



Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado



# Universidad de Buenos Aires Facultad de Ciencias Económicas Escuela de Estudios de Posgrado

---

## **CARRERA DE ESPECIALIZACIÓN EN GESTIÓN ESTRATÉGICA DE SISTEMAS Y TECNOLOGÍAS DE LA INFORMACIÓN**

---

### **TRABAJO FINAL DE ESPECIALIZACIÓN**

---

La ciencia de datos como ventaja competitiva en la  
estrategia empresarial

---

AUTOR: LOREDANA CSAMBAL

MAYO 2020

---



## RESUMEN

El presente trabajo busca analizar cómo los grandes volúmenes de datos necesitan ser analizados con las herramientas que ofrece la ciencia de datos basada en tecnología de modo que las organizaciones puedan comprender mejor a sus clientes y obtener ventajas competitivas por adelantarse a las demandas del mercado y a la competencia.

Se aborda el concepto de Big Data<sup>1</sup>, que hace referencia a los grandes volúmenes de datos, para entender por qué las organizaciones deben tener en cuenta el procesamiento de los datos que procesa día a día en su operatoria, la gobernanza de los mismos y su transformación en información de valor para la toma de decisiones.

Luego se analiza la ciencia de datos basada en tecnología y las herramientas que proporciona para hacer que los datos transaccionados por las organizaciones puedan ser relevantes para conocer mejor a los clientes y ser más competitivas.

También se detallan las aptitudes y habilidades que debe tener un científico de datos, es decir el experto en ciencia de datos, para que las empresas puedan contratar al mejor personal que pueda manipular las herramientas que ofrece la ciencia de datos.

Implementar la ciencia de datos desafía a las compañías a desarrollar una correcta gestión del cambio en todos los niveles organizaciones, operación y proyectos para poder sacar provecho de todos los datos que transacciona y transformarlos en información de calidad para la toma de decisiones.

Para abordar la incorporación de la ciencia de datos en las organizaciones, se analiza el concepto de proyecto organizacional, qué conlleva a los equipos de trabajo a una buena gestión de los proyectos basada en el PMBOK<sup>2</sup> así como la gestión del cambio en las organizaciones a través de los pasos para una correcta implementación de los cambios empresariales.

---

<sup>1</sup> Altos volúmenes, variedad y velocidad de la información.

<sup>2</sup> Project Management Body of Knowledge o en español cuerpo de conocimiento para la administración de proyectos.



Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado



El objetivo general de este trabajo consistió entonces en analizar cómo la ciencia de datos comienza a ser adoptada por las organizaciones que buscan generar un valor incremental en su estrategia empresarial en menor tiempo.

Palabras Clave: Big Data. Ciencia de Datos. Toma de decisiones. Estrategia Empresarial.



## ÍNDICE

INTRODUCCIÓN .....	5
1. FUNDAMENTACIÓN Y PLANTEAMIENTO DEL PROBLEMA .....	5
2. OBJETIVOS.....	7
3. ASPECTOS METODOLÓGICOS .....	8
MARCO TEÓRICO .....	9
EL BIG DATA .....	9
LA CIENCIA DE DATOS BASADA EN TECNOLOGÍA .....	11
LOS CAMBIOS ORGANIZACIONALES .....	13
OPERACIÓN Y PROYECTOS EMPRESARIALES .....	16
DESARROLLO .....	18
LOS GRANDES VOLÚMENES DE INFORMACIÓN .....	18
LA CIENCIA DE DATOS Y EL APRENDIZAJE AUTOMÁTICO .....	25
EL CIENTÍFICO DE DATOS .....	29
DESAFÍOS EN LA IMPLEMENTACIÓN DE LA CIENCIA DE DATOS.....	35
TRANSICIÓN EN LOS CAMBIOS ORGANIZACIONALES .....	39
GOBERNANZA DE DATOS.....	41
LA OPERACIÓN Y LOS PROYECTOS ORGANIZACIONALES .....	43
CONCLUSIONES .....	54
REFERENCIAS BIBLIOGRÁFICAS .....	59



## INTRODUCCIÓN

### 1. FUNDAMENTACIÓN Y PLANTEAMIENTO DEL PROBLEMA

Actualmente, los grandes datos ya no se utilizan solamente para las redes sociales y registros informáticos generados por computadoras. Las organizaciones comenzaron a encontrar respuestas a preguntas que en el pasado no tenían acceso a responder y los grandes datos podrán ayudar a identificar preguntas que no se pensaba que se podían hacer.

La ciencia de datos<sup>3</sup> está transformando lentamente la forma en la que las organizaciones plantean su planificación a mediano y largo plazo. Con este campo de estudio de la información surgieron herramientas que buscan analizar, comprender y modelar los datos para predecir el comportamiento de los clientes y el rumbo que deben seguir las organizaciones.

En los próximos años los procesos empresariales se verán fuertemente afectados por el avance la ciencia de datos que proporcionará un nuevo marco en la toma de decisiones de las organizaciones.

Para convertirse y permanecer competitivas, las empresas deben adoptar análisis avanzados y adaptar sus modelos de negocio, establecer equipos especializados en ciencia de datos y repensar sus estrategias generales para seguir el ritmo de la competencia.

El presente trabajo busca desarrollar la respuesta a cómo la ciencia de datos está transformando la estrategia de las organizaciones y el impacto que va a tener su desarrollo en las empresas para lograr una ventaja frente a sus competidores en el mercado, abordando el tema desde la gestión del cambio en los procesos organizacionales.

Para comprender por qué las empresas deberían ir integrando la ciencia de datos dentro del alcance de sus estructuras organizacionales, se hará foco en entender cómo funciona esta ciencia de datos apoyada en tecnología, sus comienzos, la evolución en los últimos años y sus oportunidades en el futuro.

---

<sup>3</sup> El autor entiende por ciencia de datos al campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de los datos ya sean estructurados o no estructurados. El presente trabajo se centrará en la ciencia de datos apoyada en tecnología.



Para poder desarrollar el tema, se irán respondiendo puntualmente las siguientes preguntas:

1. ¿Cómo la ciencia de datos está transformando la estrategia empresarial y el impacto que va a tener desarrollarla por las organizaciones para generar una ventaja competitiva?
2. ¿Por qué las empresas deberían ir integrando la ciencia de datos dentro del alcance de sus estructuras organizacionales?
3. ¿Qué deben tener en cuenta las organizaciones desde la gestión del cambio en los procesos organizacionales con la incorporación de la ciencia de datos?
4. ¿Cómo funciona esta ciencia de datos apoyada en tecnología? ¿Cómo fueron sus comienzos, la evolución en los últimos años y sus oportunidades en el futuro?



## 2. OBJETIVOS

El objetivo principal es poder analizar cómo la ciencia de datos comienza a ser adoptada por las organizaciones que buscan generar un valor incremental en su estrategia empresarial en menor tiempo, abordando la gestión del cambio que conlleva incorporar las herramientas de dicha ciencia en los procesos, proyectos y en la toma de decisiones.

Como primer objetivo, se detallarán los componentes de la ciencia de datos, su evolución y aplicación en las organizaciones actualmente.

En un segundo objetivo, se analizará cómo se ven afectados los procesos organizacionales y la gestión del cambio que conlleva la incorporación de la ciencia de datos en las empresas.

Por último, el tercer objetivo se orientará en analizar qué se espera a futuro de la explotación de los datos, desde el punto de vista de la operación y de los proyectos que surjan en las organizaciones.



### 3. ASPECTOS METODOLÓGICOS

Para abordar el tema del presente trabajo, se utiliza un enfoque cualitativo y el tipo de estudio es de carácter de investigación exploratoria descriptiva, haciendo foco en las organizaciones con un mercado digital establecido o por desarrollar y con un desarrollo de la administración de datos avanzado.

En el trabajo se investigaron las fuentes primarias y se realizó la lectura bibliográfica para la investigación y descripción de los temas detallados en cada uno de los objetivos que se buscan cumplir con el presente trabajo. Para ello, se hizo una descomposición de los temas que se vinculan con el avance de la ciencia de datos y la gestión del cambio en las organizaciones.

Se analizaron los siguientes temas para complementar los objetivos propuestos: las similitudes y diferencias entre operación y proyectos empresariales, ciencia de datos y procesos organizacionales, comparando cada tema con la estrategia organizacional en el pasado y cómo se está comenzando a encarar actualmente, dados los avances en materia de ciencia de datos.

También se relevaron los temas vinculados a la gestión del cambio de forma general y ocho pasos para una gestión del cambio del autor John Kotter.



## MARCO TEÓRICO

### EL BIG DATA

Big Data<sup>4</sup> es un término que describe el gran volumen de datos, estructurados y no estructurados, que las organizaciones generan y explotan día a día. Pero no es la cantidad de datos lo importante, sino lo que las organizaciones hacen con estos datos. El Big Data se analiza para obtener hallazgos que conlleven a mejores decisiones y acciones de negocios estratégicas.

Capitalizar en las oportunidades que se pueden identificar, los datos y los análisis deben considerarse como una capa del tejido empresarial digital, asumiendo un papel más activo y dinámico en el impulso de las actividades de toda la organización, no solo reflejando dónde ha estado.

El uso del término Big Data en las organizaciones tiende a referirse al análisis del comportamiento de los clientes, extrayendo valor de los datos almacenados, y formulando predicciones a través de los patrones observados.

La disciplina dedicada a los datos masivos se enmarca en el sector de las tecnologías de la información y la comunicación. Esta disciplina se ocupa de todas las actividades relacionadas con los sistemas que manipulan grandes conjuntos de datos. Las dificultades más habituales vinculadas a la gestión de estos grandes volúmenes de datos, se centran en la recolección y el almacenamiento de los mismos, en las búsquedas, las comparticiones, y los análisis, y en las visualizaciones y representaciones.

Se puede describir al Big Data como “una nueva realidad que pone al “dato” en el centro de la escena. En este siglo XXI una adecuada gestión de ellos pasa a ser uno de los factores estratégicos más importantes que tienen disponible las organizaciones” (Fernandez Blanco, Coronel, & Chinkes, 2015), haciendo hincapié en que estos datos que tienen las empresas se están transformando en un recurso indispensable para la toma de decisiones.

Existen diferentes factores con origen principal en la tecnología que comenzaron a impulsar el uso de técnicas de ciencia de datos en un conjunto amplio de sectores. Estos factores

---

<sup>4</sup> El autor entiende por Big Data a los altos volúmenes, variedad y velocidad de la información.



pueden agruparse en cuatro ejes: (1) el incremento sin precedentes del volumen y la tipología de datos disponibles, (2) la conectividad y el acceso a los datos, (3) la mejora en los algoritmos para analizar los datos y (4) el aumento de la capacidad computacional de los sistemas en las organizaciones.

Las herramientas de Big Data sirven principalmente para (Pampliega, 2017):

- Analizar y almacenar datos en el nivel más granular (datos históricos análisis).
- Permitir el procesamiento de una enorme cantidad de datos en tiempo real (internet de cosas, detección de fraude, análisis en tiempo real, etc).
- Analizar y almacenar de manera efectiva datos semiestructurados o no estructurados (minería de texto, procesamiento de imágenes, etc).
- Tratar el hardware como mercancía y escalarlo horizontalmente con el objetivo de reducir costos.

La cantidad de datos en nuestro mundo ha estado creciendo exponencialmente, y el análisis de grandes conjuntos de datos, se convertirá en una base clave de la competencia, apuntalando nuevas olas de crecimiento de la productividad, innovación y excedente del consumo. Los líderes en todos los sectores tendrán que lidiar con las implicaciones de los grandes datos, no solo con algunos gerentes orientados a los datos. El aumento del volumen y los detalles de la información capturada por las empresas, el auge de los medios multimedia, las redes sociales y el internet de las cosas impulsarán un crecimiento exponencial de los datos en el futuro previsible (Manyika & Chui, 2011).

El Big Data también ayuda a crear nuevas oportunidades de crecimiento y categorías de empresas completamente nuevas, como aquellas que agregan y analizan datos de la industria. Muchas de estas serán compañías que se sientan en medio de grandes flujos de información donde se pueden capturar y analizar datos sobre productos y servicios, compradores y proveedores, y las preferencias e intenciones de los consumidores. Es probable que los ejemplos incluyan compañías que interactúan con un gran número de consumidores que compran una amplia gama de productos y servicios, compañías que permiten cadenas de suministro globales, compañías que procesan millones de transacciones y aquellas que proporcionan plataformas para experiencias digitales de consumo.



Actualmente para poder analizar datos tan masivos como el Big Data que en general se encuentran de manera distribuida en múltiples computadoras, bases de datos, repositorios, etc mediante algoritmos basados en el aprendizaje automático existen muchas interpretaciones, variantes y técnicas (Monleón-Getino, 2015).

Las principales técnicas estadísticas (Schmidberger, 2013) que se utilizan en el análisis del Big Data y como parte del aprendizaje automático pueden resumirse como:

- **Clasificación:** consiste en asignar una clase a un determinado objeto o individuo. La salida del sistema es una “etiqueta”. Así se podría clasificar un determinado producto comercial como “bueno” o “malo” según sus características.
- **Regresión:** es una técnica que busca resolver el problema de la clasificación. La salida del sistema es un número o un vector de números reales. Se podría predecir el incremento de ventas de un determinado producto a partir de las consultas en web de un catálogo comercial.
- **Clustering (agrupamiento):** técnicas para organizar objetos o individuos en grupos que tengan sentido. Agrupación de objetos o clases que puede ser jerárquica o no jerárquica.

## **LA CIENCIA DE DATOS BASADA EN TECNOLOGÍA**

La ciencia de datos logra analizar los grandes conjuntos de datos desordenados e incompletos, para llegar a hallazgos que impulsan decisiones sobre operaciones y productos. Es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sean estructurados o no estructurados lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva (Liu, 2015).

Las técnicas de aprendizaje automático están experimentando un auge sin precedentes en el mundo empresarial y constituyen una importante palanca de transformación. Si bien estas



técnicas eran conocidas en el pasado, diversos factores están provocando que su uso sea más intensivo cuando antes era minoritario y que se extienda a otros campos cuando antes prácticamente no eran utilizadas, tanto por los elevados costos de implantación como por los escasos beneficios inicialmente esperados de su aplicación (Calvo, Guzmán, & Ramos, 2018, pág. 6).

Dada la conectividad mundial y la generación de datos sin precedentes es importante contar con habilidades, conocimientos, experiencias y técnicas en las organizaciones para procesar, analizar y visualizar de formas más inteligentes los datos en información, promoviendo así, más y mejores conocimientos de la realidad en diferentes contextos.

Como especialidad, la ciencia de datos es joven. Surgió de los campos del análisis estadístico y la minería de datos. El Data Science Journal<sup>5</sup> debutó en 2002, publicado por el Consejo Internacional para la Ciencia: Comité de Datos para la Ciencia y la Tecnología. En 2008, el título de científico de datos había surgido y el campo despegó rápidamente. Desde entonces ha habido una escasez de científicos de datos, a pesar de que cada vez más colegios y universidades han comenzado a ofrecer títulos en ciencias de datos.

La ciencia de datos es un campo que ayuda a comprender eventos o a obtener información útil simplemente revisando y analizando los datos. Los resultados de los análisis de datos se pueden utilizar para crear una decisión empresarial. Esta decisión a menudo la toma una empresa para ayudar a comprender mejor a los clientes y anticiparse a las demandas así como para encarar nuevos productos o mejorar los existentes. El objetivo de las decisiones basadas en datos se utilizan para mejorar las habilidades de toma de decisiones, principalmente en negocios.

Utiliza métodos, procesos, algoritmos y sistemas científicos para extraer valor de los datos. Los científicos de datos combinan una variedad de habilidades, incluidas estadísticas, ciencias de la computación y conocimiento del campo a estudiar, para analizar los datos recopilados de la web, teléfonos inteligentes, clientes, sensores y otras fuentes.

---

<sup>5</sup> Data Science Journal es una revista electrónica, de acceso abierto y revisada por pares, que publica artículos sobre la gestión, difusión, uso y reutilización de datos y bases de datos de investigación en todos los dominios de investigación, incluida la ciencia, la tecnología, las humanidades y las artes.



## LOS CAMBIOS ORGANIZACIONALES

La estrategia de información es un compromiso organizacional a largo plazo para la explotación de la información para mejorar el negocio y los resultados. Requiere del entendimiento y difusión en la organización de que la información obtenida en su operatoria es el punto de acceso al entendimiento de cómo se está ahora y qué tan cerca – o lejos – se está de los objetivos estratégicos.

Según Kotter (1995) “la mayoría de los esfuerzos de cambio exitosos comienzan cuando alguna persona o algún grupo empieza a observar más de cerca la situación competitiva de la empresa, su posición de mercado y desempeño financiero, y sus avances tecnológicos”, por lo que las empresas deben planificar el futuro de sus negocios digitales pensando cómo generar nuevas posibilidades para crear valor comercial a través de datos y análisis.

La transición al negocio digital requiere la intervención desde el CIO<sup>6</sup> y directores de datos hasta líderes de datos y análisis para dar un salto hacia una nueva vista de datos y análisis. Se debe superar la vieja mentalidad de mantener datos en silos, aplicaciones comerciales y análisis aplicaciones basadas en sistemas de almacenamiento de datos e inteligencia de negocios, principalmente para informes empresariales y propósitos analíticos básicos. En cambio, se necesita ver a los datos como la materia prima para cualquier decisión, y considerar que los datos provienen tanto de dentro como de fuera de la empresa.

John Kotter (1995), profesor de la Harvard Business School, desarrolló ocho pasos para realizar una gestión de cambios eficaz dentro de las organizaciones que se detallan a continuación:

1. **Crear sentido de la urgencia:** para que un cambio pueda ocurrir efectivamente, el primer paso es involucrar a las personas que van a participar en ese cambio.. Es necesario que todas las personas entiendan las razones del cambio, y la importancia de actuar inmediatamente, para que se despierte la motivación en cada una de ellas.

---

<sup>6</sup> El autor entiende por CIO (Chief information officer) al director de sistemas de información de las organizaciones



2. **Formar alianzas fuertes:** otro paso fundamental durante la gestión de los cambios es la identificación de los funcionarios y gestores que pueden servir como agentes del cambio. Estas personas se pueden elegir no sólo por el cargo que ocupan, sino también por su status, experiencia en el tema, o incluso por su importancia política.
3. **Crear una visión para el cambio:** así como las empresas utilizan la visión de futuro para orientar a los equipos en la búsqueda de los objetivos estratégicos del negocio, estas pueden aprovechar la misma herramienta para los procesos de cambio. Aquí, es necesario elaborar una visión con los principales valores relacionados al cambio, de una forma altamente concisa y objetiva, permitiendo que todos entiendan rápidamente lo que se debe hacer en el día a día para adecuarse a la nueva situación.
4. **Invertir en la comunicación:** el proceso de gestión de cambios probablemente dividirá la atención del equipo con muchas otras cuestiones prioritarias en la organización. Por eso, es fundamental dirigir los esfuerzos para que la visión del cambio se comunique debidamente a los diferentes niveles organizacionales para garantizar que todas las personas de la empresa sean conscientes de dicho cambio.
5. **Empoderar a toda la base:** algunos procesos de cambio fallan no sólo por la falta de adhesión de las personas, sino también porque la empresa no está preparada a la hora de girar la mesa. Se propone que las empresas se cercioren de eliminar debidamente las barreras humanas, técnicas y normativas, para poder adoptar las nuevas soluciones sin obstáculos.
6. **Definir metas a corto plazo:** las personas se sienten más motivadas cuando perciben algún progreso en las actividades que se ejecutan. De esta forma, otro importante paso es la creación de metas e incentivos para los equipos, a corto plazo, con el objetivo de conmemorar los objetivos alcanzados y dar recompensas para aquellos que actúen en favor del cambio.
7. **No disminuir el ritmo:** aunque las ganancias a corto plazo son atractivas en un momento inicial, los cambios verdaderos tardan cierto tiempo hasta que se absorban. Por lo tanto, las organizaciones no deben perder el foco en los cambios estructurales y más profundos, pues sólo estos serán capaces de aportar mejoras continuas para los procesos y proyectos.



8. **Hacer que el cambio sea parte de la cultura:** las empresas necesitan incorporar el cambio en la cultura del negocio. Esta iniciativa debe partir del liderazgo, mediante el refuerzo constante sobre cuáles son las expectativas del nuevo escenario, pero también en su incorporación a los sistemas formales, evitando que la nueva realidad se deje de lado por la fuerza del hábito.

Integrar el cambio en la cultura de una organización, va a preparar a los equipos de trabajo para cambios futuros y poder afrontar un proyecto como la incorporación de la ciencia de datos.

El proceso del cambio planeado implica la presencia de tres elementos muy bien delimitados (Terán, 2002):

- El **Sistema** (en el que se llevará a cabo el cambio). Que puede ser un individuo, un grupo, una comunidad, una organización, un país e incluso toda una región del mundo.
- El **Agente de Cambio** (responsable de apoyar técnicamente el proceso de cambio). Uno o varios agentes de cambio, cuya función básica consiste en proporcionar al sistema el apoyo técnico o profesional necesario para que el cambio se lleve a cabo con éxito.
- Un **Estado Deseado** (las condiciones que el sistema debe alcanzar). Un estado deseado, que define las condiciones específicas que el sistema, con la ayuda del agente de cambio, desea alcanzar.

El proceso de cambio planificado consta de 5 grandes etapas:

- Diagnóstico de la situación.
- Determinación de la situación deseada.
- Determinación de los cauces de acción a seguir.
- Ejecución de las acciones.
- Evaluación de los resultados.



## OPERACIÓN Y PROYECTOS EMPRESARIALES

La gestión de las operaciones en las organizaciones tiene como responsabilidad la supervisión, la dirección y el control de las operaciones del negocio. Las operaciones evolucionan para dar soporte al negocio en el día a día, y son necesarias para alcanzar los objetivos estratégicos y tácticos del negocio.

Los proyectos, pese a su carácter temporal, contribuyen al logro de los objetivos de la organización cuando están alineados con la estrategia empresarial. Muchas veces las organizaciones modifican su operación, productos o sistemas mediante las iniciativas que se desarrollan e implementan a través de proyectos (Guía de los fundamentos para la dirección de proyectos (guía del PMBOK) - Quinta edición, 2013, pág. 12).

El objetivo principal de un proyecto de software es desarrollar y entregar uno o más productos de trabajo aceptables dentro de las limitaciones de las características requeridas, los atributos de calidad, el alcance del proyecto, el presupuesto, los recursos, la fecha de finalización, la tecnología y otros factores. Los productos de trabajo que se entregarán (por ejemplo, código de objeto, materiales de capacitación e instrucciones de instalación) son el resultado del flujo de productos de trabajo intermedios que son producidos por y fluyen a través de los procesos de trabajo (requisitos, diseño, código fuente y escenarios de prueba) (Failey, 2009).

Dentro de un proyecto de ciencia de datos se tienen en cuenta las siguientes fases (Cady, 2017):

Primero, se debe establecer el marco del problema, es decir, entender la necesidad del negocio y definir claramente el problema. En esta fase es importante definir claramente la pregunta o las preguntas de negocio a las que la organización quiere dar respuesta incorporando la ciencia de datos en un equipo de trabajo. Es importante saber cómo sería una respuesta o resultado que se considerase como una solución. En el caso de proyectos complejos, puede ser necesario recopilar todas las necesidades en un documento.

En segundo lugar, es importante entender los datos a través de un análisis preliminar de los mismos para revisar su tamaño, calidad, si son completos o parciales, si son representativos,



si existen datos atípicos y si existen identificadores y cuáles son. De forma resumida, es importante saber si se dispone de los datos necesarios para contestar a la pregunta o preguntas que definen el problema a resolver y, por tanto, el objetivo del proyecto. En esta fase, además, se transforman de su formato original a alguno otro más apto para su tratamiento mediante herramientas de análisis. Luego se hará un análisis exploratorio, observación y graficar los datos desde diferentes perspectiva.

En tercer lugar, es necesario extraer características de los datos. En cierto modo, es una continuación de la fase anterior y pueden ser trabajadas en paralelo . Se trata de extraer números o categorías que caractericen los datos. Además, y como resultado, se transforman esos datos a formatos tabulares aptos para su tratamiento.

Como cuarta fase, se realiza el modelado y el análisis de los datos. Se realizan las clasificaciones, se aplican modelos, e incluso el aprendizaje automático. Aunque esta fase parece la más compleja del proyecto, puede terminar siendo bastante simple porque ya existen modelos preconstruidos, y potentes herramientas que hacen fácil su aplicación. Con frecuencia, se aplican varios modelos para examinar los datos desde diferentes perspectivas.

La última fase consiste en presentar los resultados del análisis de los datos. Es la fase menos técnica de todo el proyecto, pero es la más importante, porque es a través de la cual se consigue que el negocio, los interesados que encargaron resolver ciertos interrogantes, reciban sus respuestas y puedan decidir con base en ellas.



## DESARROLLO

“La información es la gasolina del siglo XXI, y la analítica de datos el motor de combustión” (Sondergaard, 2011).

### LOS GRANDES VOLÚMENES DE INFORMACIÓN

La llegada de los grandes volúmenes de datos comienza a desencadenar una expansión del espacio de almacenamiento. Como resultado, el almacenamiento se convierte en uno de los obstáculos para la mayoría de las organizaciones, que tienen que tener en cuenta crear una estructura y desarrollar una solución para almacenar datos. Una vez que las empresas encuentran una solución al problema del gran almacenamiento de datos, deben enfocarse a entender cómo pueden procesarse los datos, dando lugar a que muchas veces se comience a desarrollar la ciencia de datos en las organizaciones.

Tradicionalmente, los datos se estructuraban en un tamaño pequeño. Esto significa que no había inconvenientes si se quería analizar los datos dado que había herramientas de Inteligencia de Datos simples que se podrían usar para analizar dichos datos.

Actualmente la mayoría de los datos ya no son estructurados y son diferentes de los datos tradicionales, por lo que se necesita tener métodos avanzados de análisis de datos que provienen de diferentes fuentes, como archivos de texto, registros financieros, sensores, formularios multimedia e instrumentos. Las herramientas de Inteligencia de Datos simples no se pueden utilizar para procesar este tipo de datos como resultado de la naturaleza masiva de los datos. Por esta razón, se requieren herramientas analíticas complejas y avanzadas y algoritmos de procesamiento. Estos tipos de herramientas ayudan a las organizaciones a analizar y obtener información importante basada en los datos.

La esencia importante del Big Data no es tanto la importancia numérica de los datos sino todo lo que se puede hacer si se aprovecha el potencial y se descubren nuevas oportunidades de los grandes volúmenes de datos.

Según la empresa IBM (Barranco Fragoso, 2012) existen diferentes tipos de datos que vamos a encontrar en el Big Data. Esta empresa ha clasificado en cinco los tipos de datos de



Big Data, que aunque no es fácil de distinguir en algunos casos si pueden permitir tener una idea general.

Datos	Descripción	Ejemplos
Web and Social Media	Contenido web e información que es obtenida de las redes sociales	www, Facebook, Twitter, LinkedIn, blogs
Machine-to-Machine (M2M):	Tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún acontecimiento en particular. Se transmiten a través de redes alámbricas, inalámbricas o híbridas.	Velocidad, temperatura, presión, variables meteorológicas, variables químicas
Big Transaction Data	Incluye datos procedentes de transacciones masivas de los centros de atención telefónica, de banca, finanzas, atención a clientes, etc	Incluye registros de facturación, en telecomunicaciones los llamados registros detallados de las llamadas (Call Detail Record o CDR), etc.
Biometrics	Información biométrica. En el área de seguridad e inteligencia, los datos biométricos son sumamente importantes para los gobiernos, seguridad privada, servicios de inteligencia, policía, etc	Huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc
Human Generated	Datos digitales generados por las personas, en sentido genérico	Notas de voz, correos electrónicos, documentos electrónicos, resultados de estudios médicos, multas, etc

Tabla 1: Tipos de datos del Big-data según IBM (Barranco Frago, 2012)

El mayor desafío de la revolución del Big Data es dar sentido a toda la información generada por la vasta economía digital de hoy en día. Cuantos más datos se poseen, mayor será la oportunidad de generar informes y recomendaciones estratégicas de calidad, siempre que se puedan analizar los datos de un modo inteligente y rápido y hacer que sean prácticos gracias a la valiosa información recabada. De lo contrario, más datos pueden conllevar más problemas: datos desordenados, problemas de almacenamiento, riesgos de seguridad, equipos empresariales frustrados y personal del departamento de tecnología desbordado.

Hoy una organización depende de los datos para su propia supervivencia. Los datos se utilizan para tomar decisiones estratégicas sobre la dirección futura que tomará una organización y para eso los datos deben ser actuales y precisos. Debido a que los datos son un



activo para una empresa, se les puede dar un valor y se pueden intercambiar (Lake & Crowther, 2013).

En la práctica, son varias las personas que trabajan en equipo para construir productos de datos. Sus análisis solo serán tan buenos como lo sea el equipo responsable de la recopilación, construcción y análisis de los datos subyacentes.

Para que las organizaciones puedan utilizar los datos y transformarlos en información de valor que genere ventajas competitivas, deben contar con una eficiente gestión de los datos que recopilan de sus transacciones.

La gestión de datos una práctica que busca recopilar, mantener y utilizar datos de forma segura, eficiente y rentable. El objetivo de la gestión de datos es ayudar a las empresas a optimizar el uso de los datos dentro de los límites de las políticas y regulaciones para que puedan tomar decisiones y tomar medidas que maximicen los beneficios para la organización. Una estrategia sólida de gestión de datos se está volviendo más importante que nunca a medida que las organizaciones confían cada vez más en activos intangibles para crear valor.

La gestión de datos digitales en una organización implica una amplia gama de tareas, políticas, procedimientos y prácticas. El trabajo de gestión de datos tiene un amplio alcance, que abarca factores como la forma que se crean, accedan, almacenen y actualicen datos en un nivel de datos diversos. Además, permite proporcionar alta disponibilidad y recuperación ante desastres, usar datos en una variedad creciente de aplicaciones, análisis y algoritmos y garantizar la privacidad y seguridad de los datos.

De alguna manera, los grandes datos son exactamente lo que parecen: muchos, muchos datos. Pero los datos grandes también vienen en una variedad más amplia de formas que los datos tradicionales, y se recopilan a una alta velocidad. Un ejemplo podría ser pensar en todos los datos que llegan todos los días, o cada minuto, de una fuente de redes sociales como Facebook<sup>7</sup>. La cantidad, variedad y velocidad de esos datos son los que los hacen tan valiosos para las empresas, pero también hacen que su administración sea muy compleja.

---

<sup>7</sup> Facebook Inc. es una compañía estadounidense que ofrece servicios de redes y medios sociales en línea.



La mayoría de los desafíos en la gestión de datos actuales se derivan del ritmo más rápido de los negocios y la creciente proliferación de datos. La variedad, la velocidad y el volumen de datos en constante expansión disponibles para las organizaciones los empujan a buscar herramientas de administración más efectivas para mantenerse al día.

Las organizaciones capturan, almacenan y utilizan cada día más datos. Para mantener tiempos de respuesta máximos en este nivel en expansión, las organizaciones necesitan monitorear continuamente el tipo de preguntas que responde la base de datos y cambiar los índices a medida que cambian las consultas, sin afectar el rendimiento.

Las regulaciones de cumplimiento son complejas y multijurisdiccionales, y cambian constantemente. Las organizaciones deben poder revisar fácilmente sus datos e identificar cualquier cosa que se encuentre bajo requisitos nuevos o modificados. En particular, la información de identificación personal debe ser detectada, rastreada y monitoreada para cumplir con las regulaciones de privacidad global cada vez más estrictas.

En el nuevo mundo de la gestión de datos, las organizaciones almacenan datos en múltiples sistemas, incluidos almacenes de datos y lagos de datos no estructurados que almacenan cualquier información en cualquier formato en un único repositorio. Los científicos de datos de una organización necesitan una manera de transformar rápida y fácilmente los datos de su formato original en la forma, el formato o el modelo que necesitan para una amplia gama de análisis.

La integración de Big Data trabaja diferentes tipos de datos y los transforma para que puedan ser consumidos. La gestión de Big Data almacena y procesa datos en un lago de datos o en un almacén de datos de manera eficiente, segura y confiable, a menudo utilizando el almacenamiento de objetos.

Las empresas pueden utilizar el Big Data para mejorar y acelerar el desarrollo de productos, el mantenimiento predictivo, la experiencia del cliente, la seguridad, la eficiencia operativa y mucho más. A medida que los grandes datos crecen, también lo hacen las oportunidades en las empresas y ser más competitivas.

Aprender a capitalizar los datos de las organizaciones comprende la recopilación de la información necesaria para producir o mejorar un bien o un servicio. No se puede construir



sobre información que no se ha capturado, por lo que toda información que no es recopilada de la actividad de una empresa no puede capitalizarse.

Lo que quiere una organización es satisfacer las necesidades de sus clientes, por lo que, al conocerlos mejor, se puede adecuar de forma más precisa la oferta y hacerlo también de forma segmentada. Ésta es una tendencia de los últimos años, en los que se tiende a personalizar cada vez más el producto o servicio que venden a grupos cada vez más pequeños y delimitados.

La capacidad de realizar esta segmentación no es realmente una ventaja competitiva, ya que esto se va extendiendo y se encuentra al alcance de muchas. La ventaja competitiva real es tener acceso a toda esa información sobre la que trabajar y diseñar las estrategias sobre las que se trabajarán a corto, medio y largo plazo.

El correcto procesamiento de los datos no sólo posibilita un mejor tratamiento del producto o servicio a la venta, sino que también ofrece otros beneficios a la hora de expandir el negocio si es necesario. Si conocemos el perfil del target principal, se pueden analizar estos datos para encontrar perfiles similares en otras zonas en las que abrir nuevas tiendas, sucursales o franquicias.

En sí, los datos son un bien no rival, es decir, pueden ser usados simultáneamente por distintas partes sin que la cantidad de datos disponibles para el resto se vea afectada. Por ejemplo, es tecnológicamente posible que todos los investigadores en el campo de la medicina utilicen el stock agregado de datos médicos de pacientes al mismo tiempo. Debido a esta no rivalidad, el intercambio de flujos de datos puede comportar enormes beneficios para la sociedad.

Poder extraer valor a los datos otorga importantes ventajas competitivas. Los datos, por sí solos, no tienen ningún valor: el reto es convertir la información en valor. En otras palabras, contar con datos de millones de interacciones no sirve de nada si esta información no puede utilizarse para conocer mejor al consumidor o usuario y saber qué necesita o cómo mejorar la experiencia del cliente. No obstante, convertir la información en valor requiere de unas capacidades específicas. Incluye, entre otros, una infraestructura adecuada para almacenar y procesar los datos, experiencia en análisis de datos, o contar con talento especializado.



Dado que los datos pueden otorgar importantes ventajas de información respecto a competidores, las empresas no tienen incentivos para compartir con terceros los datos que han acumulado. En este contexto, la información puede concentrarse, de forma desproporcionada, entre un número relativamente reducido de grandes empresas.

Asimismo, la explotación conjunta de efectos de red y de grandes cantidades de información puede ampliar la posición de dominio del mercado de algunas empresas. Ello explica, por ejemplo, por qué las grandes empresas tecnológicas tienen un gran potencial para explotar grandes cantidades de datos. En particular, cuantos más usuarios tiene una plataforma digital, más atractivo es para otros usuarios registrarse y operar en esa plataforma (llamado efecto de red). A medida que la organización acumula más información sobre los clientes, está en disposición de mejorar los productos y servicios, y atraer todavía a más de ellos pudiendo ampliar así su ventaja competitiva sobre empresas rivales y su posición dominante en el mercado.

A pesar de que hay distintas maneras de extraer valor, el uso responsable de los datos debe estar presente en todas ellas. El uso ético y transparente de los datos es un ámbito que está centrando cada vez más la atención de consumidores y de la sociedad en general, especialmente tras varios casos de mal uso de datos personales y a medida que se generaliza el uso de la información digital y de modelos de inteligencia artificial por parte de empresas. En este contexto, asegurar que las empresas operan con criterios éticos y que se garantizan los derechos individuales es vital para mantener la confianza de la sociedad en los servicios digitales.

Otro aspecto igualmente crítico es la aplicación responsable de las técnicas de inteligencia artificial a los datos. En concreto, las máquinas suelen ser las encargadas de analizar, a partir de algoritmos creados por programadores informáticos, grandes cantidades de datos (puesto que se están considerando muchas más dimensiones de las que una mente humana podría concebir). A esto se le llama modelo de aprendizaje automático y permite a las empresas extraer valor de los datos de forma automatizada y escalable (por ejemplo, mediante la identificación de patrones). No obstante, las implicaciones éticas de estas técnicas son complejas, puesto que si son usadas de forma incorrecta pueden llegar a perpetuar sesgos o prejuicios presentes en los datos sobre los que se basan estos modelos.



Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado



Por ejemplo, en la aplicación de técnicas de inteligencia artificial a procesos de selección de personal, si en los datos históricos hay una infrarrepresentación de mujeres, el algoritmo podría estar sesgado en contra de dicho colectivo a la hora de buscar candidatos. Debido a ello, es importante conocer los sesgos existentes en las bases de datos usadas, corregirlos a la hora de diseñar los algoritmos que ejecutan las máquinas e incorporar consideraciones éticas en el uso de dichos algoritmos.



## LA CIENCIA DE DATOS Y EL APRENDIZAJE AUTOMÁTICO

Un entorno de ciencia de datos puede ayudar a una organización a saber qué datos tiene y luego hacerlos utilizables. Este entorno permite a los científicos de datos crear, probar y evaluar automáticamente modelos que se utilizan para encontrar datos, y luego transformarlos para que sean utilizables y valiosos para la organización. Con una plataforma centralizada, los científicos de datos pueden trabajar en un entorno colaborativo utilizando sus herramientas de código abierto favoritas, con todo su trabajo sincronizado por un sistema de control de versiones.

Una de las ventajas de utilizar la ciencia de datos en las organizaciones es que puede ayudarlas a tomar algunas decisiones de negocio importantes. En el pasado, las empresas debían confiar en su experiencia y conocimiento sobre el mercado y su industria para determinar si estaban tomando decisiones inteligentes. Puede ser que aquellas organizaciones que han estado en la industria por mucho tiempo puedan seguir basándose en la experiencia pero podrían estar perdiendo información importante que puede ayudarlas a ser más competitivas.

Otra ventaja de la implementación de la ciencia de datos disminuye la subjetividad en la toma de decisiones, estas no serán basadas en creencias, suposiciones o gustos personales sino en datos reales e interpretativos, que se analizan en el momento exacto.

La ciencia de datos, y especialmente el aprendizaje automático, son excelentes para resolver los problemas complejos y ricos en datos. Las tendencias pueden ser fácilmente descubiertas por la ciencia de datos, obteniendo como resultado tendencias del mercado, nuevas ideas y problemas que aún no son atendidos.

Junto con los avances en la tecnología de almacenamiento, que hacen posible cada vez más almacenar cantidades tan grandes de datos a un costo relativamente bajo, ya sea en almacenes de datos comerciales, laboratorios de investigación científica o en otros lugares, se ha dado cuenta cada vez más de que dichos datos contienen conocimiento oculto que puede ser almacenado para el crecimiento o el declive de una empresa. Sin embargo, los enormes volúmenes involucrados significan que la mayoría de estos datos simplemente se almacenan, para nunca ser examinados de una manera más que superficial, si es que lo hacen. Se ha dicho



con razón que el mundo se está convirtiendo en "rico en datos, pero pobre en conocimiento". La tecnología de aprendizaje automático tiene el potencial de resolver el problema de la marejada de datos que inunda a organizaciones, gobiernos e individuos (Bramer, 2016).

Una de las tecnologías protagonistas de la “revolución digital” es el Aprendizaje Automático, es decir, máquinas alimentadas por algoritmos capaces de autoabastecerse y aprender de los nuevos datos que acumulan. Gracias a este aprendizaje, pueden diagnosticar problemas o generar modelos predictivos de comportamiento. En otras palabras, las máquinas van incorporando datos que se transforman en conocimiento gracias al que sabrán reaccionar ante situaciones futuras sin necesidad de intervención humana.

El dato es la materia prima del aprendizaje automático y, como tal, cuantos más genere una organización, más información para alimentar los algoritmos y mayor probabilidad de desarrollar modelos predictivos eficaces.

A continuación se detallan los ámbitos o departamentos de una empresa que puede cambiar con el Aprendizaje automático:

### **Ciberseguridad: predicción de ataques y fraudes**

La ciberseguridad es una de las principales preocupaciones de las organizaciones y una de las disciplinas en las que el aprendizaje automático puede ser de mayor utilidad. Los nuevos antivirus y sistemas de protección recurren al aprendizaje automático para, mediante algoritmos inteligentes que incorporan datos de comportamiento, predecir ataques y amenazas o detectar si la conducta de un dispositivo es la habitual o presenta indicios de malware.

También son muchas las empresas que, especialmente los bancos, utilizan el aprendizaje automático en sus políticas antifraude, pues los algoritmos pueden descubrir prácticas fraudulentas en función del comportamiento de los clientes.

### **Atención al cliente: chatbots que adaptan las respuestas a los clientes**

La automatización, reducción de costos o prestación de servicio 24/7 son solo algunas de las ventajas de utilizar chatbots para gestionar las conversaciones entre empresa y usuario en el entorno digital.



Estos sistemas aprenden de la interacción con los clientes, detectando patrones y analizando posibles errores. Como consecuencia, pueden resolver las incidencias que vayan apareciendo a medida que la empresa evoluciona y responder a cada cliente en función de las necesidades de las organizaciones. Además, obtienen datos que mejoran el conocimiento del consumidor por parte de la empresa.

### **Comercio electrónico: recomendación de productos y ofertas personalizadas**

Un ámbito en el que el Machine Learning<sup>8</sup> tiene potencial y capacidad de influencia en el cliente, ese es el comercio electrónico.

Del comportamiento del usuario en una plataforma del comercio electrónico se pueden obtener infinidad de datos sobre gustos, necesidades y patrones de compra, por lo que aprender de esos hábitos deriva en la generación de ofertas personalizadas muy segmentadas, la recomendación de productos complementarios con elevada probabilidad de compra o la predicción de la demanda en cada momento o época del año.

Se trata por lo tanto de un ganar-ganar para empresas y clientes. Para las primeras, porque optimizan al máximo el stock y pueden obtener información muy útil del cliente; para los segundos, porque disfrutan de una mejor y más satisfactoria experiencia de compra al recibir directamente impactos con productos ajustados a sus necesidades, sin tener que perder tiempo en buscarlos en el catálogo.

### **Toma de decisiones y optimización de procesos internos**

Gracias a los datos sobre el funcionamiento interno de la empresa, el Machine Learning puede automatizar la gestión del inventario o predecir las necesidades y posibles fallos del ciclo de producción antes de que sucedan, optimizando su funcionamiento.

### **Recursos Humanos: detección automática de perfiles idóneos**

Podría resultar paradójico que para el reclutamiento de recursos humanos se utilice el aprendizaje automático de las máquinas, lo cierto es que cada vez más empresas recurren a sistemas de inteligencia artificial o aprendizaje automático para la captación de empleados.

---

<sup>8</sup> Aprendizaje Automático



En concreto, estos sistemas son capaces de establecer los criterios de selección que más se ajustan a los requerimientos de las empresas, hacer una primera selección y calificación de candidatos o generar cuestionarios para entrevistas desarrollados por algoritmos.

### **Logística: organización basada en datos**

Mediante la utilización de datos logísticos y el aprendizaje automatizado, es posible implementar sistemas de reconocimiento de objetos en almacenes o realizar predicciones sobre las condiciones meteorológicas o de tráfico para encontrar las mejores rutas, favoreciendo la organización interna y la experiencia del cliente en la recepción de sus mercancías.

En resumen, implementar una estrategia de aprendizaje automático en una empresa puede alterar el funcionamiento habitual de muchos departamentos, si bien el resultado puede merecer el esfuerzo: reducción de costos, más eficacia y mayor productividad.



## EL CIENTÍFICO DE DATOS

Davenport y Patil en su artículo “Científico de datos: el trabajo más sexy del siglo XXI” (2012), definieron por primera vez el concepto de científico de datos y con ello generaron una gran revolución. Además de definir quién es un científico de datos, presentan un decálogo para encontrar el científico de datos correcto, explican cuáles son los intereses del profesional y de los cuidados que deberán tener las empresas para conservarlos.

En términos generales, el científico de datos combina estadística, matemáticas, programación y solución de problemas, con la captura de datos de forma ingeniosa y la capacidad de mirar las cosas de manera diferente (encontrar patrones), además de hacer las actividades propias de limpieza, preparación e integración de datos (Monnappa, 2017).

Los científicos de datos, ingenieros de datos, desarrolladores y analistas de negocios cualificados suponen figuras transformadoras en los negocios modernos. Son el corazón de la economía del big data. No solo es que estén diseñando nuevos sistemas; sino que también salen en defensa de nuevas fuentes de datos y nuevos modos de utilizar esos datos. Por supuesto, el departamento de Tecnología de la Información aún sigue teniendo que construir el sistema, pero los profesionales de la ciencia de datos son los que ayudan a los departamentos a colaborar para resolver problemas y acelerar la innovación.

Los tipos de programación más profundos, como los ingenieros de datos y los desarrolladores de interfaces aportan la idoneidad a la ciencia de datos. Los matemáticos y científicos de datos utilizan algoritmos estadísticos para encontrar patrones en los datos. A lo largo de todo el proceso, es necesario que todo esté orientado a conseguir los resultados empresariales deseados, bajo la dirección del analista de negocios. Al trabajar juntos, el equipo de científicos de datos puede adelantarse a los desafíos y problemas actuales para crear nuevas oportunidades y posibilidades para el futuro.

Los científicos poseen una rara combinación de talentos: gestionan una variedad de responsabilidades y conjuntos de habilidades que abarcan las matemáticas, la estadística, la especialización sectorial, las comunicaciones y muchos más. Básicamente, el trabajo de un científico de datos es buscar patrones ocultos. Lo consiguen aplicando avanzadas técnicas de



análisis que incluyen (pero no se limitan a) aprendizaje automático, modelización, estadística y visualización.

Un buen científico de datos no se dedicará a los problemas empresariales, se concentrará en los problemas que poseen mayor relevancia para la organización. A menudo los científicos de datos construyen modelos para predecir resultados o para descubrir patrones subyacentes; su plan consiste en producir conocimiento práctico que pueda utilizarse para mejorar los resultados futuros. Experimentan continuamente desplegando nuevos modelos predictivos, reglas empresariales y lógica de organización en aplicaciones basadas en la próxima mejor acción.

Los científicos de datos se distinguen por su gran perspicacia empresarial, además de por su capacidad para comunicar los resultados tanto a los líderes del negocio como a los de Tecnología de la Información de un modo que pueda influir en la manera en la que una organización se enfrenta a un desafío empresarial. El científico de datos a menudo se convierte en el enlace entre el departamento de Tecnología de la Información y los ejecutivos de más alto nivel. Por lo tanto, deberá ser capaz de hablar ambos "idiomas" y comprender la jerarquía de los datos; no puede ser solo un experto en datos. Esto también significa que los científicos de datos poseen una sólida comprensión del negocio además de la convicción a la hora de apoyar sus descubrimientos frente a la oposición.

Un científico de datos cualificado explora y examina datos procedentes de múltiples y dispares fuentes. Estos analizarán atentamente todos los datos entrantes con el objetivo de enlazar nueva información con datos históricos para encontrar una relación o tendencia que ofrezca una ventaja competitiva crucial o aborde un apremiante problema empresarial.

No se dedican simplemente a recopilar y comunicar los datos; los examinan desde muchas perspectivas, determinan lo que significan y entonces recomiendan modos de aplicar los descubrimientos. Necesitan asegurarse de que sus indagaciones son correctas y deben ser capaces de apoyar sus conclusiones con modelos sólidos y datos fiables. A menudo se espera que un científico de datos presente recomendaciones a los equipos de gestión y dirección.



Los deberes de un científico de datos pueden incluir el desarrollo de estrategias para analizar datos, preparación de datos para el análisis, explorar, analizar y visualizar datos, construir modelos con datos utilizando lenguajes de programación, y la implementación de modelos en aplicaciones.

El científico de datos no trabaja solo. De hecho, la ciencia de datos más efectiva se realiza en equipos. Además de un científico de datos, este equipo podría incluir un analista de negocios que define el problema, un ingeniero de datos que prepara los datos y cómo se accede a ellos, un arquitecto de TI<sup>9</sup> que supervisa los procesos y la infraestructura subyacentes, y un desarrollador de aplicaciones que implementa modelos o resultados del análisis en aplicaciones y productos.

El rol que un científico de datos juega dentro de una empresa depende en gran medida de la configuración de ésta y de la orientación analítica de su organigrama. Puede suceder que esta vocación por el análisis avanzado sea más bien reciente y el data scientist contratado tenga el honor de inaugurar su propio departamento, marcando el camino a seguir para quienes llegarán después, con quienes formará el equipo de científicos de datos que cubrirá todas las necesidades de la empresa en materia de Big Data.

El modelado estadístico y el aprendizaje automático se están convirtiendo en elementos esenciales para convertirse en un científico de datos. El factor diferenciador para quienes trabajan en el campo es qué tan bien pueden comunicar sus hallazgos de una manera simple, pero accionable.

---

<sup>9</sup> Tecnología de la Información

En el siguiente cuadro se hace un resumen de las habilidades que deben tener los científicos de datos:

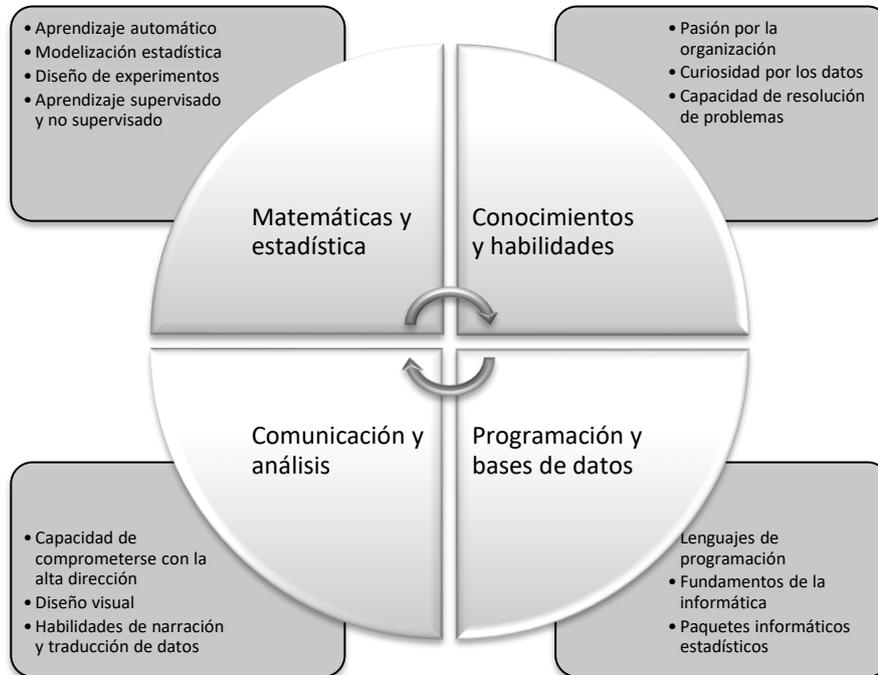


Gráfico 1. Elaboración propia

Las organizaciones están utilizando equipos de ciencia de datos para convertir los datos en una ventaja competitiva al refinar productos y servicios. Por ejemplo, las empresas analizan los datos recopilados de los centros de atención telefónica para identificar a los clientes que probablemente abandonen, por lo que el marketing puede tomar medidas para retenerlos. Las empresas de logística analizan los patrones de tráfico, las condiciones climáticas y otros factores para mejorar las velocidades de entrega y reducir los costos. Las compañías de atención médica analizan los datos de las pruebas médicas y los síntomas informados para ayudar a los médicos a diagnosticar enfermedades antes y tratarlas de manera más efectiva.

Muchas compañías han hecho de la ciencia de datos una prioridad y están invirtiendo mucho en ella. Como resultado, los líderes de datos y análisis implementan cada vez más capacidades de autoservicio para crear una cultura basada en datos en toda la organización. Esto significa que los usuarios comerciales pueden aprender más fácilmente a usar y beneficiarse de



herramientas eficaces de análisis y de Inteligencia de Negocio<sup>10</sup>, impulsando resultados comerciales favorables en el proceso.

Los resultados de los algoritmos y modelos sólo son efectivos si ayudan a resolver el problema correcto en el contexto adecuado. Esto significa trabajar de la mano con las partes interesadas para identificar y refinar la declaración del problema y la hipótesis al comienzo del proceso y mantenerlos involucrados durante todo el flujo de trabajo. Y al final del flujo de trabajo, significa saber cómo comunicar los resultados a los socios de negocios de manera relevante y accionable.

El proceso de analizar y actuar sobre los datos es iterativo en lugar de lineal, pero así es como fluye el trabajo para un proyecto de modelado de datos:

1. Definir un proyecto y sus posibles resultados.
2. Desarrollar un entorno de trabajo, asegurándose de que los científicos de datos tengan las herramientas adecuadas, así como el acceso a los datos correctos y otros recursos, como la potencia de cálculo.
3. Cargar los datos en el entorno de trabajo.
4. Preparación, análisis, exploración y visualización los datos.
5. Crear, entrenar y validar modelos para que funcionen según sea necesario.
6. Implementar modelos en producción.

Los trabajadores de datos quieren y necesitan que los datos y las acciones estén en el mismo lugar. En lugar de realizar análisis en un silo y actuar en otro, cualquier persona que trabaje con datos debería poder permanecer en el contexto de sus procesos y flujos de trabajo de negocio. Las plataformas de inteligencia de negocios están satisfaciendo esta necesidad al fusionarse con las operaciones, los flujos de trabajo y los procesos centrales del negocio a través de funcionalidades como análisis móvil, análisis incorporado, etc. Como resultado, los análisis accionables aceleran el proceso de toma de decisiones tanto para roles técnicos como no técnicos.

---

<sup>10</sup> El autor entiende por Inteligencia de Negocio al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitectura técnicas que están enfocados a la administración y creación de conocimiento, a través del análisis de los datos existentes en una organización.



Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado



Estas funcionalidades permiten a los trabajadores de datos analizar datos y realizar acciones después de encontrar información, todo en el mismo lugar. Un ejemplo de convergencia entre la información y la acción es el análisis incorporado. El análisis incorporado incluye datos e información que se encuentran allí donde las personas ya están trabajando para que no tengan que navegar hacia otra aplicación o servidor compartido.



## DESAFÍOS EN LA IMPLEMENTACIÓN DE LA CIENCIA DE DATOS

La ciencia de los datos se compone de cuatro elementos que se detallan en el siguiente cuadro:

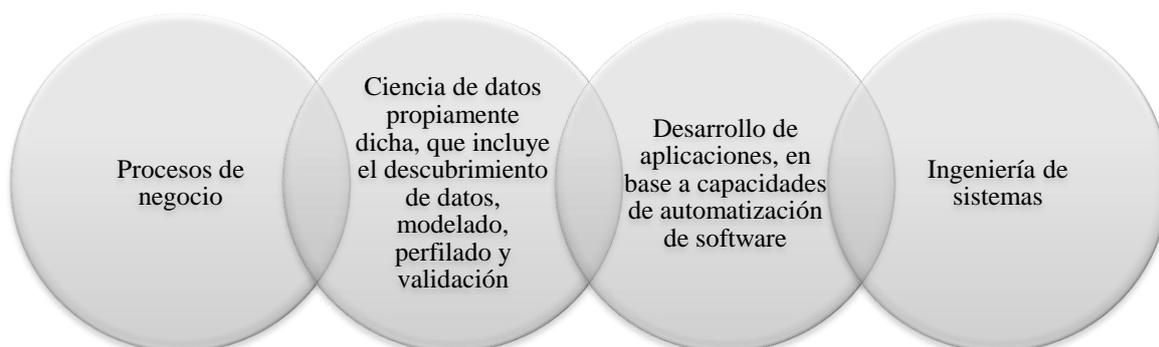


Gráfico 2. Elaboración propia

A pesar de la promesa de la ciencia de datos y las enormes inversiones en equipos de ciencia de datos, muchas compañías no se están dando cuenta del valor total de sus datos. En su carrera por contratar talentos y crear programas de ciencia de datos, algunas organizaciones pueden experimentar flujos de trabajo de equipo ineficientes, con diferentes personas que utilizan diferentes herramientas y procesos que no funcionan bien en conjunto. Sin una administración central disciplinada, los ejecutivos podrían no ver un retorno total de sus inversiones.

Los científicos de datos no pueden trabajar de manera eficiente. Debido a que el acceso a los datos debe ser otorgado por un administrador de Tecnología de la Información, los científicos de datos a menudo tienen que esperar mucho los datos y los recursos que necesitan para analizarlos. Una vez que tienen acceso, el equipo de ciencia de datos puede analizar los datos utilizando herramientas diferentes y posiblemente incompatibles. Puede llevar semanas, o incluso meses, implementar los modelos en aplicaciones útiles.



Los gerentes de negocios están demasiado alejados de la ciencia de datos. Los flujos de trabajo de la ciencia de datos no siempre se integran en los procesos y sistemas de toma de decisiones empresariales, lo que dificulta que los gerentes de negocios colaboren con conocimiento con los científicos de datos. Sin una mejor integración, a los gerentes de negocios les resulta difícil entender por qué lleva tanto tiempo pasar del prototipo a la producción, y es menos probable que respalden la inversión en proyectos que perciben como demasiado lentos.

Por mucho que las empresas automaticen procesos de datos, sin importar qué tan grande sea el conjunto de datos o por más inteligentes que sean los cálculos, si no se pueden comunicar los hallazgos en la organización, no se podrá lograr un impacto con los análisis. Este es el poder de la visualización de datos. La visualización de datos es un lenguaje y un estándar para que los analistas sepan cómo transmitir información a los responsables de la toma de decisiones de una manera práctica y fácil de entender. Esta habilidad, combinada con la capacidad de los analistas para compartir los pasos que tomaron para descubrir información a partir de los datos, suele definirse como "contar historias con datos".

Contar historias con datos es un elemento crítico del proceso de análisis. Y la cultura cambiante en el lugar de trabajo, el análisis está refinando la definición de contar historias con datos. A medida que las organizaciones crean culturas de análisis, los métodos de los analistas para contar historias con datos se basan cada vez más en fomentar una conversación en torno a los datos y menos en argumentar a favor de una conclusión en particular. Estas culturas analíticas también están fomentando los esfuerzos de alfabetización de datos dirigidos a enseñar a las personas a comprender los datos en su dimensión real y a participar en la conversación analítica, desde el momento del descubrimiento hasta la decisión de negocios resultante.

En muchos productos uno de los diferenciadores puede ser el uso de datos. En este sentido identificar fuentes de datos diferenciales y poder valorar su utilidad en etapas tempranas del proyecto es clave para la viabilidad de las ideas generadas. Por ejemplo, se puede tener la idea de que la información en portales públicos de compra venta de pisos podría ser utilizada para mejorar la valoración del precio de una vivienda, que hasta ahora era generada basada en datos internos.



Algo similar ocurre con la disponibilidad de nuevas herramientas o algoritmos para el tratamiento de datos. En ocasiones, el acceso a un nuevo algoritmo puede hacer viable, técnica o económicamente, una idea.

Como parte del entendimiento de una idea, el análisis de los datos del mercado nos va a ayudar a definir mejor la oportunidad (público objetivo, segmentación, modelo de negocio, etc.) y a acotar el trabajo a realizar en el resto de planos.

En esta etapa es necesario profundizar en las implicaciones de usar las fuentes de datos que se hayan identificado y la forma en la que se van a usar. En este caso hay que tener en cuenta tanto aspectos legales, de privacidad y éticos. Es el momento de identificar posibles sesgos en los datos y considerar si el producto agravaría esos sesgos.

También si es capaz de satisfacer a todos los usuarios objetivo, o por el contrario se requieren mecanismos o acciones especiales para que la solución elegida sea inclusiva. Por ejemplo, en aplicaciones que incluyen reconocimiento de voz se ha encontrado que la baja representación de ejemplos femeninos o de voces no nativas suele implicar un peor calidad en el reconocimiento para estos colectivos.

En un siguiente nivel, se busca encontrar en las diferentes soluciones para la creación o mejora del producto o servicio, así como todo lo necesario como para poder empezar a experimentar y aprender con algo tangible.

El objetivo es definir las características del producto y en el caso específico de que requieran funcionalidades basadas en el uso y explotación de datos, abordar cómo se podrían construir. A grandes rasgos las aportaciones que se pueden hacer desde el equipo de datos para contribuir a la solución se agrupan en los bloques que se detallan a continuación.

En el primer bloque se encuentra el análisis exploratorio de datos. Se trata en este caso de responder preguntas de negocio, por ejemplo, mediante los datos de uso del producto. Estos análisis pueden ayudar a definir una estrategia de producto, escoger una solución o priorizar el problema en el que enfocarse.

En el segundo bloque se realiza la visualización y resumen de los datos. Muchos productos basados en datos incluyen alguna forma de visualización de los propios datos



generados por sus usuarios, por ejemplo una aplicación destinada a medir nuestra actividad física. Es labor del equipo de datos buscar una visualización eficaz y adecuada.

En el tercer bloque se realiza una predicción a través de los datos analizados. Es la capacidad para inferir el valor de una o varias variables de interés para nuestra aplicación. Puede tratarse, entre otras, de la categoría de un objeto (por ejemplo, una imagen), el valor numérico de una cantidad medible (por ejemplo, el precio de un pasaje de avión) o las películas en las que puede estar interesado un usuario, como en el caso de un recomendador.

En el cuarto bloque, se hace el diseño de ciclos de realimentación de datos. Cualquier funcionalidad basada en datos tiende a obtener mejores resultados si dispone de más datos, así que parte de la labor del equipo de datos es definir mecanismos que permitan recogerlos y transformarlos en mejores predicciones o visualizaciones de mayor exactitud.

En el plano de la implementación es donde se genera un incremento de producto de forma que se entrega valor al cliente o beneficiario. Aquí el científico de datos tiene la oportunidad de contribuir en el impacto esperado con dicho incremento, aprovechando lo trabajado en todos los planos anteriores y de forma conjunta con el resto del equipo, para así entregarlo a los correspondientes beneficiarios o clientes finales.

Se asume que gran parte del trabajo en entendimiento del negocio y en la decisión del enfoque analítico se ha llevado a cabo en los planos etapas previas de entendimiento e ideación. Para la disciplina de datos este plano abarca desde el trabajo técnico de la ingesta de datos, la validación de estos mismos, exploración, construcción del modelo predictivo (o de una visualización) hasta la posterior validación de sus resultados y en el mejor de los casos el despliegue de la funcionalidad y la infraestructura adecuada. Dejamos aparte la medición de los resultados del proyecto, y por tanto la obtención de feedback, como parte los niveles transversales.



## TRANSICIÓN EN LOS CAMBIOS ORGANIZACIONALES

El desarrollo organizacional ayuda a los líderes a abordar y adoptar el cambio desde la perspectiva de que el cambio es una oportunidad y no una amenaza. Casi todas las demandas de un cambio provienen del exterior de la organización –de las dependencias gubernamentales, los competidores, las nuevas tecnologías, los clientes, las fuerzas del mercado y la sociedad en general.

En ocasiones, las demandas para el cambio provienen del interior de la organización – un nuevo ejecutivo, productos o servicios obsoletos, una nueva dirección estratégica, una rentabilidad en disminución, o una fuerza laboral crecientemente diversa. Es necesario comprender un cambio y el cambio planificado para comprender el desarrollo organizacional.

El escenario empresarial actual, en el que la competencia es altísima y evoluciona rápidamente, obliga a mejorar la capacidad de respuesta de las organizaciones. Para ello, lo más importante es disponer de la mejor información que permita tomar las mejores decisiones a todos los niveles (estratégico, táctico y operacional). Esto implica apoyar a la dirección con la información adecuada (pasada, presente y futura) y anticipar y responder proactivamente a las necesidades de los cliente.

La evolución de las tecnologías está dirigiendo a las empresas a explotar todo el conocimiento que los activos de datos pueden ofrecer para lograr los resultados que esperados. Aprovechar tecnologías emergentes como Internet de las Cosas, Inteligencia Artificial y Machine Learning permite enfocarse en los resultados de mayor valor.

Estas tecnologías aportan las siguientes funcionalidades clave:

1. Visibilidad. La capacidad de recopilar y conectar datos que previamente estaban en silos y reconocer patrones invisibles.
2. Enfoque. Simular el impacto de opciones potenciales y direccionar recursos escasos hacia las áreas de máximo impacto.
3. Agilidad. Responder más rápido a los cambios del mercado o negocio y orientar los procesos negocio hacia las áreas demandadas.



En este punto es cuando la cultura empresarial se encuentra con que debe cambiar. La ciencia del datos implica a todos los niveles, desde operaciones a estrategia corporativa y necesita de la implicación de todos los actores asociados. Es entonces cuando aparecen los primeros problemas, ya que estamos hablando de procedimientos novedosos y la visión de negocio de dichos procedimientos es difícil de vislumbrar.

Una vez superada la fase de temor a la transformación digital, y cuando todos los niveles de negocio se han dado cuenta de la necesidad de convertirse en una empresa inteligente, es cuando surge la dificultad de iniciar este proceso de transformación.

Se necesitan analizar las necesidades reales de este proceso de transformación y diseñar una línea de trabajo que permita, por un lado, proporcionar resultados reales a corto plazo y, por otro, evolucionar en el nivel de conocimiento organizacional.



## GOBERNANZA DE DATOS

Un punto importante a tener en cuenta en el cambio organizacional para incorporar la ciencia de datos es la gobernanza de datos. Hay muchas empresas que tienen la información confidencial en juego. La cantidad de riesgo con tales datos generalmente compensa los esfuerzos y gastos gastados en el gobierno de datos.

La gobernanza de datos se debe principalmente a requisitos legales y reglamentarios; aunque una regla de gobernanza también puede ser cualquier política que la organización quiera practicar. Esta gobernanza establece dónde se almacenarán categorías específicas de datos y codifica los métodos de protección de datos principalmente, como la seguridad de la contraseña o el cifrado. También dicta la copia de seguridad de los datos y define los derechos de acceso y la destrucción de los datos archivados después de un tiempo adecuado. Las organizaciones también pueden establecer objetivos de gobierno para monitorear la calidad de los datos y acceder a los silos que contienen ciertos datos aislados.

El paso principal para la implementación exitosa de los programas de gobierno de datos es la creación de un equipo y la asignación de responsabilidades para los activos de datos a las personas de la organización. Estos propietarios de datos son responsables de la calidad de datos resultante y deben respaldar los procesos e iniciativas de calidad de datos para la organización.

Es un deber del equipo de datos asegurarse de que las iniciativas de gobierno de datos estén de acuerdo con los requisitos y necesidades del negocio. El gobierno de datos da la impresión de estar alineado con el departamento de tecnología de la organización, pero en realidad, se requiere que esté estrechamente relacionado con el negocio en sí para permitir a los interesados el acceso seguro a toda la información que necesitan para tomar decisiones adecuadas que dependen de los datos.

Además, se requiere el desarrollo de procesos de datos. Consiste en definiciones de almacenamiento de datos, su movimiento, cambios, acceso y seguridad. Los procesos de monitoreo, auditoría y control también necesitan ser trabajados principalmente para el cumplimiento en industrias altamente reguladas.



El gobierno de datos es la orquestación formal de personas, procesos y tecnología para permitir que una organización aproveche sus datos como un activo empresarial. Cumple una función crítica en los negocios para respaldar el cumplimiento normativo, pero también es crucial para garantizar una comprensión común de los activos de datos organizacionales en una empresa.

Los componentes clave incluyen el glosario empresarial, los diccionarios de datos, el linaje de datos y la gestión de metadatos, todos los cuales informan a los usuarios sobre la fuente, el uso, las relaciones y las definiciones relacionadas con los datos, incluidos los términos comerciales, los atributos y las dependencias. También asigna propietarios y administradores de datos a los activos de datos para aumentar la responsabilidad y permitir el acceso a los recursos de datos, fomentando el uso de datos. Y puede tener un impacto directo en la calidad de los datos al proporcionar monitoreo, dimensiones de calidad y puntajes.

La gobernanza de datos se trata de aumentar la comprensión de los datos en una empresa y fomentar la colaboración para aprovechar al máximo sus activos de datos. No debería residir en el departamento de Tecnología de Información como una empresa técnica; debe centrarse en el negocio para impulsar la utilización de datos.

Los usuarios comprometidos que confían y entienden los datos tienen muchas más probabilidades de utilizar el análisis para buscar inteligencia y conocimiento. El buen gobierno de los datos crea usuarios educados que pueden confiar en los activos de datos y con frecuencia recurrirán a los análisis para ayudar a resolver problemas comerciales y descubrir problemas y oportunidades invisibles.

Los elementos de gobernanza también hacen posible la ciencia de datos. La gestión de metadatos es la supervisión de "datos sobre datos", un aspecto importante de un enfoque integral de gobernanza de datos. Pero también es la clave para el análisis de datos para obtener información valiosa del negocio, y el análisis predictivo es a su vez la clave para obtener una ventaja competitiva y una visión de los resultados y eventos futuros.



## LA OPERACIÓN Y LOS PROYECTOS ORGANIZACIONALES

Muchos de los problemas que las organizaciones pueden encontrar en proyectos de software son causados por dificultades de gestión y liderazgo (es decir, planificación, estimación, medición, control, comunicación, coordinación y gestión de riesgos) en lugar de problemas técnicos (es decir, análisis, diseño, codificación y prueba). Estas dificultades surgen de múltiples fuentes; algunos pueden ser controlados por los equipos de proyectos y otros no. Los factores que no se pueden controlar se denominan restricciones, que son limitaciones impuestas por agentes externos en algunos o todos los dominios operativos, requisitos operativos, requisitos del producto, alcance del proyecto, presupuesto, recursos, fecha de finalización y tecnología de plataforma.

Las organizaciones que busquen una ventaja competitiva a través de los datos se enfrentan una serie de desafíos a la hora de querer incorporar la ciencia de datos dentro de la operación y procesos empresariales.

Los principales retos que enfrentan las empresas son: datos carentes de calidad, falta de talentos (científicos de datos), políticas y controles internos, carencia de claridad en las preguntas que queremos responder, datos no democratizados, abandono del proyecto por desuso, altos costos.

Para poder encarar estos desafíos, las organizaciones deben contar con una estrategia clara y un proyecto definido orientado a la incorporación de la ciencia de datos dentro de la operación organizacional.

Aunque la ciencia de datos basada en tecnología es una disciplina que puede aplicarse a todo tipo de campos, sectores, negocios e investigaciones, lo cierto es que, dejando aparte las diferencias particulares dependientes del campo de aplicación y situaciones específicas organizacionales, en sus líneas básicas, los proyectos siguen una operatoria muy parecida, en una serie de fases que suelen repetirse en distintos proyectos de implementación de ciencia de datos en las empresas.

Es más difícil distinguir la diferencia entre un éxito modesto y la excelencia. De hecho, en la ciencia de los datos pueden parecer muy similares durante tal vez un año. Después de



varios años, sin embargo, una excelente estrategia producirá de manera evidente resultados más valiosos que aquella que no fue debidamente analizada.

Tanto las estrategias mediocres como las excelentes comienzan con una serie de experimentos e inversiones que conducen a proyectos de datos. Después de unos años, algunos de estos proyectos funcionan y están en camino a la producción.

En la estrategia mediocre, uno o dos de estos proyectos pueden incluso tener un retorno de inversión positivo para el negocio. Normalmente, estos proyectos son algún tipo de automatización para ahorrar costos, o aplicar el aprendizaje automático a un proceso existente para mejorar su eficiencia o rendimiento. Esto se parece mucho al éxito, y puede ser suficiente, pero se pueden perdiendo las ventajas únicas de una excelente estrategia de datos.

En una excelente estrategia, se desarrollan más proyectos de datos y su desarrollo es más rentable. Además, el proceso de construcción de los primeros proyectos inspira nuevas ideas de proyectos. En una excelente estrategia, los proyectos incluyen automatización y mejoras de eficiencia y rendimiento, pero también incluirán proyectos e ideas para la generación de nuevos ingresos y negocios completamente nuevos impulsados por sus activos de datos únicos. Los equipos de datos trabajan bien juntos, se basan en el trabajo de los demás y colaboran sin problemas con sus socios comerciales. Hay una visión clara de cómo puede verse el futuro de la empresa impulsado por el aprendizaje automático, y todos están trabajando juntos para lograrlo.

La elaboración de una estrategia de datos requiere muchos interlocutores en la mesa de trabajo, incluidos los expertos en datos, el liderazgo tecnológico y los expertos en negocios y en la materia.

Muchas organizaciones deciden estrictamente qué proyectos de datos perseguir, que solo es una fórmula para la estrategia de datos mediocre. La gerencia identifica un conjunto de proyectos que le gustaría ver construido y crea el diagrama de de priorización: un eje representa el valor de un proyecto dado para el negocio y el otro eje representa su complejidad o costo de desarrollo estimados.



A cada proyecto se le asigna un lugar en el cuadro, y la administración asigna los recursos limitados de la empresa a los proyectos que creen que serán los más económicos y que tengan el mayor valor comercial.

Esto no es incorrecto, pero tampoco es óptimo. Una excelente estrategia de datos se mueve más allá de una evaluación directa de cada proyecto de forma aislada para considerar algunas dimensiones adicionales.

Primero, una excelente estrategia de datos incluye un núcleo organizativo bien coordinado. Se basa en una inversión en tecnología centralizada y en valores predeterminados bien seleccionados y coordinados para la arquitectura de las aplicaciones de datos. Esta centralización de valores predeterminados permite que cada aplicación tome decisiones diferentes si es necesario, al tiempo que mantiene la máxima compatibilidad en toda la organización y la flexibilidad en el tiempo de forma predeterminada.

Por ejemplo, cada línea de negocios tenía una pila de tecnología diferente y un grupo de Tecnología de la Información independiente, lo que conllevaba desafíos relacionados con la integración de datos que ya existían y diferentes arquitecturas para todas las inversiones futuras.

En segundo lugar, una excelente estrategia de datos es específica a corto plazo y flexible a largo plazo. En el mercado es posible entender acerca de cómo serán las capacidades de aprendizaje automático del mañana, pero menos acerca de cómo serán las capacidades requeridas en el corto plazo. Del mismo modo, el panorama empresarial se está transformando, lo que lleva a una nueva competencia y nuevas oportunidades. Las organizaciones que participan en ciclos de planificación de cinco años pueden perder las oportunidades que surgen mientras tanto. Una estrategia excelente es aquella que es adaptable y se considera un documento vivo.

Las mejores estrategias son fuertes en convicción direccional, pero flexibles en los detalles. Desea saber dónde quiere terminar, pero no necesariamente predefinir cada paso que necesita tomar para llegar allí.

Finalmente, una excelente estrategia de datos tiene en cuenta una visión clave: los proyectos de ciencia de datos no son independientes entre sí. Con cada proyecto completado,



sea exitoso o no, se crea una base para construir proyectos posteriores más fácilmente y a un menor costo.

Este enfoque hace que los proyectos de mayor valor, aquellos que quizás hubieran parecido demasiado ambiciosos, se vean menos como un empuje agresivo y costoso. En cambio, revela que tales proyectos pueden ser más eficientes y más seguros para proceder que otros proyectos de menor valor que parecían atractivos en un análisis ingenuo.

Dicho de otra manera, una excelente estrategia de datos reconoce que los proyectos son incompatibles entre sí, y que los costos de los proyectos cambian con el tiempo a la luz de otros proyectos emprendidos (y también de las nuevas tecnologías). Esto permite una planificación más precisa y puede ampliar las capacidades de la organización más de lo esperado. Puede revisar este proceso de planificación trimestralmente, lo que está en línea con la rapidez con la que las tecnologías de aprendizaje automático están cambiando.

Actualmente el mundo está en un momento de desarrollo del aprendizaje automático, inteligencia artificial y datos donde la tecnología no se comercializa y no es del todo obvio dónde invertir. Las empresas con excelentes estrategias de datos tendrán más probabilidades de elegir bien.

La ciencia relacionada con los datos es capaz de generar productos que la gente quiere y valora. Y esto no ocurre sólo en compañías de Internet sino también en cadenas de tiendas físicas que son capaces de utilizar los datos para optimizar cada aspecto de sus operaciones de venta de productos.

Dada la importancia de la ciencia de datos, resulta esencial plantearse qué pueden conseguir los equipos de científicos de datos en tres niveles:

1. Valor que aportan a la organización con sus descubrimientos e innovaciones.
2. Integración con otros departamentos y contribución a sincronizarlos con el equipo de Tecnología de la Información.
3. Aumento de la efectividad de las estrategias de ciencia de los datos, que tendrá mucho que ver con la construcción de equipos de data science.

A la hora de crear equipos de científicos de datos de alto rendimiento hace falta entender que, gran parte de su potencial generador de valor, radica en la variedad de roles que pueden



asumir estos profesionales. Entre ellos destacan los relacionados con los roles que se describen a continuación:

**Inteligencia empresarial:** el apoyo de sus conclusiones a la toma de decisiones empresarial es crítico. Los equipos de científicos de datos consiguen, en base al estudio y monitorización de diferentes métricas ayudar a los responsables del negocio a entender mejor las circunstancias en cada momento. Conocer las herramientas necesarias para llevar a cabo esta labor, avanzar al ritmo que lo hace la organización y ganar en sofisticación resulta esencial para proporcionar información de calidad. En este proceso, los científicos de datos se encargan de investigar las fuentes de datos existentes y de combinarlas con otras procedentes del exterior de la compañía para facilitar una mejor comprensión del panorama competitivo, priorizar la estrategia y la táctica y proporcionar claridad sobre las hipótesis que pueden surgir durante la planificación estratégica.

**Análisis de productos y marketing:** hay muchas organizaciones que son capaces de diseñar aplicaciones que interactúan directamente con los clientes, utilizando datos para explicar y mostrar una proposición de valor de un servicio o producto. Esta estrategia resulta más económica, dado el rápido descenso del costo de la computación, y permite a las organizaciones utilizar algoritmos y técnicas numéricas comunes para probar la efectividad de sus lanzamientos. Algunos de los más exitosos tienen que ver con productos que facilitan la introducción en otros productos, productos que proporcionan contenido altamente personalizado o que ayudan a impulsar la propuesta de valor de la empresa.

**La lucha contra amenazas, protección ante ataques y prevención del fraude:** la seguridad de datos es uno de los cometidos de los equipos de data scientist en la organización. Empezando por recopilar la información adecuada, es posible identificar las fuentes de riesgo potencial y mitigar o evitar sus efectos. Para una protección efectiva son importantes las habilidades de los científicos de datos y su integración con el área de negocio.

Y no sólo estas tres áreas, sino que la monitorización y el mantenimiento de los sistemas de almacenes y bases de datos, así como todas las cuestiones relativas a la arquitectura e ingeniería de datos son también competencias del científico de datos.

La clave para crear equipos de científicos de datos de primera categoría es evitar su aislamiento, tanto del grupo técnico dedicado a operaciones tradicionales, como del área de



negocio; fomentar la heterogeneidad en lo que a habilidades se refiere y elegir a esos individuos que destacan por su curiosidad.

Para formar un grupo potente de científicos de datos es preciso encontrar perfiles interdisciplinarios pero capaces de trabajar en equipo, con ganas de aprovechar las oportunidades que presentan los datos y la confianza para lograrlo juntos.

A continuación se detallan una serie de fases a tener en cuenta para implementar un departamento de ciencia de datos en una organización.

En primer lugar, se debe evitar que los clientes dirijan la integración de la ciencia de datos basada en tecnología. Puede ser probable que los clientes impliquen en ayudar a implementar qué tipo de funcionalidades debe tener la herramienta. No obstante, el abuso de estas prácticas conlleva que las empresas encargadas en suministrar las herramientas sobrepasen el presupuesto y los plazos, y que a la larga acaben con un sistema muy complejo que resulta demasiado difícil de utilizar. Se debe entender que ciertas partes de las empresas no aprecian ni entiendan la parte tecnológica de la herramienta. Para mejorar, debe asegurarse de que ha seleccionado cuidadosamente a los representantes de los clientes finales e invierte mucho tiempo en comprender sus expectativas y preocupaciones

En segundo lugar, se deben realizar comprobaciones y seguimientos complejos y completos de la herramienta después de implementar. Debido a que una solución de Business Intelligence implica nuevos procesos, aplicaciones y flujos de trabajo para las empresas, es obvio que es imprescindible realizar un gran volumen de comprobaciones antes de que dicha solución pueda ser implementada. Es recomendable que la empresa encargada de suministrar la herramienta plane las comprobaciones en el momento en el que se completa el proyecto.

Como tercera fase, se debe hacer una elección adecuada de la Tecnología de Datos. Muchas empresas que suministran herramientas de Inteligencia de Datos al implementarlas utilizan tecnología de datos obsoleta como un sistema de planificación de recursos empresariales. Las prácticas de negocio están cambiando a gran velocidad, y las herramientas de inteligencia de negocio necesitan ser actualizadas y lo bastante versátiles para que las empresas se sientan cómodas a la hora de replicar sus procesos cambiantes en sus procesos de Inteligencia de Datos. Se debe seleccionar una tecnología que ofrezca visualizaciones de datos, dashboards de arrastrar y soltar y que ofrezca accesibilidad móvil.



En cuarto lugar, es importante poner atención a los problemas que surgen para integrar los datos. Las aplicaciones de Inteligencia de Datos sirven para proporcionar una gran capacidad de análisis, procesamiento y presentación de datos a las empresas. Por eso hay que prestar atención a la hora de introducir datos limpios y en los formatos correctos. Todo error en los informes financieros y en el cumplimiento de las reglas puede conllevar a consecuencias graves para la reputación de una empresa.

Por último se debe cuidar el soporte a las empresas después de implementar ciencia de datos en la compañía. La preselección de los principales grupos de clientes, la elaboración de planes de formación detallados y la impartición de la formación son esenciales.

Es importante que las organizaciones se planteen, previo a comenzar un proyecto de incorporación de ciencia de datos basada en tecnología, cómo será la estrategia para implementarla, es decir, si se plantea como la incorporación de un equipo de trabajo permanente o uno temporal que tome algunos análisis de datos ad hoc y se disuelva posteriormente.

Embarcarse en la ciencia de datos y el análisis predictivo requiere una comprensión clara de cómo se va a presentar, mantener y ampliar aún más la iniciativa en términos de estructura de equipo. Recomendamos considerar tres estructuras básicas de equipo que coincidan con las diferentes etapas de la adopción del aprendizaje automático.

### **Estructura centrada en Tecnología de la Información**

A veces, contratar científicos de datos no es una opción, y debe aprovechar el talento que ya está en la empresa. La función principal de análisis y liderazgo sería un "traductor de negocios", generalmente denominado director de análisis o director de datos. El último término gradualmente se vuelve redundante a medida que la mayoría de los procesos de datos se modifican hacia el análisis predictivo. Esta persona debe ser capaz de liderar la iniciativa.

A continuación se detallan los beneficios de la estructura centrada en Tecnología de la Información:

1. Se aprovechan las nuevas inversiones con los recursos de Tecnología de la Información existentes.
2. La infraestructura informática es proporcionada y mantenida por un servicio externo.



3. Los especialistas internos pueden recibir capacitación para desarrollar aún más el potencial de análisis predictivo.
4. La gestión de silos cruzados se reduce ya que todas las operaciones se llevan a cabo dentro del departamento de Tecnología de la Información.
5. Menos tiempo de comercialización para tareas de aprendizaje automático relativamente simples que requieren uno o algunos modelos.

Las desventajas de la estructura centrada en Tecnología de la Información son:

1. Métodos limitados de aprendizaje automático y procedimientos de limpieza de datos que proporcionan estos servicios
2. Se deben considerar los costos de la capacitación, prueba y predicción del modelo. Esto conlleva la incertidumbre del costo eventual por predicción ya que el número de iteraciones necesarias no puede estimarse por adelantado

### **Estructura integrada**

Con la estructura integrada, un equipo de ciencia de datos se enfoca en la preparación de conjuntos de datos y la capacitación de modelos, mientras que los especialistas de Tecnología de la Información se hacen cargo de las interfaces y la infraestructura que soportan los modelos implementados. La combinación de experiencia en aprendizaje automático con recursos de Tecnología de la Información es la opción más viable para operaciones de aprendizaje automático constantes y escalables.

A diferencia del enfoque centrado en Tecnología de la Información, el método integrado requiere contar con un científico de datos experimentado en un equipo y un esfuerzo de reclutamiento elaborado de antemano. Esto garantiza una mejor flexibilidad operativa en términos de técnicas disponibles. Además de servicios de extremo a extremo y aún limitados, puede aprovechar las herramientas y bibliotecas de aprendizaje automático más profundas.

Beneficios de aplicar una estructura integrada:

1. Aprovechar los recursos e inversiones de Tecnología de la Información existentes.
2. Los científicos de datos se centran en la innovación utilizando todo el potencial de las aplicaciones de aprendizaje automático personalizadas y como servicio.



3. Se puede comenzar con uno o dos científicos de datos, luego se capacita y aborda a más expertos locales.
4. Se pueden utilizar combinaciones de modelos personalizados que producen predicciones mejores o más amplias.

Desventajas de aplicar una estructura integrada:

1. Se requiere infraestructura informática en caso de uso personalizado de aprendizaje automático.
2. La gestión entre equipos divididos requiere un esfuerzo organizacional considerable.
3. Importantes inversiones en adquisición de talento de ciencia de datos.
4. Compromiso y retención de talentos en ciencia de datos.

### **Departamento especializado en ciencia de datos**

Para reducir el esfuerzo de gestión y crear un marco de aprendizaje automático que lo abarque todo, puede ejecutar todo el flujo de trabajo de aprendizaje automático dentro de un departamento independiente de ciencia de datos. Este enfoque implica un costo más alto. Todas las operaciones, desde la limpieza de datos y la capacitación de modelos hasta la creación de interfaces frontend<sup>11</sup>, son realizadas por un equipo dedicado de ciencia de datos.

No significa necesariamente que todos los miembros del equipo deben tener una formación en ciencia de datos, sino que deben adquirir infraestructura tecnológica y habilidades de gestión de servicios.

Un modelo de estructura especializado ayuda a abordar tareas complejas de ciencia de datos que incluyen la investigación, el uso de múltiples modelos de aprendizaje automático adaptados a diversos aspectos de la toma de decisiones o múltiples servicios respaldados por el aprendizaje automático. En el caso de grandes organizaciones, los equipos especializados en ciencia de datos pueden complementar diferentes unidades de negocios y operar dentro de sus campos específicos de interés analítico.

Beneficios de desarrollar un departamento especializado de ciencia de datos:

---

<sup>11</sup> El autor entiende por frontend a la parte de un sitio web que interactúa con los usuarios, es decir, que está del lado del cliente.



1. Gestión centralizada de la ciencia de datos y mayor capacidad de resolución de problemas.
2. Se aprovecha todo el potencial de las aplicaciones de aprendizaje automático personalizadas y como servicio.
3. Se pueden resolver problemas complejos de predicción que requieren una investigación profunda o construir fábricas de modelos segmentados (que operan automáticamente en diferentes segmentos y unidades de negocios).
4. Es posible establecer un área de juegos de ciencia de datos con todas las funciones para fomentar la innovación.
5. Mayor potencial de escalabilidad.

Desventajas de desarrollar un departamento especializado de ciencia de datos:

1. Se debe construir y mantener una infraestructura computacional compleja.
2. Conlleva grandes inversiones en adquisición de talento en ciencia de datos.
3. Implica un gran compromiso y retención de talentos en ciencia de datos.

Previo a establecer un proyecto de ciencia de datos los equipos que estarán invocados en incorporar dicha ciencia basada en tecnología deben tener en cuenta que todo proyecto organizacional tiene una serie de pasos que pueden llevarlo al éxito o al fracaso, dependiendo del foco que se haya puesto.

Primero hay que definir metas concretas. Como en todo proyecto, las primeras preguntas que se plantean son ¿Qué se quiere conseguir al final del proyecto? ¿Para qué la organización se embarca en este proceso e se invierte dinero y esfuerzo en él? Estos objetivos deben estar bien marcados y, sobre todo, bien comunicados. Es clave que se compartan las metas, se analicen entre todo el equipo, y que se pregunte también a las personas que ejecutarán cada una de las actividades del proyecto.

En segundo lugar, se debe establecer una estrategia. En esta fase se identifican las actividades principales del plan de proyecto, orientado a la meta que anteriormente se ha valorado y comunicado.



Luego hay que definir y asignar responsabilidades. Es muy importante es definir las funciones y actividades adecuadas como asignar la persona correcta para su ejecución. Por ello, es conveniente elaborar una lista de todos los colaboradores que formarán parte del proyecto.

Cuando hay que organizar un equipo de trabajo, se deben tener en cuenta una serie de requisitos. Uno de ellos, será formar equipos multidisciplinarios. Cada uno de sus miembros posee unas singularidades que lo hacen necesario en el proyecto.

Posteriormente, es necesario confeccionar el boceto inicial del plan a seguir. Puede ser en la misma reunión inicial. Se comparten con todo el equipo los primeros pasos para poner en marcha el proyecto.

Un punto importante es poder identificar los parámetros que medirán el éxito del proyecto. ¿Qué es lo que se considerará como un triunfo? Hay que cuantificar cada logro y establecer los hitos de proyecto. Además hay que definir cuál es el camino crítico del proyecto y cuáles son los cuellos de botella. Al planificar cada actividad y relacionarlas entre sí, se genera una ruta crítica o camino crítico. Lo que quiere decir que si alguna de las actividades de esta ruta sufre una demora, todo el proyecto se retrasaría, siendo susceptible el fracaso del mismo.

En los proyectos es necesario hacer un correcto análisis de riesgos, es decir, listar los riesgos que pueden distorsionar el proyecto y ver cómo pueden ser mitigados Sólo así podrán minimizarse rápidamente.

Para iniciar un proyecto con éxito es conveniente diseñar un buen ejemplo de plan de comunicación. Lo que quiere decir que se establecerán las reuniones periódicas, los responsables en notificar regularmente al cliente y encargarse de que todos tengan voz y voto en el proyecto. Además, se deben asignar los recursos necesarios para llevar a cabo el proyecto. Todos deben tener acceso a las herramientas que van a necesitar en sus actividades.



## CONCLUSIONES

Tradicionalmente los negocios se han dirigido por la experiencia y teoría que acumulaban los directivos de las empresas. Por esto el conocimiento de los empresarios y el perfil formativo del resto de empleados en cada sector era un aspecto muy valorado dentro de grandes corporaciones.

En el mercado actual, cambiante, dinámico e innovador estos conocimientos que provienen de experiencias profesionales del pasado son cada vez más difíciles de valorar, porque en nada se parecen a los nuevos modelos de negocio que actuales.

¿Qué ha sucedido? La respuesta está en los datos. Desde que comenzó la era digital, se comenzó a guardar registro de todas las transacciones de las organizaciones; y la consecuencia inevitable de esto es la generación constante de una gran cantidad de datos.

Actualmente estamos en la era de los grandes datos, por lo que el desafío de las organizaciones no es obtener datos sino obtener los datos correctos y usar computadoras para aumentar su conocimiento del dominio e identificar patrones que no vieron o que no pudieron encontrar previamente.

El Big Data es el resultado de recopilar información en su nivel más granular: es lo que se obtiene cuando las organizaciones instrumentan un sistema y guardan todos los datos que dicha instrumentación puede reunir.

El concepto de Big Data solía significar datos que una sola máquina no podía manejar. Ahora los grandes datos se han convertido en un concepto de moda que significa todo lo relacionado con el análisis o la visualización de datos.

Actualmente el Big Data se ha expandido para significar una situación en la que la logística de almacenamiento, procesamiento o análisis de datos ha superado las capacidades operativas tradicionales de las organizaciones. Dicho de otra forma, ahora las organizaciones tienen demasiados datos para almacenar de manera efectiva o computar de manera eficiente utilizando métodos tradicionales. Esto también puede incluir tener muy poco tiempo para procesar los datos para usar la información para tomar decisiones.



Además, Big Data significa utilizar todos los datos disponibles en sus modelos, no solo muestras, es decir, indica la capacidad de usar conjuntos de datos completos en lugar de solo segmentos como en años anteriores.

La ciencia de datos basada en tecnología es uno de los campos de estudio capaz de evolucionar y revolucionar todos los departamentos de una empresa por su habilidad para reducir los tiempos y mejorar la eficiencia a varios niveles.

Los datos por sí solos no ofrecen a priori ningún valor, ni a las empresas ni a la sociedad. En ocasiones, incluso, ni siquiera son legibles o comprensibles a simple vista, y es necesaria una transformación previa de los mismos para poder interpretarlos.

El ciencia de datos involucra métodos científicos, procesos y modelos para llevar a cabo esta extracción de valor. Incluye campos de análisis como la analítica descriptiva, la estadística, la minería de datos o el aprendizaje automático. La ciencia de datos, desde algo básico como un análisis descriptivo hasta algo evolucionado como la implementación de modelos predictivos, permite a las organizaciones obtener información de valor procedente de dichos datos, detectar patrones, y conseguir así ventajas competitivas, identificar nuevas oportunidades de negocio y mejorar la experiencia de los usuarios.

La ciencia de datos se encarga de extraer, analizar e interpretar los datos provenientes de diversos dispositivos como lo son los smartphones, las aplicaciones móviles, los sistemas de punto de venta, entre otros. La finalidad es convertir dichos datos en información valiosa y relevante que ayude a que la empresa crezca mucho más al mejorar por completo la experiencia del consumidor mediante estrategias que le ofrezcan lo que el desea.

Además, la ciencia de datos permite que las empresas estén en una posición única donde pueden cuantificar y determinar cuál o cuáles son las estrategias que están funcionando, las que no y el por qué.

Hay que tener en cuenta que se debe ejecutar un buen análisis de datos para que las organizaciones puedan ofrecer la mejor experiencia al cliente, de lo contrario o en caso de que la empresa decida dejar de lado los datos que se le indican, podría perder clientes.



La realidad es que no es necesario tener grandes volúmenes de datos para poder extraer valor. Lo importante es la calidad de la información que se obtiene de ellos tras aplicar técnicas de la ciencia de datos y eso depende principalmente de dos aspectos clave:

- la relevancia de la información en el problema que se quiere resolver,
- y la limpieza de los datos para que el ruido no interfiera en la búsqueda de patrones reales.

A medida que las organizaciones recopilan volúmenes de datos cada vez más grandes enfrentan el desafío de extraer valor de ellos. Para eso, hay posiciones que son cruciales, como el científico de datos, que es uno de los profesionales más demandados actualmente en todo el mundo.

El científico de datos no es meramente un nuevo nombre para la gente con grandes habilidades analíticas como a algunos pueda parecer, sino más bien una designación para la gente que, además de capacidad analítica, tiene un amplio conocimiento de las diversas áreas importantes para lidiar con los problemas complejos de la vida real de la ciencia de los datos. Tratar con el tamaño, la velocidad, la diversidad y la complejidad de los modernos conjuntos de datos requiere mucho más que ser experto en estadística y herramientas de software comercial.

Un científico de los datos debe saber sobre representación de datos, ingeniería del software, diseño de visualización de datos, y procesos de los negocios, al igual que ser capaz de contar la “historia de los datos” basándose en rigurosos análisis cuantitativos.

El desarrollo organizacional ayuda a las organizaciones a abordar y adoptar el cambio desde la perspectiva de que el cambio puede ser una oportunidad y no una amenaza. Las demandas de un cambio pueden provenir del exterior de la organización así como del interior de la misma.

Todo proyecto de software, grande o pequeño, simple o complejo, debe mantener la visión del proceso (la hoja de ruta del proyecto) y la visión del producto (los objetivos del producto) de principio a fin; de lo contrario, es fácil perder de vista la visión y los objetivos en medio de las actividades laborales diarias de un proyecto.



La gestión del desarrollo y la modificación de los proyectos se logra descomponiendo las actividades de trabajo de alto nivel en actividades de trabajo de nivel inferior de manera jerárquica. Las actividades laborales de nivel más bajo sujetas a la planificación de la gestión y la responsabilidad se denominan tareas; por lo tanto, las actividades son agregaciones de tareas y actividades subordinadas.

Para que las organizaciones puedan desarrollar la ciencia de datos y el análisis predictivo se requiere de un proyecto de cómo se va a presentar, mantener y ampliar aún más la iniciativa para la estructura de un equipo dedicado a los análisis de datos. Se pueden considerar tres estructuras básicas de equipo que adaptables con las etapas de la adopción de la ciencia de datos:

- **Estructura centrada en Tecnología de la Información:** sirve cuando no es una posibilidad contratar científicos de datos y se debe aprovechar el talento que ya está en la organización. La función principal de análisis y liderazgo sería un "traductor de negocios".
- **Estructura integrada:** un equipo de ciencia de datos se enfoca en la preparación de conjuntos de datos y la capacitación de modelos, mientras que los especialistas de TI se hacen cargo de las interfaces y la infraestructura que soportan los modelos implementados.
- **Departamento especializado en ciencia de datos:** se ejecuta todo el flujo de trabajo de aprendizaje automático dentro de un departamento independiente de ciencia de datos. Este enfoque puede implicar un costo más elevado.

Previo a establecer un proyecto de ciencia de datos se debe tener en cuenta que todo proyecto organizacional tiene una serie de pasos que pueden llevarlo al éxito o al fracaso.

- Primero hay que definir metas concretas.
- En segundo lugar, se debe establecer una estrategia.
- Luego hay que definir y asignar responsabilidades.
- En cuarto lugar hay que formar equipos multidisciplinarios.
- Posteriormente, es necesario confeccionar el boceto inicial del plan a seguir.
- Se deben identificar los parámetros que medirán el éxito del proyecto.
- Cuantificar los riesgos.



Universidad de Buenos Aires  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado



En resumen, la ciencia de datos ofrece a las organizaciones un conjunto de herramientas para que puedan sacar información de calidad de los datos que recopila a través de las transacciones con sus clientes.

Incorporarla requiere de un plan de trabajo que incluye la gestión del cambio en la organización, seguir los pasos adecuados para administrar proyectos, la contratación de personal idóneo y adecuar a los equipos de trabajo que llevan la operatoria de la empresa.



## REFERENCIAS BIBLIOGRÁFICAS

- Kotter, J. (1995). *Liderar el cambio: por qué fracasan los intentos de transformación*. Boston: Harvard Business Review.
- Fernandez Blanco, M. L., Coronel, L., & Chinkes, E. (2015). *Big Data: El Dato en un Rol Estratégico*. Buenos AIRES.
- Liu, A. (2015). *Data Science and Data Scientist*. IBM.
- Institute, P. M. (2013). *Guía de los fundamentos para la dirección de proyectos (guía del PMBOK) - Quinta edición*. Pensilvania: Instituto de Gerenciamiento de Proyectos.
- Calvo, J., Guzmán, M., & Ramos, D. (2018). *Aprendizaje Automático: una pieza clave en la transformación de los modelos de negocio*. España: Management Solutions.
- Cady, F. (2017). *The Data Science Handbook*. Wiley.
- Sondergaard, P. (2011). Gartner Symposium/ITxpo. *Gartner Symposium/ITxpo*.
- Davenport, T. H., & Patil, D. (2012). *Científico de datos: el trabajo más sexy del siglo XXI*. Harvard Business Review.
- Monnappa, A. (2017). *Data Science vs. Big Data vs. Data Analytics*. Obtenido de <https://www.simplilearn.com/>: [https://www.simplilearn.com/authors/avantika-monappa?source=frs\\_detailsPage](https://www.simplilearn.com/authors/avantika-monappa?source=frs_detailsPage)
- Schmidberger, M. (2013). *Big-data Analyses with R*. Sandringham.
- Monleón-Getino, A. (02 de 12 de 2015). *El impacto del Big-data en la Sociedad de la Información. Significado y utilidad*. Obtenido de <https://revistas.ucm.es/>: [https://doi.org/10.5209/rev\\_HICS.2015.v20.n2.51392](https://doi.org/10.5209/rev_HICS.2015.v20.n2.51392)
- Barranco Frago, R. (18 de 06 de 2012). *¿Qué es Big Data? Todos formamos parte de ese gran crecimiento de datos*. Obtenido de <https://www.ibm.com/>: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/que-es-big-data-pdf.pdf>
- Lake, P., & Crowther, P. (2013). *Concise Guide to Databases*. Londres: Springer.
- Bramer, M. (2016). *Principles of Data*. Londres: Springer.
- Pampliega, J. M. (2017). *Introduction to Big Data*. ITBA.
- Failey, R. (2009). *Managing and Leading Software Projects*. Canadá: Wiley.
- Terán, R. (2002). *Desarrollo organizacional y gestión del cambio*. Gestipolis.
- Manyika, J., & Chui, M. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Gardiner, C. (2016). *La revolución de los datos*.