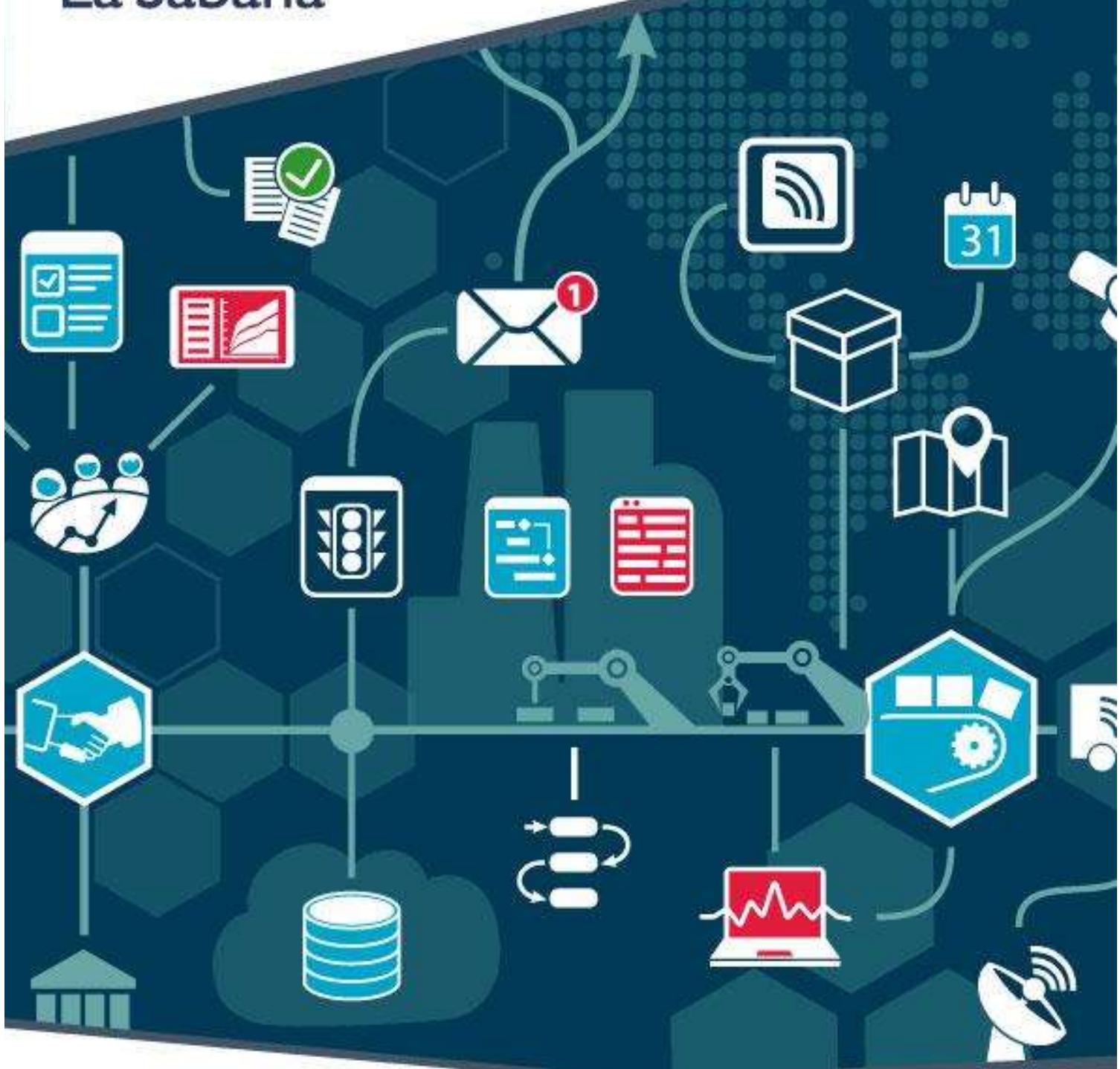




Universidad de  
**La Sabana**

Analítica y Big Data



En este módulo se establece la importancia que actualmente tiene el Big Data para la toma de decisiones en las empresas, así como las diferencias entre Big Data y Data Science y los diferentes perfiles profesionales o roles que en este momento se están demandando no solo en Colombia sino también en Latinoamérica y el mundo.

1

Por lo tanto, se conocerán conceptos generales sobre el Big Data y las distintas herramientas que existen para aplicarlo en diferentes proyectos que se estén desempeñando en las empresas y lo más importante la descripción global para que el participante vea en qué parte del universo Big Data está o hacia dónde se quiere mover.

## Tabla de contenido

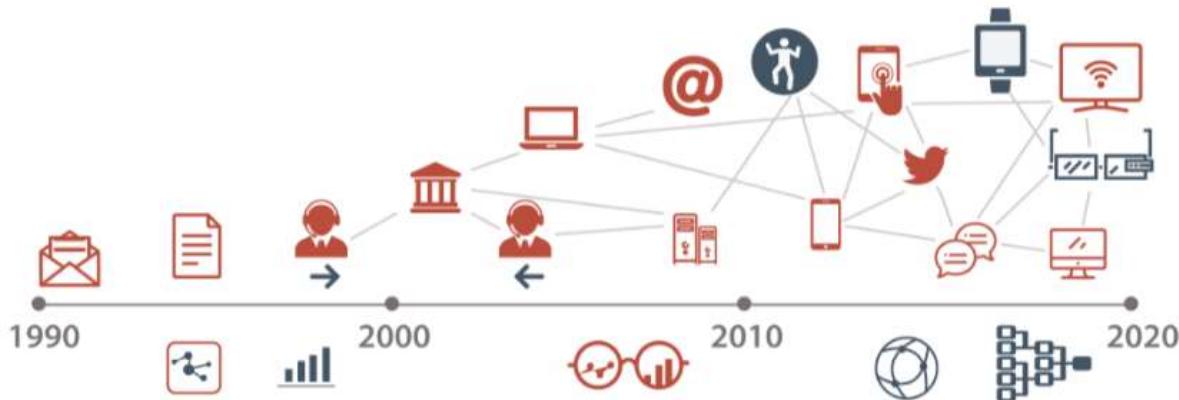
1. Introducción a la analítica del negocio.....	2
1.1. Diferencias entre Big Data y Data Science .....	4
1.2. Las V de Big Data.....	5
1.3. Arquitectura de datos en Big Data.....	9
1.4. Herramientas de analítica y Big Data.....	11
2. Introducción a Machine Learning .....	12
3. Proceso Crisp-DM.....	15
4. Casos de negocio en Big Data .....	18
5. Analítica y Big Data en Latam y Colombia.....	20
6. Demandas en cargos de Big Data.....	23
Referencias.....	26



## 1. Introducción a la analítica del negocio

Dentro del campo analítico es importante para las compañías de hoy tener un análisis de la información que producen en cada momento. Es por eso que es importante definir el proceso analítico, el cual empieza con la multiplicidad de datos.

2



Mientras que para 1990 solo se tenía información de textos planos, ya en el 2000 se empieza a tener información de sistemas CRM, donde se evidencia que se involucra al cliente con la relación de productos. Es decir, a medida que se avanza en el tiempo se empieza a tener información de internet y el computador genera información de una manera exponencial.

Por ejemplo, para el 2010 se tiene la explosión de las redes sociales y ya para el 2018 mediante la información generada por máquinas, sensores, tablets, entre otros, hacen que el Big Data tenga suma importancia en todas las fases; para posteriormente entrar en una era analítica donde todas las decisiones se toman en función de lo que necesita un usuario o un ente en tiempo real.



De esta manera, se entra en un mundo muy inmenso que a donde se mire todo está generando datos como, por ejemplo, al momento de hacer una compra con la tarjeta de crédito se puede evidenciar qué tipo de tienda visité, el monto, el día, incluso relacionar el día con el clima y saber la ubicación geográfica de la tienda.

3

En fin, todas estas variables y muchas más son relevantes y le servirán al banco emisor de la tarjeta, a la tienda donde se compró, a la franquicia de la tarjeta y muchos más. Como se evidencia cada situación genera un sinfín de información que debe analizarse de manera precisa.

La siguiente es una encuesta que encierra distintos usos del Big Data:



Imagen 1. Encuesta de los usos del Big Data

Fuente: (IBM Institute for Business Value and Saïd Business School, University of Oxford, 2014)

**Nota:** se pidió a los encuestados que de las opciones facilitadas eligieran hasta dos descripciones de la visión que tenían sus empresas de Big Data. Las opciones se han abreviado y las elecciones se han normalizado para alcanzar el 100%. Total de encuestados: 1.144.



## 1.1. Diferencias entre Big Data y Data Science

Big Data	Data Science
Big Data es la capacidad de procesamiento y lectura de distintas fuentes de datos.	Capacidad analítica medida en distintas fases.
Trabaja fuentes de datos en términos de almacenamiento. Storage es el término con el cual se relacionan las características de guardado de información.	Realiza análisis de datos en términos descriptivos, predictivos y prescriptivos.
Transformación: Capacidad de transformar los datos que no tienen una adecuada lectura rápida.	Realiza transformaciones analíticas usando algoritmos que definen particiones diferenciales al negocio.
Dispone de datos consumibles a los usuarios.	Analiza los datos consumibles y realiza predicciones sobre los mismos.

4

Tabla 1. Diferencias de Big Data y Data Science

En conclusión, el Big Data y Data Science son dos áreas que se complementan en gran medida. El primero dispone los datos y tiene las herramientas para hacerlo, y el segundo analiza toda la data que está disponible desde el Big Data.



## 1.2. Las V de Big Data



5

### Volumen



### Variedad

- Tipos de datos

**Estructurados:** son aquellos datos que tienen una estructura definida, por ejemplo, los datos que leemos en una hoja de Excel donde las filas son los clientes y las columnas son las variables demográficas.



### Ejemplo:

	A	B	C	D	E
1	Producto	Enero	Febrero	Marzo	Abril
2	Producto 1	25,318	16,815	25,755	11,720
3	Producto 2	21,564	20,800	16,059	17,174
4	Producto 3	28,362	29,802	10,670	21,885
5	Producto 4	23,856	16,919	17,386	18,979
6	Producto 5	18,617	11,363	11,649	20,052
7	Producto 6	27,518	27,498	19,450	29,478
8	Producto 7	14,115	29,544	22,327	15,680
9	Producto 8	11,972	14,375	13,809	23,585

Tabla 2. Ejemplo de datos estructurados

**Semiestructurados:** son aquellos datos donde se debe hacer algún tipo de transformación, con el fin de darle una mejor estructura con unas técnicas básicas.

**Ejemplo:** texto en donde se tienen distintas palabras que se publican en twitter.

- Estoy en desacuerdo con el gobierno nacional en aprobar más presupuesto para la guerra que para la educación.
- El gobierno desaprueba el presupuesto para la educación.

Para los dos tweets anteriores podemos darle una estructura de la siguiente manera:



twitt	desacuerdo	gobierno	presupuesto	educación	guerra	desaprueba
1	1	1	1	1	1	0
2	0	1	1	1	0	1

Tabla 3. Ejemplo de datos semiestructurados

7

En donde la palabra “guerra” aparece en la frase 1 y no en la 2. La matriz anterior muestra con un 1 cuando la palabra existe en la frase y 0 cuando no. De esta manera, los textos tienen una estructura que se puede contar o analizar con técnicas de text analytics.

**No estructurados:** son datos que tienen una estructura definida, pero se tienen que emplear técnicas más avanzadas para poder darle una estructura más definida.

**Ejemplo:** un audio por sí mismo no es analizable. Pero si se tienen 10000 audios en un call center, lo primero que se debe hacer es transformar el audio a texto y se tendrían datos como semiestructurados, para luego aplicar la técnica mencionada anteriormente y darle una estructura.

Datos estructurados	Datos semiestructurados	Datos no estructurados
<b>Fichas de clientes</b> Fecha de nacimiento Nombre Dirección Transacciones en un mes Puntos de compra	<b>Correos electrónicos</b> <i>Parte estructurada:</i> destinatario, receptores, tema <i>Parte no estructurada:</i> cuerpo del mensaje	<b>Persona a persona</b> Comunicaciones en las redes sociales <b>Persona a máquina</b> Dispositivos médicos Comercio electrónico Ordenadores, móviles <b>Máquina a máquina</b> Sensores, dispositivos GPS Cámaras de seguridad

Tabla 4. Tipos de datos en el paradigma Big data

Fuente: (Jiménez, Carlos Maté)



## Velocidad

Real time

Velocidad en consumo de datos y análisis de datos.

8

Flujo de datos masivo y continuo.

Ritmo en que los datos de entrada fluyen desde las diversas fuentes como procesos de negocio, máquinas y sensores, redes sociales, dispositivos móviles, etc.

## Veracidad

Sesgo, el ruido y la alteración de datos.

Los datos que se almacenan y extraen son directamente relacionados y significativos al problema que se trata de analizar.

Los datos deben ser verídicos.

## Validez

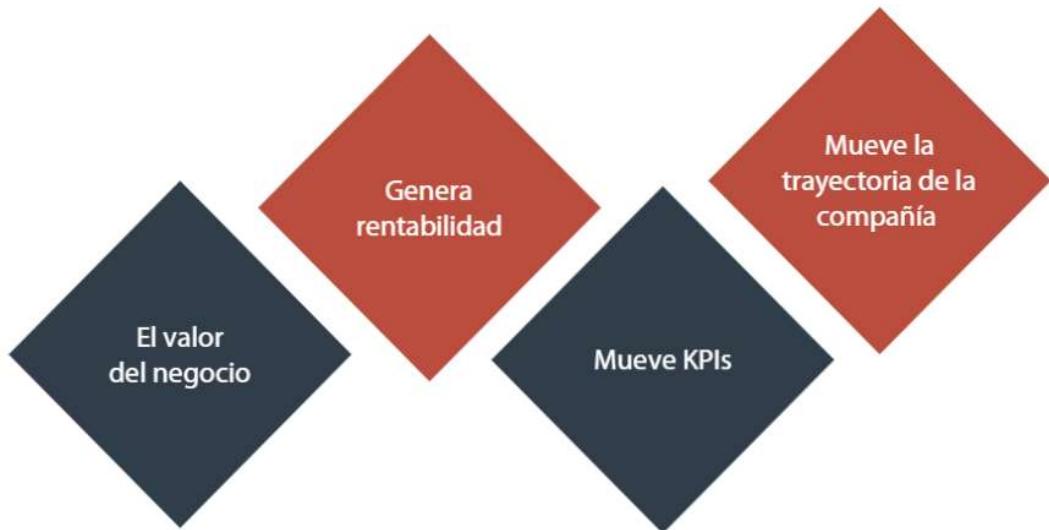
Los datos tienen validez

Son robustos

Data cleaning



## Valor



### 1.3. Arquitectura de datos en Big Data

La arquitectura en Big Data es importante en función de todos los usuarios que entran a interactuar con ella. El gráfico siguiente muestra los distintos pasos en esta arquitectura:

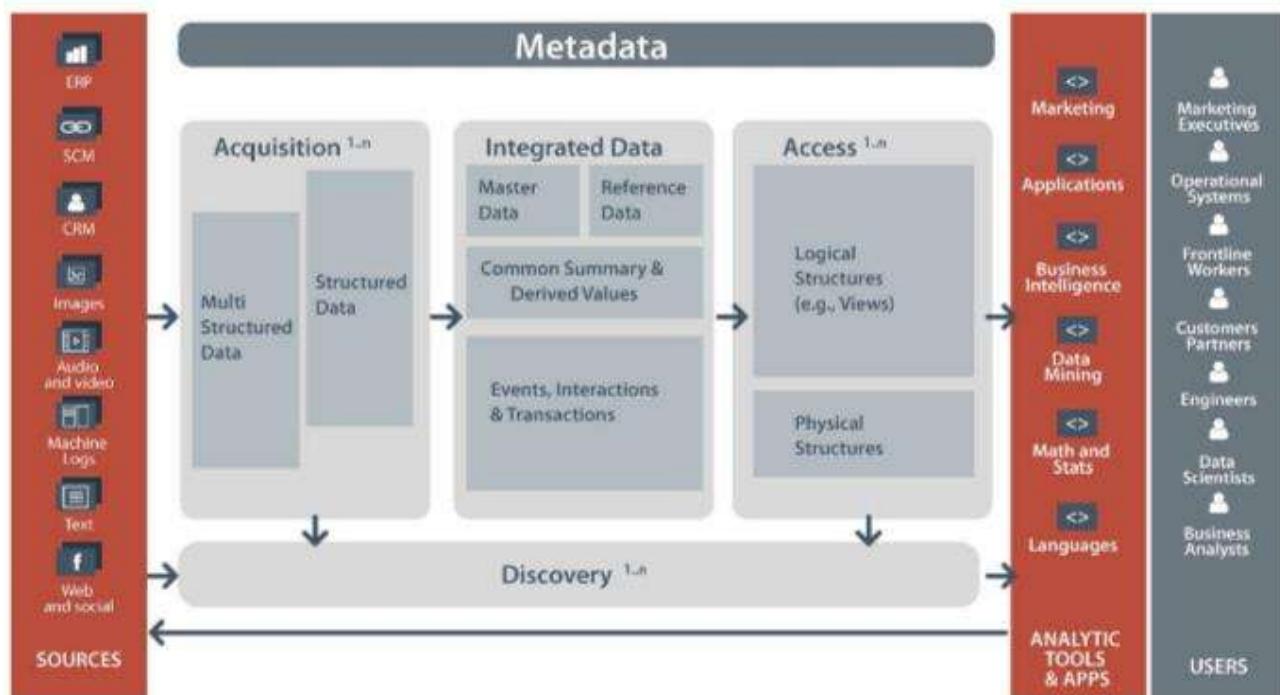


Imagen 2. Arquitectura de datos Big Data



En el lado izquierdo están las fuentes de datos. Allí se tienen desde las estructuradas hasta las no estructuradas o más complejas, en la parte inferior.

En la parte derecha en gris se encuentran los usuarios y más arriba se encuentran las personas que tienen una lectura más básica de los datos; a medida que se baja se encuentran perfiles que son más complejos a la hora de acceder y entender la data como los Data Scientist y los Business Analysts.

10

En la parte derecha, al lado de los usuarios se encuentran las técnicas. En la parte superior están las técnicas de marketing y asociadas a tener una visualización o tableros de control de una manera rápida. En cambio, en la parte inferior, están las técnicas estadísticas o matemáticas que permiten la predicción de datos.

Finalmente, en la parte central está la división de las tres partes de la arquitectura en Big Data.

1. Adquisición: cómo se adquieren las distintas fuentes de datos, ya sean estructuradas o no.
2. Transformación:
  - Distintas transformaciones o formas de ver los datos.
  - Tablas agrupadas o resumidas.
  - Crear variables a partir de otras existentes.
  - Crear data marts o accesos especiales.
  - Crear capas de datos.



### 3. Acceso:

- Dar acceso a las transformaciones vía tableros de control.
- Acceso de tablas consumibles en aplicaciones.
- Acceso a la data para modelar y entregar predicciones y resultados sobre las mismas.

11

## 1.4. Herramientas de analítica y Big Data

En la analítica de datos existe un sinfín de herramientas para ejecutar los distintos análisis. Aunque algunos factores como el precio, la dificultad de programación, el soporte o simplemente la comodidad, han hecho que las herramientas “sofisticadas” para analítica avanzada no sean las preferidas por los DSc de “a pie”.

Estas son algunas herramientas de uso analítico, de uso predictivo, de storage y de visualización:

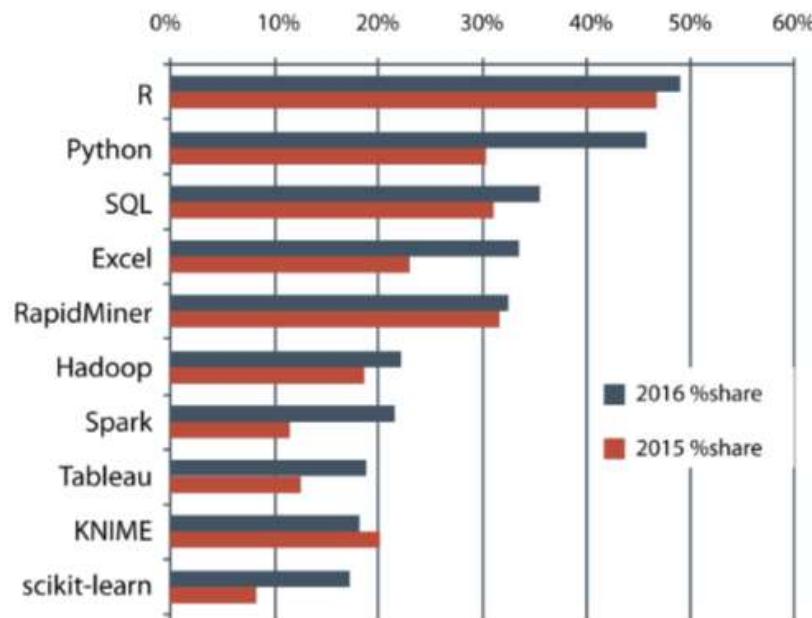


Imagen 3. Herramientas de analítica

Fuente: (Piatetsky, 2016)



Este módulo se centrará en el uso analítico a través de R, la cual es una herramienta open source que ofrece una infinidad de análisis.

## 2. Introducción a Machine Learning

12

**Machine Learning** es una disciplina científica del ámbito de la Inteligencia Artificial, que crea sistemas que aprenden automáticamente. En este contexto, aprender quiere decir identificar patrones complejos en millones de datos. Es decir que la máquina que realmente aprende es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. (González, 2014)

Del mismo modo, automáticamente en este contexto implica que estos sistemas se mejoran de forma autónoma con el tiempo y sin intervención humana.

Existen dos divisiones principales del Machine Learning que se verán a continuación:



Aprendizaje  
supervisado



Aprendizaje  
no supervisado



**Aprendizaje supervisado:** en principio consiste en aprender de un conjunto de datos conocidos para realizar predicciones sobre un conjunto de datos con información desconocida.

La información a predecir puede ser para tipo de datos categóricos o continuos, en los cuales se pueden usar algoritmos de clasificación o de regresión según corresponda.

Ejemplos: Un caso de uso en un banco es predecir cuál es el cliente que es más propenso a... En el caso de un producto en una empresa de telecomunicaciones será el cliente más probable de retiro de la compañía.

Los algoritmos más usados son:

- Regresión lineal y logística
- Regresión no lineal
- Árboles de decisión
- Máquinas de vectores de soporte (SVM)
- Naïve Bayes
- Análisis discriminante
- Redes neuronales

**Aprendizaje no supervisado:** no cuenta con información a priori como en el supervisado, por lo cual no tiene un conjunto de datos de prueba.

Ejemplo: en retail es importante decidir cuáles categorías de compra están relacionadas entre ellas e inferir si un usuario compró un producto y cuál es el segundo producto más probable que se va a comprar.

Los algoritmos más usados son:





Una visualización de los algoritmos y distintos usos del Machine Learning, de acuerdo a las clasificaciones que se describieron anteriormente son:

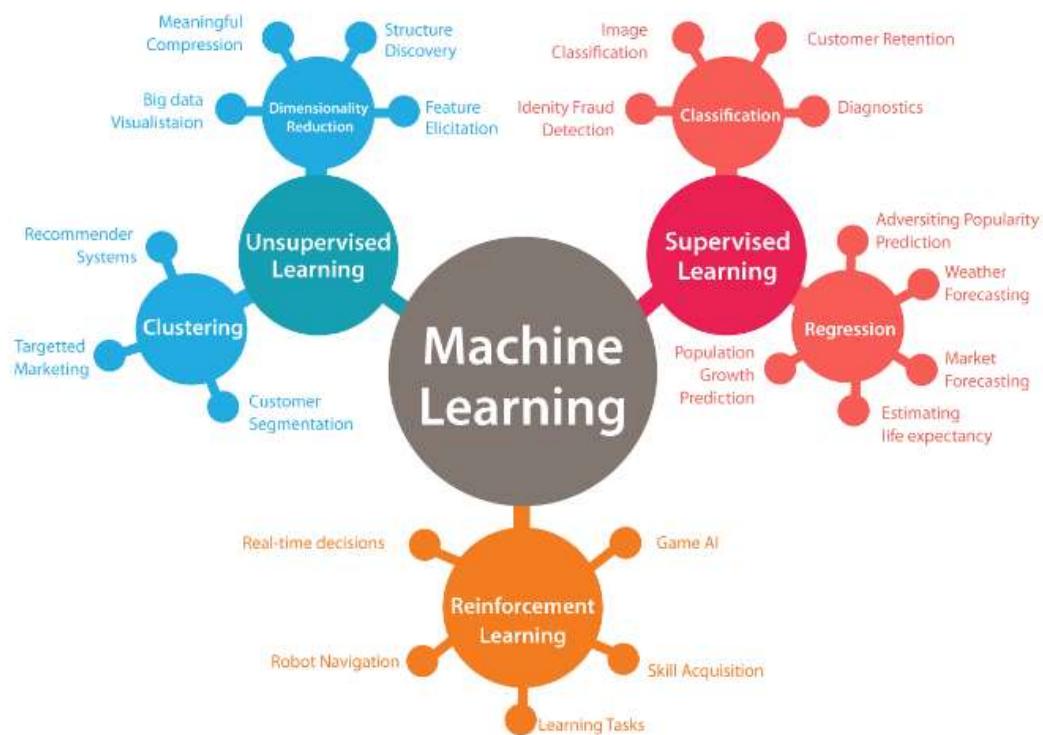


Imagen 4. Machine Learning  
Fuente: (Alogos, 2018)

### 3. Proceso Crisp-DM

Es una metodología para dar una organización más definida a proyectos analíticos, la cual se basa en seguir seis pasos para lograr obtener de forma exitosa y efectiva un proyecto. A continuación, se muestra un esquema de la metodología:

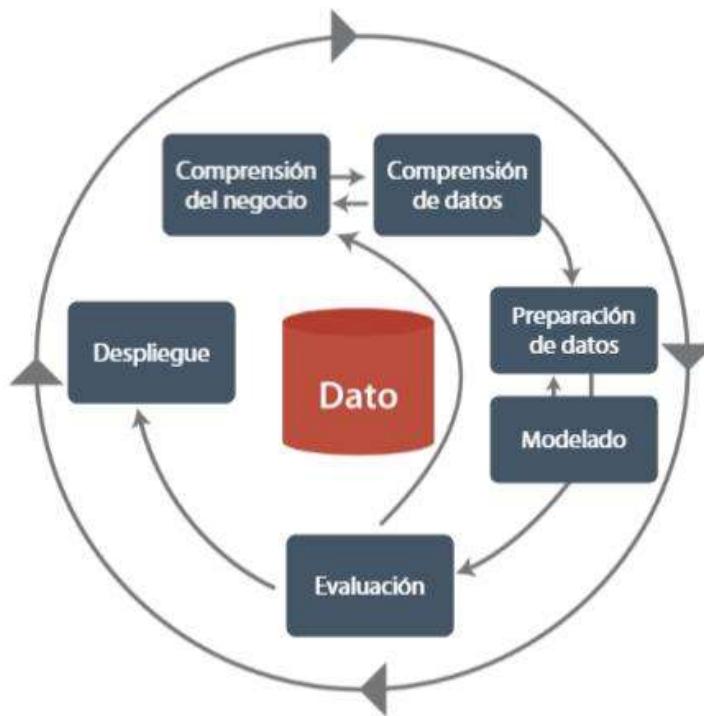


Imagen 5. Cross industry standard process for data mining

Fuente: (Pechenizkiy, Puuronen, & Tsymbal, 2008)

Cada una de las fases anteriores se componen de:

1. **Comprensión del negocio:** esta fase inicial se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial, realizando lo siguiente:



- a. **Determinar los objetivos del proyecto** conociendo los antecedentes del negocio y los objetivos principales.
- b. **Evaluar la situación** determinando las condiciones iniciales con las que el proyecto empieza.
- c. **Determinar los objetivos de la minería de datos** conociendo el problema en específico al cual aplicar las técnicas de modelamiento.
- d. **Desarrollar el plan del proyecto** conociendo el *time-to-market* del modelo y las especificaciones iniciales para su desarrollo.

**2. Comprensión de datos:** esta fase comienza con una colección inicial de datos y procesos con actividades, para familiarizarse con los datos.

- a. **Recolección de fuentes** conociendo las tablas de las cuales se dispone y el posible uso que pueden tener dentro del desarrollo del proyecto.
- b. **Descripción de los datos** conociendo la definición *funcional* de los datos, su definición teórica y utilidad.
- c. **Exploración de datos** generando reportes de exploración, los cuales contienen los análisis descriptivos de las tablas.
- d. **Verificación de la calidad de los datos** generando reportes de calidad de datos.

**3. Preparación de datos:** esta fase cubre todas las actividades para construir el conjunto de datos.

- a. **Generación de tabla única:** se realiza luego del entendimiento o comprensión de datos.
- b. **Selección de datos:** se definen reglas para selección de variables. Proceso descriptivo y univariado de datos.
- c. **Limpieza de datos:** se realiza a partir de técnicas como imputación de valores perdidos, outliers, valores atípicos, cotas de datos, etc.



- d. **Construcción de datos:** se genera con la ayuda de nuevas variables por concepto de negocio.
  - e. **Integración de datos:** se realiza por medio de la combinación de bases y/o variables (variables convulucionadas).
- 4. Modelado:** en esta fase se seleccionan y aplican varias técnicas de modelado, además de calibrar los parámetros para obtener óptimos resultados.
- a. **Selección de la técnica metodológica** evaluando los métodos estadísticos más adecuados a la naturaleza del problema.
  - b. **Generación del diseño de prueba** esbozando el proceso de desarrollo del modelo.
  - c. **Construcción del modelo** realizando el desarrollo del mismo, estimación de parámetros y descripción del modelo.
  - d. **Desempeño del modelo** evaluando el rendimiento del mismo, realizar revisión de los parámetros obtenidos, calibración del modelo, entre otros.
- 5. Evaluación:** en esta etapa en el proyecto se ha construido un modelo (o modelos) que parece tener gran calidad, desde una perspectiva de análisis de datos.
- a. **Evaluación de resultados:** realizar la evaluación del performance del modelo en una base ajena a la de desarrollo.
  - b. **Revisión del proceso:** realizar la evaluación de los supuestos y metodología utilizada.
  - c. **Determinar los próximos pasos:** plantear una evaluación posvalidación.
- 6. Despliegue:** esta fase depende de los requerimientos, pudiendo ser simple como la generación de un reporte, o compleja como la

17



implementación de un proceso de explotación de información que atraviese a toda la organización.

- a. **Despliegue de desarrollo:** se debe establecer las reglas de su implementación, usos y limitaciones.
- b. **Monitoreo:** realizar evaluación continua del modelo, según necesidad y uso, para determinar posibles calibraciones que se deban realizar al modelo, que permita detectar deficiencias o problemas con su desempeño.
- c. **Documentación y reporte final.**

18

Existen varias metodologías para el análisis de información. Además de la CRIPS.DM existe otra metodología llamada SEMMA, para ello consulte el material de lecturas y recursos del módulo.

## 4. Casos de negocio en Big Data

### El caso de las Fintech

La definición básica de una Fintech es la combinación de tecnología y compañías de servicios financieros. Al mezclar las dos cosas se forma una nueva compañía que tiene una gran manera de explotación de información.

- El uso cada vez más intensivo de las TI en la vida cotidiana y en el desarrollo de negocios (y principalmente en la industria financiera) ha llevado a la generación de **industrias híbridas** para acomodarse a las nuevas tendencias mundiales.
- Las **FinTech** son el resultado de esta adaptación sistémica.



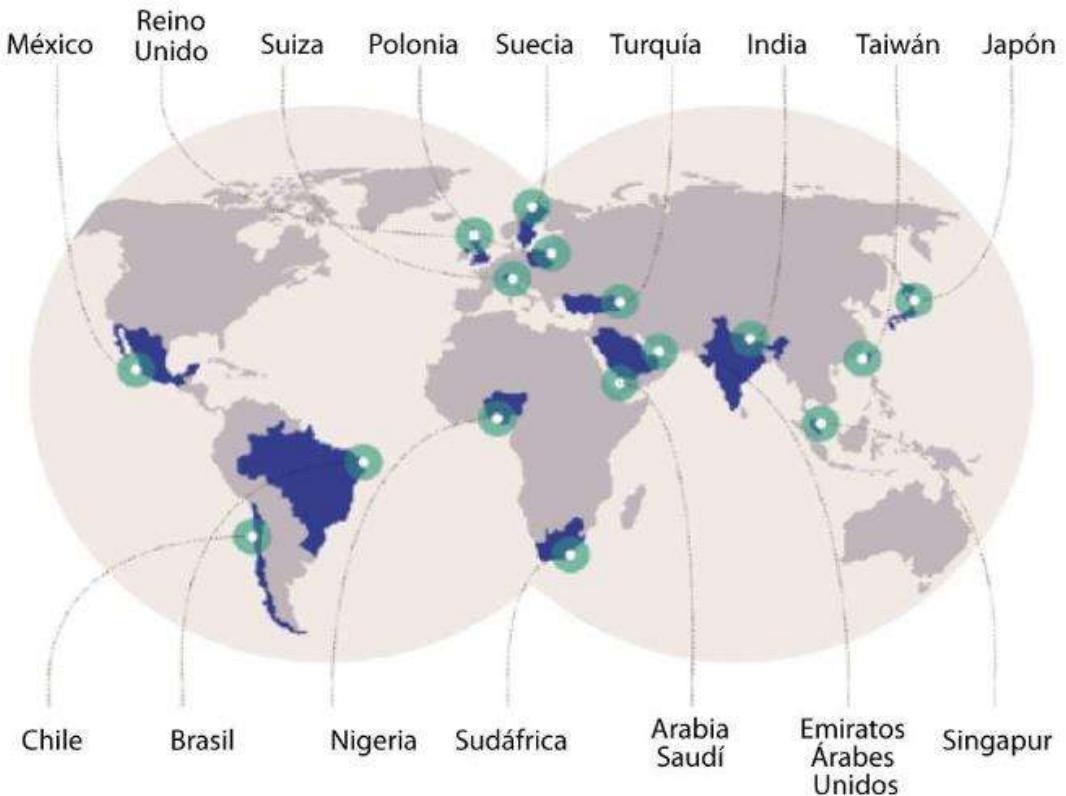


Imagen 6: Las Fintech como tendencia mundial

Por ejemplo, no es sorprendente que el mecanismo de pago para el sector ***retail*** se esté modificando. Estamos en la era de los:



## El caso de los Blockchain

Blockchain es un tipo de libro de registros, de tal manera que está distribuido para mantener un registro permanente de datos transaccionales; es decir que se mantenga a prueba de manipulaciones.

20

Este caso de aplicación hace que las compañías que manejan monedas virtuales se establezcan como funcionales obteniendo los siguientes beneficios:

### ¿Para qué sirve la tecnología Blockchain?

Aspectos/actividades empresariales que mejorarán gracias a blockchain



Imagen 7. ¿Para qué sirve la tecnología Blockchain?

Fuente: IBM Institute for Business Value citado por (Moreno, 2018)

## 5. Analítica y Big Data en Latam y Colombia

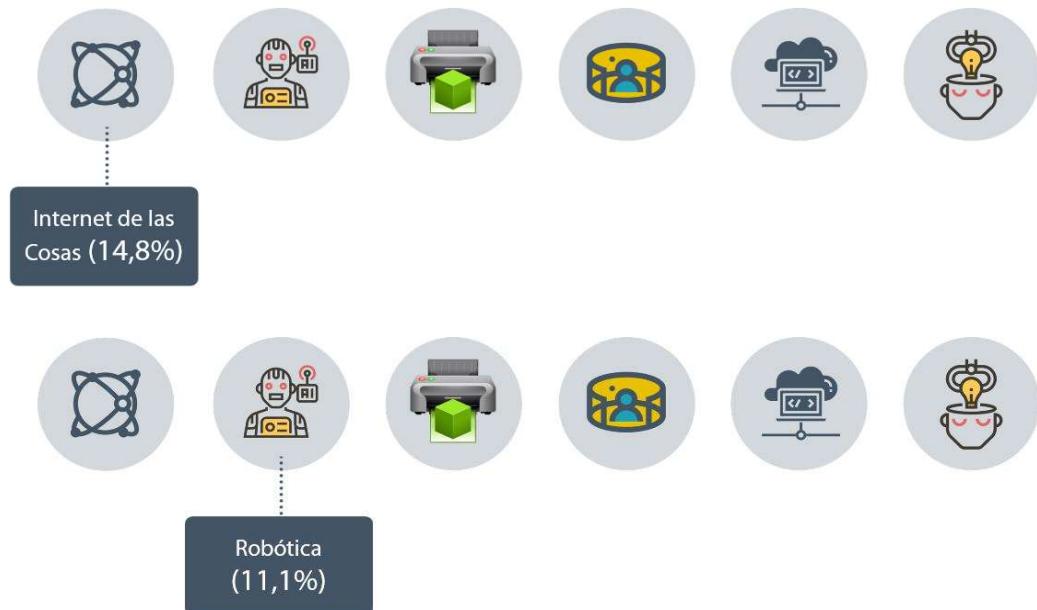
*"Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Gartner, 2012).*

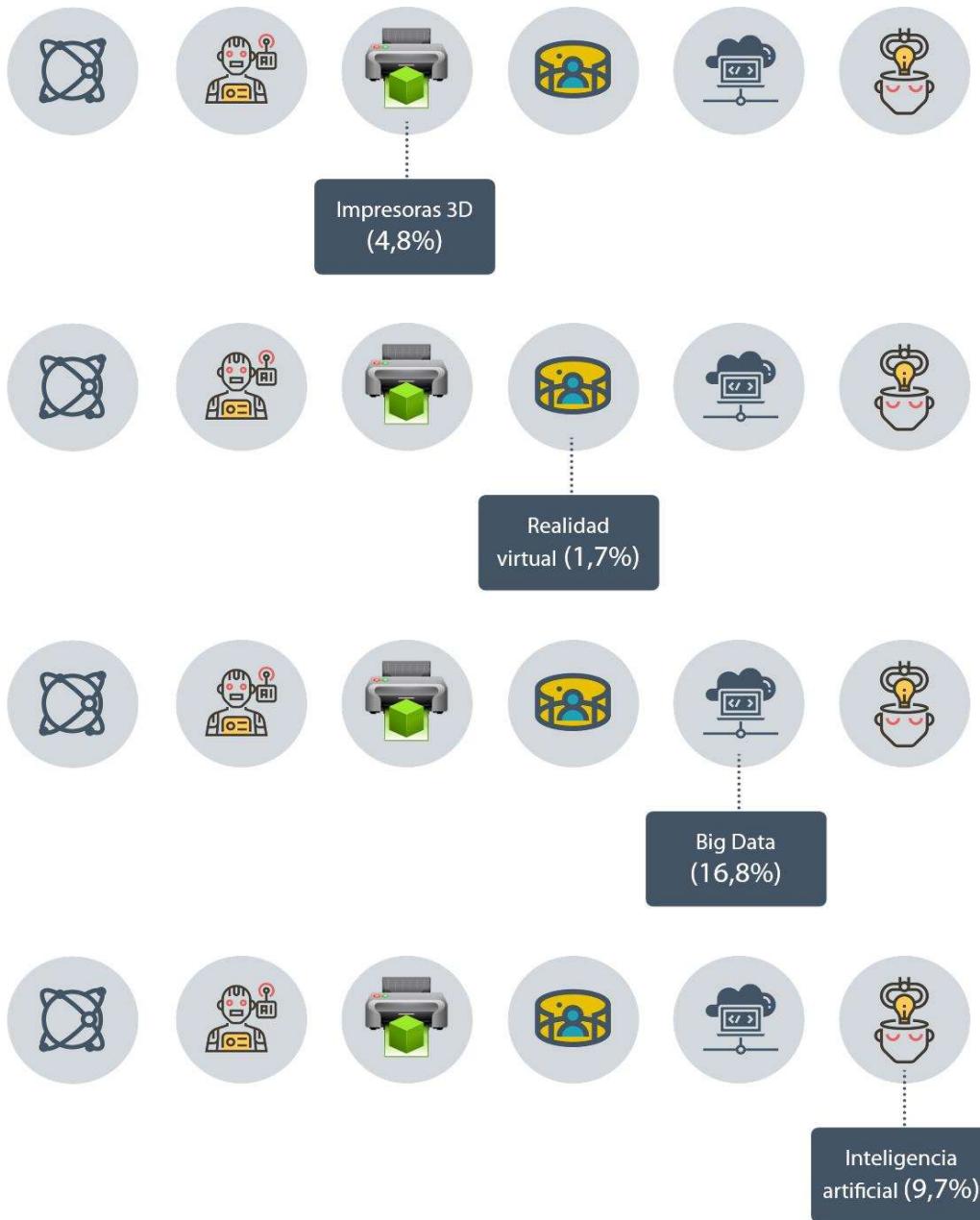


- El aumento de datos no estructurados representa un reto para todo proyecto en analytics. Pero también, quizás sea de las fuentes de información más ricas para su explotación.
- Este uso (de datos estructurados y no estructurados) es diferenciable por tipo de sector y fines por los cuales se utilizan.
- Cada vez que una persona compra algún producto y comparte algún video o texto, registra un evento en su celular, tiene algún sensor de deporte o salud, está creando un sinfín de datos que deben ser analizables.
- Para Latinoamérica su auge de crecimiento es importante.

Para septiembre del año 2017, la firma de investigación de mercado Gartner, Frost & Sullivan informaba de un crecimiento cercano al 40% en adquisición de soluciones e implementación de herramientas para análisis de Big Data para los próximos 4 años.

Por ejemplo, para Colombia, según el Observatorio de Economía Digital, en 2017 las empresas grandes en el país presentaron una penetración muy baja en:





De este modo el análisis de información para Latinoamérica resulta ser importante en términos de aprovechar las distintas tecnologías para analizar la información a fin de tener un impacto considerable dentro de sus compañías.



## 6. Demandas en cargos de Big Data

La demanda del Big Data a nivel mundial está en crecimiento en diferentes áreas, tal y como se observa en la siguiente gráfica:

23



Imagen 8. Benefitting from Big Data

Fuente: Booz & Company (2012) citado por (Tejedor, 2016)

La comunidad de expertos en tecnología de negocios de Wikibon, clasificó el mercado mundial de Big Data en 18.300 millones de dólares en 2014, pronosticando que crecerá a una tasa anual del 14,4%, para alcanzar los 92,2 mil millones de dólares en 2026. De esta manera la demanda de cargos de distintos perfiles es importante para afrontar este reto de analizar datos.

A continuación, se presenta el top 12 de los cargos en big data según (Kumar, 2016):



## 1. CIENTÍFICO DE DATOS

Esta persona utiliza sus capacidades analíticas y técnicas para extraer información significativa de los datos.

**Salario:** \$ 65,000 - \$ 110,000

## 2. INGENIERO DE DATOS

Asegura un flujo ininterrumpido de datos entre servidores y aplicaciones y también es responsable de la arquitectura de datos.

**Salario:** \$ 60,0945 - \$ 124,635

## 3. INGENIERO BIG DATA

Construye los diseños creados por los arquitectos de soluciones. Desarrolla, mantiene, prueba y evalúa soluciones de big data dentro de las organizaciones.

**Salario:** \$ 100,000 - \$ 165,000

## 4. CIENTÍFICO MACHINE LEARNING (APRENDIZAJE DE MÁQUINAS)

Trabaja en la investigación y desarrollo de algoritmos que se utilizan en sistemas adaptativos. Construye métodos para predecir sugerencias de productos y pronósticos de demanda, y exploran el big data para extraer patrones automáticamente.

**Salario:** \$ 78,857 - \$ 124,597

## 5. ESPECIALISTA EN ANÁLISIS DE NEGOCIOS

Un especialista en análisis de negocios apoya varias iniciativas de desarrollo, asiste en las actividades de prueba y en el desarrollo de scripts de prueba, realiza investigaciones para comprender problemas de negocios y desarrolla soluciones prácticas y rentables para los problemas.

**Salario:** \$ 50,861 - \$ 94,209



## 6. DESARROLLADOR DE VISUALIZACIÓN DE DATOS

Diseña, desarrolla y proporciona soporte de producción de visualizaciones de datos interactivas utilizadas en toda la empresa. Posee una mente artística que conceptualiza, diseña y desarrolla visualizaciones gráficas / de datos reutilizables y utiliza un sólido conocimiento técnico para implementar estas visualizaciones utilizando las últimas tecnologías.

**Salario:** \$ 108,000 - \$ 130,000

## 7. INGENIERO EN INTELIGENCIA DE NEGOCIOS (BI)

Tiene experiencia en análisis de datos y para configurar herramientas de informes, consultar y mantener almacenes de datos. Es una persona práctica con el big data y adopta un enfoque basado en datos para resolver problemas complejos.

**Salario:** \$ 96,710 - \$ 138,591

## 8. ARQUITECTO DE SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS (BI)

Crea soluciones rápidamente para ayudar a las empresas a tomar decisiones sensibles, tiene habilidades sólidas de comunicación y análisis, pasión por la visualización de datos y un impulso para la excelencia y la automotivación.

**Salario:** \$ 107,000 - \$ 162,000

## 9. ESPECIALISTA DE INTELIGENCIA DE NEGOCIOS (BI)

Es responsable de respaldar un marco de inteligencia de negocios para toda la empresa. Esta posición requiere pensamiento crítico, atención al detalle y habilidades de comunicación efectivas.

**Salario:** \$ 77,969 - \$ 128,337

## 10. ADMINISTRADOR DE ANALÍTICA



Es responsable de la configuración, diseño, implementación y soporte de la solución de análisis de datos o la herramienta de BI. Se le exige específicamente que analice grandes cantidades de información recopilada a través de la actividad transaccional.

**Salario:** \$ 83,910 - \$ 134,943

26

## 11. INGENIERO MACHINE LEARNING (APRENDIZAJE DE MÁQUINAS)

La "salida" final del ingeniero de Machine Learning es el software que funciona, y su "audiencia" para esta salida consiste en otros componentes de software que se ejecutan de manera autónoma con una supervisión humana mínima. Las decisiones las toman las máquinas y afectan el comportamiento de un producto o servicio.

**Salario:** \$ 96,710 - \$ 138,591

## 12. ESTADÍSTICO

Recopila datos numéricos y luego los muestra, y ayuda a las empresas a dar sentido a los datos cuantitativos, a detectar tendencias y hacer predicciones.

**Salario:** \$ 57,000 - \$ 80,110

## Referencias

Alogos. (2018). Introducción a Machine Learning. *Alogos*, <http://alogos.es/introduccion-machine-learning/>.

González, A. (1 de julio de 2014). *¿Qué es Machine Learning?* Obtenido de Cleverdata: <https://cleverdata.io/que-es-machine-learning-big-data/>

IBM Institute for Business Value and Saïd Business School, University of Oxford. (2014). Analytics: The Real-world Use of Big Data. <https://es.slideshare.net/IBMBDA/analytics-the-realworld-use-of-big-data-34457738>.



Jiménez, Carlos Maté. (s.f.). Big data. Un nuevo paradigma de análisis de datos. *Revista de la Asociación de Ingenieros del ICAI*.

Kumar, S. (19 de octubre de 2016). *Top 12 interesting careers to explore in big data [Infographic]*. Obtenido de [bigdata-madesimple.com: https://bigdata-madesimple.com/top-12-interesting-careers-to-explore-in-big-data/](https://bigdata-madesimple.com/top-12-interesting-careers-to-explore-in-big-data/)

Moreno, G. (10 de enero de 2018). *¿Para qué sirve la tecnología blockchain?* Obtenido de Statista: <https://es.statista.com/grafico/10644/para-que-sirve-la-tecnologia-blockchain/>

Pechenizkiy, M., Puuronen, S., & Tsymbal, A. (17 de agosto de 2008). Does Relevance Matter to Data Mining Research? En *Data Mining: Foundations and Practice* (págs. 251-275). Obtenido de [https://www.researchgate.net/publication/225609704\\_Does\\_Relevance\\_Matter\\_to\\_Data\\_Mining\\_Research](https://www.researchgate.net/publication/225609704_Does_Relevance_Matter_to_Data_Mining_Research)

Piatetsky, G. (2016). Top 10 most popular tools in 2016 . *KDnuggets*, <https://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html>.

Tejedor, L. (4 de noviembre de 2016). *Chema Alonso y el Big Data: "De la Universidad deben salir líderes con mente de ingenieros capaces de transformar el mundo"*. Obtenido de Telefonica.com: <https://www.telefonica.com/es/web/negocio-responsable/articulo/-/blogs/chema-alonso-y-el-big-data-de-la-universidad-deben-salir-lideres-con-mente-de-ingenieros-capaces-de-transformar-el-mundo->



**Módulo**  
**Analítica y Big Data**

**Autor**  
Oscar Ayala

**Universidad de La Sabana**  
**Forum E-learning**

Versión 1  
Copyright 2019. Todos los derechos reservados

