

# Big data, algoritmos e Inteligencia Artificial

Claves para directivos



ETHALO

# Contenido

1. Oportunidad del programa
2. Público objetivo
3. Requisitos para poder realizar el programa
4. Estructura del programa
5. Metodología empleada

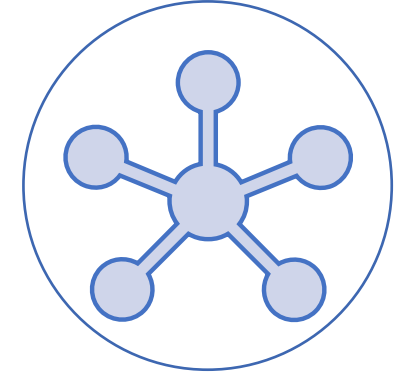
## Los equipos directivos hoy necesitan...



Conocer las **posibilidades** y limitaciones del uso de datos y modelos en los procesos de gestión.



Poder **dirigir los esfuerzos** de los equipos especializados en datos y modelos, asegurando un entendimiento común (**lingüa franca**).



**Integrar** los modelos y datos en los procesos, optimizando sus resultados y **controlando los riesgos** asociados.

**Gestionar el big data y los equipos de científicos de datos para aportar valor a la organización**

## Una situación que debemos anticipar



Sobre esta problemática ya tenemos antecedentes.

**En el mundo del desarrollo del software situaciones semejantes han tenido un enorme coste para las empresas.**

- Los programadores tradicionalmente se han quejado de que los usuarios piden cosas imposibles y sin conocimiento;
- En el otro lado, los usuarios acaban pensando que los programadores no hacen bien su tarea, pues tienen que asumir limitaciones en el producto final sin entender qué ha impedido su desarrollo.

Este programa quiere **facilitar esta tarea en el ámbito del big data y la modelización, creando una capacidad de diálogo natural** entre los gestores de negocio y los especialistas del dato, que maximice el valor del trabajo coordinado entre ambos.

Para ello, este programa aborda los componentes clave...

...desde una visión integrada, incluyendo aspectos que permiten **visualizar posibilidades**, conocer los **procesos de construcción e implantación** de modelos y **asegurar el control** de los riesgos y la **transformación** de la organización:

0

Introducción a la ciencia del dato y los modelos  
Posibilidades del Big Data y los modelos de decisión

1

Los datos como  
materia prima

2

La modelización

3

Toma de decisiones  
automáticas con  
modelos

4

Riesgo de modelo.  
Validación y control

5

Hacia una organización data driven: nuevas formas de trabajar.  
Claves éticas y de gobierno.

¿A quién va dirigido?

- A **directivos responsables de los negocios o su primera línea**, que quieran introducir las ventajas del big data y los modelos automáticos de decisión en sus procesos de negocio.
- **Incluye responsables de funciones internas**, como finanzas o recursos humanos, que, de igual manera, quieren mejorar sus procesos mediante la incorporación de estas técnicas procedentes del ámbito de datos y modelos.
- El curso aporta el **conocimiento necesario para entender las posibilidades y requerimientos** asociados a las soluciones basadas en el aprovechamiento del big data, saber proponer estas soluciones a la alta dirección, **y poder interactuar con los especialistas en la materia** y lograr resultados economizando tiempo, errores y otros costes asociados.
- El programa aborda una visión técnica, pero bajo la premisa de las necesidades específicas de un directivo para su negocio, **transformando la complejidad de la terminología y procesos asociados en conceptos asequibles**, y asegurando que existen las bases suficientes para desarrollar cualquier solución en este ámbito en colaboración con los equipos de datos y desarrollo de modelos.

## ¿Qué debe saber el asistente al programa?

- El programa está destinado a directivos que **no tienen necesariamente conocimientos en la materia**.
- Aunque abordará temas de índole estadística y tecnológica, se hará desde la perspectiva de lo que un directivo necesita saber para interactuar con los especialistas, revelando los aspectos esenciales y traduciéndolos en conceptos manejables para un directivo no especialista.
- Todos los temas se abordarán bajo la premisa de cumplir un **triple objetivo**:
  - Asegurar **que se adquieren los conceptos básicos** de cada tema sin asumir conocimientos previos, usando un **formato divulgativo**, sin perder la precisión necesaria.
  - **Facilitar la visualización**, e incluso la ejecución de casos reales de uso, de forma que el asistente pueda **desmitificar la complejidad** tradicionalmente atribuida a estas materias. Por ejemplo, los directivos podrán automatizar la descarga de datos desde una web a través de Excel, hacer un modelo predictivo o descubrir patrones en un set de datos.
  - **Identificar los puntos clave (*check points*)** que deben asegurar en su interrelación con los equipos especialistas y que garanticen un diálogo fluido y eficiente.
- Por tanto, sólo requiere del asistente una mentalidad abierta a integrar un conjunto de conceptos esenciales para aprovechar el big data y la modelización en su plan de negocio.

# Estructura del programa

Profundizaremos en los **elementos**, en ocasiones **intangibles, necesarios para sacar el máximo provecho** de la ciencia de los datos y los modelos en las organizaciones. Los aspectos **culturales, organizativos y de gobierno** imprescindibles para ello, así como los impactos éticos y reputacionales.

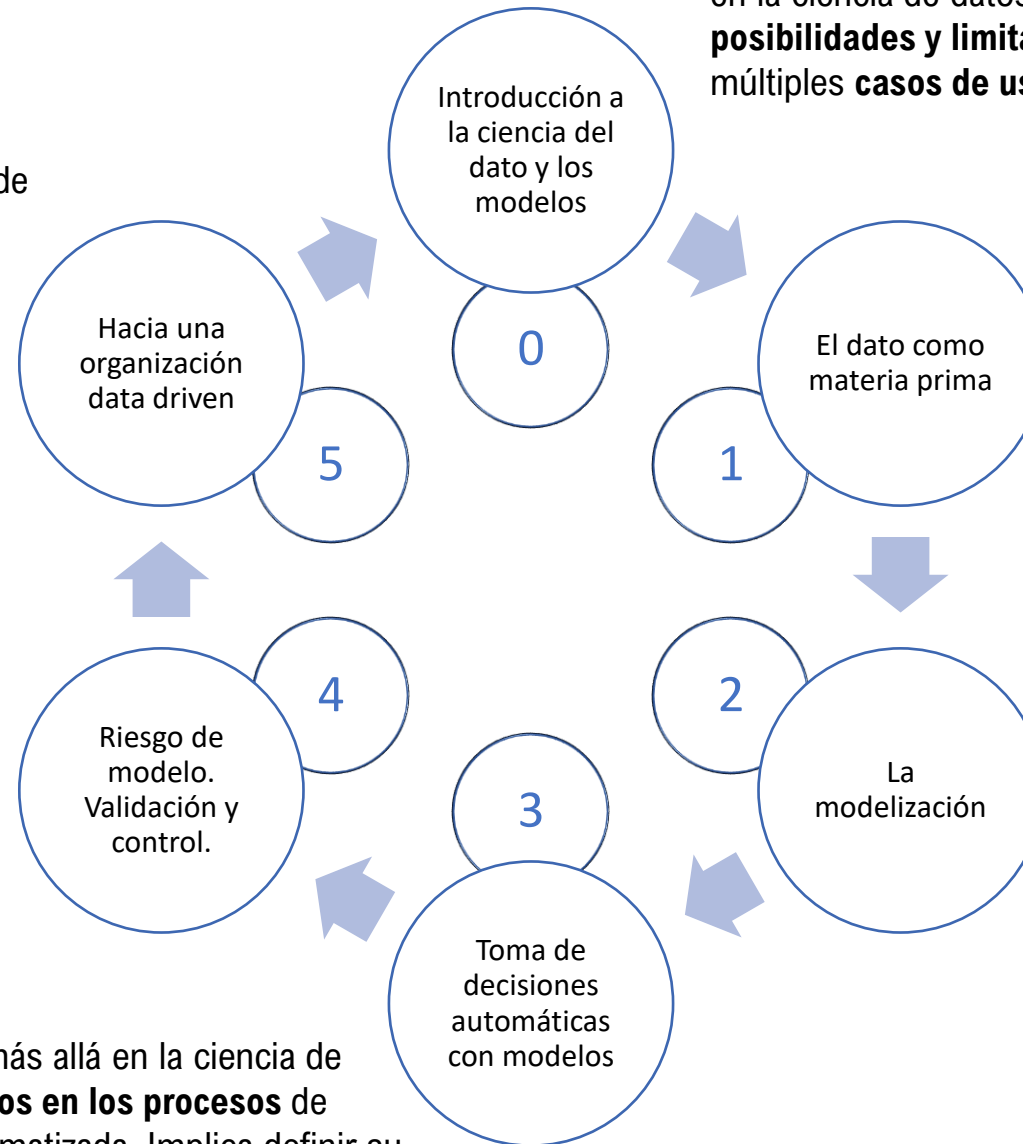
Toda acción conlleva un **riesgo**. Identificar el riesgo que estamos dispuestos a asumir (apetito de riesgo), la forma de medirlo, los **mecanismos de control** necesarios y las necesidades de **mantenimiento** y actualización asociados al uso de modelos son los objetivos de este tema.

**Los modelos de decisión** son un paso más allá en la ciencia de datos. Es la **integración de sus resultados en los procesos** de negocio para la toma de decisiones automatizada. Implica definir su funcionamiento con precisión, entender sus consecuencias y asegurar su gestión dinámica.

Estructuraremos los conocimientos del directivo en la ciencia de datos y entenderemos sus **posibilidades y limitaciones** a través de múltiples **casos de uso**.

El dato es un activo que se da por supuesto y la falta de atención sobre el mismo suele generar la mayor parte de los problemas en el desarrollo de modelos útiles. Aprenderemos los **aspectos clave en el tratamiento del dato y la infraestructura asociada**.

El objetivo de este tema es dotar al directivo de los conceptos necesarios para poder **dialogar con el científico de datos y acompañarle en el proceso de construcción de modelos**, asegurando que así cumple los objetivos buscados y que la calidad y capacidad del modelo es la adecuada.



## Estructura del programa

- Cada tema está estructurado en diferentes **módulos suficientemente autónomos** entre sí para **configurar un programa a medida según las necesidades** específicas. Para ello, se han configurado tres niveles de conocimiento:

### Contenido troncal

- Los módulos pertenecientes al contenido troncal dotan al participante de una visión integrada y amplia de los conceptos necesarios para entender el mundo del big data y los modelos.
- Se tratan todos los **temas relevantes**, con el nivel de profundidad necesario para comprender los puntos clave y los riesgos asociados.

Hasta  
24h

### Contenido experto

- El contenido experto facilita al participante un nivel de **profundidad mayor en temáticas específicas**.
- Por tanto, se consideran módulos de valor añadido para el directivo.
- **Son módulos independientes** entre si.

Hasta  
+16h

### Contenido complementario

- El contenido complementario incluye materias adicionales, que pueden tratarse de forma separada de las materias troncales y expertas y que abordan **temas particulares de interés y actualidad**.
- Por su naturaleza, son módulos que **pueden abordarse sin haber realizado ningún modulo experto**.

Hasta  
+48h

## 0. Introducción: posibilidades del Big Data y la modelización

- Introducción a los contenidos: diferenciación de conceptos clave
- ¿Qué es un modelo?
- Ejemplos aplicados de modelos
- La utilidad de los modelos
- El camino para llegar a integrar un modelo en los procesos de gestión
- Roles clave

Identificar el sentimiento de marca en la web

Conocer el valor y liquidez de un inmueble por su localización y servicios disponibles

Identificar a qué clientes o no clientes llamar y a qué hora es más adecuada

Precios y ofertas en función del tiempo, lugar, cliente...

Tener disponibles los precios a los que vende la competencia y cómo se mueven en el tiempo

Valorar automáticamente el riesgo que supone financiar las compras de nuestros clientes

Identificar localizaciones para instalaciones de energía solar



# 1. El dato como materia prima

- La importancia del dato
- Tipos de datos
- El proceso de captura de datos. Una visión de proceso y negocio
- *Web scrapping, text mining y otras técnicas específicas*
- IoT y otras fuentes de datos
- El diccionario de datos
- El gobierno de los datos
- Apuntes sobre la protección de datos
- El control de calidad de los datos
- Almacenamiento de los datos y herramientas de explotación
- Tratamiento de los datos

¿Qué es el web scrapping?

¿Cómo optimizar la recogida de datos en mi proceso de gestión del cliente?

¿Qué es Python?

¿Cómo asegurar que los datos usados son fiables?

¿Se pueden tratar textos o imágenes para extraer datos?

¿Qué fuentes de datos externas existen?

## 2. Modelización

- Tipologías de modelos
- La definición del objetivo
- Segmentación de datos
- Supuestos de construcción
- Elección de variables y complejidad
- Algoritmos de modelización
- Niveles de conservadurismo
- Herramientas de modelización
- Visualización del modelo
- Machine learning

¿Cómo se eligen las variables relevantes de un modelo?

¿Puedo construir un modelo que me anticipe la intención de compra de un cliente en base a su visita a mi web?

¿Cuántos datos necesito para construir un buen modelo?

¿Qué importancia tiene la correlación de variables?

¿Para hacer el modelo, que tengo que saber del comportamiento de mis clientes?

### 3. Automatización de decisiones

- Integración en el proceso
- Criterios mínimos
- Puntos de corte
- Reglas de decisión
- Zona Gris
- Estrategias *champion-challenger* y otros procesos de mejora de los modelos de decisión
- El gobierno de los modelos de decisión
- El rol del negocio

¿Una vez que tengo el modelo, puede resolver cualquier situación?

¿Qué es un árbol de decisión?

¿Durante cuánto tiempo es válido un modelo?

## 4. Riesgo de modelo y control

- Métricas clave. Poder predictivo
- Estabilidad
- *Backtesting*
- Validación de modelos
- Modelos de caja negra
- El modelo de control y el cuadro de mando
- Gobierno y líneas de defensa
- El rol de la alta dirección y del consejo en el uso de modelos

¿Cómo puedo saber si un modelo es bueno o malo?

¿Qué son las tres líneas de defensa?

¿Qué es el índice de Gini o el KS de un modelo?

¿El consejo debe conocer los modelos que se usan en el negocio?

## 5. Hacia una organización *data driven*

- ¿Qué es una organización *data driven*?
- Metodología Agile aplicada a la implantación de modelos
- Gestión y motivación del equipo de científicos de datos
- I+D
- La documentación como valor añadido
- Errores habituales en el proceso de gestión de los modelos
- El entorno legal vigente y en camino
- Ética en la sociedad del dato

¿Cuáles son las tendencias actuales en big data?

¿Qué aporta la documentación al proceso?

¿Qué puede aportar agile en el desarrollo de modelos?

¿Cómo retener a los científicos de datos?

¿Qué condiciones son necesarias para ser una organización *data driven*?

## Metodología del programa

- El programa **combina**, según la naturaleza de cada módulo, **diferentes aproximaciones formativas**:
  - La **exposición** del tema por parte del profesor para aquellos aspectos más conceptuales, que requieren que el alumno asimile conceptos con explicaciones sencillas y claras, adaptadas a su negocio y su perfil profesional. Este formato permite aclarar dudas y asegurar que se alcanzan los mínimos que luego facilitan la ejecución de las prácticas.
  - Mediante **casos** se abordarán aquellas partes del programa que pretenden mostrar las diferentes posibilidades de aplicación práctica de los conceptos más relevantes. Se verán casos que muestran las posibilidades que se pueden lograr con determinadas acciones así como otros casos que ayudarán a visualizar los riesgos y errores más frecuentes.
  - Las **prácticas** están destinadas a que el asistente experimente en primera persona algunas de las acciones que normalmente realizará un científico de datos, lo que permitirá, por un lado, desmitificar su complejidad, al tiempo que valorar la importancia de algunas de estas acciones para lograr un buen resultado.
  - Los **debates** están orientados, a partir de casos reales o teóricos, a ayudar al participante a valorar los pros y contras de determinadas opciones, de forma que puedan integrar los conceptos aprendidos en sus propias decisiones de negocio.
  - Las **masterclass** son sesiones especializadas, donde en ocasiones participará un formador ad hoc que dará una visión integrada de un tema específico, incluyendo ejemplos reales.
  - Las **auto evaluaciones** a través de dinámicas rápidas de juego.

# *Boosting a healthful data driven society*



[www.ethalo.es](http://www.ethalo.es)

☰ [info@ethalo.es](mailto:info@ethalo.es)

☎ 615 90 31 61