



GUÍA PRÁCTICA SOBRE MACHINE LEARNING

Índice

Introducción	3
Aplicación del machine learning y su impacto en el negocio	4
Dos aspectos fundamentales que no debes olvidar en un proyecto de machine learning	6
La calidad del dato es determinante	
¿Tiene sentido de negocio?	
Unbox the black box o cómo interpretar los algoritmos de machine learning	10
La aplicación del Machine learning por sector	13
Sector Retail: El Customer Intelligence Journey, el resultado de aplicar Machine Learning en la relación con los clientes	
Sector Banca: El Big Data nos desvela las verdaderas relaciones entre empresas	
Sector Travel & Leisure: Estudio de transformación digital y machine learning	

Introducción



La irrupción de nuevas técnicas de modelación adaptadas y asociadas al gran aumento en la capacidad de computación de los ordenadores están posibilitando la adopción y ejecución de los algoritmos de **machine learning** para resolver muchos problemas de negocio de forma más eficiente y con mayor acierto que con técnicas anteriores.

Su aplicación ofrece un amplio abanico de posibilidades que facilitan la resolución de dificultades, permitiendo abordar nuevas soluciones para crear un paradigma en que la captación, transformación y el uso de los datos dirigen las decisiones estratégicas de negocio desde el más alto nivel.

Aplicación del *machine learning* y su impacto en el negocio

Los conceptos técnicos de *machine learning* se llevan rápidamente a la práctica en el negocio para conseguir un retorno. A continuación nombramos algunos ejemplos de cómo esta aplicación del machine learning impacta realmente en los negocios:

- En **publicidad programática**, permite hacer campañas mejor dirigidas y con una tasa de conversión mayor. Para conseguirlo, hay que aprovechar los datos de las redes sociales (intereses, likes, artículos, opiniones), determinar los atributos relevantes para la campaña y aprender del perfil y del producto/servicio anunciado para establecer las características que aumentan la tasa de acierto para cada usuario.
- Proporciona una mejor **recomendación** sobre los productos y una personalización de la oferta. Para ello, es necesario basarse en el histórico de transacciones, opiniones, frecuencia de compra, gustos, etc. del usuario y enriquecer la información con datos de usuarios similares en cuanto a perfil de consumo y perfil sociodemográfico.
- Se puede **controlar el fraude** ya que mediante el histórico de transacciones de clientes se establecen modelos que estiman la probabilidad de que una transacción sea fraudulenta, basándose en transacciones similares como en comportamientos de clientes gemelos.

- Permite una **optimización de pricing**. Con el fin de establecer una tasa de interés acorde a cada cliente en concreto, se estudia el histórico de tasas aceptadas/rechazadas y se cruza dicha información con su posición financiera y con datos macroeconómicos y sociodemográficos que ayuden a establecer las correlaciones entre las variables. Con las redes de relacionamiento bancario se puede tener una mejor aproximación del resultado.
- A través de **Real-time Analytics** consigue obtener una respuesta inmediata ante un disparador externo. La información puede provenir de sensores, reacción ante un evento de compra, incidencia en una planta de generación de energía, etc. Si, además, se tiene acceso al seguimiento de los resultados, se pueden modificar/reentrenar los algoritmos semiautomáticamente para adecuarse a nuevas condiciones.
- Determina qué **campañas, medios, mensajes y ofertas** han tenido un mayor impacto, teniendo en cuenta factores externos (competencia, calendario), tendencias y estacionalidades intrínsecas al negocio, con lo que se determina el efecto real de la campaña.
- También permite concluir el comportamiento de clientes similares a través de la **segmentación de clientes**, haciendo que la gestión de los mismos sea más eficiente y personalizada. En estos casos, se utiliza información sociodemográfica, pero puede ser enriquecida con variables que aporten información sobre el grado de madurez digital de los clientes, así como de su sofisticación financiera. La información de redes sociales permite afinar los segmentos finales y las acciones comerciales.
- En el campo de la energía, estima el consumo y la **demandas energéticas** para aprovisionar o redimensionar la red de abastecimiento. Además del análisis de la propia serie temporal histórica, se deben tener en cuenta información externa, como eventos, actividad turística, climatología, etc. Establecer una probabilidad de fallo para los componentes críticos en la red y realizar estimaciones de crecimiento de la zona, precio de la energía, etc.

Las aplicaciones son interminables pero para que un proyecto tenga éxito siempre debemos tener en cuenta la calidad del dato y el sentido de negocio de la solución.

Dos aspectos fundamentales que no debes olvidar en un proyecto de machine learning

Machine Learning, Artificial Intelligence, Big Data... Son términos que hoy en día se repiten constantemente en cualquier segmento de negocio, debido sobre todo a que las nuevas tecnologías están impulsando una transformación digital en las compañías, **el volumen de datos generados crece exponencialmente y los avances tecnológicos están posibilitando capacidades de almacenamiento, cálculo, desarrollo e implantación de procesos muy superiores a los de hace años.**

Existe una gran cantidad de bibliografía sobre todo lo que concierne a algoritmia, figuras de mérito y metodología de modelación del machine learning (ML), pero, para aplicar todo ese marco teórico a soluciones de negocio accionables y medibles, se deben ejecutar una serie de fases que, aunque no sean complejas o parezca que no aportan valor, son fundamentales a la hora de ejecutar el proyecto con éxito. Hablamos de la **calidad del dato** y el **sentido de negocio de la solución**, que por lo general son obviadas o ni siquiera tenidas en cuenta en la planificación del proyecto.

La calidad del dato es determinante

Esta será la fase que defina toda la solución. Es imposible (o muy poco probable) que se encuentre una solución satisfactoria si nuestros datos no tienen la calidad suficiente. El nuevo paradigma analítico nos capacita para tener **acceso a una cantidad enorme de datos**, con diferentes fuentes y estructuras, formato, temporalidades y niveles, que hacen que la labor de construir un entorno analítico de modelación (datos estructurados, limpios y optimizados) del que se alimentarán los algoritmos de ML sea una fase compleja y tediosa pero necesaria y obligatoria. **Si a la entrada tenemos datos erróneos, a la salida tendremos resultados erróneos.**



Para ello, recomendamos, en primer lugar, **centrarse en los datos que a priori ayuden a resolver el problema**. Obviamente debemos contar con la experiencia y talento adquirido en proyectos anteriores para estimar qué variables son las que, en una primera iteración, son fundamentales para llegar a nuestros objetivos y realizar las primeras mediciones de resultados. A medida que vayamos avanzando, podemos ir enriqueciendo nuestro universo de variables.

En segundo lugar, **asegurarse de que los datos sean correctos**. De cada posible variable, debemos eliminar errores, ruido, conflictos, sesgos; y desarrollar tareas automatizadas que incluyan la limpieza, estandarización, optimización y reestructuración del dato para que el algoritmo pueda ejecutarse de manera automatizada. Por lo general, es mejor **premiar la calidad a la cantidad** de datos a la hora de desarrollar el proyecto.

Estas tareas, aunque pueden resultar de valor escaso, sobre todo para los usuarios finales que van a llevar a cabo la aplicación de la solución al problema de negocio, poseen una influencia crucial en el proyecto.

¿Tiene sentido de negocio?

Básicamente, los proyectos basados en modelos ML para predecir/clasificar un evento pueden categorizarse, según si el resultado conlleva o no una acción externa.



En el primer caso, si el evento a predecir es la **fuga de clientes**, el modelo identifica clientes propensos a ello, pero se debe hacer una acción (comercial en este caso) para evitar que ocurra. En el segundo caso, si el evento es la detección de fraude, en principio poco importa “por qué” ocurre, simplemente se busca maximizar la capacidad de acertar.

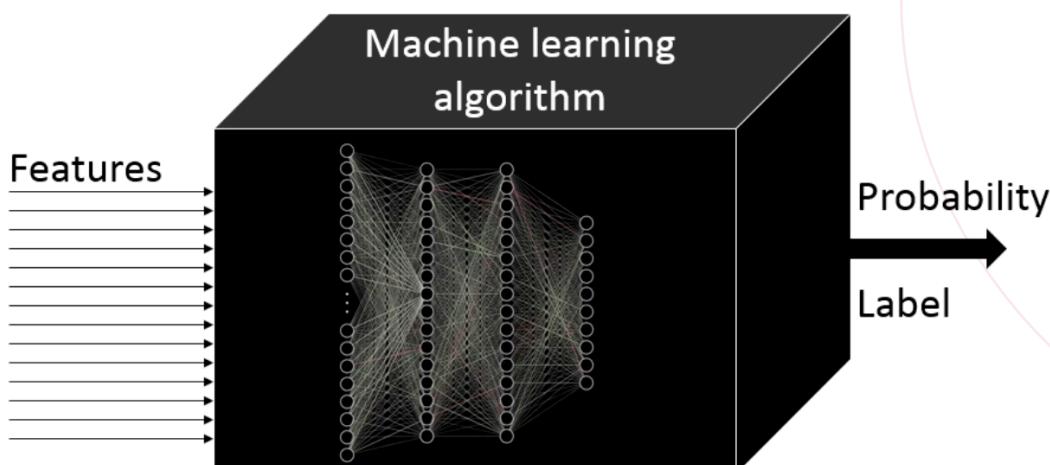
Si el proyecto requiere de una **acción externa**, debemos certificar que los perfiles, conclusiones e inferencias a las que ha llegado el modelo ML concuerdan con las premisas de negocio. La validación del sentido de negocio de la solución debe hacerse posterior y completamente abstraída de la fase de modelación. En el extremo y dependiendo de cada problema en cuestión, nos puede compensar un modelo ML que, aunque acierte menos, sea más interpretable y accionable en términos de negocio, ya que a futuro producirá mayor retorno de la inversión.

En conclusión, el desarrollo satisfactorio de un proyecto ML aplicado a un problema de negocio real tiene más fases que la modelación y elección del algoritmo ganador. En general, no es suficiente con tomar una gran cantidad de datos, dejar que sea la máquina la que aprenda automáticamente y poner en producción la solución obtenida. La experiencia nos justifica que si se quieren tener soluciones sólidas, accionables y estables se deben tener en cuenta aspectos tanto anteriores como posteriores a la elección del algoritmo ML ganador.

Unbox the black box o cómo interpretar los algoritmos de machine learning

El uso de algoritmos de machine learning para resolver problemas (en los que principalmente se requiera obtener una predicción o estimación de que un evento ocurra) se ha extendido de tal manera que ya prácticamente todas las corporaciones que gestionan clientes poseen modelos para adelantarse a los eventos clave que definen su negocio. Sin embargo, **los resultados analíticos y la aplicación estratégica al negocio suelen hablar idiomas diferentes** y no siempre se consiguen los resultados esperados (en forma de retorno de la inversión) por la dificultad de aunar estas dos áreas.

Por ello, nos centremos precisamente en cómo aplicar los resultados de los algoritmos de machine learning a aquellas soluciones de negocio



que requieran una palanca de acción, esto es, no solo se requiere tener una estimación del evento a predecir (generalmente comportamientos de cliente asociados a retención, cross/up selling, satisfacción, estrategia de e-mail marketing), sino también qué acción realizar para incentivar o evitar dicho comportamiento (qué hacer para retenerlo, para potenciar la atracción de nómina, para maximizar el click-through rate, etc.)

Dentro de machine learning existen multitud de algoritmos (cada uno con diversos hiperparámetros a ajustar) que proveen de diferentes resultados en términos de acierto ante un mismo fenómeno que queremos predecir. En esta fase de selección del modelo ganador, pueden entrar varios criterios. Dependiendo de los recursos disponibles suelen ser: acierto, complejidad, tiempo de cómputo, etc. siendo en general una ponderación de todos ellos la que determine el modelo a poner en producción.

Hoy día el hardware y las librerías de machine learning permiten la ejecución de modelos computacionales complejos con un alto nivel de abstracción, que suelen presentar una estimación muy sólida respecto al evento que queremos predecir, pero dicha abstracción implica una pérdida de control y entendimiento respecto a lo que ocurre dentro de la máquina, la cual básicamente está recalculando en cada iteración una matriz de pesos que relaciona las entradas con la salida.

Por ello, describir (incluso matemáticamente) el proceso de aprendizaje de la máquina es muy complejo, ya que pasa por varios estados hasta llegar al final del proceso. **En todo algoritmo de machine learning existe una intrínseca deficiencia de transparencia en lo que está haciendo la máquina** (aprende automáticamente, sin supervisión), pero precisamente esa falta de transparencia en el aprendizaje automático permite que puedan llegar a resultados óptimos ante problemas complejos que un ser humano no podría alcanzar.

No obstante, ello no debe implicar tener una fe absoluta en los resultados ni dejar en un segundo plano la interpretación de los mismos. Por lo general, **modelos que puedan ser explicados en términos de negocio suelen ser más estables en el tiempo y presentar unos mejores retornos de la inversión**. Por ello, para poder esclarecer y entender cómo “piensa” la máquina se proponen las siguientes técnicas:

- Probar la máquina ya entrenada usando valores test de entrada con el valor conocido de la salida para ver cómo la variación de la estimación y del error depende de qué valores de entrada y en

qué medida. Con ello podríamos comprobar si variables sociodemográficas como edad o ingresos son relevantes y en qué grado sobre la salida.

- Extraer las reglas que la máquina ha codificado. No es necesario intentar interpretar la máquina según los pesos de las conexiones, ya que mediante los árboles de decisión se puede obtener una aproximación representativa de la máquina. Para lo cual, se debe tomar el output que provee la máquina (no el output real) como target a la hora de ejecutar el árbol de decisión. Este proceso nos permite obtener una “visualización” de las reglas que permiten obtener insights y comprobar qué features son relevantes a la hora de realizar predicciones
- Del mismo modo, se puede ejecutar una segmentación con el top X% de registros según la estimación que provee la máquina, y comprobar cuáles son y en qué medida influyen las features a la hora de encontrar grupos homogéneos entre sí y heterogéneos entre los registros de entrada. A pesar de que pueda haber comportamientos mezclados, si que se tienen las características principales del conjunto de datos que la máquina decide que son más propensos al evento
- Se puede entrenar la máquina con un conjunto reducido de features y comprobar la degradación de los resultados comparándolo con el dataset entero. De esta manera, se tiene una aproximación de la importancia de cada feature e incluso comprobar si la información que aporta puede ser inferida a través de otras features. De igual manera, se puede entrenar la máquina con features a las que hemos añadido una perturbación (generalmente ruido gaussiano) y de nuevo ver cómo se han degradado las prestaciones en función de a qué variables hemos añadido el ruido
- Por último, existen librerías y software propio que realizan una selección automática del mejor algoritmo y sus hiperparámetros, preveyendo de la importancia de cada feature, con lo que se puede estimar cuáles son las más determinantes

De esta manera, además de tener otro criterio mediante el cual elegir al modelo ganador (la interpretabilidad del mismo) podemos definir palancas de acción según qué features son las que determinan el comportamiento de los clientes, y realmente poder accionar modelos de caja negra mediante campañas en modo prescriptivo.

La aplicación del *machine learning* por sector

Sector Retail: El Customer Intelligence Journey, el resultado de aplicar Machine Learning en la relación con los clientes

El **Customer Journey** se ha convertido en el vehículo de comunicación ideal entre las compañías y los consumidores. Gracias a la tecnología actual los retailers pueden poner el conocimiento del cliente al servicio del negocio, utilizando la inteligencia analítica partiendo de toda la información del cliente obtenida a través del call center, las encuestas, el CRM operacional, las RRSS o los programas de fidelización.

En Cognodata aplicamos Machine Learning en el Customer Journey con el fin de **aportar inteligencia en los procesos de relación con los clientes**. Como resultado tenemos lo que conocemos como “**Customer Intelligence Journey**” y podemos destacar tres fases dependiendo de donde se produzcan los momentos de contacto.

1. Antes de la compra

En esta primera fase es fundamental la monitorización del momento de las interacciones del cliente, las consideraciones demográficas y los hábitos de consumo. Según el Decálogo del Ecommerce en España 2017 de IAB Spain, los compradores utilizan las siguientes fuentes para informarse:

- un 53% a través de la Web de la marca
- un 43% por amigos y familiares
- un 41% a través de foros y blogs
- un 26% por redes sociales.



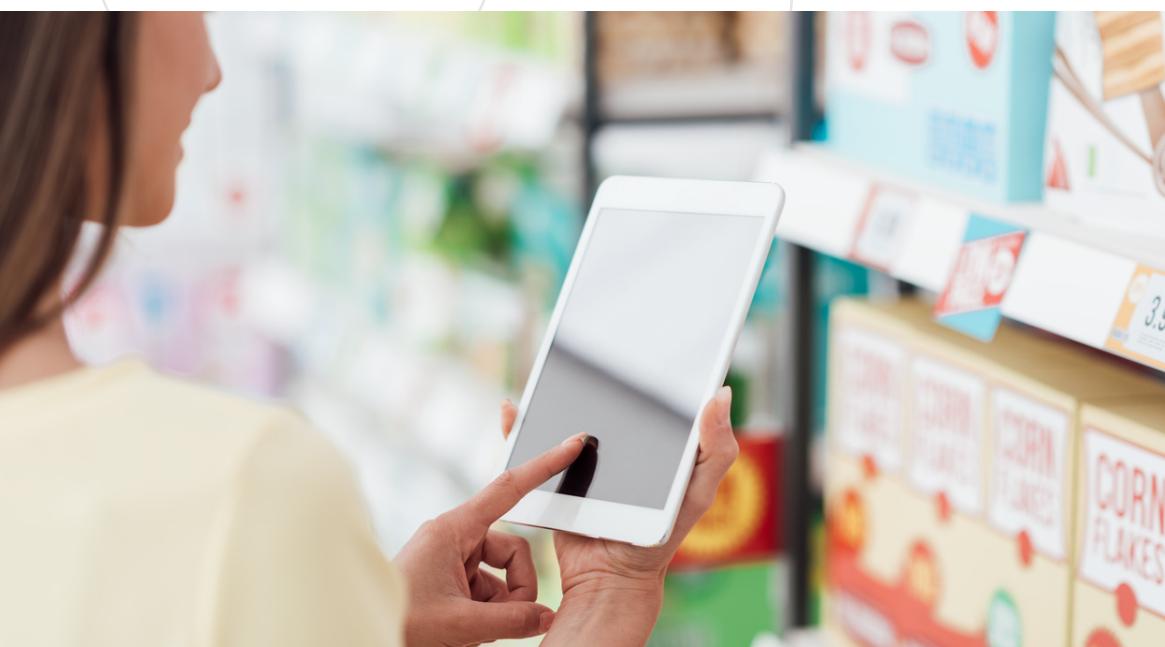
2. Durante la compra.

Es la **fase crítica** tanto en el acto de la compra como en la entrega del producto.

En la compra, las promociones inteligentes impactan en el punto de contacto, ubicación y momento del cliente en el *customer journey*. Las más efectivas son aquellas que apuntan a los **impulsos tanto**:

- **Personales:** nacimiento de bebé, vuelta al cole, cumpleaños, etc.
- **Negocio:** rebajas, blackfriday, etc.
- **Meteorológicos:** tarde de lluvia...
- **Festivos:** Reyes Magos, Día del Padre, etc.

En cuanto a la entrega, según el informe de IAB Spain, los consumidores valoran más las **características del envío** (79%) que la confianza (77%) o el precio (53%). Además 1 de cada 3 clientes estaría dispuesto a pagar por recibir antes su producto.



3. Despues de la compra.

Los retailers no deben olvidar a sus clientes una vez finalizada la compra. Deberán **acompañarles para entender mejor su comportamiento** a través de sus gustos o exigencias no resueltas; sus hábitos, con las rutinas o costumbres y realizar un seguimiento de ubicación monitorizando el geoposicionamiento del cliente.



En definitiva, el **machine learning** nos posibilita crear una mejor experiencia de cliente y acompañarlo, a partir de un inmenso número de datos e informaciones del cliente a lo largo de su Customer Journey y de las tiendas. La incorporación de las nuevas variables de dónde está el cliente (geoposicionamiento) y el momento donde se encuentra (*real-time*) son claves para potenciar el acierto de las acciones con el cliente, aplicando la inteligencia analítica a lo largo de su *Customer Intelligence Journey*.

Sector Banca: El Big Data nos desvela las verdaderas relaciones entre empresas

Teniendo en cuenta la cantidad de movimientos bancarios que realiza una persona física a través de su entidad principal, si pasamos a situarnos en la óptica de un cliente empresarial, ¿cuántos miles o cientos de miles de movimientos al cabo de un mes realiza una gran compañía? Pagos de nómina y gastos a trabajadores, pagos a proveedores, cobros de servicios o productos a clientes, liquidaciones de contratos, etc.



Los **algoritmos de machine learning** ayudan a identificar, medir y entender las dependencias entre empresas a través del análisis de todos sus movimientos, las llamadas redes económicas. Por otro lado, las **soluciones de Big Data** impulsan y mejoran los análisis predictivos sobre estas redes. Podemos ir incluso más lejos y establecer relaciones causa – efecto entre empresas que no se relacionan de manera directa, sino a través de compañías intermedias.

Algunas de las **utilidades** más interesantes para la banca son:

- **Anticiparse a las necesidades de los clientes analizando los movimientos de las carteras de empresas:** identificar necesidades de productos muy específicos y ofrecerlos en tiempo récord.
- **Identificar clientes potenciales a través de los movimientos:** clientes potenciales concretos y su grado de vinculación con la compañía, pudiendo mejorar las actividades comerciales de desarrollo y retención de los clientes.
- **La optimización de riesgos:** analizando las relaciones económicas de las empresas, se pueden optimizar los scores de riesgo tradicional o anticiparse a contagios en todo el sistema.

En definitiva, apostamos por usar algoritmos de Big Data y técnicas de machine learning para descifrar lo que denominamos redes económicas, entendiéndolas como todas las relaciones entre empresas que implican un flujo monetario, así como las relaciones entre empresas y personas físicas. Descifrar todas esas relaciones del tejido empresarial equivaldría a entender una economía con mucha mayor profundidad, generando un conocimiento de incalculable valor.

Sector Travel & Leisure: Estudio de transformación digital y machine learning

Cognodata ha realizado un estudio entre las empresas punteras del sector Travel & Leisure sobre los **cambios que se están produciendo en las organizaciones para la gestión del cliente en entornos digitales, aprovechando las nuevas oportunidades tecnológicas como el machine learning**. Para la realización del estudio ha contado con la colaboración de diferentes empresas pertenecientes a los sectores hotelero, juego, agencias de viaje, cine, museos y aerolíneas.

El estudio muestra los parámetros en los que se deberían focalizar las empresas en base a los cambios del mercado y entornos digitales.



Por un lado, la estrategia empresarial debe aprovechar las nuevas oportunidades generadas por la situación actual de crecimiento y expansión en dos vías:

La inversión:

1. Una inversión enfocada a la **mejora de objetivos y beneficios** en herramientas, como el *machine learning*, para trabajar información masiva de cliente y actualizada a través de redes sociales, big data y CRM.
2. Inversión para **generar servicios** con el mejor *customer experience*, precio y calidad que garanticen una mejora en la atención al cliente, mayor valor añadido y una mejora en la reputación.
3. Por último, inversión en la **organización integral y en la especialización** genera nuevos profesionales y nuevos departamentos. Paralelamente, es necesaria una gestión integral y transversal donde todos los departamentos se tengan en cuenta.

Los nuevos nichos:

Por otro lado, una estrategia de cliente enfocada en la innovación que permita un conocimiento integral del cliente. Esto genera en primer lugar, un mayor grado de conocimiento y en segundo lugar, mayor periodicidad y profundidad analítica, ambas variables se retroalimentan.

A su vez, y junto con las herramientas adecuadas como *machine learning*, Big Data, redes sociales, CRM o gestor de compras, se generan eficientes campañas y planes de fidelización. En cuanto a las campañas, se han convertido en hiper-personalizadas gracias a la innovación y al poder que ha cobrado el cliente mientras que los programas de fidelización están ahora centrados en *customer experience*, precio y calidad.

cognodata,

Travel and Leisure

Estudio sobre la gestión de cliente en entornos digitales, donde se han incluido diferentes empresas pioneras del sector hotelero, juego, agencias de viaje, cine, museos y aerolíneas.

El sector en la actualidad

Tendencia del sector

Tendencia	Porcentaje
Negativa	0%
Estable	20%
Positiva	80%

Tendencia positiva **80%**
Sector en expansión **70%**

Cambios significativos en las estructuras organizativas y tecnológicas **90%**

- Creación de nuevos departamentos y perfiles
- Demanda de más recursos
- Necesidad de hacer mayor foco en experiencia al cliente

Factores

Atención Cliente, Valor añadido, Precio

Desafíos

Digitalización, Reputación, Calidad

Crecimiento del turismo en los últimos años

- Crecimiento **27%**
- Incremento de ingresos **50%**
- Nuevos nichos de mercado

Ecológico
Gastronómico
Tipos de turismo con mayor expansión a corto plazo

Gestión de clientes

Gestión principalmente estratégica

Foco	Porcentaje
Operativa	10%
Estratégica	55%
Funcional	35%

Inversión

Inversión como alta y con tendencia positiva **80%**

Recursos en los que más invierten

Personal

Herramientas

Poder de decisión del Cliente

4 puntos sobre 5

¿Quién toma las decisiones estratégicas?

- 35% Un solo departamento
- 45% Dos o tres departamentos
- 20% Todos o casi todos los departamentos.

cognodata,

Travel and Leisure

Las empresas han cobrado conciencia de la importancia del cliente

Trabajando en...

Estrategias enfocadas en un experiencia cliente más especializada y un aumento de servicio.

Conocer al cliente

Más del **62,5%** conocen mejor a sus clientes, con un porcentaje mayor al **35%**

Periodicidad de los análisis

Periodicidad	Porcentaje
Semanal	40%
Mensual	20%
Trimestral	20%
Anual	15%

90% Socio demográfica Preferencias y gustos

Herramientas

Web, Mail, RSS

Herramientas más utilizadas para la obtención de conocimiento y posterior análisis

BIG DATA
CRM

Conciencia de la necesidad de poseer buena información y herramientas para manejarla, por ello demandan:

Campañas

Campañas de marketing segmentadas **63%**

Tipo de campaña	Porcentaje
Cliente	37%
Producto	50%
Servicio	13%

Estudio de campañas

Admiten necesitar una mejora del análisis **33%**

Recogida de información:

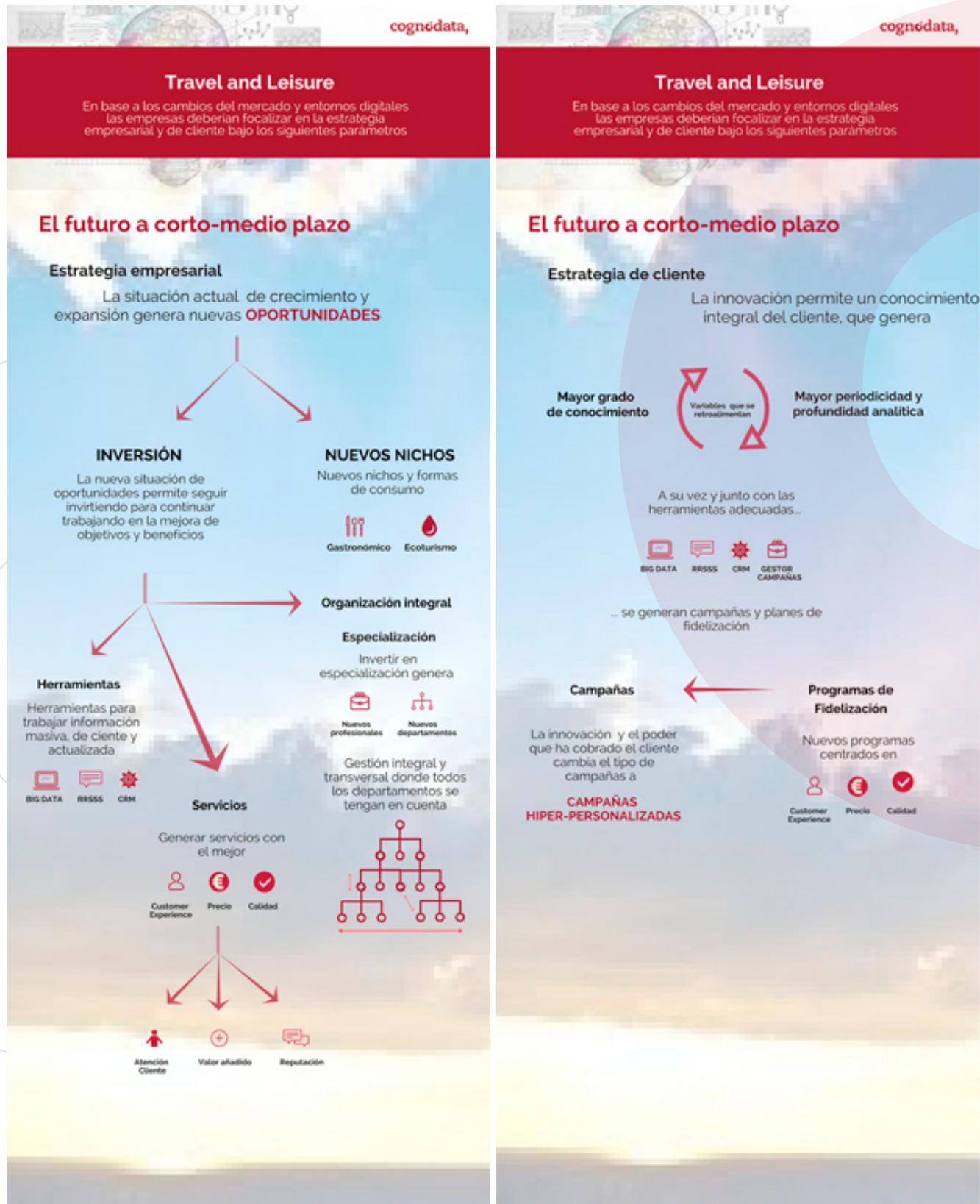
Web, Mail, RSS

Fidelización

No tienen Plan de Fidelización **40%**

Factores que debe percibir el cliente para ser fidelizado

Experiencia cliente, Precio, Calidad



cognodata,



www.cognodata.com
