% del total si nos centramos en empresas de más de 250 empleados tal y como podemos comprobar en la tabla Nº1.

	Total	De 10 a 49 empleados	De 50 a 249 empleados	De más de 250 empleados
<b>Total Empresas</b>				
I.1 % de empresas que analizaron Big Data	11,24	9,25	18,80	31,01
I.1.A % de empresas que analizaron Big Data por tipo de fuente: datos de la propia empresa con sensores o dispositivos inteligentes	33,62	29,79	38,02	53,75
I.1.B % de empresas que analizaron Big Data por tipo de fuente: datos por geolocalización a partir de dispositivos portátiles	49,69	53,55	42,13	39,44
I.1.C % de empresas que analizaron Big Data por tipo de fuente: datos generados por medios sociales	44,18	45,23	41,49	43,43
I.1.D % de empresas que analizaron Big Data por tipo de fuente: otras fuentes de Big Data	26,89	23,57	31,18	42,80
I.2.A % de empresas cuyo análisis de Big Data lo hicieron sus propios empleados	68,96	64,35	76,65	85,55
I.2.B % de empresas cuyo análisis de Big Data lo hicieron proveedores externos	46,41	48,93	40,87	41,69
I.3 % empresas con formación específica en Big Data	18,75	13,91	24,15	44,59
I.4 % empresas con formación en Big Data impartida por proveedores externos	15,43	11,16	20,36	37,82

Tabla Nº 1: Análisis de Big Data en empresas españolas. Datos en %. Fuente: INE

Ahora bien, ¿qué entendemos por Big Data? Hay muchas posibles definiciones y ninguna parece obtener un consenso suficiente. Partiendo de la definición dada por Gartner (2012) que define Big Data como “*gran volumen, gran velocidad y gran variedad de información y datos que necesitan nuevas e innovadoras formas para su procesado con el fin de obtener nuevos y valiosos patrones que nos permitan mejorar la toma de decisiones*”, en la propia definición van implícitas las características que ha de tener el Big

<sup>1</sup> <http://www.ine.es/jaxi/Datos.htm?path=/t09/e02/a2017-2018/I0/&file=02013.px>

Data que son las conocidas como V's del Big Data: volumen, velocidad y variedad. Algunos autores (Gandomi & Haider, 2015) han añadido tres más: veracidad, valor y variabilidad, si bien esta última no es mayoritariamente reconocida por la literatura científica.

Es necesario profundizar un poco en cada una de las características. Por **volumen** entendemos la cantidad de datos a tratar al estar hablando de varios Terabytes (TB= $10^{12}$  bytes) e incluso Petabytes (PB= $10^{15}$  bytes) o más. Aquí se parte de la idea que varios ordenadores son más rápidos que el más rápido de los servidores y además tal cantidad de datos ha de estar físicamente en más de un ordenador por lo que herramientas habituales de la gestión empresarial que utilizan datos como recurso principal como CRM (Customer Relationship Management) o ERP (Enterprise Resource Planning) utilizadas en local tienden a desaparecer y hay que empezar a pensar en otro tipo de arquitecturas para soportar tal volumen de datos.

Otra característica importante es la **variedad** y, por lo tanto, la heterogeneidad de los datos. Esto nos obliga a usar otro tipo de bases de datos. Las "antiguas" bases de datos relacionales funcionan muy bien con estructuras claras y definidas de los datos (datos homogéneos, normalmente numéricos), pero no funcionan cuando los datos no tienen una estructura predefinida (datos heterogéneos). Lo mismo podemos almacenar un tuit, que una imagen, un documento, una reseña en una red social o un vídeo, entre otros.

Asimismo, la proliferación de dispositivos digitales con conexión a internet (o a redes de intranet) genera datos a tal **velocidad** que hay que estar preparados para analizarlos en tiempo real si queremos sacar el máximo provecho de ellos.

Por otra parte, hay que tener en cuenta que hay que tratar con datos inciertos y dudosos como, por ejemplo, el sentimiento hacia una marca en una red social, por lo que hay que usar herramientas adecuadas para constatar la **veracidad** de estos. También hay que saber tratar con la **variabilidad** ya que los datos se pueden generar en diferentes momentos con picos muy altos de diferente periodicidad y todo lo expuesto anteriormente ha de generar **valor**, que suele ir correlacionado con el volumen de datos analizado.

Resumiendo, el Big Data tiene que dar respuesta a todos los retos que plantea el ciclo de vida de los datos: la captación de los propios datos, el procesamiento y su gestión (Akerkar, 2014; Zicari, 2014). Cada etapa plantea sus propias dificultades. Las características de los datos (volumen, variedad, velocidad, veracidad, volatilidad, valor y visualización) plantean problemas en su recopilación. Las técnicas relacionadas con la captura de datos, su integración en bases de datos, cómo filtrarlos, transformarlos y obtener los resultados significan un proceso complejo. La gestión de la información no puede obviar aspectos como la privacidad, seguridad y otros aspectos éticos.

Estos retos los podemos agrupar en dos grandes procesos o subdivisiones del Big Data: 1) la propia gestión de los datos como por ejemplo, y entre otros, su adquisición, grabación y almacenamiento y 2) la analítica de datos (D. Agrawal, Bernstein, & Bertino, 2011). El ámbito de la ingeniería se centra en el primer gran bloque mientras que el segundo bloque es el que nos interesa desde el punto de vista del marketing. A partir de ahora, cada vez que hablemos de Big Data, estaremos hablando de Big Data Analytics (BDA).

Los beneficios, aplicaciones y utilidades que esta tecnología puede aportar a las empresas son numerosos (Watson, 2019; Wedel & Kannan, 2016a), sobre todo a la hora de tomar decisiones basadas en datos (McAfee & Brynjolfsson, 2012). El adoptar técnicas de Big Data incluso mejora la percepción que tienen los usuarios sobre los beneficios que esta tecnología puede ofrecerles (Verma, Bhattacharyya, & Kumar, 2018), ayudando a las empresas a innovar (Wright, Robin, Stone, & Aravopoulou, 2019).

No obstante, la implantación de técnicas y herramientas de Big Data en las empresas actuales supone tener que sortear ciertos obstáculos como el desconocimiento de la herramienta, el miedo a la tecnología, la resistencia a los cambios en la organización, la desconfianza sobre su rendimiento real,... sin olvidar, como señalan Yaqoob et al. (2016), que la propia tecnología tiene sus limitaciones de cara a ser empleada por las empresas.

Además, la revisión de la literatura científica muestra un mayor interés por aspectos más técnicos relacionados con el desarrollo del Big Data: aplicaciones, minería de datos, analítica de datos, modelado estadístico, estudios predictivos y experimentación (Sivarajah, Kamal, Irani, & Weerakkody, 2016) y sin embargo, son escasos los estudios

que ven la herramienta desde la perspectiva de la integración y el uso dentro de la empresa como instrumento para obtener información para la toma de decisiones: adopción de la analítica de Big Data en empresas (Kwon, Lee, & Shin, 2014; Verma et al., 2018; Brünink, 2016; Rahman, 2016).

El BDA podrá utilizarse para optimizar muchos procesos y para mejorar la producción, pero donde provocará un auténtico cambio es en la forma de procesar y utilizar la información para la gestión del marketing ya que podremos mejorar nuestra toma de decisiones permitiendo a las empresas (Fan, Lau, & Zhao, 2015): (1) estudiar cómo seleccionar las fuentes de datos apropiadas para cada objetivo de marketing; (2) analizar cómo seleccionar y utilizar los métodos de análisis de datos apropiados; (3) preguntar cómo integrar diferentes fuentes de datos para estudiar complejos problemas de marketing; (4) investigar cómo tratar la heterogeneidad de las fuentes; (5) examinar cómo balancear inversiones entre las diferentes técnicas de inteligencia de marketing; y (6) implementar mejoras conforme van desarrollándose nuevas tecnologías asociadas al Big Data.

Además de todas estas mejoras, hay que añadir que todo el software necesario para el uso y explotación del BDA es de código libre por lo que el precio de las licencias no es un obstáculo para implementarlas en ningún tipo de empresa.

## Objetivos de la tesis doctoral

Realizado este planteamiento, si el BDA permite mejorar los procesos de captación de información y permite conocer mejor a los clientes al obtener conocimiento de los propios datos de la empresa y de otras fuentes y, además, casi todo el software es de código abierto y por lo tanto libre de pagar licencias, ¿cómo es posible que no lo estén usando todas las empresas en la actualidad?

La presente tesis doctoral pretende analizar los factores que afectan a la adopción y uso de la tecnología Big Data por las empresas para la toma de decisiones. Con ello, poder establecer pautas y recomendaciones para su implantación más eficaz en las empresas.

Con ese fin, adaptaremos el modelo de aceptación de tecnología, la teoría unificada de aceptación y uso de la tecnología, UTAUT (Venkatesh, Morris, Davis, & Davis, 2003) al que añadiremos factores inhibidores y otros antecedentes referidos al contexto de la adopción de Big Data.

Aprovecharemos que vamos a usar el modelo UTAUT, para realizar una revisión en la literatura científica de los distintos modelos de adopción de tecnología para obtener de su adaptación una mejor explicación de los factores que afectan a la adopción del Big Data en las empresas.

## Estructura de la tesis doctoral

La estructura que vamos a seguir con esta tesis por compilación de artículos consta de seis capítulos. Un capítulo 0 de introducción, en el que se enmarca toda la investigación y se fijan los objetivos de la misma. Seguiremos con el capítulo 1 en el que se hará una introducción tanto al Big Data Analytics (BDA), como a la adopción de tecnología dentro del contexto de gestión y toma de decisiones en marketing.

Los tres capítulos centrales (2, 3 y 4) recogen con cada uno de los trabajos publicados fruto de la investigación desarrollada para la realización de esta tesis doctoral.

El capítulo 2, titulado “Extendiendo el modelo UTAUT para evaluar los factores que afectan la adopción del Big Data en empresas españolas”, publicado como capítulo en el libro *Nuevos Horizontes del Marketing y de la Distribución Comercial*, editado por la Cátedra Ramón Areces de Distribución Comercial de la Universidad de Oviedo (ISBN: 978-84-8367-622-6), páginas 181-200, en el que con una primera muestra de 131 empresas españolas comprobamos la menor significación de algunas de las variables del modelo original y que otras nuevas variables que extienden el modelo como el riesgo percibido aportan y lo mejoran.

El capítulo 3, “Factors affecting the adoption of Big Data Analytics in companies” publicado en la *Revista de Administração de Empresas (RAE)* Vol. 59, num. 6, en 2019 (JCR-IF 2018= 0’701), propone extender el modelo UTAUT con una variable más, la

resistencia al uso. En este artículo vemos cómo evoluciona el UTAUT ya que el riesgo percibido deja de tener un efecto significativo con la muestra ampliada de empresas y que la resistencia al uso toma un papel importante como inhibidor de la adopción del BDA. Asimismo, comprobamos que la muestra se comporta de forma diferente en función del nivel de madurez en la implantación de Big Data Analytics dentro de sus compañías, y que este nuevo modelo obtiene mejores resultados en cuanto al poder explicativo que el modelo del capítulo 2.

“Acceptance and use of Big Data Techniques in services companies” es el título del capítulo 4, publicado en el *Journal of Retailing and Consumer Services*, Vol 52, January, en 2020 (JCR-IF 2018= 3'585) en el que incluimos, además de las variables previamente mencionadas, una nueva, el coste de oportunidad, resultando significativo y proporcionando un mayor poder explicativo del modelo extendido. Nuevamente, esta vez con una muestra de 199 empresas de servicios, mejoramos la capacidad predictiva del UTAUT y obtenemos nuevas variables significativas mientras otras, como la expectativa de esfuerzo dejan de serlo.

Por último, en el capítulo 5 exponemos las conclusiones de la investigación, así como sus limitaciones y futuras investigaciones a partir de los resultados obtenidos en esta tesis doctoral. En cuanto a las futuras investigaciones expondremos con brevedad la investigación realizada sobre la aceptación y uso de aplicaciones basadas en Big Data por parte de consumidores finales y las conclusiones previas obtenidas de este nuevo proyecto investigador.

## Referencias

- Agrawal, D., Bernstein, P., & Bertino, E. (2011). Challenges and Opportunities with Big Data 2011-1. Proceedings of the VLDB Endowment, 1–16. Retrieved from <http://dl.acm.org/citation.cfm?id=2367572%5Cnhttp://docs.lib.purdue.edu/cctech/1/>
- Akerkar, R. (2014). Analytics on big aviation data: Turning data into insights. *International Journal of Computer Science and Applications*, 11(3), 116–127.
- Brünink, L. (2016). Cross-Functional Big Data Integration: Applying the Utaut Model. University of Twente (The Netherlands).
- Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, 2(1), 28–32. <https://doi.org/10.1016/j.bdr.2015.02.006>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gartner. (2012). What Is Big Data? - Gartner IT Glossary - Big Data. Retrieved May 22, 2017, from <http://www.gartner.com/it-glossary/big-data/>
- Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, 34(3), 387–394. <https://doi.org/10.1016/j.ijinfomgt.2014.02.002>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data. The management revolution. *Harvard Business Review*, 90(10), 61–68. <https://doi.org/10.1007/s12599-013-0249-5>
- Rahman, N. (2016). Factors Affecting Big Data Technology Adoption, 0–29. Retrieved from <http://pdxscholar.library.pdx.edu/studentsymposium%5Cnhttp://pdxscholar.library.pdx.edu/studentsymposium/2016/Presentations/10>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2016). Critical analysis of Big Data

- challenges and analytical methods. *Journal of Business Research*, 70, 263–286.  
<https://doi.org/10.1016/j.jbusres.2016.08.001>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478.  
Retrieved from <http://www.jstor.org/stable/30036540>
- Verma, S., Bhattacharyya, S. S., & Kumar, S. (2018). An extension of the technology acceptance model in the big data analytics system implementation environment. *Information Processing and Management*, 54(5), 791–806.  
<https://doi.org/10.1016/j.ipm.2018.01.004>
- Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, 44, 364–379. <https://doi.org/10.17705/1CAIS.04421>
- Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6), 97–121. <https://doi.org/10.1509/jm.15.0413>
- Wright, L. T., Robin, R., Stone, M., & Aravopoulou, D. E. (2019). Adoption of Big Data Technology for Innovation in B2B Marketing. *Journal of Business-to-Business Marketing*, 00(00), 1–13. <https://doi.org/10.1080/1051712X.2019.1611082>
- Zicari, R. (2014). Big data: Challenges and opportunities. *Big Data Computing*.  
<https://doi.org/doi:10.1201/b16014-5>



# CAPÍTULO 1

## Big Data y Adopción de Tecnología

## Big Data Analytics (BDA)

El Big Data ha supuesto un cambio radical en la forma de obtener y analizar los datos. Con el Big Data cambia la manera en que las empresas pueden obtener gran cantidad de información valiosa, cómo gestionar esta gran cantidad de datos, casi en tiempo real y cómo usar esta información para lograr una gran ventaja competitiva (Sivarajah et al., 2016).

Las empresas pueden con el Big Data recolectar gran cantidad de datos de sus clientes, pero necesitan poder analizarlos e interpretarlos para obtener de ellos la máxima información, lo que denominamos Big Data Analytics (BDA). Utilizar BDA significa el desarrollo de nuevas capacidades y habilidades en el personal responsable de la información en las organizaciones. Los denominados científicos de datos (*Data Scientist*), habitualmente con formación en ingeniería, estadística y conocimientos del mundo empresarial, son pieza clave en las empresas de cara a implantar y utilizar BDA. Este tipo de perfiles en las empresas permitirá analizar e interpretar los datos generados por la compañía y los obtenidos de otras fuentes accesibles y cambiarán la toma de decisiones en las empresas, basadas ahora en datos y no tanto en experiencias de los responsables de la compañía (McAfee & Brynjolfsson, 2012).

En este entorno, las analíticas de Marketing tienen una especial relevancia ya que proporcionan métricas más potentes, efectivas y eficientes para tomar estas decisiones aunque todavía no esté totalmente claro qué tipo de analítica funciona mejor para qué tipo de problemas y de datos disponibles así como los procedimientos necesarios a implementar en las empresas para poder sacar el máximo jugo de estas métricas (Wedel & Kannan, 2016a).

Hay diferentes métodos analíticos para obtener esas métricas que podemos agrupar tal y como se muestra en la figura N° 1:

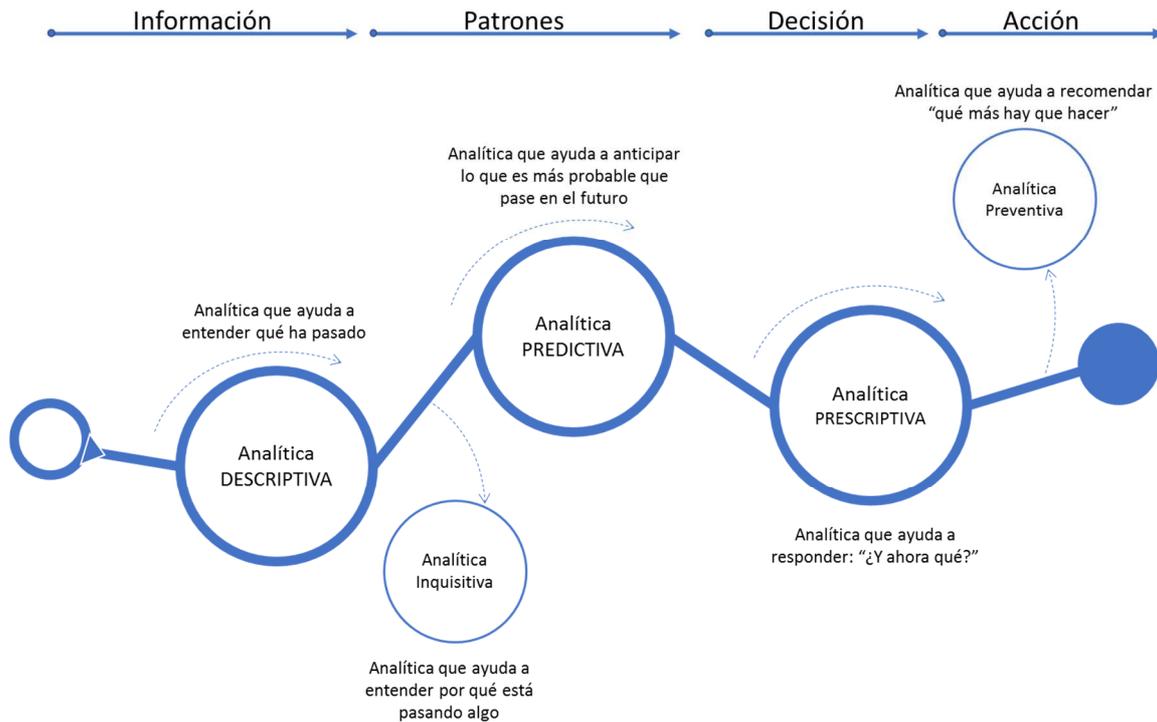


FIGURA Nº 1: Clasificación de los diferentes tipos de analítica (Sivarajah, Kamal, Irani & Weerakkody, 2016)

Y que si lo llevamos al ámbito del marketing, las podemos resumir en tres grandes bloques (Wedel & Kannan, 2016a): 1) analítica que nos permita optimizar el gasto en nuestro marketing mix; 2) analítica que nos permita personalizar el marketing mix; y 3) analítica que nos permita mejorar la seguridad de los datos y privacidad de los clientes.

Para todo esto no sólo podremos usar los datos generados por la propia compañía sino otros muchos más provenientes de fuentes diversas. Como Señala Dull (2015), el clásico *data warehouse* o "almacén de datos" de la empresa está evolucionando hacia los *data lakes* ("lago de datos") donde se almacena toda esta ingente cantidad de datos en crudo usando las aplicaciones de Big Data como HDFS, Hadoop o almacenamiento en la nube como el *Amazon Simple Storage Service*, Amazon S3.

En la tabla Nº 2 podemos ver una comparación entre los *data warehouses* y los nuevos *data lakes* a muy alto nivel. Esto está cambiando rápidamente y las fronteras entre uno y otro se están diluyendo hacia una única base de datos dentro de la compañía, que será parte de otra más grande y, quizás, virtual con muchos más componentes interconectados (Watson, 2019).

Características	Data warehouses	Data Lakes
Datos	Datos relacionales obtenidos de las aplicaciones de operaciones y de negocio	Datos relacionales y datos del Internet de las cosas, datos de webs, aplicaciones móviles, etiquetas RFID y social media
Extracción, Limpieza y Transformación / Integración de Datos	Grabación, limpieza y transformación de los datos; modelo de datos al escribir	Grabación y entrada con ficheros nativos; modelo de datos se obtiene al leer
Analíticas	Queries, informes, cuadro de mandos y visualización de datos	Descubrimiento de datos, análisis predictivo y machine learning (o aprendizaje automático)
Usuarios	Usuarios finales, analistas y científicos de datos	Programadores, analistas senior y suficientemente formados y científicos de datos

Tabla Nº 2: Comparativa Datawarehouse y DataLakes

Continuando con los datos, una de las grandes ventajas que aporta la Analítica Big Data es que soporta tanto los datos estructurados como los no estructurados permitiéndonos, por tanto, aplicarla a los siguientes grandes ámbitos:

- Analítica de textos
- Analítica de audio
- Analítica de video e imágenes
- Analítica de medios sociales
- Analítica de datos (estructurados)
- Inteligencia Artificial

Todos estos ámbitos no son excluyentes entre sí ya que pueden usarse técnicas de Inteligencia Artificial para distinguir fotografías o reconocimiento de voz, aparte de patrones de comportamiento o cualquier otro fenómeno del que se dispongan datos. De hecho, son complementarios ya que la mayor parte de las veces se usan conjuntamente, aunque cada uno de ellos tenga algoritmos específicos y optimizados para una tarea en concreto.

La cantidad de métodos y algoritmos que podemos encuadrar dentro de la Analítica Big Data es muy amplia, aunque podemos clasificarlos según propusieron Rehman, Chang, Batool, & Wah (2016), clasificación a la que añadimos el aprendizaje por refuerzo y que podemos ver en la tabla Nº3.

Tipo	Método
Aprendizaje automático	Aprendizaje supervisado
	Aprendizaje no supervisado
	Aprendizaje semi-supervisado
	Aprendizaje por refuerzo
	Aprendizaje profundo ( <i>Deep Learning</i> )
Minería de Datos	Clasificación
	Reglas de asociación
	Análisis de regresión
Métodos estadísticos	Estadística descriptiva
	Estadística Inferencial

TABLA Nº 3: Métodos de análisis de datos para Big Data

Siendo el proceso global de la Analítica Big Data el mostrado en la figura Nº2 (Rehman, Khan, & Batool, 2016):

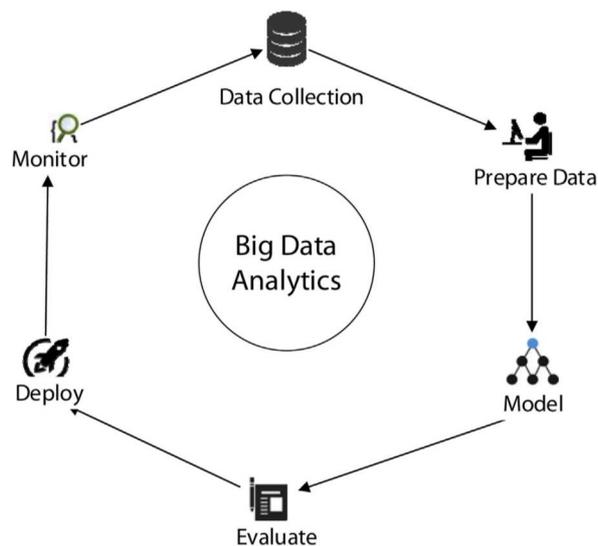


FIGURA Nº 2: Proceso de la Analítica Big Data

En proceso de ciclo completo empieza en la recolección de datos de múltiples y variadas fuentes, seguido de su limpieza y preparación (a veces ésta es la tarea que más tiempo consume), confección del modelo con el que vamos a hacer las simulaciones, clasificaciones o predicciones y su entrenamiento para pasar a evaluar la bondad de su ajuste con el mundo real. Si predice o clasifica correctamente (o con un porcentaje muy aceptable) se pone en ejecución o funcionamiento en el mundo real donde vamos monitorizando el rendimiento del modelo. Si el rendimiento no es el esperado, volvemos a empezar el ciclo ajustando los parámetros de los modelos o bien usando nuevos

modelos que sean capaces de predecir o clasificar mejor con los datos disponibles. En cualquier caso, este proceso es siempre recursivo ya que los datos están continuamente cambiando y por lo tanto, es posible que un modelo que antes funcionó bien, en cierto tiempo deje de funcionar y haya que volver a reiniciar el proceso completo.

Este proceso también variará en función de si queremos hacer un modelo meramente descriptivo, predictivo o prescriptivo (Lavallo, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Para un modelo meramente descriptivo podemos usar técnicas estadísticas clásicas mientras que para modelos predictivos y prescriptivos necesitamos recurrir a técnicas de *Machine Learning* o de Minería de Datos (Tibishirani, Friedman, & Hastie, 2009). En un principio, se utilizaron técnicas de análisis multivariante clásicas (Joseph F Hair, Anderson, Tatham, Black, & others, 1999) como la regresión y la clasificación, aunque a niveles mucho más avanzados gracias al Big Data. A continuación, haremos un breve repaso de todas estas técnicas en los siguientes epígrafes tomando como base la clasificación de la tabla Nº 3.

## Machine Learning o Aprendizaje Automático

Hay muchas definiciones, pero de las que mejor se ajustan a cómo podríamos definir el Aprendizaje Automático es de la forma que ya hizo Arthur Samuel en 1959: “Ciencia (y arte) de programar ordenadores para que puedan aprender de los datos”. Este aprendizaje puede ser: supervisado, no supervisado, semi-supervisado, por refuerzo y profundo, tal y como veremos en los siguientes sub-epígrafes.

La mayoría de estos algoritmos están implementados en dos de los lenguajes de programación más usados en esta disciplina como son *R* y *Python* (aunque también lo están siendo en nuevos lenguajes como *Julia* y otros menos nuevos como *Matlab*, *Java*, etc.). Ambos tienen librerías como *scikit-learn* (Pedregosa et al., 2011), *tensorflow* (Abadi et al., 2016) o *keras* (Chollet & others, 2018) con los algoritmos actualizados y de libre acceso mantenidos por la comunidad científica, por lo que su uso es tremendamente fácil para cualquiera que sepa un mínimo sobre programación.

## Aprendizaje Supervisado

El aprendizaje supervisado es la técnica más usada para predecir y clasificar futuros casos ya que estos algoritmos se entrenan con datos etiquetados. El algoritmo “aprende” cuál ha de ser el resultado y una vez entrenado ya es capaz de ofrecer el resultado para nuevos datos.

La cantidad de algoritmos para clasificar o predecir futuras instancias de datos es grande. Destacamos, entre ellos, los siguientes con algunos de los autores que más han aportado a su desarrollo:

- Regresión y todas sus variantes como logística, splines, ridge, lasso, GAM (R. Huang & Xu, 2015; Jun, Lee, & Ryu, 2015; Núñez, Steyerberg, & Núñez, 2011).
- Árboles de decisión (Breiman, 1984; Quinlan, 1993).
- Bayes (Meyer et al., 2014; Roever et al., 2006).
- Máquinas de Vectores de Soporte (Cortes & Vapnik, 1995).
- K-Nearest Neighbors (Deng, Zhu, Cheng, Zong, & Zhang, 2016; Wang, 2011).
- Redes Neuronales (Lecun, Bengio, & Hinton, 2015) o Aprendizaje Profundo (*Deep Learning*).

El Deep Learning (Bengio, Courville, & Vincent, 2012; Conneau, Schwenk, Barrault, & Lecun, 2016; G. Hinton, Vinyals, & Dean, 2015; LeCun & Bengio, 1995) constituye hoy la herramienta más poderosa para el Procesado Natural del Lenguaje y la Visión por Computador, áreas donde son ampliamente usadas. Las veremos un poco más en profundidad en un epígrafe posterior.

Por norma general, para entrenar estos algoritmos se divide el conjunto de datos disponible en 2 subconjuntos, uno para entrenamiento y otro para validación. Se entrena con el primer conjunto y luego se testea con el de validación. Comparamos la salida obtenida con la esperada y ése será el error que habrá que minimizar. Hay muchas más técnicas para evaluar la bondad de ajuste del modelo entre las que podemos destacar las de validación cruzada (Xu et al., 2018) y la de re-muestreo (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2005).

Con la finalidad de conseguir disminuir aún más el error obtenido en los algoritmos anteriores, surgieron nuevos modelos de aprendizaje conocidos como *ensemble learning* o aprendizaje conjunto. Básicamente, consiste en usar diferentes algoritmos (o el mismo con diferentes parámetros y diferentes inicializaciones) y hallar una “media” de los resultados, demostrándose que esta “media” es siempre mejor que el resultado de un único algoritmo. Esta “media” es calculada de diversas formas, desde una media aritmética hasta el resultado más votado, pasando por medianas y varias alternativas más. Los principales métodos de aprendizaje conjunto son:

- **Bagging:** se construyen múltiples modelos del mismo tipo usando diferentes submuestras (Bootstrap) del conjunto de entrenamiento. El modelo más utilizado es el Random Forest (que usa árboles de decisión) y es uno de los modelos que mejores resultados obtiene (Leo Breiman, 2001).
- **Boosting:** al igual que el caso anterior, se construyen múltiples modelos del mismo tipo, pero ahora el modelo 2 trata de mejorar las predicciones con más error del modelo 1 y así sucesivamente hasta que ya no se mejore más (Freund, Schapire, & Abe, 1999; Friedman, 2001). Actualmente, se están alcanzando excelentes resultados en clasificación con estos métodos, como el AdaBoost o Gradient Boosting.
- **Stacking:** en este caso un modelo supervisor es el que combina de la mejor manera posible los múltiples modelos generados (Sikora & others, 2015).

Todos estos algoritmos se usan cuando presuponemos que todas las muestras son independientes, pero podría suceder que en algunos casos unas muestras dependieran de muestras anteriores. En esos casos habrá que recurrir a otros algoritmos que conforman lo que se conoce como Aprendizaje Supervisado Secuencial, entre los que podemos destacar:

- Hidden Markov Models (HMM). El más relevante de todos y por ello del que hay más algoritmos implementados (Helske & Helske, 2019; Spedicato, 2017).
- Sliding Window.
- Recurrent Sliding Window.
- Maximum Entropy Markov Models (MEMM).

- Conditional Random Fields.
- Graph Transformer Networks.

La principal debilidad del aprendizaje supervisado es que no siempre es sencillo disponer de datos etiquetados en grandes cantidades así como tampoco son fáciles de usar en entornos con datos altamente cambiantes.

## Aprendizaje No Supervisado

En este tipo de aprendizaje los datos no están etiquetados y los algoritmos tratarán de aprender de ellos. Tratarán de buscar características relevantes o patrones por los que clasificarlos. Este tipo de aprendizaje son mayoritariamente usados en clustering (o agrupamiento), reducción de la dimensionalidad de los datos, detección de anomalías y aprendizaje de reglas de asociación (Géron, 2019).

Los principales algoritmos para cada uno de estos usos son:

- **Clustering:** Muy útiles para segmentar y para encontrar características comunes dentro del conjunto de datos. Los algoritmos más conocidos son K-Media , K-Mediodes o PAM, DBSCAN y Análisis Jerárquico de Clusters HCA (Amini, Wah, & Saboohi, 2014; Tibishirani et al., 2009).
- **Detección de Anomalías:** Como su propio nombre indica, estos algoritmos se usan para buscar datos que no siguen ningún patrón establecido o claramente se diferencian del resto. Los algoritmos más usados son los Bosques de Aislamiento (Isolation Forest) y Máquinas de Soporte Vectorial de una clase (Ruff et al., 2018).
- **Visualización y Reducción de la Dimensionalidad:** Cuando tenemos una gran cantidad de datos con muchas variables es muy difícil “ver” nada. Para ello, se usan técnicas para reducir a dos dimensiones y agrupar por las variables más significativas (o conjunto de ellas) y así poder mostrar una visualización rápida de los datos y sus agrupaciones. También usamos la reducción de dimensiones para reducir el conjunto de datos a una cantidad inferior sin perder mucha información y convertirlos en más manejable tanto para su procesamiento mediante la máquina como para su posterior interpretación humana. Entre sus principales

algoritmos tenemos el Análisis de Componentes Principales, el KernelPCA, y el Análisis Factorial (Mika et al., 1999; Yang, Zhang, Frangi, & Yang, 2004).

- **Aprendizaje de Reglas de Asociación:** Se usan para buscar sucesos que ocurren en común dentro del conjunto de datos. Por ejemplo, si tenemos un supermercado con todos los datos que genera, podemos descubrir si los que compran pañales y cerveza también compran palomitas. Los algoritmos más conocidos son ECLAT (Zaki, 2000) y A Priori (R. Agrawal, Imielinski, & Swami, 1993).

## Aprendizaje Semi-Supervisado

Estos modelos son capaces de tratar tanto con datos etiquetados como no etiquetados. Para aprender usan datos etiquetados, y luego se van adaptando a los datos sin etiquetar que les puedan llegar, mejorando por tanto los problemas de los algoritmos supervisados puros con el rápido cambio en los datos. También puedo buscar primero o encontrar patrones y luego etiquetarlos. Por ejemplo, cuando tienes un álbum de fotos y aparecen personas, primero he de ver que alguien sale en las fotos 3, 1, 4 y otro sale en la 15, 19 y 27 para lo que usaría un algoritmo no supervisado y luego manualmente podría ir etiquetando a esas personas y seguir buscando en todo el álbum (o nuevos) de fotos.

Otra modalidad es la de usar varios elementos no supervisados como las máquinas restringidas de Boltzman (RBM) y agruparlas para formar una red profunda (DBN) en la que sí usará salidas esperadas (datos etiquetados) para ajustar toda la pila de RBMs Hinton, Osindero, & Teh, 2006). En esta parte podríamos añadir también el Aprendizaje basado en Grafos y el Aprendizaje Heurístico.

## Aprendizaje Por Refuerzo

Este tipo de aprendizaje es completamente distinto a los anteriormente señalados. Aquí no aprende un modelo sino un “agente”. Este agente recibirá premios por acciones bien realizadas y penalización por acciones mal realizadas. Así, irá aprendiendo solo hasta encontrar la mejor estrategia de aprendizaje. Por ejemplo, si usamos un agente para

aprender a jugar a Go (Silver et al., 2016), juego parecido a las damas, el agente irá aprendiendo en base a buenas y malas jugadas para maximizar el resultado final. El caso de Go es el paradigma del Reinforcement Learning o Aprendizaje por Refuerzo ya que Google DeepMind creó la inteligencia artificial llamada AlphaGo para ganar al campeón del mundo (Google, 2016). Esto supuso el desarrollo inicial de este tipo de aprendizaje, mucho más complejo que los anteriores y todavía con muchas menos aplicaciones comerciales.

### Aprendizaje Profundo (*Deep Learning*)

Los modelos de aprendizaje profundo son una representación jerárquica de los modelos supervisados y no supervisados. Funcionan muy bien con grandes cantidades de datos con múltiples dimensiones y por eso, son la mejor elección para Big Data. Funcionan muy bien con textos, audios, imágenes y videos y, a pesar de haber pasado por varias etapas de uso y posterior desuso (Nilsson, 2009), en la actualidad, dado que tenemos suficiente capacidad de procesamiento están en pleno apogeo y con múltiples aplicaciones en gran cantidad de ámbitos (Ahmad, Farman, & Jan, 2019; Yuan, He, Zhu, & Li, 2019).

Este aprendizaje está inspirado en las neuronas humanas y en cómo están conectadas, aunque las neuronas artificiales han ido alejándose de su parecido biológico y han evolucionado hacia fórmulas matemáticas cada vez más complejas (Géron, 2019). Una neurona artificial no es más que una unidad de cálculo a la que llegan varias entradas, se ponderan, se suman y si superan un determinado umbral, activará la salida exactamente igual a una biológica. La neurona artificial más básica es la conocida como el **perceptrón** (Rosenblatt, 1958):

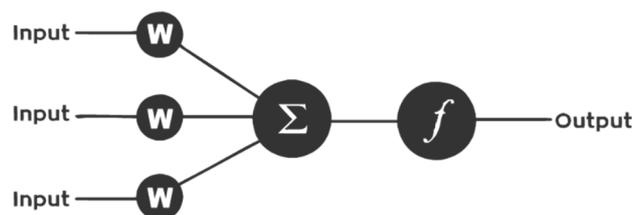


FIGURA Nº 3: Perceptrón

En la figura N°3 podemos comprobar que las diferentes entradas se ponderan con una serie de pesos  $w$ , que posteriormente se suman y llegan a la función de activación representada por una  $f$  y que, si pasa de un determinado umbral, activará la neurona. La función de activación clásica, que se parece más a la que tienen las neuronas biológicas, es la sigmoide, aunque en la actualidad hay muchas más funciones de activación.

Una red neuronal estará compuesta por una o varias capas de neuronas como la anteriormente descrita. Por norma general, suele haber una capa de entrada que se hace cargo de los datos que vamos a usar para entrenar nuestra red neuronal y una capa de salida en la que obtendremos el resultado esperado o a la que le hemos de decir qué queremos obtener, entrenando así la red. Pero, también es habitual situar capas intermedias que mejoran el rendimiento de la red ya que van aprendiendo o abstrayendo características clave de los datos. Estas capas intermedias o profundas son las que dan el nombre a este tipo de aprendizaje (*Deep Learning*). Podemos ver un perceptrón o **red multicapa** en la figura N°4:

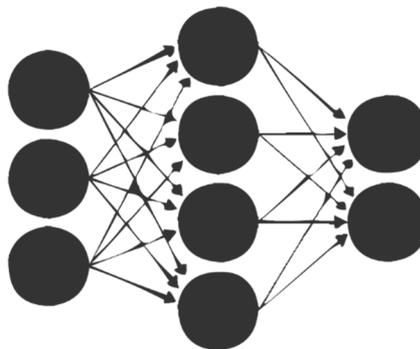


FIGURA N° 4: Perceptrón Multicapa

Cuantas más capas profundas tenga la red mayor será su capacidad de aprendizaje aunque también será más difícil de entrenar ya que hacen falta muchos datos y otra serie de problemas asociados a esta tipología de la red (Ba, Kiro, & Hinton, 2016). A esta forma de apilar las capas, donde cada perceptrón o unidad de cálculo de una capa se conecta con todos los de la siguiente capa, se le llama **red densa o totalmente conectada** y fueron las primeras redes neuronales artificiales.

Básicamente, entrenar una red neuronal consiste en encontrar los pesos de cada conexión de tal forma que la entrada se corresponda con la salida esperada. A este proceso se le denomina aprendizaje o entrenamiento y funciona de la siguiente manera: 1) partimos de una serie de datos etiquetados por lo que sabemos el resultado que hemos de obtener; 2) la diferencia entre lo obtenido y lo que deberíamos obtener nos sirve para retroalimentar la red (*“backpropagation”*) e ir recalculando los pesos de nuevo; y 3) este proceso hacia adelante y hacia atrás se repite hasta minimizar el error entre la salida real y la esperada. Por supuesto, este proceso de entrenamiento tiene muchos matices y las funciones matemáticas usadas son altamente complejas (Chollet & others, 2018; G. Hinton et al., 2012; Lecun et al., 2015).

Este tipo de redes es muy utilizada para clasificaciones binarias como multi-clase e incluso para realizar regresiones para predecir valores.

Posteriormente, se buscaron nuevas aplicaciones más específicas como el reconocimiento de imágenes en las que las redes densas estaban llegando a un límite. Así fue como surgieron las **redes convolucionales** (Conneau et al., 2016; Lecun et al., 2015; Szegedy et al., 2014) en las que hay nuevas capas llamadas convolucionales y que transforman la entrada aplicando una convolución matemática que servirá para extraer propiedades. Cuantas más capas convolucionales, más características será capaz de obtener la red, características que luego servirán para clasificar con una red densa tal y como podemos ver en la figura N°5.

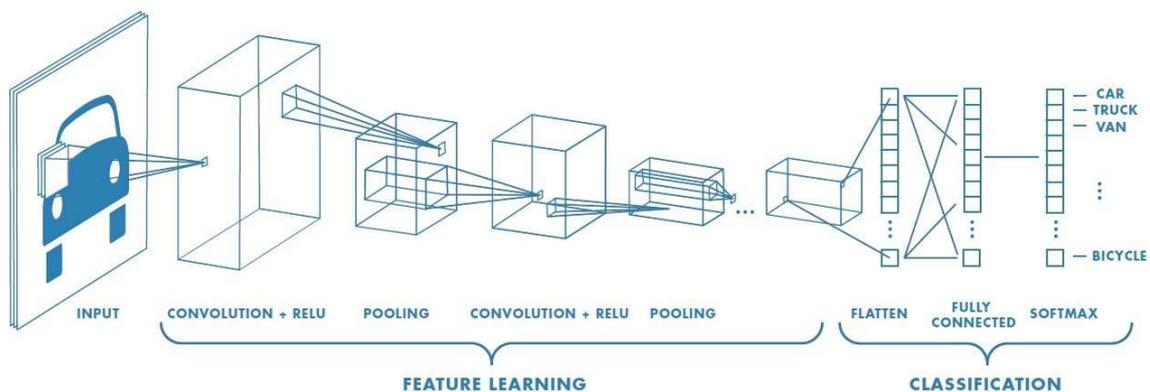


FIGURA N° 5: Red convolucional

El paso siguiente supuso la aparición de las **redes recurrentes**, o redes con memoria. En estas redes se almacenan datos que se pueden usar para propagarlos también por la red

y que podrán ser usados en momentos posteriores. Cada unidad de procesamiento ahora es mucho más compleja que un perceptrón y en función de la profundidad (o número de celdas) será capaz de usar más o menos datos. Estas redes se sitúan en contextos cercanos, como es el caso de predecir una palabra en una frase “las estrellas están en el...”. También, en los casos en los que la red necesita más información, es decir, un contexto más amplio como podría ser, “yo nací en España y por lo tanto puedo hablar...”. Aunque la red sabe que es un idioma es posible que no sepa cuál y por eso se necesita un mayor número de capas (Graves, Fernández, Gomez, & Schmidhuber, 2006; Liwicki, Graves, Bunke, & Schmidhuber, 2007; Sukhbaatar, Szlam, Weston, & Fergus, 2015; Xiong et al., 2016). Un ejemplo de estas redes lo vemos en la figura N°6.

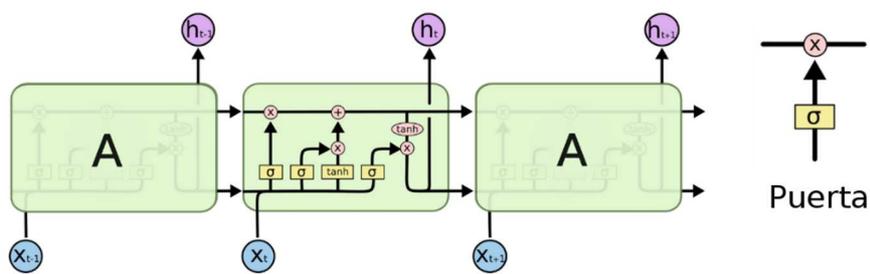


FIGURA N° 6: Red Recurrente

Hoy día, sigue siendo muy costoso, computacionalmente hablando, entrenar una red neuronal, sobre todo si tiene muchas capas. Afortunadamente, hay muchas redes ya pre-entrenadas disponibles de libre acceso como, por ejemplo, *Imagenet*<sup>2</sup> y que puedo reutilizar mediante una técnica conocida como **Transfer Learning** o “transferencia de aprendizaje” e incluso muchos modelos de redes ya pre-entrenados y listos para usar con nuestros modelos de datos en los frameworks más populares como *Keras* (Chollet & others, 2018).

## Minería de Datos

La minería de datos es el proceso de descubrir patrones en grandes volúmenes de datos. La distinción de esta disciplina con el aprendizaje automático o Machine Learning es cada

<sup>2</sup> <http://www.image-net.org/>

vez más sutil y prácticamente inexistente a día de hoy. De hecho, este concepto se ha generalizado a todo sistema que apoye a la toma de decisiones incluyendo Inteligencia Artificial, Aprendizaje Automático o cualquier sistema de Inteligencia de Negocio haciendo cada vez más difícil la distinción de estas técnicas de las de Machine Learning (Nilsson, 2009; Wei, Wang, Rui, & Chen, 2016). Quizás el único matiz diferente es que en la Minería de Datos podemos usar modelos de Machine Learning para clasificar (cualquiera de los vistos en el apartado anterior) como modelos clásicos que simplemente realizan una transformación algebraica sobre los datos como el Análisis Factorial del modelo (Hair et al., 1999).

Otro matiz que los diferencia es en lo relativo a las reglas de asociación, donde también son válidos todos los algoritmos comentados en el epígrafe anterior sobre Machine Learning, pero fijándonos ciertos soportes mínimos y un umbral mínimo de confianza para que podamos empezar a sacar conclusiones, por ejemplo usando un análisis a priori (Chen & Lin, 2014).

## Métodos estadísticos

Estos métodos clásicos siguen siendo útiles para hacer no sólo una previsualización de los datos sino también análisis preliminares que nos permiten evaluar las características de los datos que hay que analizar. La estadística nos ayuda a resumir los datos y sacar conclusiones de ellos, así como a generalizar estas conclusiones a toda la población. Estas técnicas clásicas siguen estando en pleno vigor ya que nos ayudan a evaluar los posibles métodos de aprendizaje estadístico que podremos usar después con el conjunto de datos.

La estadística es muy útil para investigar una muestra y luego poder generalizar los resultados a la población al completo, por eso una de sus principales aplicaciones son las ciencias sociales, y en concreto el marketing. Ya se empezaron a aplicar técnicas estadísticas como el ANOVA (análisis de la varianza) para investigación en marketing a principios del siglo pasado (Ferber, 1949) aunque el gran desarrollo de estas técnicas para

investigación en marketing se produjo gracias a la iniciativa de la Fundación Ford y el Instituto Harvard de Matemáticas para su aplicación en los Negocios (Bartels, 1988).

La estadística se divide en dos grandes áreas:

- **Estadística descriptiva.** Su uso para resumir los datos, detectar “outliers” o datos extremos, varianzas y una larga serie de parámetros que, en definitiva, nos permiten obtener una visión global del conjunto de datos y de sus variables. Media, mediana, moda, desviación típica, frecuencias, frecuencias acumuladas, análisis de regresión, ANOVA, análisis factorial, correlaciones, contrastes de hipótesis, histogramas, diagramas de caja y bigotes y otras muchas herramientas clásicas se usan para realizar un análisis previo del tipo de datos del que se dispone y que nos permitirán elegir el modelo a utilizar (Hair, Anderson, Tatham, Black, & others, 1999).
- **Estadística inferencial.** Técnicas que nos ayudan a aplicar lo descubierto en una muestra al conjunto de la población bajo unas determinadas premisas (Luque-Martínez, 2014).

## Aplicaciones al Marketing

¿Y cómo podríamos aplicar todo lo anteriormente visto al marketing? El paradigma actual del marketing (López, 2009; Tavira & Estrada, 2015), nos centra la disciplina en las relaciones con el cliente, la orientación al mercado y en la necesidad de tener un exhaustivo conocimiento de los consumidores y de la competencia. Hablamos por ello, de la necesidad de información que sirva para tomar decisiones (McAfee & Brynjolfsson, 2012). Se trata pues de ver cómo incorporar todos los procesos vistos en los epígrafes previos a esta toma de decisiones y cómo estos pueden ayudar en las decisiones sobre el marketing-mix ampliado a 5 P’s (Fan et al., 2015). Recogemos la propuesta que relaciona Big Data y marketing-mix en la figura Nº7.

	Personas (People)	Producto (Product)	Promoción (Promotion)	Precio (Price)	Distribución (Place)
Datos	<ul style="list-style-type: none"> <li>• Demográficos</li> <li>• Redes Sociales</li> <li>• Clientes</li> <li>• Navegación</li> <li>• Encuestas</li> </ul>	<ul style="list-style-type: none"> <li>• Características de producto</li> <li>• Categorías de producto</li> <li>• Revisiones de clientes</li> <li>• Encuestas</li> </ul>	<ul style="list-style-type: none"> <li>• Datos promocionales</li> <li>• Encuestas</li> </ul>	<ul style="list-style-type: none"> <li>• Datos transaccionales</li> <li>• Encuestas</li> </ul>	<ul style="list-style-type: none"> <li>• Redes Sociales con geolocalización</li> <li>• Encuestas</li> </ul>
Método	<ul style="list-style-type: none"> <li>• Agrupamiento</li> <li>• Clasificación</li> </ul>	<ul style="list-style-type: none"> <li>• Asociación</li> <li>• Agrupamiento</li> <li>• Modelado de tópicos</li> </ul>	<ul style="list-style-type: none"> <li>• Regresión</li> <li>• Asociación</li> <li>• Filtrado colaborativo</li> </ul>	<ul style="list-style-type: none"> <li>• Regresión</li> <li>• Asociación</li> </ul>	<ul style="list-style-type: none"> <li>• Regresión</li> <li>• Clasificación</li> </ul>
Aplicación	<ul style="list-style-type: none"> <li>• Segmentación clientes</li> <li>• Perfiles de clientes</li> </ul>	<ul style="list-style-type: none"> <li>• Ontología de producto</li> <li>• Reputación de producto</li> </ul>	<ul style="list-style-type: none"> <li>• Análisis de marketing promocional</li> <li>• Sistemas de recomendación</li> </ul>	<ul style="list-style-type: none"> <li>• Análisis de estrategia de precios</li> <li>• Análisis de la competencia</li> </ul>	<ul style="list-style-type: none"> <li>• Publicidad geolocalizada</li> <li>• Análisis dinámico de comunidades</li> </ul>

FIGURA Nº 7: Marco de trabajo con el marketing mix para Big Data

De esta forma, BDA podrá utilizarse para ayudar a tomar las decisiones de marketing y permitir a las empresas (Fan et al., 2015; Watson, 2019; Wedel & Kannan, 2016a): 1) estudiar cómo seleccionar las fuentes de datos apropiadas para cada objetivo de marketing; 2) analizar cómo seleccionar los métodos de análisis de datos apropiados; 3) preguntar cómo integrar diferentes fuentes de datos para estudiar complejos problemas de marketing; 4) investigar cómo tratar la heterogeneidad de las fuentes; 5) examinar cómo balancear inversiones entre las diferentes técnicas de inteligencia de marketing; y 6) ir implementando mejoras conforme van desarrollándose nuevas tecnologías asociadas al Big Data.

## Software Asociado

Casi todas estas técnicas y algoritmos están implementados en la mayoría de los lenguajes de programación por lo que están fácilmente disponibles. Los dos lenguajes más usados para la Ciencia del Dato en la actualidad son R y Python ya que gozan de una amplia comunidad científica y disponen de una ingente cantidad de librerías y documentación para la aplicación de todos estos algoritmos. Estos lenguajes de programación, así como las librerías, son de código libre y de uso gratuito. De hecho, ya hay distribuciones completas de estos dos lenguajes con todas las librerías de Machine

Learning y Ciencia del Dato precargadas y listas para su uso multiplataforma (PC, Mac y Linux) completamente gratis, como Anaconda<sup>3</sup>.

Asimismo, la parte más técnica del Big Data, como la captura de datos de diversas fuentes, el almacenamiento y el procesamiento paralelo, está basado en software de la fundación Apache<sup>4</sup> como Hadoop, Spark, HDFS, Bases de datos no relacionales y otros, también de código libre.

## Adopción de Tecnología

Para que una nueva tecnología (como lo es el Big Data) tenga éxito, es determinante su fase de adopción. Por ello, el proceso de adopción de una nueva tecnología es una de las áreas más maduras dentro de la literatura de los sistemas de información. Se han desarrollado bastantes modelos, que consistentemente han explicado una varianza superior al 40% (Venkatesh et al., 2003) en cuanto a la intención individual de usar una tecnología.

Los modelos de aceptación de tecnología provienen del ámbito de la psicología, la sociología y los sistemas de información (Westenberg, 2014). Otros modelos, como el de difusión de la innovación (Rogers, 1995), provienen del campo empresarial e incluso hay autores (Moore & Benbasat, 1996) que han integrado modelos de aceptación provenientes del ámbito de las ciencias sociales como el TRA (Fishbein & Ajzen, 1975) con modelos de difusión de la tecnología provenientes del ámbito empresarial.

En nuestra investigación nos centramos en los modelos de adopción y uso de tecnología como base para la construcción de nuestros modelos teóricos propuestos para evaluar la adopción del Big Data Analytics (BDA) en las empresas. Para ello, haremos un breve repaso histórico de los modelos más relevantes desarrollados hasta llegar al UTAUT (Venkatesh et al., 2003), modelo de adopción de tecnología por parte de las empresas, y su desarrollo posterior para usuarios, el UTUAT2 (Venkatesh, Thong, & Xu, 2012).

---

<sup>3</sup> <https://www.anaconda.com/distribution/>

<sup>4</sup> <https://apache.org/>

## Teoría de la acción razonada (TRA)

La teoría de la Acción Razonada (Fishbein & Ajzen, 1975) viene del campo de la psicología social y básicamente dice que cuando un individuo tiene la intención de realizar algo, lo realizará.

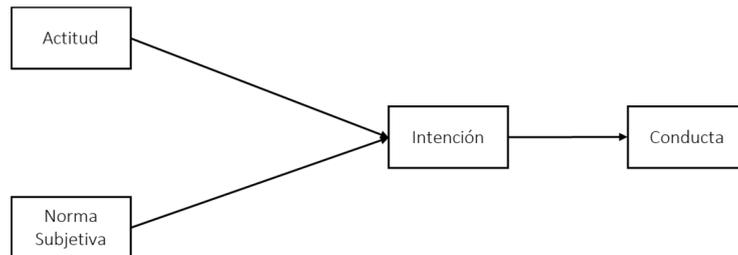


FIGURA Nº 8: Modelo de la TRA

La teoría tiene tres constructos principales que son la actitud y la norma subjetiva como antecedentes de la intención. La intención es el determinante de la conducta. Tanto la actitud como la norma subjetiva (influencia social) son una suma de creencias por lo que cambios externos raramente provocarán cambios en ellas hasta que cambien las creencias.

La principal limitación que se hace de este modelo, es que hay más factores que afectan al comportamiento de un individuo y no sólo los conscientes (Vallerand et al., 1992).

## Teoría del comportamiento planeado (TPB)

Esta teoría (Ajzen, 1991) se desarrolló como una extensión de la TRA para intentar solucionar su principal limitación. Para ello se añadió el constructo Control del Comportamiento Percibido (dificultad percibida para realizar la acción) como antecedente tanto de la intención, como de la conducta. Por ejemplo, dados dos individuos con la misma intención de hacer algo. El primero tiene alguna limitación temporal, monetaria, o de cualquier otro tipo, mientras que el segundo no las tiene. En estas circunstancias, es más probable que el segundo realice la acción que el primero.

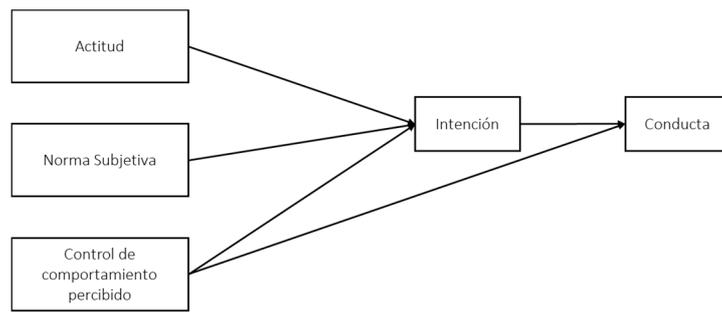


FIGURA Nº 9: Modelo de la TPB

Ha sido un modelo ampliamente usado y que ofrece buenos resultados todavía (Hassan, Shiu, & Parry, 2016; Steinmetz, Knappstein, Ajzen, Schmidt, & Kabst, 2016), aunque modelos posteriores han logrado mejores resultados (Abraham & Sheeran, 2003; Montano & Kasprzyk, 2015).

### Modelo de aceptación de la tecnología (TAM)

La primera propuesta del Modelo de Aceptación de la Tecnología-TAM ( Davis, 1986) fue una modificación de la TRA (Fishbein & Ajzen, 1975), específicamente adaptada a la aceptación de tecnologías, esto es, en la aceptación y uso de los sistemas de información, mientras que la TRA era sobre conductas genéricas.

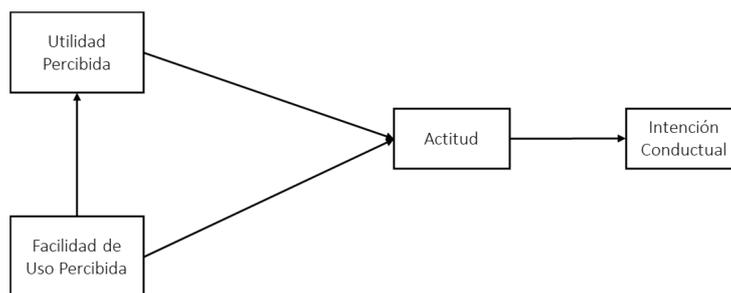


FIGURA Nº 10: Modelo TAM

Los principales constructos eran la Utilidad Percibida (percepción subjetiva de un usuario sobre la tecnología concreta) y la Facilidad de Uso Percibida (el grado de esfuerzo esperado para el uso de dicha tecnología) como antecedentes de la Actitud hacia dicha tecnología, que a su vez será el único antecedente de la Intención de Uso. Este modelo

fue reformulado ( Davis, 1989) para añadir variables externas como diseño del sistema o estructura organizativa como ya formularon también Fishbein y Ajzen (1975). También se añadió la Utilidad Percibida como antecedente de la Intención de Uso y como variable dependiente el Uso Actual del Sistema tal y como podemos ver en la Figura N°11.

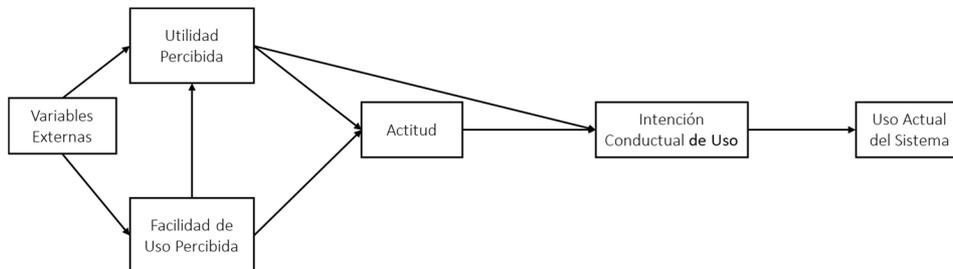


FIGURA N° 11: TAM reformulado

Este modelo fue, durante muchos años, la referencia (y aún hoy se sigue usando) para la investigación en aceptación de tecnología (Vogelsang, Steinhüser, & Hoppe, 2013; Westenberg, 2014) ya que unía las variables cognitivas de anteriores teorías psicológicas, con las variables afectivas como determinantes de la aceptación y uso de nuevas tecnologías.

## TAM2

El TAM siguió evolucionando hasta la propuesta del TAM2 (Venkatesh & Davis, 2000) que ahonda en las razones por las que un individuo percibía un sistema como útil. Formulaba la Utilidad Percibida en términos de Influencia Social (no presente en el TAM original) y procesos cognitivos, como la imagen, la relevancia dada a la tarea y la calidad del resultado obtenido al usar la tecnología.

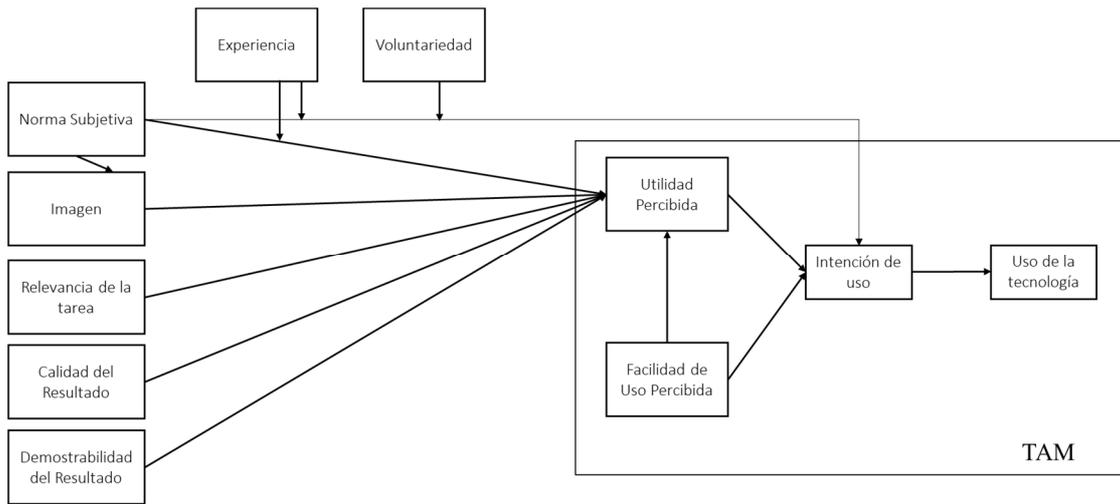


FIGURA N° 12: TAM2

Su principal limitación se encuentra en la mayor complejidad del modelo que le hacía perder su parsimonia original.

### TAM3

El TAM3 (Venkatesh & Bala, 2008) se desarrolló a partir del TAM2 ampliando los antecedentes de la Facilidad de Uso Percibida. Para ello usa las creencias del individuo (factores de anclaje) y su evolución o cambio a lo largo del tiempo (factores de ajuste). Esta ampliación extiende también las variables moderadoras de Experiencia y Voluntariedad a más relaciones tal y como podemos observar en la Figura N°13.

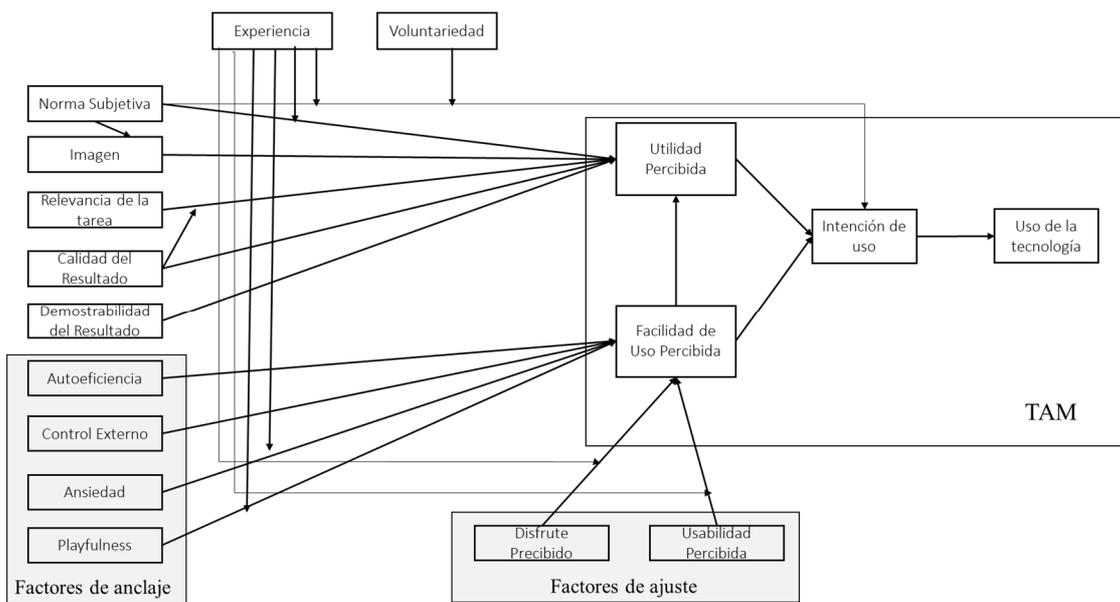


FIGURA N° 13: TAM3

## Modelo del uso de ordenadores personales (MPCU)

El Modelo de Uso de Ordenadores Personales (Thompson, Higgins, & Howell, 1991) se basó también en la TRA (Fishbein & Ajzen, 1975) y en la Teoría de la Conducta Humana (Triandis, 1979) aunque fue desarrollado pensando que los investigadores habían usado de forma incorrecta las teorías anteriores como la TRA para la aceptación y uso de ordenadores personales.

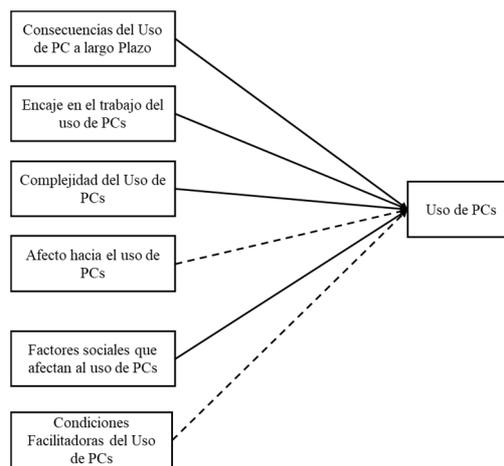


FIGURA Nº 14: Modelo del Uso de PCs

## Teoría cognitiva social

La nueva y actualizada Teoría Cognitiva Social (Compeau, Higgins, & Huff, 1999) se desarrolló sobre la previa Teoría Cognitiva Social (Bandura, 1986) pero adaptándola al uso de ordenadores y tiene como novedad la continua retroalimentación entre el entorno, la conducta y las percepciones mientras que las teorías anteriores se basaban solo en relaciones causa-efecto.

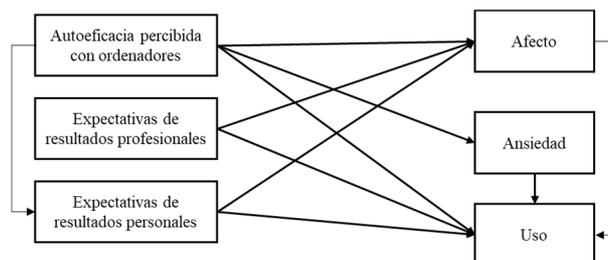


FIGURA Nº 15: Modelo de la Teoría Cognitiva Social

## Teoría unificada de la aceptación y uso de la tecnología (UTAUT)

El desarrollo de tantos modelos explicativos llevó a Venkatesh, Morris, Davis, & Davis (2003) a construir un modelo que unificase los anteriores procurando aprovechar las virtudes de cada uno y fue así como se desarrolló la teoría unificada de la aceptación y uso de la tecnología- UTAUT (Venkatesh et al., 2003). Después de hacer una selección de entre todos los modelos desarrollados, tomaron como referencia los 8 modelos que mejores resultados conseguían y que eran: la Teoría de la Acción Razonada, Teoría de la Difusión de la Innovación, Teoría del Comportamiento Planeado, Modelo de aceptación de Tecnologías, Modelo Motivacional, Teoría Cognitiva Social, Modelo de uso del PC y un modelo que hibridaba el TAM y la TRA.

No sólo se sintetizó lo mejor de los modelos anteriores, sino que también se cambió la muestra a organizaciones empresariales (los anteriores modelos habían sido testados en su mayoría con estudiantes) y se usó en más de un periodo de tiempo dentro de una misma organización, lo que permitió descubrir variables moderadoras, y en distintos niveles del organigrama organizativo para evitar el sesgo que puede suponer que los jefes puedan obligar a adoptar cierta tecnología a sus colaboradores y subordinados.

Este modelo tiene tres grandes constructos que son los antecedentes de la Intención de Uso y otro constructo como antecedente del Uso además de la relación final de la Intención de Uso sobre el Uso, tal y como podemos ver en la Figura Nº 16<sup>5</sup>.

---

<sup>5</sup> Para evitar perder matices con la traducción y puesto que es el modelo que vamos a usar en nuestra investigación, vamos a dejar los nombres originales de los constructos.

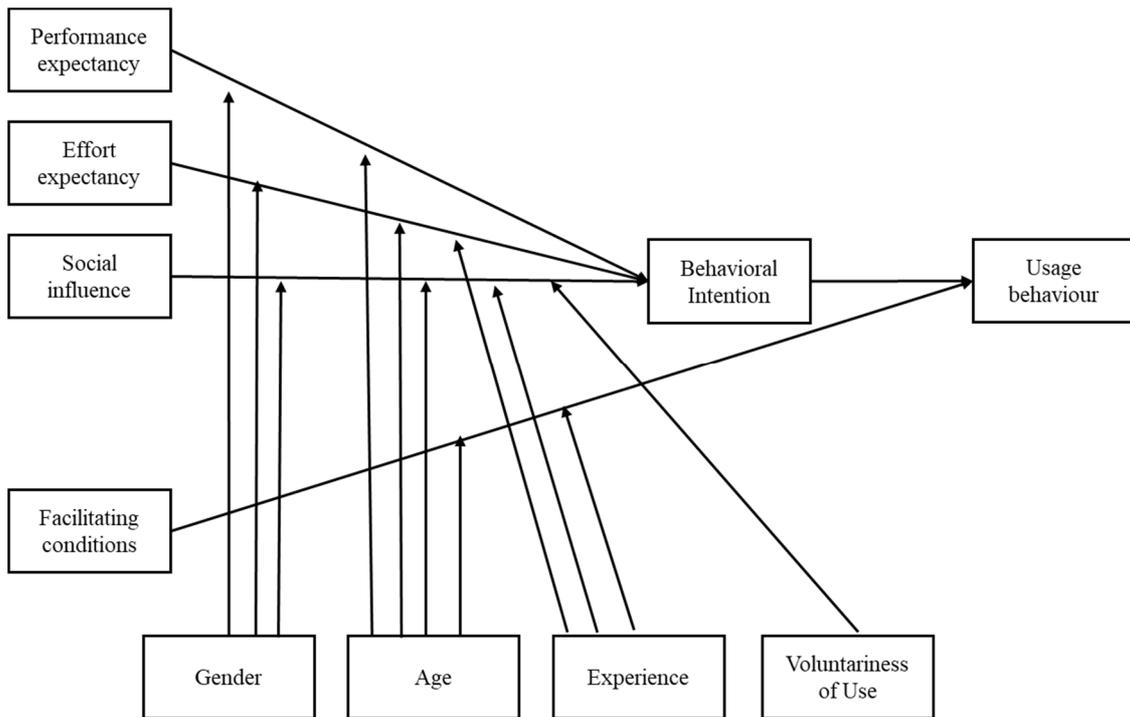


FIGURA N° 16: Modelo UTAUT

Los principales constructos del modelo que actúan como antecedentes son:

**Performance Expectancy**, que es lo que esperamos obtener al aplicar la nueva tecnología. Tiene sus precedentes en la utilidad percibida, la motivación extrínseca y el encaje en el trabajo y está moderada en su relación con la Intención de Uso por la edad y el género.

**Effort Expectancy**, que es la facilidad de uso de la nueva tecnología, basada en las precedentes de la facilidad de uso percibida y facilidad de uso, siendo moderada su relación con la Intención de Uso, por el género, la edad y la experiencia.

**Social Influence**, que es el grado en el que el individuo percibe que es importante para los demás el estar usando esa tecnología. Está basada, en la norma subjetiva, los factores sociales y la imagen. El efecto moderador, en la relación de la Influencia Social sobre la Intención de Uso, viene dado por el género, la edad, la experiencia y la voluntariedad de uso.

**Facilitating Conditions**, que es el grado en el que el individuo cree que la organización y la infraestructura técnica y humana de la empresa facilitan el uso de la nueva tecnología. Está basada en el control del comportamiento percibido, las condiciones facilitadoras y la compatibilidad, siendo moderada su relación con el Uso por la edad y la experiencia.

## UTAUT2

El UTAUT fue desarrollado para explicar la adopción y uso de nuevas tecnologías por parte de las empresas, pero todavía quedaba pendiente la adopción de tecnologías por el usuario final. Eso llevó al desarrollo del UTAUT2 (Venkatesh, Thong, & Xu, 2012) para poder explicar cómo los consumidores finales adoptan nuevas tecnologías.

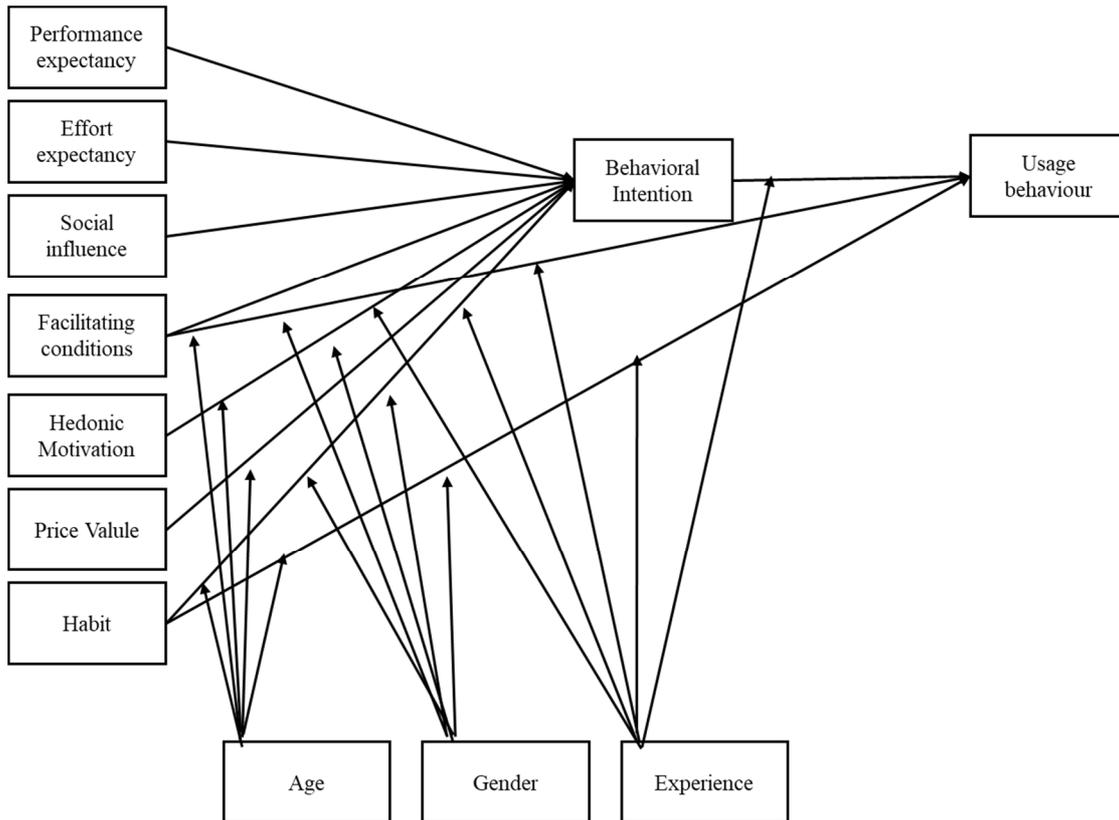


FIGURA Nº 17: Modelo UTAUT2

Los constructos originales cambiaron también ligeramente para ser adaptados a los usuarios finales en vez de empresas y se incluyeron nuevos constructos como la motivación hedónica, el valor percibido por el precio pagado y el hábito de uso. Se eliminaron variables como la voluntariedad (se entiende que en el consumidor final no hay obligación a usar una tecnología) y se añadió una nueva relación entre las Condiciones Facilitadoras y la Intención de Uso además de la ya existente con el uso.

Aunque aceptamos que UTAUT es un modelo de aceptación de tecnologías ampliado que nos permitiría analizar la aceptación y uso por parte de las empresas del BDA, entendemos que otras variables, provenientes de la revisión de la literatura en trabajos

que utilizan modelos de adopción de tecnología, pueden ser útiles para proponer una extensión del UTAUT que mejore la capacidad explicativa de nuestro modelo propuesto en su contrastación en las empresas españolas analizadas en la muestra. Estas variables son: riesgo percibido, resistencia al uso y coste de oportunidad que son presentadas en epígrafes sucesivos.

## Nuevas variables

En la revisión de la literatura realizada se han usado muchas más variables no incluidas en los modelos revisados en el anterior epígrafe. Conforme han ido evolucionando los modelos de adopción de tecnología hemos visto que se han dejado de usar algunas variables, otras se han ampliado y han sido absorbidas por nuevas variables y algunas otras se han usado únicamente para ampliar un modelo y mejorar su capacidad explicativa.

Después de una profunda revisión de todas estas posibles variables a usar para mejorar la capacidad explicativa de nuestro modelo, hemos seleccionado tres variables:

- Riesgo Percibido
- Resistencia al Uso
- Coste de Oportunidad

Que usaremos para ampliar nuestro modelo de adopción con la finalidad de aumentar su capacidad predictiva.

## Riesgo percibido

Los trabajos de diferentes autores (Featherman & Pavlou, 2003; D. J. Kim, Ferrin, & Rao, 2008; Lee & Song, 2013; Martins, Oliveira, & Popovič, 2014) han estudiado la influencia del riesgo percibido en la intención de adoptar una nueva tecnología.

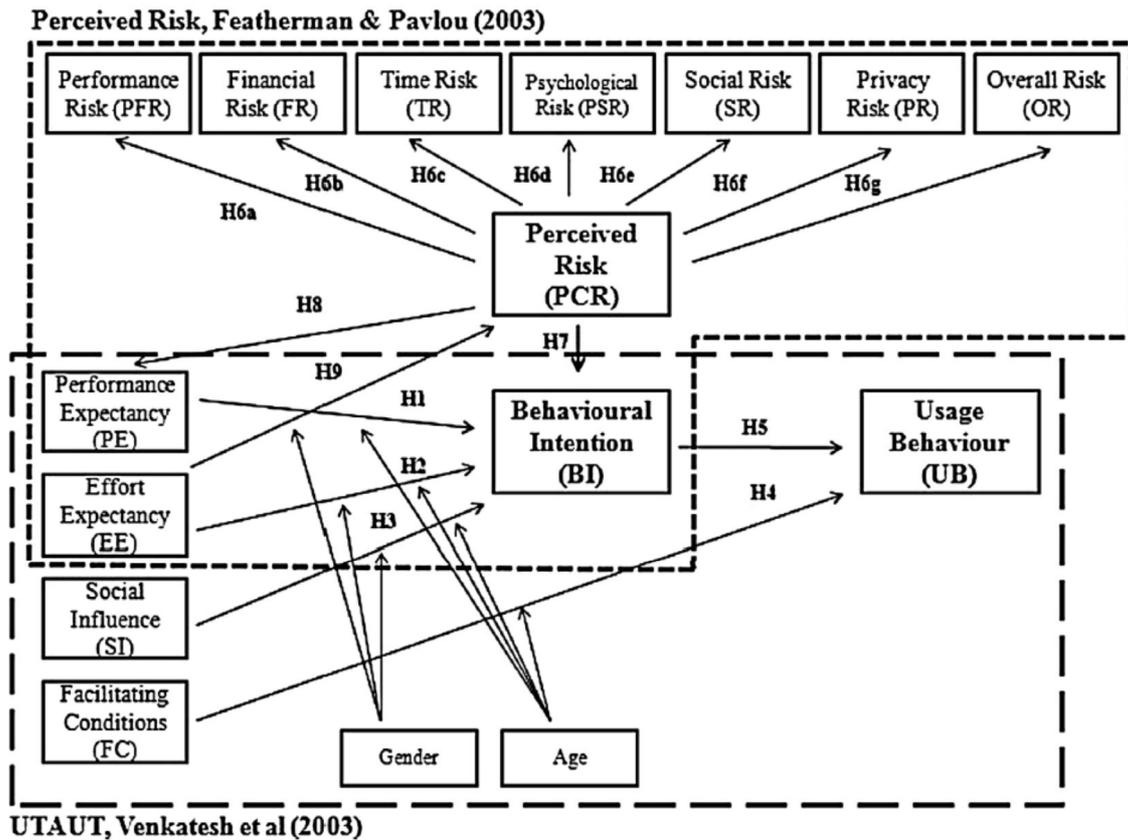


FIGURA Nº 18: Riesgo Percibido y UTAUT (Martins et al., 2004)

De hecho, en el modelo de la Figura Nº18 llega a explicar mucha más varianza que en el UTAUT original sin esta variable añadida.

### Resistencia al uso

La resistencia al uso es la oposición o reacción negativa de un usuario a una nueva tecnología (Gibson, 2004; H.-W. Kim & Kankanhalli, 2009; Lapointe & Rivard, 2017) y podemos encontrar su definición en la Teoría del Status Quo (Samuelson & Zeckhauser, 1988). Algunos estudios la han relacionado como un inhibidor de la intención de uso (Hsieh, 2015) así como también una relación negativa entre la resistencia y el uso de la nueva tecnología (Bhattacharjee & Hikmet, 2007; Poon et al., 2004). También, hay estudios que han integrado la Resistencia al Uso con la TPB (Hsieh, 2015) tal y como se ve en la Figura Nº 19, aunque hay poca literatura integrándola con el UTAUT. Por todos

estos motivos pensamos que esta variable deberíamos tenerla en cuenta a la hora de mejorar el UTAUT.

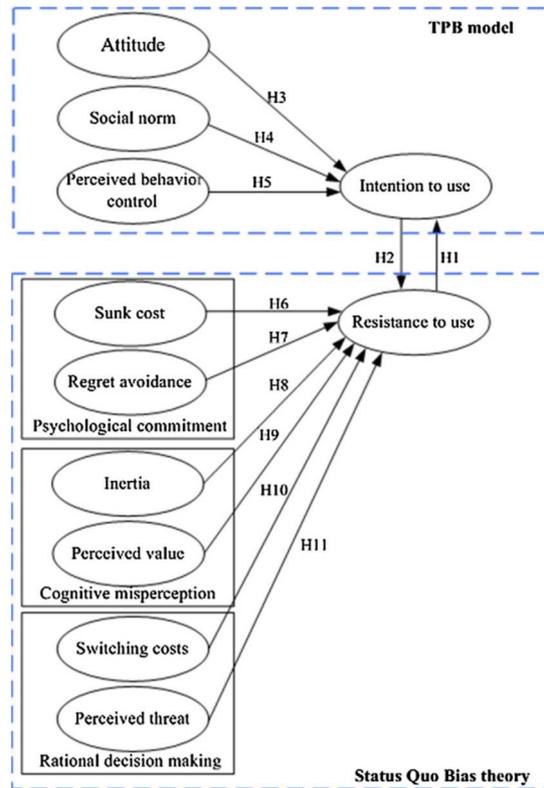


FIGURA Nº 19: Resistencia al Uso integrada en la TPB (Hsieh, 2015)

## Coste de oportunidad

Podríamos definir coste de oportunidad como la pérdida potencial en la búsqueda de un cierto resultado a la hora de implementar una nueva tecnología. ¿Qué pasa si no implemento técnicas de BDA en mi empresa? Aunque hay muy poca literatura al respecto (Lu, Hsu, & Hsu, 2005), sí que hay autores que han puesto en valor la importancia del coste de oportunidad a la hora de tomar una decisión de compra (Zikmund & Scott, 1974), y por ello pensamos que también puede ser importante a la hora de tomar una decisión sobre adopción de esta tecnología por las empresas.

## Modelos propuestos

Después de revisar todos estos modelos y sus variables más importantes, así como su evolución a lo largo del tiempo, pensamos que UTAUT, aunque altamente testado para diferentes tecnologías como salud a través de internet (Bozan, Parker, & Davey, 2016; Duarte & Pinho, 2019), comercio electrónico (Escobar-Rodríguez & Carvajal-Trujillo, 2014), banca por internet (Tarhini, El-Masri, Ali, & Serrano, 2016), juegos online (Ramírez-Correa, Rondán-Cataluña, Arenas-Gaitán, & Martín-Velicia, 2019) o telefonía 3G (Y. L. Wu, Tao, & Yang, 2007), entre otras y que podría ser mejorado extendiéndolo y ampliando con las variables señaladas con la finalidad de conseguir una mayor varianza explicada. La mayoría de los modelos consigue explicar una varianza en el entorno del 40% (Westenberg, 2014), pero las nuevas tecnologías son mucho más complejas, difíciles de entender y quizás de implantar. Queremos descubrir cuáles son los factores que afectan a dicha adopción y hacerlo mediante una extensión del UTAUT para que se adapte mejor a estas nuevas tecnologías (BDA) para las empresas.

Para ello, en esta tesis doctoral por compendio empezamos añadiendo el riesgo percibido (PR) al UTAUT tradicional tal y como se ve en la Figura N°20:

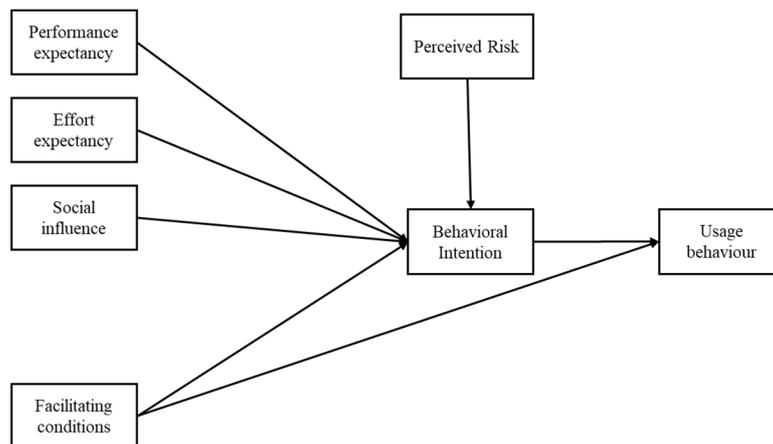


FIGURA N° 20: UTAUT más Riesgo Percibido

Los resultados después de ejecutar el modelo con y sin el riesgo percibido demostraron que el riesgo percibido era significativo y que mejoraba la capacidad predictiva del modelo, para cuya contrastación usamos una muestra inicial de 131 empresas. Vistos estos resultados, investigamos nuevas variables que se adaptaran al escenario de

adopción de Big Data Analytics y añadimos la Resistencia al Uso (*Resistance to Use*) tal y como podemos ver en la Figura N°21:

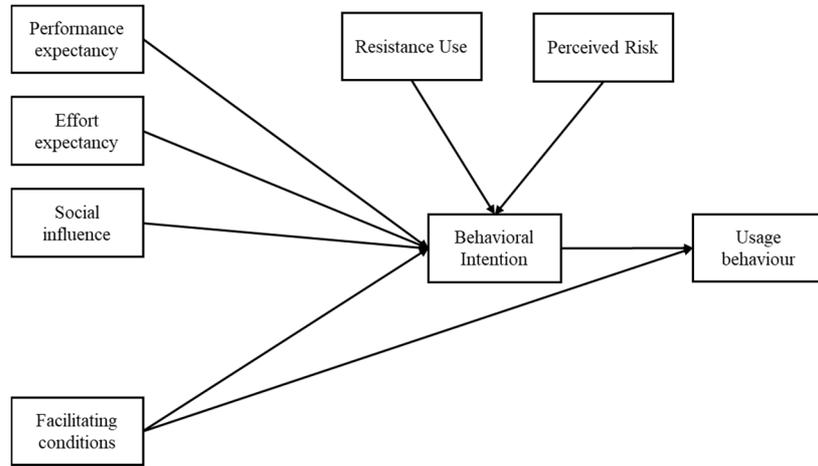


FIGURA N° 21: UTAUT más Riesgo Percibido y Resistencia al Uso

Una vez más conseguimos mejorar los resultados y conseguir explicar más del 48% de la varianza. Descubrimos nuevas variables significativas como RU y PR y a su vez descubrimos que variables clásicas como la EE iban perdiendo significatividad. Esto nos animó a dar un paso más con una muestra ampliada de empresas de servicios y auxiliares (199 empresas) y a tratar de conseguir mejorar aún más los datos obtenidos con un nuevo modelo en el que incluíamos otra nueva variable, el coste de oportunidad (OC) y posibles relaciones entre las variables propuestas tal y como podemos ver en la Figura N° 22.

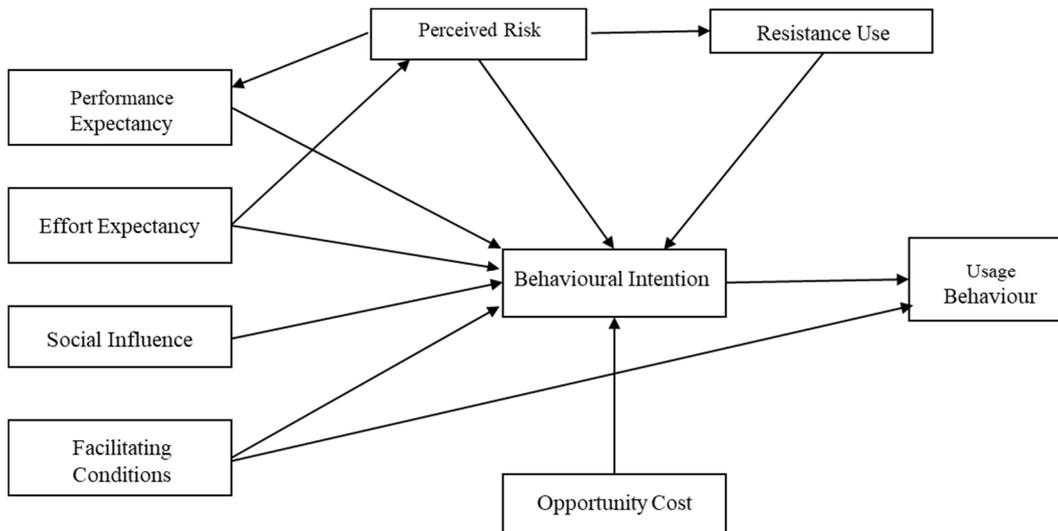


FIGURA N° 22: UTAUT más Riesgo Percibido, Resistencia al Uso y Coste de Oportunidad

Una vez más, conseguimos mejorar la capacidad predictiva de este nuevo UTAUT y confirmar que variables como la EE ya no eran significativas por lo que para nuevas tecnologías como el Big Data Analytics, los modelos propuestos mejoran el clásico UTAUT. Profundizaremos más en los resultados en el último capítulo de esta tesis doctoral. En este último capítulo justificaremos todas las hipótesis propuestas y su significatividad, así como las hipótesis que no resultaron significativas con las muestras de empresas empleadas.

## Referencias

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others. (2016). Tensorflow: A system for large-scale machine learning. In 12th Symposium on Operating Systems Design and Implementation ((pp. 265–283).
- Abraham, C., & Sheeran, P. (2003). Implications of goal theories for the theories of reasoned action and planned behavior. *Current Psychology*, 22(3), 264-280. <https://doi.org/10.1007/s12144-003-1021-7>.
- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data, SIGMOD '93 (pp. 207–216). Washington, D.C., USA.
- Ahmad, J., Farman, H., & Jan, Z. (2019). Deep learning methods and applications. In *Deep Learning: Convergence to Big Data Analytics* (pp. 31–42). Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Amini, A., Wah, T. Y., & Saboohi, H. (2014). On density-based data streams clustering algorithms: A survey. *Journal of Computer Science and Technology*, 29(1), 116–141.
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer Normalization. <https://doi.org/10.1038/nature14236>
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ, 1986.
- Bartels, R. (1988). *The History of Marketing Thought* (3d edition). Columbus: Publishing Horizons.
- Bengio, Y., Courville, A., & Vincent, P. (2012). Representation Learning: A Review and New Perspectives, (1993), 1–30. <https://doi.org/10.1145/1756006.1756025>
- Bhattacharjee, A., & Hikmet, N. (2007). Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. *European Journal of Information Systems*, 16(6), 725–737.

<https://doi.org/10.1057/palgrave.ejjs.3000717>

Bozan, K., Parker, K., & Davey, B. (2016). A closer look at the social influence construct in the UTAUT Model: An institutional theory based approach to investigate health IT adoption patterns of the elderly. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2016-March, 3105–3114. <https://doi.org/10.1109/HICSS.2016.391>

Breiman, L. (1984). *Algorithm cart. Classification and Regression Trees*. California Wadsworth International Group, Belmont, California.

Breiman, Leo. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Chen, X.-W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. *IEEE Access*, 2, 514–525.

Chollet, F., & others. (2018). *Keras: The python deep learning library*. Astrophysics Source Code Library.

Compeau, D., Higgins, C. A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 145–158.

Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2016). Very Deep Convolutional Networks for Text Classification. *Neurocomputing*, 3(1), 130–142. <https://doi.org/10.1007/s13218-012-0198-z>

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. [https://doi.org/10.1016/S0305-0483\(98\)00028-0](https://doi.org/10.1016/S0305-0483(98)00028-0)

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Unpublished Doctoral dissertation, MIT Sloan School of Management, Cambridge, M.A. <https://doi.org/oclc/56932490>

Deng, Z., Zhu, X., Cheng, D., Zong, M., & Zhang, S. (2016). Efficient kNN classification algorithm for big data. *Neurocomputing*, 195, 143–148.

- Duarte, P., & Pinho, J. C. (2019). A mixed methods UTAUT2-based approach to assess mobile health adoption. *Journal of Business Research*, 102, 140–150.
- Dull, T. (2015). Marketers ask: Isn't a data lake just the data warehouse revisited? Retrieved August 17, 2019, from <https://blogs.sas.com/content/customeranalytics/2015/08/28/isnt-a-data-lake-just-the-data-warehouse-revisited/>
- Escobar-Rodríguez, T., & Carvajal-Trujillo, E. (2014). Online purchasing tickets for low cost carriers: An application of the unified theory of acceptance and use of technology (UTAUT) model. *Tourism Management*, 43, 70–88. <https://doi.org/10.1016/j.tourman.2014.01.017>
- Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, 2(1), 28–32. <https://doi.org/10.1016/j.bdr.2015.02.006>
- Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human Computer Studies*, 59(4), 451–474. [https://doi.org/10.1016/S1071-5819\(03\)00111-3](https://doi.org/10.1016/S1071-5819(03)00111-3)
- Ferber, R. (1949). *Statistical Techniques in Market Research*. New York: McGraw Hill Book Company.
- Fishbein, M., & Ajzen, I. (1975). Belief Attitude, Intention and Behavior. An Introduction to Theory and Research. *Philosophy & Rhetoric*, 10(2), 130–132.
- Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771–780), 1612.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.
- Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems*. New York: O'Reilly Media.
- Gibson, C. F. (2004). CENTER FOR INFORMATION SYSTEMS of Management

- Massachusetts Institute of Technology IT-enabled Business Change : An Approach to Understanding and Managing Risk September 2004 Research Article : a completed research article drawing on one or more CISR resea. Group.
- Google. (2016). The History of AlphaGo. Retrieved August 20, 2019, from <https://deepmind.com/research/case-studies/alphago-the-story-so-far>
- Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification. In Proceedings of the 23rd international conference on Machine learning - ICML '06 (Vol. 32, pp. 369–376). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1143844.1143891>
- Hair, J. F., Anderson, R. E., Tatham, R. L., Black, W. C., & others. (1999). *Análisis multivariante*. Prentice Hall: Madrid.
- Hassan, L. M., Shiu, E., & Parry, S. (2016). Addressing the cross-country applicability of the theory of planned behaviour (TPB): A structured review of multi-country TPB studies. *Journal of Consumer Behaviour*, 15(1), 72–86.
- Helske, S., & Helske, J. (2017). Mixture hidden Markov models for sequence data: the seqHMM package in R. ArXiv Preprint ArXiv:1704.00543.
- Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2005). Bootstrap methods and permutation tests. *Introduction to the Practice of Statistics*, 5, 1–70.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., ... Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition. *Ieee Signal Processing Magazine*, (November), 82–97. <https://doi.org/10.1109/MSP.2012.2205597>
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>
- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network, 1–9. <https://doi.org/10.1063/1.4931082>
- Hsieh, P. J. (2015). Healthcare professionals' use of health clouds: Integrating technology

- acceptance and status quo bias perspectives. *International Journal of Medical Informatics*, 84(7), 512–523. <https://doi.org/10.1016/j.ijmedinf.2015.03.004>
- Huang, R., & Xu, W. (2015). Performance evaluation of enabling logistic regression for big data with R. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2517–2524. <https://doi.org/10.1109/BigData.2015.7364048>
- Jun, S., Lee, S. J., & Ryu, J. B. (2015). A divided regression analysis for big data. *International Journal of Software Engineering and Its Applications*, 9(5), 21–32. <https://doi.org/10.14257/ijseia.2015.9.5.03>
- Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, 44(2), 544–564. <https://doi.org/10.1016/j.dss.2007.07.001>
- Kim, H.-W., & Kankanhalli, A. (2009). Investigating user resistance to implementation: a status quo bias perspective. *MIS Quarterly*, 33(3), 2009. <https://doi.org/Article>
- Lapointe, & Rivard. (2017). A Multilevel Model of Resistance to Information Technology Implementation. *MIS Quarterly*, 29(3), 461. <https://doi.org/10.2307/25148692>
- Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21–32.
- LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(April 2016), 255–258. <https://doi.org/10.1109/IJCNN.2004.1381049>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, J.-H., & Song, C. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: An ...*, 41(3), 587–598. <https://doi.org/10.2224/sbp.2013.41.4.587>
- Liwicki, M., Graves, a, Bunke, H., & Schmidhuber, J. (2007). A novel approach to on-line

- handwriting recognition based on bidirectional long short-term memory networks. Proc. 9th Int. Conf. on Document Analysis and Recognition, 367--371. <https://doi.org/10.1.1.139.5852>
- López, J. F. C. (2009). Del marketing transaccional al marketing relacional. *Entramado*, 5(1), 6–17.
- Lu, H.-P., Hsu, C.-L., & Hsu, H.-Y. (2005). An empirical study of the effect of perceived risk upon intention to use online applications. *Information Management & Computer Security*, 13(2), 106–120. <https://doi.org/10.1108/09685220510589299>
- Luque- Martínez, T. (2014) . *Técnicas de Análisis de datos em Investigación de Mercados*. Madrid: Ed. Pirámide.
- Martins, C., Oliveira, T., & Popovič, A. (2014). Understanding the internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. *International Journal of Information Management*, 34(1), 1–13. <https://doi.org/10.1016/j.ijinfomgt.2013.06.002>
- McAfee, A., & Brynjolfsson, E. (2012). Big Data. The management revolution. *Harvard Business Review*, 90(10), 61–68. <https://doi.org/10.1007/s12599-013-0249-5>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2014). e1071: Misc functions of the Department of Statistics (e1071), TU Wien. R Package Version, 1(3).
- Mika, S., Schölkopf, B., Smola, A. J., Müller, K.-R., Scholz, M., & Rätsch, G. (1999). Kernel PCA and de-noising in feature spaces. In *Advances in neural information processing systems* (pp. 536–542).
- Montano, D.E., & Kasprzyk, D. (2015). Theory of Reason Action, Theory of Planned Behavior, and the Integrated Behavioral Model. *Health Behavior: Theory, Research and Practice*, 70(4), 231.
- Moore, G. C., & Benbasat, I. (1996). Integrating Diffusion of Innovations and Theory of Reasoned Action models to predict utilization of information technology by end-users. *Diffusion and Adoption of Information Technology* SE - 10, 132–146.

[https://doi.org/10.1007/978-0-387-34982-4\\_10](https://doi.org/10.1007/978-0-387-34982-4_10)

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence. A history of Ideas and Achievement*. Cambridge: Cambridge University Press.

<https://doi.org/10.1017/CBO9780511819346>

Núñez, E., Steyerberg, E. W., & Núñez, J. (2011). Estrategias para la elaboración de modelos estadísticos de regresión. *Revista Española de Cardiología*, 64(6), 501–507.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.

Poon, E. G., Blumenthal, D., Jaggi, T., Honour, M. M., Bates, D. W., & Kaushal, R. (2004). Overcoming barriers to adopting and implementing computerized physician order entry systems in U.S. hospitals. *Health Affairs*, 23(4), 184–190. <https://doi.org/10.1377/hlthaff.23.4.184>

Quinlan, J. R. (1993). *C4. 5: Programs for Machine Learning*. Morgan Kaufmann.

Ramírez-Correa, P., Rondán-Cataluña, F. J., Arenas-Gaitán, J., & Martín-Velicia, F. (2019). Analysing the acceptance of online games in mobile devices: An application of UTAUT2. *Journal of Retailing and Consumer Services*, 50, 85–93.

Rehman, M. H., Khan, A. ur R., & Batool, A. (2017). Big data analytics in mobile and cloud computing environments. In *Decision Management: Concepts, Methodologies, Tools, and Applications* (Vol. 3–4, pp. 1478–1496). <https://doi.org/10.4018/978-1-5225-1837-2.ch068>

Rehman, M. H. U., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, 36(6), 917–928. <https://doi.org/10.1016/j.ijinfomgt.2016.05.013>

Roeber, C., Raabe, N., Luebke, K., Ligges, U., Szepannek, G., Zentgraf, M., ... SVMlight, S. (2006). *The klaR package*. Department of Statistics, University of Dortmund.

- Rogers, E. M. (1995). *Diffusion of innovations*. Macmillian Publishing Co.  
<https://doi.org/citeulike-article-id:126680>
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65, 386.
- Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., ... Kloft, M. (2018). Deep one-class classification. In *International Conference on Machine Learning* (pp. 4393–4402).
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7–59. <https://doi.org/10.1007/BF00055564>
- Sikora, R., & others. (2015). A modified stacking ensemble machine learning algorithm using genetic algorithms. In *Handbook of Research on Organizational Transformations through Big Data Analytics* (pp. 43–53). IGI Global.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2016). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Spedicato, G. A. (2017). Discrete time Markov chains with R. *The R Journal*, 9(2), 84–104.
- Steinmetz, H., Knappstein, M., Ajzen, I., Schmidt, P., & Kabst, R. (2016). How effective are behavior change interventions based on the theory of planned behavior? *Zeitschrift Für Psychologie*.
- Sukhbaatar, S., Szlam, A., Weston, J., & Fergus, R. (2015). End-To-End Memory Networks, 1–9. <https://doi.org/v5>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Arbor, A. (2014). Going Deeper with Convolutions, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Tarhini, A., El-Masri, M., Ali, M., & Serrano, A. (2016). Extending the UTAUT model to understand the customers' acceptance and use of internet banking in Lebanon: A

- structural equation modeling approach. *Information Technology & People*, 29(4), 830–849. <https://doi.org/10.1108/ITP-02-2014-0034>
- Tavira, E. G., & Estrada, E. M. R. (2015). Marketing relacional: valor, satisfacción, lealtad y retención del cliente. Análisis y reflexión teórica. *Ciencia y Sociedad*, 40(2), 307–340.
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: toward a conceptual model of utilization. *MIS Quarterly*, 125–143.
- Tibishirani, R., Friedman, J., & Hastie, T. (2009). The elements of statistical learning: Data mining, inference, and prediction. Retrieved from <http://www-stat.stanford.edu/ElemStatLearn>
- Triandis, H. C. (1979). Values, attitudes, and interpersonal behavior. In Nebraska symposium on motivation.
- Vallerand, R. J., Deshaies, P., Cuerrier, J.-P., Pelletier, L. G., et al, & Robert J. Vallerand, Paul Deshaies, L. G. P. and C. M. (1992). Ajzen and Fishbein's theory of reasoned action as applied to moral behavior: A confirmatory analysis. *Journal of Personality and Social Psychology*, 62(1), 98–109. <https://doi.org/10.1037/0022-3514.62.1.98>
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39, 273–315.
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478. Retrieved from <http://www.jstor.org/stable/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance And Use of Technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.1017/CBO9781107415324.004>

- Vogelsang, K., Steinhüser, M., & Hoppe, U. (2013). A qualitative approach to examine technology acceptance.
- Wang, X. (2011). A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality. In The 2011 International Joint Conference on Neural Networks (pp. 1293–1299).
- Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, 44, 364–379. <https://doi.org/10.17705/1CAIS.04421>
- Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6), 97–121. <https://doi.org/10.1509/jm.15.0413>
- Wei, T., Wang, C., Rui, Y., & Chen, C. W. (2016). Network Morphism. Retrieved from <http://arxiv.org/abs/1603.01670>
- Westenberg, R. (2014). Managing Customer Innovation in the Financial Services Industry. A technology Acceptance Perspective.
- Wu, Y. L., Tao, Y. H., & Yang, P. C. (2007). Using UTAUT to explore the behavior of 3G mobile communication users. *IEEM 2007: 2007 IEEE International Conference on Industrial Engineering and Engineering Management*, 199–203. <https://doi.org/10.1109/IEEM.2007.4419179>
- Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., ... Zweig, G. (2016). Achieving Human Parity in Conversational Speech Recognition. <https://doi.org/10.1109/TASLP.2017.2756440>
- Xu, L., Fu, H.-Y., Goodarzi, M., Cai, C.-B., Yin, Q.-B., Wu, Y., ... She, Y.-B. (2018). Stochastic cross validation. *Chemometrics and Intelligent Laboratory Systems*, 175, 74–81.
- Yang, J., Zhang, D. D., Frangi, A. F., & Yang, J. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*.

Zaki, M. J. (2000). Scalable Algorithms for Association Mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3), 372–390.

Zikmund, W. G., & Scott, J. E. (1974). A multivariate analysis of perceived risk self-confidence and information sources. *ACR North American Advances*.



## CAPÍTULO 2

Extendiendo el modelo  
UTAUT para evaluar los  
factores que afectan la  
adopción del Big Data en  
empresas españolas

---

Investigación publicada como capítulo de libro: Cabrera-Sánchez J.P. y Villarejo-Ramos, A.F. (2018). Extendiendo el Modelo UTAUT para evaluar los factores que afectan la adopción del Big Data en empresas españolas. En J.A. Trespalacios, R. Vázquez, E. Estrada y C. González (Coord.) *Nuevos horizontes del marketing y de la distribución comercial*. Ed. Cátedra Fundación Ramón Areces de Distribución Comercial, pp. 181-200.

---

## Resumen

La implantación de técnicas de Big Data es una realidad en la toma de decisiones de marketing. Este desarrollo cuestiona qué factores afectan a las empresas en la aceptación de estas tecnologías. En respuesta, extendemos el modelo de aceptación de tecnologías UTAUT al Big Data y añadimos el riesgo percibido como inhibidor. Modelizamos mediante ecuaciones estructurales con Smart-PLS sobre una muestra de empresas que utilizan Big Data. Los resultados muestran la mayor predicción de la extensión del UTAUT planteada y el menor efecto de variables propuestas en el modelo original.

PALABRAS CLAVE: Big Data, intención de uso, UTAUT, riesgo percibido, PLS

## Abstract

The widespread implementation of Big Data techniques is usual in marketing decision making. This development leads us to question about the factors that affect the acceptance of this technology by companies. To answer this question, we extend the UTAUT technology acceptance model to the Big Data context adding the perceived risk as inhibitor. We use modelling on structural equations by Smart-PLS on a survey of companies that use Big Data. The results highlight that the extending model imply better prediction that original UTAUT model.

KEYWORDS: Big Data, intention behaviour, UTAUT, perceived risk, PLS

## CAPÍTULO 3

# Factors affecting the adoption of Big Data Analytics in companies

Investigación publicada como artículo de revista: Cabrera-Sánchez, J-P-, Villarejo-Ramos, A.F. (2019): Factors Affecting the Adoption of Big Data Analytics in Companies. *Revista de Administração de Empresas*, 59 (6) (in press) (JCR-IF 2018= 0'701).

---

## Abstract

With the total quantity of data doubling every two years, the low price of computing and data storage, make Big Data analytics (BDA) adoption desirable for companies, as a tool to get competitive advantage. Given the availability of free software, why have some companies failed to adopt these techniques? To answer this question, we extend the unified theory of technology adoption and use of technology model (UTAUT) adapted for the BDA context, adding two variables: resistance to use and perceived risk. We used the level of implementation of these techniques to divide companies into users and non-users of BDA. The structural models were evaluated by partial least squares (PLS). The results show the importance of good infrastructure exceeds the difficulties companies face in implementing it. While companies planning to use Big Data expect strong results, current users are more skeptical about its performance.

**KEYWORDS:** Big Data, intention behavior, unified theory of acceptance and use of technology, resistance use, perceived risk.

## Resumo

Com a quantidade total de dados duplicando a cada dois anos, o baixo preço da computação e do armazenamento de dados tornam a adoção de análises de Big Data (BDA) desejável para as empresas, como aquelas que obterão uma vantagem competitiva. Dada a disponibilidade de software livre, por que algumas empresas não adotaram essas técnicas? Para responder a essa pergunta, estendemos a teoria unificada de adoção e uso de tecnologia (UTAUT) adaptado para o contexto do BDA, adicionando duas variáveis: resistência ao uso e risco percebido. Usamos a nível da implementação da tecnologia para dividir as empresas em usuários e não usuários de técnicas de BDA. Os modelos estruturais foram avaliados por partial least squares (PLS). Os resultados mostram que a importância de uma boa infraestrutura excede as dificuldades que as

empresas enfrentam para implementá-la. Enquanto as empresas que planejam usar Big Data esperam resultados fortes, os usuários atuais são mais céticos em relação ao seu desempenho.

**PALAVRAS-CHAVE:** Big Data, intenção de usar, teoría unificada de adoção e uso de tecnología, resistência ao uso, riesgo percebido.

## Resumen

Con la cantidad total de datos duplicándose cada dos años, el bajo precio de la informática y del almacenamiento de datos, la adopción del análisis Big Data (BDA) es altamente deseable para las empresas, como un instrumento para conseguir una ventaja competitiva. Dada la disponibilidad de software libre, ¿por qué algunas empresas no han adoptado estas técnicas? Para responder a esta pregunta, ampliamos la teoría unificada de la adopción y uso de tecnología (UTAUT) adaptado para el contexto BDA, agregando dos variables: resistencia al uso y riesgo percibido. Utilizamos el grado de implantación de estas técnicas para dividir las empresas entre: usuarias y no usuarias de BDA. Los modelos estructurales fueron evaluados con Partial Least Squares (PLS). Los resultados muestran que la importancia de una buena infraestructura excede las dificultades que enfrentan las empresas para implementarla. Mientras que las compañías que planean usar BDA esperan muy buenos resultados, las usuarias actuales son más escépticos sobre su rendimiento.

**PALABRAS CLAVE:** Big Data, intención de uso, teoría unificada de la adopción y uso de tecnología, resistencia al uso, riesgo percibido.



## CAPÍTULO 4

# Acceptance and use of Big Data Techniques in services companies

Investigación publicada como artículo de revista: Cabrera-Sánchez, J.P.; Villarejo-Ramos, A.F. (2020), Acceptance and Use of Big Data Techniques in Services Companies. *Journal of Retailing and Consumer Services*, 52 (January). <https://doi.org/10.1016/j.jretconser.2019.101888> (JCR-IF 2018= 3'585).

---

## Abstract

Companies able to take advantage of the information coming from the use of Big Data will have a competitive advantage by being able to make decisions based on greater knowledge of customers and competition. Besides, the access to the software for the treatment of this great amount of data is free. So, the objective of this paper is to study the level of acceptance and use of these technologies, Big Data techniques, by services companies. To analyse the intention and use it extends the acceptance technologies model- Unified Theory of Acceptance and Use of Technology (UTAUT) - to the context of Big Data techniques, incorporating the effect on it of three new variables: resistance to use, perceived risk and opportunity cost. The structural model was evaluated using partial least squares structural equation modelling (PLS-SEM) with an adequate global fit. The verification is carried out with a sample of 199 Spanish services companies, and its main results are the strong effect of the facilitating conditions on the intention and use of Big Data, as well as the direct effect of the opportunity cost and the resistance to use on the intention, and the indirect inhibiting effect of the perceived risk through the resistance to use on intention behaviour.

**KEYWORDS:** Big Data, intention behaviour, UTAUT, use resistance, perceived risk, opportunity cost, services companies.

# CAPÍTULO 5

## Conclusiones, implicaciones, limitaciones y futuras líneas de investigación

## Conclusiones

El Big Data se ha convertido en uno de los términos tecnológicos más recurrentes cuando hablamos de marketing analítico. Esta herramienta, de enorme potencial para la toma de decisiones en las empresas, aun estando todavía en su fase incipiente de implantación, ya muestra grandes potencialidades de desarrollo, especialmente aplicables al marketing cuando hablamos de Analítica Big Data-BDA (Watson, 2019; Wedel & Kannan, 2016).

Para la correcta implantación de esta herramienta, es necesario el desarrollo del software que permita su aplicación, recabando información y analizando los datos relevantes para las empresas y para su toma de decisiones. El desarrollo de este software está en constante evolución (Chollet & others, 2018; Lecun, Bengio, & Hinton, 2015), gracias al trabajo desarrollado, principalmente, por universidades y centros de investigación de la mano de una comunidad que cree firmemente en el código libre y que ha desarrollado lenguajes de amplio uso y libre distribución como Python y R.

Los modelos de adopción de tecnología, ampliamente desarrollados por la literatura científica, nos servirán de base teórica para la construcción de nuestro modelo propuesto para la aceptación y uso del Big Data por parte de las empresas españolas que han constituido la muestra en nuestra investigación. Nuestra intención de dar un paso más en el desarrollo de dichos modelos para conseguir mejorar los resultados de un modelo inicial y un mejor ajuste del mismo, como han pretendido investigaciones precedentes (Dwivedi, Rana, Jeyaraj, Clement, & Williams, 2017; Venkatesh, Thong, & Xu, 2016). Con la extensión del modelo y la inclusión de nuevas variables, que mejoran su capacidad predictiva y ajuste, hemos alcanzado un mayor conocimiento sobre los verdaderos factores que afectan la adopción y el uso de las técnicas y herramientas de Big Data en las empresas españolas.

Los resultados de los tres modelos ampliados muestran una mejora en la capacidad predictiva del UTAUT, observada al incluir variables inhibitoras y facilitadoras de la intención de uso del Big Data. Volvemos en este capítulo a presentar, de forma resumida, los resultados obtenidos tras las sucesivas extensiones del modelo inicial UTAUT que han dado lugar a los tres modelos ampliados que se recogen en los capítulos 2, 3 y 4 de la presente tesis doctoral. Para facilitar la lectura de los datos, las relaciones significativas

se dibujarán en negrita y en líneas discontinuas las hipótesis no soportadas. En las Figuras A, B y C vemos estos resultados analizados en los capítulos que soportan las tres publicaciones de esta tesis por compendio.

En el capítulo 2, el modelo es ampliado con una variable, el Riesgo Percibido, que actúa como inhibidora de la intención de uso. Incluimos también una relación que actúa como facilitadora del uso del Big Data que es la relación entre las condiciones facilitadoras (*Facilitating conditions*) y la intención de uso (*Usage Behaviour*) siguiendo, en este caso, la propuesta recogida en el UTAUT2 (Venkatesh, Thong, & Xu, 2012), testado inicialmente para usuarios finales tecnológicos y no para empresas como en nuestro caso.

En la Figura A podemos ver los principales resultados del modelo (UTAUT más riesgo percibido (*Perceived risk*) en el que la relación entre riesgo percibido e intención de uso resulta significativa. Sin embargo, la expectativa de esfuerzo (*Effort expectancy*) sobre la misma intención de uso no alcanza p-valores significativos. Las condiciones facilitadoras en su relación con la intención de uso y con el uso han resultado muy determinantes y significativas, por encima de relaciones que se mostraban muy influyentes en el molde original como la influencia de las expectativas de resultado (*Performance expectancy*) y la influencia social (*Social Influence*) sobre la intención de uso de la tecnología. Esto tiene importancia a la hora de proponer implicaciones profesionales para las empresas que deben prestar una mayor atención a facilitar la implantación del Big Data, con herramientas que faciliten su uso, y condiciones que capaciten a los empleados para la utilización del BDA.

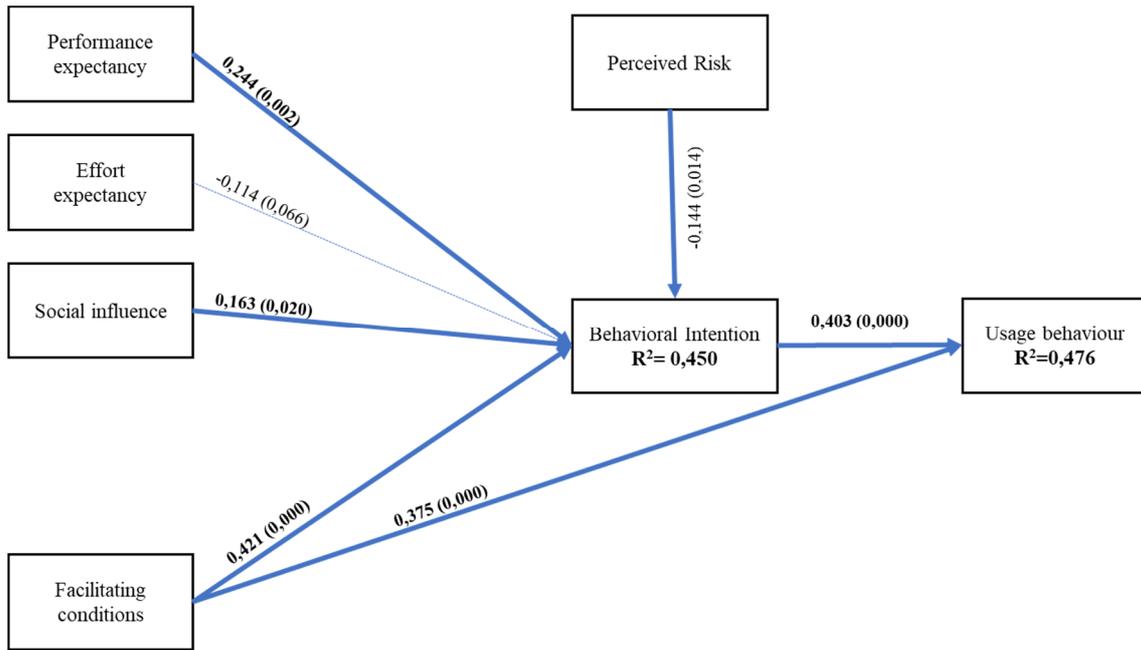


FIGURA A: UTAUT más Riesgo Percibido

La revisión de la literatura, nos permitió encontrar nuevas variables que fueron incluidas en el modelo presentado en la segunda publicación (capítulo 2) del compendio. Estas variables han permitido seguir mejorando y alcanzar mayor poder explicativo al modelo de adopción de la tecnología Big Data en las empresas. Al incluirlas en el modelo, y ser testado con una muestra mayor de empresas, los resultados volvieron a mejorar tal y como podemos ver en la Figura B. Si con el modelo de la Figura A explicamos un 45% de la intención de uso y un 47,6% del uso, el modelo con el riesgo percibido y la resistencia al uso (*Resistance use*) consigue incrementar los R cuadrado hasta un 56% para la intención de uso y al 48,3% en el caso del uso. La nueva variable incluida, resistencia al uso, resultó ser significativa en su relación con la intención de uso, perdiendo peso y significación el efecto negativo del riesgo percibido sobre la intención de uso, lo que pudimos explicar, en el tercer modelo, con un efecto indirecto del riesgo percibido a través de la resistencia al uso. De nuevo, la variable condiciones facilitadoras era la que más información aportaba a la intención de uso del Big Data para la muestra ampliada de empresas.

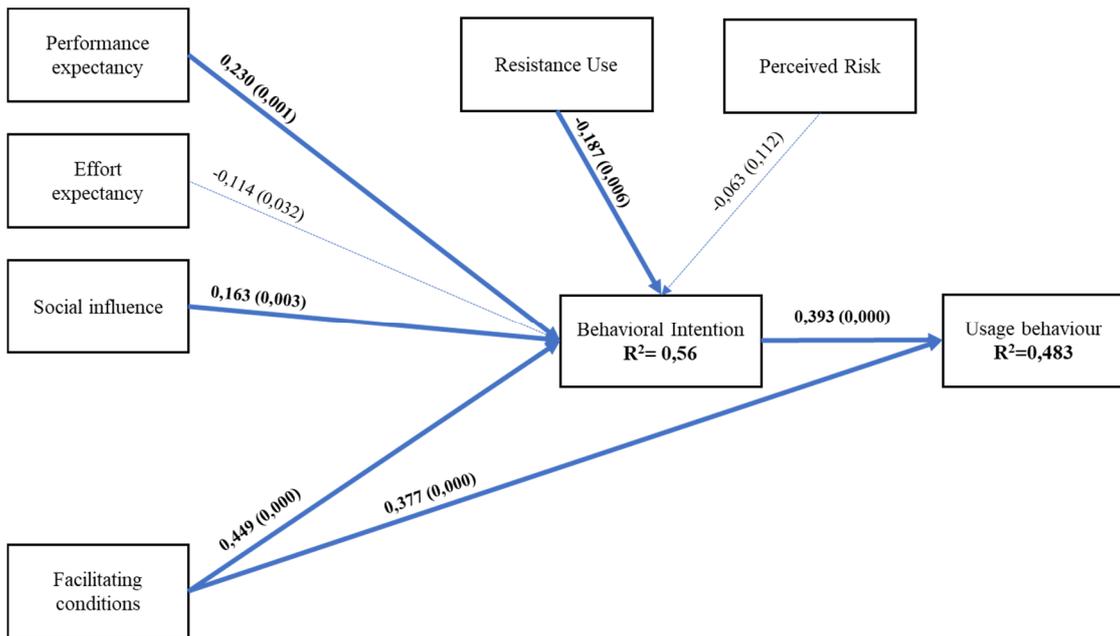


FIGURA B: UTAUT más Riesgo Percibido y Resistencia al Uso

El tercero de los modelos que nos llevó a la tercera publicación del compendio (capítulo 4) incluye las relaciones del riesgo percibido sobre la resistencia al uso, para mediar su efecto sobre la intención de uso, y del riesgo sobre la expectativa de esfuerzo. Además incluimos una nueva variable, el coste de oportunidad (*Opportunity cost*) influyendo de forma negativa sobre la intención de uso. Con este nuevo modelo (Figura C) volvimos a mejorar los resultados obtenidos. Conseguimos subir el R cuadrado de la intención de uso del 56% al 58'3% y confirmamos la influencia indirecta del riesgo percibido sobre la intención de uso, a través de la resistencia al uso.

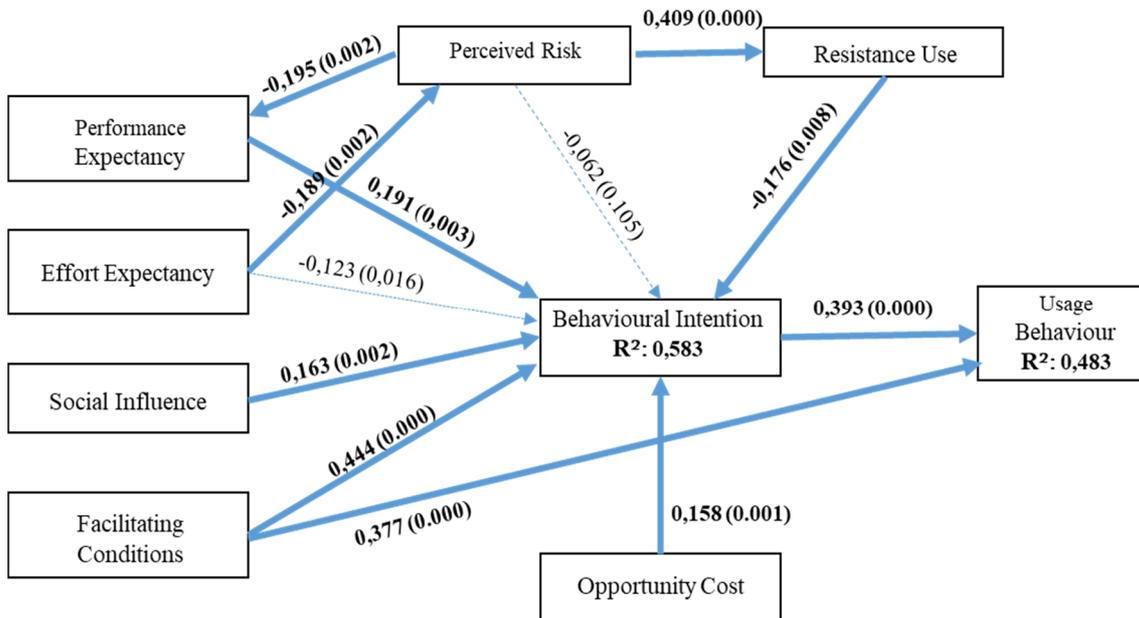


FIGURA C: UTAUT más Riesgo Percibido, Resistencia al Uso y Coste de Oportunidad

La relación de la expectativa de esfuerzo con la intención de uso resultó significativa pero en el sentido contrario al planteado. Esto se debe al efecto supresor (Falk & Miller, 1992) que realizan de la variable condiciones facilitadoras sobre la Intención de Uso. El efecto supresor ocurre cuando hay información duplicada, es decir, la información de las expectativa de esfuerzo ya está contenida en la relación condiciones facilitadoras con la intención de uso (relación no contemplada en el UTAUT original). Por otra parte, también fue significativa la relación de la nueva variable incluida, coste de oportunidad, sobre la intención de uso del Big Data, mostrando que los encuestados consideran un coste de oportunidad no implantar técnicas de Big Data en sus empresas por las consecuencias negativas que esto podría acarrearles.

## Implicaciones teóricas

La inclusión de nuevas variables ha mejorado la capacidad predictiva del UTAUT, que a su vez nos va a permitir conocer qué es lo que más afecta a la intención de uso las herramientas de Big Data dentro de las empresas. Esta intención de uso viene determinada de forma positiva por:

- La percepción de obtener buenos resultados implementando estas tecnologías (*Performance Expectancy*) ya mostrada en estudios anteriores (Chin, Kim, & Lee, 2013).
- El efecto positivo de usarlas por el efecto de la influencia que otras empresas que lo consideran importante (*Social Influence*), resultado también mostrado con anterioridad (Bozan, Parker, & Davey, 2016).
- El hecho de que la empresa proporcione las condiciones adecuadas para implantar la tecnología Big Data mediante, software, formación y capacitación, así como el fomento de su uso (Alharbi, 2014), tiene un efecto elevado sobre la intención de usarla.

Y de forma negativa por:

- El riesgo percibido a usar una nueva tecnología (Lee & Song, 2013), influyendo, directa o indirectamente a través de la resistencia al uso, sobre la intención de usar Big Data en las empresas.
- La resistencia al uso de una nueva tecnología (Polites & Karahanna, 2012), en este caso del Big Data, supone un freno importante para su implantación por parte de las empresas de la muestra.

Por su parte, y en línea con trabajos anteriores (Hsieh, 2015), los factores facilitadores tienen más peso que los factores inhibidores sobre la intención de uso del Big Data por parte de las empresas.

Destacamos, asimismo, que la expectativa de esfuerzo (*Effort Expectancy*) no afecta a la intención de uso ya que la relación no es significativa y de sentido contrario al planteado en el modelo UTAUT original. Aunque en las primeras fases de nuestra investigación dimos como explicación plausible de esta falta de significación y efecto contrario el hecho de que los ejecutivos de las empresas entrevistados podrían haber pensado que algo extremadamente fácil de usar, no debía aportar mucho valor y por ello, la intención de uso no se viera afectada por la facilidad de uso. Posteriormente, con el desarrollo y ampliación de nuestra investigación, explicamos esta situación, como hemos señalado en el resumen del tercer modelo, a través del efecto supresor descrito por Falk y Miller (1992).

El nivel de implantación de la tecnología, esto es, el hecho de que las empresas usen técnicas de Big Data o que estén empezando a usarlas, frente a las empresas que no las utilizan, ha significado un diferente comportamiento en el contraste de los distintos modelos, jugando un papel mucho más relevante la variable condiciones facilitadoras en las empresas que tienen experiencia en el uso de las herramientas y técnicas de Big Data.

## Implicaciones prácticas

Los resultados obtenidos invitan a pensar que toda nueva tecnología tiene su propia curva de aprendizaje y que este periodo de aprendizaje o curva de adopción influye cada vez menos si se esperan unos grandes resultados con su uso. Del mismo modo, no se pierde nada por probar si se dispone de hardware y técnicos formados para usarla. Si los factores que actúan como facilitadores tienen más peso que los factores inhibidores, es necesario un mayor nivel de información por parte de los desarrolladores de herramientas de BDA por conseguir informar y convencer a las empresas que existen estas herramientas, que son de acceso libre y de que su uso será más sencillo y eficiente cuanto mayor sea la infraestructura de hardware y y el nivel de conocimiento de los técnicos que la estén explotando, pues será más que sencilla su adopción venciendo así la resistencia al uso de cualquier nueva tecnología(Erevelles et al., 2016).

Para ello, también recomendaríamos realizar campañas de comunicación orientadas a transmitir las potencialidades que ofrece el Big Data, mejora en la toma de decisiones, conocimiento de los clientes, relaciones con los mercados, mostrándolo a través de casos de éxito de empresas similares o del mismo sector. Las empresas deberían explotar también el hecho de que sus empleados perciben como algo bueno el usar estas técnicas de Big Data y el coste de oportunidad que conlleva si no implementan estas técnicas con relación al hecho de que lo haga la competencia.

## Limitaciones de la investigación

Aunque el UTAUT es un modelo muy testado y con resultados muy contrastados (Venkatesh et al., 2016), su aplicación a las tecnologías más actuales, como es el Big Data, ha demostrado su potencial de mejora con la extensión del modelo incorporando nuevas variables y redefiniendo el efecto de algunas de las que conformaron la versión original. Esto nos lleva a considerar que el moldeo deberá reformularse en el futuro, incluyendo nuevas variables como las propuestas en esta tesis doctoral, u otras nuevas y explorando también nuevas variables moderadoras diferentes a las del UTAUT original por si surgen nuevas interacciones.

La limitación del tamaño de la muestra, 131 empresas en una primera investigación y su ampliación a 199, así como el contexto geográfico limitado a empresas españolas, preferentemente de servicios, supone, sin duda, la necesidad de testar el modelo final ampliado en muestras internacionales que pudiesen ser observadas para testar posibles diferencias interculturales o debidas a diferencias en el nivel de desarrollo de los países, o por los sectores de actividad presentes en la muestra. La ampliación del tamaño de la muestra y su carácter más global permitirían realiza análisis multi-grupo, como el realizado en el capítulo 4, donde pudieran confirmarse las diferencias encontradas e incluso apreciar diferencias en el comportamiento entre otros grupos.

## Futuras líneas de investigación

Con la finalidad de resolver algunas de las limitaciones de esta investigación, proponemos buscar nuevas variables que consigan aumentar la capacidad explicativa del modelo, así como posibles agrupaciones con la finalidad de buscar segmentos con diferentes comportamientos a la hora de adoptar esta nueva tecnología, ya sea por diferencias sectoriales, geográficas, dimensión de las empresas o nivel de desarrollo del país, entre otras.

Como desarrollo de esta línea de investigación iniciada con la presente tesis doctoral proponemos trasladar la adopción de estas tecnologías basadas en el Big Data con relación al uso que realizan los usuarios finales. Bien es cierto que el cliente final no usa técnicas de Big Data (al menos, conscientemente) y por ello, usaremos el término más genérico de aplicaciones de Inteligencia Artificial, que sí que usan las técnicas de Big Data como base (Herrera Triguero, 2014).

Esta propuesta investigadora, que se encuentra en fase avanzada, ha seguido un proceso de investigación similar al utilizado como base de este trabajo: 1) elegir un modelo de adopción de tecnología que se adapte mejor al caso de usuarios finales; 2) buscar mediante revisión bibliográfica y seleccionar variables con las que mejoran la capacidad predictiva del modelo; 3) seleccionar una muestra adecuada de usuarios finales de aplicaciones de inteligencia artificial; 4) evaluar los resultados del modelo/s propuestos; y 5) establecer los criterios a de una posible segmentación a posteriori de la muestra con la finalidad de chequear posibles segmentos con diferentes comportamientos a la hora de adoptar aplicaciones de inteligencia artificial con base tecnológica en el Big Data.

La propuesta inicial del modelo a aplicar nos lleva a proponer el UTAUT2 (Venkatesh et al., 2012) para realizar esta investigación con clientes finales que incluye variables como la motivación hedónica, el precio/valor percibido, o el hábito. La revisión de la literatura en adopción de tecnologías a usuarios finales (Dubey & Gunasekaran, 2015; Owusu Kwateng, Osei Atiemo, & Appiah, 2019; Ramírez-Correa et al., 2019; Shaw & Sergueeva, 2019) así nos lo recomienda. En el siguiente paso, buscamos referencias que avalen la inclusión de nuevas variables que mejoran la capacidad explicativa del modelo de origen, como hicimos en la investigación con empresas, y seleccionamos dos variables nuevas: confianza (*Trust*) basada en los trabajos anteriores con muestras de consumidores (Bock et al., 2005; Schoorman, Mayer, & Davis, 2007; I. L. Wu & Chen, 2005; Zhou, 2012) y miedo a la tecnología (*Technology Fear* ()), derivada del concepto ampliamente estudiado de la ansiedad tecnológica (Guo, Sun, Wang, Peng, & Yan, 2013; Niemelä-Nyrhinen, 2007).

Los resultados preliminares, obtenidos con muestras amplias aunque preferentemente de estudiantes universitarios, dan unos resultados satisfactorios con relación al poder explicativo y la capacidad predictiva del modelo ampliado del UTAUT 2. Han sido testados

para diferentes aplicaciones de inteligencia artificial, como los recomendadores de compra online, comprobando la importancia que tiene para el uso de estas aplicaciones el efecto de la confianza y la necesidad de superar el miedo tecnológico para adoptarlas.

Por otro lado, al usar la confianza y el miedo a la tecnología como variables para realizar una segmentación a posteriori (Becker et al., 2013), encontramos segmentos con comportamientos significativamente diferentes. Con esta segmentación podríamos enriquecer más las implicaciones teóricas y prácticas al establecer formas de comportamiento diferenciado entre usuarios de aplicaciones de inteligencia artificial y podríamos adecuar la comunicación a cada uno de estos segmentos, según aquellos usuarios que muestren mayores o menores niveles de miedo tecnológico y/o de confianza y seguridad al usar estas aplicaciones.

## Referencias

- Alharbi, S. T. (2014). Trust and acceptance of cloud computing: A revised UTAUT model. Proceedings - 2014 International Conference on Computational Science and Computational Intelligence, CSCI 2014, 2(Mm), 131–134. <https://doi.org/10.1109/CSCI.2014.107>
- Becker, J.-M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats. MIS Quarterly, 37(3), 665–694. <https://doi.org/10.25300/MISQ/2013/37.3.01>
- Bock, Zmud, Kim, Lee, Bock, G.-W., Zmud, R. W., ... Lee, J.-N. (2005). Behavioral Intention Formation in Knowledge Sharing : Examining the Roles of Extrinsic Motivators, Social-Psychological Forces , and Organizational Climate. MIS Quarterly, 29(1), 87–111. <https://doi.org/10.2307/25148669>
- Bozan, K., Parker, K., & Davey, B. (2016). A closer look at the social influence construct in the UTAUT Model: An institutional theory based approach to investigate health IT adoption patterns of the elderly. Proceedings of the Annual Hawaii International Conference on System Sciences, 2016-March, 3105–3114. <https://doi.org/10.1109/HICSS.2016.391>
- Chin, W. W., Kim, Y. J., & Lee, G. (2013). Testing the Differential Impact of Structural Paths in PLS Analysis: A Bootstrapping Approach. Springer Proceedings in Mathematics and Statistics, 56, 221–229. [https://doi.org/10.1007/978-1-4614-8283-3\\_15](https://doi.org/10.1007/978-1-4614-8283-3_15)
- Chollet, F., & others. (2018). Keras: The python deep learning library. Astrophysics Source Code Library.
- Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analy...: EBSCOhost. Industrial and Commercial Training, Vol. 47 174-181. <https://doi.org/10.1108/ICT-08-2014-0059>
- Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2017). Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT):

- Towards a Revised Theoretical Model. *Information Systems Frontiers*, 1–16.  
<https://doi.org/10.1007/s10796-017-9774-y>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.  
<https://doi.org/10.1016/j.jbusres.2015.07.001>
- Falk, R. F., & Miller, N. B. (1992). *A Primer for Soft Modeling*. University of Akron Press. Akron, Ohio (USA).
- Guo, X., Sun, Y., Wang, N., Peng, Z., & Yan, Z. (2013). The dark side of elderly acceptance of preventive mobile health services in China. *Electronic Markets*, 23(1), 49–61.  
<https://doi.org/10.1007/s12525-012-0112-4>
- Heinssen Jr., R. K., Glass, C. R., & Knight, L. A. (1987). Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior*, 3(1), 49–59. [https://doi.org/10.1016/0747-5632\(87\)90010-0](https://doi.org/10.1016/0747-5632(87)90010-0)
- Herrera Triguero, F. (2014). *Inteligencia artificial, inteligencia computacional y Big Data*. Universidad de Jaén - Servicio de publicaciones e intercambio. Jaén (Spain).
- Hsieh, P. J. (2015). Healthcare professionals' use of health clouds: Integrating technology acceptance and status quo bias perspectives. *International Journal of Medical Informatics*, 84(7), 512–523.  
<https://doi.org/10.1016/j.ijmedinf.2015.03.004>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.  
<https://doi.org/10.1038/nature14539>
- Lee, J.-H., & Song, C. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: An ...*, 41(3), 587–598.  
<https://doi.org/10.2224/sbp.2013.41.4.587>
- Niemelä-Nyrhinen, J. (2007). Baby boom consumers and technology: shooting down stereotypes. *Journal of Consumer Marketing*, 24(5), 305–312.  
<https://doi.org/10.1108/07363760710773120>
- Owusu Kwateng, K., Osei Atiemo, K. A., & Appiah, C. (2019). Acceptance and use of

- mobile banking: an application of UTAUT2. *Journal of Enterprise Information Management*, 32(1), 118–151.
- Polites, G. L., & Karahanna, E. (2012). Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. *MIS Quarterly*, 36(1), 21–42. <https://doi.org/71154897>
- Ramírez-Correa, P., Rondán-Cataluña, F. J., Arenas-Gaitán, J., & Martín-Velicia, F. (2019). Analysing the acceptance of online games in mobile devices: An application of UTAUT2. *Journal of Retailing and Consumer Services*, 50, 85–93.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integrative Model of Organizational Trust: Past, Present, and Future. *Academy of Management Review*, 32(2), 344–354. <https://doi.org/10.5465/amr.2007.24348410>
- Shaw, N., & Sergueeva, K. (2019). The non-monetary benefits of mobile commerce: Extending UTAUT2 with perceived value. *International Journal of Information Management*, 45, 44–55.
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance And Use of Technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.1017/CBO9781107415324.004>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. *Journal of the Association for Information Systems*, 17(5), 328–376.
- Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, 44, 364–379. <https://doi.org/10.17705/1CAIS.04421>
- Wedel, M., & Kannan, P. K. (2016). Marketing Analytics for Data-Rich Environments. *Journal of Marketing*, 80(6), 97–121. <https://doi.org/10.1509/jm.15.0413>
- Wu, I. L., & Chen, J. L. (2005). An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study. *International Journal of Human*

Computer Studies, 62(6), 784–808. <https://doi.org/10.1016/j.ijhcs.2005.03.003>

Zhou, T. (2012). Examining Location-Based Services Usage From the Perspectives of Unified Theory of Acceptance and Use of Technology and Privacy Risk. *Journal of Electronic Commerce Research*, 13(2), 135–144. [https://doi.org/10.1007/978-3-531-92534-9\\_12](https://doi.org/10.1007/978-3-531-92534-9_12)